Lifting with Inner Functions of Polynomial Discrepancy

Yahel Manor
Department of Computer Science, University of Haifa, Israel

Or Meir
Department of Computer Science, University of Haifa, Israel

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.

2012 ACM Subject Classification Theory of computation → Communication complexity; Theory of computation → Oracles and decision trees

Keywords and phrases Lifting, communication complexity, query complexity, discrepancy

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2022.26

Category RANDOM

Funding *Yahel Manor*: Supported by the Israel Science Foundation (grant No. 716/20).
Or Meir: Partially supported by the Israel Science Foundation (grant No. 716/20).

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 of g on a single coordinate. Both questions received much attention in the communication complexity literature, see, e.g., [24, 13, 23, 7, 31, 21, 3, 22, 25, 2, 20, 33, 5, 4].

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 $g\colon \{0,1\}^b \times \{0,1\}^b \to \{0,1\}$ be functions (where $\mathcal{O}$ is some arbitrary set). The block-composed function $f \circ g^n$ is the function that corresponds to the following task: Alice gets $x_1, \ldots, x_n \in \{0,1\}^b$, Bob gets $y_1, \ldots, y_n \in \{0,1\}^b$, 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 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

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

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. [28, 16, 30, 26, 14, 27, 11, 12]), the separation of partition number and deterministic communication complexity [17], proving lower bounds on data structures [10], and an application to the famous log-rank conjecture [19], 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 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. [9, 8]. 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 [29], 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. [9] and Wu et. al. [35] proved a lifting theorem for the case where g is the inner product function. The work of [9] 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 [18] proved a lifting theorem with the inner function g being the index function. In addition, Göös et. al. [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. [8] 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 [8], 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.** *Let* $g : \{0,1\}^b \times \{0,1\}^b \to \{0,1\}$ *be a function, and let* U, V *be independent random variables that are uniformly distributed over* $\{0,1\}^b$*. Given a combinatorial rectangle* $R \subseteq \{0,1\}^b \times \{0,1\}^b$*, 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* $R \subseteq \{0,1\}^b \times \{0,1\}^b$*.*

Informally, the main theorem of [8] says that if $\mathrm{disc}(g) = 2^{-\Omega(b)}$ 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 $D^{\mathrm{cc}}(g) \geq R^{\mathrm{cc}}(g) \geq \Omega(b)$, and therefore the latter result is equivalent to Equation (1).

The research agenda of [8]

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, [8] conjectured that lifting theorems should hold for every inner function g that has a sufficiently large information cost $\mathrm{IC}(g)$.

▶ **Conjecture 2** (special case of [8, Conj. 1.4])**.** *There exists a constant* $c > 0$ *such that the following holds. Let* $f : \{0,1\}^n \to \mathcal{O}$ *and* $g : \{0,1\}^b \times \{0,1\}^b \to \{0,1\}$ *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, [8] 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 [6]. Motivated by this consideration, [8] 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 [8] proves this conjecture for the special case where $\Delta(g) = \Omega(b)$.

Our result

In this work, we prove the latter conjecture of [8] in full, by waiving the limitation of $\Delta(g) = \Omega(b)$ from their result. We note that a full proof can be found in the full version that will be published later. 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 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* $g : \{0,1\}^b \times \{0,1\}^b \to \{0,1\}$ *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 $\beta' = \beta + 2^{-\Delta(g)/50}$.

▶ Remark 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 [31] implies that for every function g such that $\Delta(g) \geq c$ for some universal constant c, it holds that $R^{cc}(g^n) = \Omega(n \cdot \Delta(g))$. Our main theorem can be viewed as a generalization of that result.

▶ Remark 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 gadgets of smaller input length than is currently known which would be considered a significant breakthrough.

▶ Remark 6. In order to facilitate the presentation, we restricted our discussion on the previous work 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., [32, 34, 16, 19, 12, 1]).

▶ Remark 7. We note that Definition 1 is in fact a special case of the common definition of discrepancy. The general definition refers to an arbitrary distribution μ over $\{0,1\}^b \times \{0,1\}^b$. The *discrepancy of g over μ* is defined similarly to Definition 1 except that the random variables U, V are distributed according to μ rather than the uniform distribution.

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 [8] and extend their main technical lemma. We now describe this argument in more detail, focusing on the main lemma of [8] and our extension of that lemma. For simplicity, we focus on the deterministic setting, but the proof in the randomized setting follows similar ideas. In this paper, due to space constraints, the simulation argument is omitted. Only the proof of the main lemma is presented in the paper.

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 $(x, y) \in (\{0,1\}^b)^n \times (\{0,1\}^b)^n$ 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 $(\{0,1\}^b)^n \times (\{0,1\}^b)^n$. 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 [8]

Suppose now that as part of the process described above, we would like to 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 $X_{[n]\setminus F}$ and $Y_{[n]\setminus F}$ to lose their density.

In order to resolve this issue, [8] 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 $X_{[n]\setminus F}$ and $Y_{[n]\setminus F}$ is maintained after the conditioning on $g^I(X_I, Y_I) = z_I$. Therefore, [8] design their decision tree such that before every 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 [8]. 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 [8] requires that $\Delta(g) = \Omega(b)$, and that our goal is to waive that requirement. Unfortunately, it turns out that main lemma of [8] fails when $\Delta(g)$ is very small relatively to b. In fact, the full version provide an example in which *all* the values in the support of X are sparsifying. In such a case, it is simply impossible to remove the sparsifying values.

In short, unlike [8], 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 $X_{[n]\setminus F}$ 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 [8]: 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 [8].

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 [8] 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 [8] 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, it follows from the averaging argument that 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 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 "dangerous" values from the supports of these variables. Our main lemma says that almost all values are safe.

There are two types of "dangerous" values: non recoverable values are values that might lead to a violation of the structure of X, Y (as per density, defined in [8]); non almost uniform values are values for which $g^I(x_I, Y_I)$ is not close enough to uniform and therefore might cause the simulation to leak too much information about X and Y. Additionally, the assumption that $g^I(x_I, Y_I)$ is close to uniform allow the simulation to assume that $X \mid g^I(X_I, Y_I) = z_I$ is close to X even when X, Y are not structured. We first define those notions formally and then compere between those notions and the notions from [8].

▶ **Definition 8** (dangerous values). *Let $\varepsilon, \alpha \geq 0$. Let Y be a random variable and ρ a restriction. Let $x \in \Lambda^n$, and let $\sigma_Y > 0$ be such that $Y_{\text{free}(\rho)}$ is σ_Y-sparse. We say that x is* almost uniform *if for any set $I \subseteq \text{free}(\rho)$ and an assignment $z_I \in \{0,1\}^I$ it holds that*

$$\Pr\left[g^I(x_I, Y_I) = z_I\right] \in 2^{-|I|}\left(1 \pm 2^{-\frac{\Delta}{10}}\right).$$

We say that x is (ε, α)-recoverable *if for all $I \subseteq \text{free}(\rho)$ and z_I the following holds: exist event $\mathcal{E}$ such that $\Pr\left[\mathcal{E} \mid g^I(x_I, Y_I) = z_I\right] \geq 1 - 2^{-\alpha\Delta}$ and the random variable*

$$Y_{\text{free}(\rho)-I} \mid \mathcal{E} \text{ and } g^I(x_I, Y_I) = z_I$$

is $(\sigma_Y + \varepsilon)$-sparse. We say that x is (ε, α)-safe *if it is both almost uniform and (ε, α)-recoverable. Almost uniform, recoverable, and safe values of Y are defined analogously.*

The notion of "dangerous" (Alternatively, not safe) in this paper is closely connected to the definition presented in [8]. We will now discuss the differences and the reasons for the changes. The first type of "dangerous" values, that is non recoverable values, are connected to notion of sparsifying from [8]. Any non recoverable value is sparsifying, but the converse is false. Both definitions regard the sparsity of the random variable $Y_{[n]-I} \mid g^I(x_I, Y_I) = z_I$, if this variable is not dense then it sparsifying. We suggest to "recover" $Y_{[n]-I} \mid g^I(x_I, Y_I) = z_I$ by conditioning it on high-probability event that make this random variable sparse enough. If such option is viable we say that x is recoverable. As show in the full version, using the original definition of sparsifying in the setting of $\Delta \ll b$ can lead to the marking all values x as dangerous, and therefore the weakening is required. Regrading the second type of "dangerous" values, the definition of almost uniform is strictly stronger than the definition of non leaking. Both definition regard the values of $\Pr\left[g^I(x_I, Y_I) = z_I\right]$, almost uniform bound the value tightly both from above and bellow while non leaking bound only from bellow. The definition of almost uniform allow us to get tight connection between $\Pr[\mathcal{E}]$ and $\Pr\left[\mathcal{E} \mid g^I(x_I, Y_I) = z_I\right]$ as can be seen in proof of correctness of the randomized theorem in the full version, where leaking is not sufficient for the analysis of the recovering process.

We turn to state our main lemma.

▶ **Proposition 9** (Main Lemma). *Let $\varepsilon \geq \frac{5}{c}$, $\alpha > \frac{1}{c}$, $\gamma > 0$, and let X and Y be independent (ρ, τ)-structured random variables. Let $\sigma_X, \sigma_Y > 0$ be such that $X_{\text{free}(\rho)}$ is σ_X-sparse, and $Y_{\text{free}(\rho)}$ is σ_Y-sparse. If $\sigma_X + 2\sigma_Y \leq \frac{9}{10} - \frac{22}{c} - \gamma - \alpha$. Then*

$$\Pr_{x \sim X}[x \text{ is not } (\varepsilon, \alpha)\text{-safe}] \leq 2^{-\gamma \cdot \Delta}.$$

In the rest of this section, we prove the main lemma. Let $\varepsilon, \alpha, \sigma_X, \sigma_Y$ be as in the lemma. Additionally, we let X, Y to be independent (ρ, τ)-structured random variables such that $X_{\text{free}(\rho)}$ is σ_X-sparse, and $Y_{\text{free}(\rho)}$ is σ_Y-sparse and let $\tau \stackrel{\text{def}}{=} \sigma_X + \sigma_Y$. We note that we do not assume that $\sigma_X + 2\sigma_Y \leq \frac{9}{10} - \frac{22}{c} - \gamma - \alpha$ in the following lemmas, and some other requirements are used instead. For simplicity, we assume that $\text{fix}(\rho) = \emptyset$ and $\text{free}(\rho) = [n]$ (otherwise, we can restrict our attention to the coordinates in $\text{free}(\rho)$). The first step of the proof is to upper bound the probability that X takes a non almost uniform value.

▶ **Proposition 10.** *Let $\gamma > 0$. Additionally assume that $\tau \leq \frac{9}{10} - \frac{11}{c} - \gamma$. The probability that X takes a value that is not almost uniform is at most $2^{-\gamma \cdot \Delta}$.*

Proposition 10 is proved in Section 2.1 below. We now introduce the definition sparsifying, informally, a value x is (ε, t)-sparsifying with respect to y_J if in the distribution $Y_J \mid g^I(x_I, Y_I) = z_I$ the value $y_J$violets the $(\sigma_Y + \varepsilon)$-sparsity of Y. While all sparsifying values

violate the $(\sigma_Y + \varepsilon)$-sparsity of Y, some of them violate it more strongly than others, and this is measured by the additional parameter t. The next steps of the proof use the following notion.

▶ **Definition 11.** *Let $x \in \Lambda^n$, $J \subseteq [n]$, and $y_J \in \Lambda^J$. We say that x is* (ε, t)-sparsifying *for y_J if there exist $I \subseteq [n]$ and $z_I \in \Lambda^I$ such that*

$$\Pr\left[Y_J = y_J \mid g^I(x_I, Y_I) = z_I\right] > 2^{(\sigma_Y+\varepsilon)\cdot\Delta\cdot|J|+t-b\cdot|J|}.$$

Informally, a value x is not recoverable if it is sparsifying for many *y_J's, whereas a value x is sparsifying according to the terminology of [8] if it is sparsifying for some y_J.*

The following proposition upper bounds the probability that X takes a sparsifying value for specific value y_J, and is proved in Section 2.2.

▶ **Proposition 12.** *Let $\gamma > \frac{2}{c}$. Additionally assume that $\varepsilon \geq \frac{5}{c}$ and $\tau \leq 1 - \frac{14}{c} - \gamma$. Then, for every $J \subseteq [n]$ and for every $y_J \in \Lambda^J$, the probability that X takes an almost uniform value x that is (ε, t)-sparsifying for y_J is at most $2^{-\gamma\cdot\Delta\cdot\frac{c\cdot\varepsilon}{2}\cdot|J|-2t}$.*

Proposition 12 is essentially a more refined version of the analysis in [8]. An important point about this proposition is that it gives a stronger bound for larger values of t. In contrast, the analysis [8] does not consider the parameter t and gives the same upper bound for all values x. In the final part of the proof, which is described in Section 2.3, we derive the main lemma from Propositions 10 and 12.

To prove the propositions in this section we will use the following lemma from [8]

▶ **Lemma 13** (see, e.g., [8, Cor. 2.13])**.** *Let $\gamma, \lambda > 0$ and let $S \subseteq [n]$. If it holds for X, Y that*

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

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

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

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

2.1 Proof of Proposition 10

In this section we prove Proposition 10, following the ideas of [8]. Essentially, the proof uses the fact that X and Y have low sparsity together with the discrepancy of g to argue that with high probability the random variable X_S takes a value x_S such that all parities $g^{\oplus S}(x_S, Y_S)$ are relatively unbiased. Then, the proof uses the latter claim together with the Vazirani lemma to conclude that the random strings $g^I(x_I, Y_I)$ are almost uniform.

▶ **Proposition 10.** *Let $\gamma > 0$. Additionally assume that $\tau \leq \frac{9}{10} - \frac{11}{c} - \gamma$. The probability that X takes a value that is not almost uniform is at most $2^{-\gamma\cdot\Delta}$.*

Proof. We start by observing that for every $x \in \Lambda^n$, if it holds that $\mathrm{bias}(g^{\oplus S}(x_S, Y_S)) \leq 2^{-\frac{\Delta}{10}}\cdot(2n)^{-|S|}$ for every non-empty set $S \subseteq [n]$, then x is almost uniform. Indeed, let $x \in \Lambda^n$ be a value that satisfies the above condition, and let $I \subseteq [n]$. Then, by applying the first variant of Vazirani's lemma to the random variable $g^I(x_I, Y_I)$, it holds that

$$\Pr\left[g^I(x_I, Y_I) = z_I\right] \in \left(1 \pm 2^{-\frac{\Delta}{10}}\right)\cdot 2^{-|I|}$$

for every $z_I \in \{0,1\}^I$. It follows 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 $\text{bias}(g^{\oplus S}(x_S, Y_S)) \leq 2^{-\frac{\Delta}{10}} \cdot (2n)^{-|S|}$ for a specific set $S \subseteq [n]$. Fix a non-empty set $S \subseteq [n]$. By assumption, it holds that

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

By applying Lemma 13 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

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

Next, by taking the union bound over all non-empty sets $S \subseteq [n]$, it follows that the probability that there exists some non-empty set S with $\text{bias}(g^{\oplus S}(x_S, Y_S)) > 2^{-\frac{\Delta}{10}} \cdot (2n)^{-|S|}$ is at most

$$\begin{aligned} &\sum_{S \subseteq [n] : S \neq \emptyset} 2^{-\gamma\Delta - 1} \cdot \frac{1}{n^{|S|}} && \text{(binomial like bound)} \\ &< 2^{-\gamma\Delta - 1} \cdot 2 \\ &= 2^{-\gamma\Delta}. \end{aligned}$$

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

2.2 Proof of Proposition 12

In this section, we prove Proposition 12 using a refined version of the analysis of [8]. The proof consists of three main steps: first, we use Bayes' formula to reduce the task of upper bounding the probability of sparsifying values into the task of 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 $g(x_I, Y_I)$. Finally, we upper bound the biases of $g(x_I, Y_I)$ using the low deficiency of X and Y and the discrepancy of g. We start by formally defining skewing values, and then prove their connection to sparsifying values.

▶ **Definition 14.** *Let $J \subseteq [n]$ and let $y_J \in \Lambda^J$. Let $e(y_J)$ be the real number such that*

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

We note that this number is non-negative as we assume Y is σ_Y-sparse. We say that x is (ε, t)-skewing *for y_J if there exist $I \subseteq [n] - J$ such that*

$$D_\infty\left(g^I(x_I, Y_I) \mid Y_J = y_J\right) > \varepsilon \cdot \Delta \cdot |J| + e\left(y_J\right) + t - 1$$

▶ **Proposition 15.** *Let $x \in \Lambda^n$, $J \subseteq [n]$, and $y_J \in \Lambda^J$. If x is (ε, t)-sparsifying for y_J and is almost uniform then x is (ε, t)-skewing for y_J.*

Proof. The proof is straightforward and been omitted due to space constraints. The proof can be found in the full version of this paper. ◀

We now formally define biasing values and connect 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 $g^{\oplus S}(x_S, Y_S)$ is too high.

▶ **Definition 16.** *Let $J \subseteq [n]$ and let $y_J \in \Lambda^J$. We say that x is* (ε, t)-biasing *for y_J if there exists a set $S \subseteq [n] - J$ such that $|S| \geq c \cdot \varepsilon \cdot |J| + \frac{t+e(y_J)-2}{\log n}$ and*

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

▶ **Proposition 17.** *Let $x \in \Lambda^n$, let $J \subseteq [n]$, and let $y_J \in \Lambda^J$. If x is not (ε, t)-biasing for y_J then x is not (ε, t)-skewing for y_J.*

Proof. The proof is omitted due to space constraints and can be found in the full version of this paper. ◀

We finally prove Proposition 12, restated next.

▶ **Proposition 12.** *Let $\gamma > \frac{2}{c}$. Additionally assume that $\varepsilon \geq \frac{5}{c}$ and $\tau \leq 1 - \frac{14}{c} - \gamma$. Then, for every $J \subseteq [n]$ and for every $y_J \in \Lambda^J$, the probability that X takes an almost uniform value x that is (ε, t)-sparsifying for y_J is at most $2^{-\gamma \cdot \Delta \cdot \frac{c \cdot \varepsilon}{2} \cdot |J| - 2t}$.*

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 (ε, t)-biasing for y_J. Indeed, if x is a value that is not (ε, t)-biasing for y_J, then by Proposition 17 it is not (ε, t)-skewing for y_J, and then by Proposition 15 it cannot be both (ε, t)-sparsifying for y_J and almost uniform. It remains to upper bound the probability that x is (ε, t)-biasing for y_J.

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

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

for some non-empty fixed set $S \subseteq [n] - J$ such that $|S| \geq c \cdot \varepsilon \cdot |J| + \frac{t+e(y_J)-2}{\log n}$. Let S be such a set. In order to upper bound the latter probability, we use Lemma 13, which in turn requires us to upper bound the deficiencies $D_\infty(X_S)$ and $D_\infty(Y_S | Y_J = y_J)$. By assumption, we know that $D_\infty(X_S) \leq \sigma_X \cdot \Delta \cdot |S|$. We turn to upper bound $D_\infty(Y_S | Y_J = y_J)$. For every $y_S \in \Lambda^S$, it holds that

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

It follows that

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

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

$$e(y_J) \leq \log n \cdot |S| + 2 \leq 3 \cdot \log n \cdot |S| \leq \frac{3}{c} \cdot \Delta \cdot |S| \,.$$

It follows that

$$\begin{aligned}
&D(X_S) + D_\infty(Y_S \mid Y_J = y_J) \\
\leq&(\sigma_X + \sigma_Y + \frac{3}{c}) \cdot \Delta \cdot |S| \\
\leq&(1 - \frac{11}{c} - \gamma) \cdot \Delta \cdot |S| && (\sigma_X + \sigma_Y \leq 1 - \frac{14}{c} - \gamma) \\
=&\left(\Delta - \frac{7\Delta}{c} - \gamma\Delta - \frac{4\Delta}{c}\right) \cdot |S| \\
\leq&\left(\Delta - 7 - \gamma\Delta - 3\log n - 1\right) \cdot |S| \,.
\end{aligned}$$

Now, by applying Lemma 13 with $\gamma = \gamma\Delta + 2\log n$ and $\lambda = \log n + 1$, it follows that the probability that X takes a value x such that

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

is at most

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

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,ε)-biasing for y_J is at most

$$\begin{aligned}
&\sum_{S\subseteq[n]:|S|\geq c\cdot\varepsilon\cdot|J|+\frac{t+e(y_J)-2}{\log n}} 2^{-\gamma\cdot\Delta\cdot|S|} \cdot \frac{1}{n^{|S|+1}} \\
\leq&\sum_{S\subseteq[n]:|S|\geq\left(\frac{c\cdot\varepsilon}{2}+2\right)|J|+\frac{t-2}{\log n}} 2^{-\gamma\cdot\Delta\cdot|S|} \cdot \frac{1}{n^{|S|+1}} && (e(y_J)\geq 0, \varepsilon \geq \frac{5}{c}) \\
\leq&\, 2\cdot 2^{-\gamma\cdot\Delta\cdot\left(\left(\frac{c\cdot\varepsilon}{2}+2\right)|J|+\frac{t-2}{\log n}\right)} \cdot \frac{1}{n} && \text{(binomial like bound)} \\
\leq&\, 2^{-\gamma\cdot\Delta\cdot\left(\frac{c\cdot\varepsilon}{2}+2\right)|J|} \cdot 2^{-2t+\left(\frac{\gamma\cdot\Delta\cdot 2}{\log n}\right)} && (\gamma \geq \frac{2}{c} \geq \frac{2\log n}{\Delta}, n \geq 2) \\
\leq&\, 2^{-\gamma\cdot\Delta\cdot\frac{c\cdot\varepsilon}{2}\cdot|J|-2t} \cdot 2^{\gamma\cdot\Delta\cdot\left(\frac{2}{\log n}-2\right)} && (|J|\geq 1) \\
\leq&\, 2^{-\gamma\cdot\Delta\cdot\frac{c\cdot\varepsilon}{2}\cdot|J|-2t} && (\text{n}\geq 2)
\end{aligned}$$

as required. ◀

2.3 Proof of the Main Lemma from Propositions 10 and 12

In this section, we derive the main lemma from the previous propositions. The difficult part is to prove an upper bound on the probability of non-recoverable values x, which is essentially equivalent to proving the following statement:

- There are very few values x that are sparsifying for many values y_J.

Proposition 12 essentially tells us the following statement:

- For every y_J, there are very few values x that are sparsifying for it.

It is tempting to try to deduce the first statement from the second statement via an averaging argument. However, there is a significant obstacle here: in the first statement, when we say "for many values y_J", we count the values y_J with respect to a distribution that depends on x. This complication renders a naive averaging argument impossible. In order to overcome this obstacle, we consider all the pairs (x, y_J) such that x is (ε, t)-sparsifying for y_J, 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 12 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 start by defining the notion of a "light" value of X, which is a value x that is *not* sparsifying for many values y_J for a particular set $J \subseteq [n]$. The term "light" is motivated by the intuitive idea that the relevant values y_J are "heavy" in terms of their probability mass, so a "light" value x is one that does not make many values y_J "heavy". We show that "light" values of x are recoverable, intuitively this is true as one can remove all the relevant values y_J that cause the sparsity by condition on high the probability event of not choosing any of them. We then consider x that are not light with respect to some specific J. We proceed by bounding the probability of x that are not light with respect to single J, and complete the proof by taking union bound over all J.

▶ **Definition 18.** *Let $x \in \Lambda^n$ and let $J \subseteq [n]$. For every set $I \subseteq [n] - J$ and a value $z_I \in \{0,1\}^I$, we denote by*

$$\mathcal{H}_{x,J,I,z_I} \stackrel{\text{def}}{=} \left\{ y_J \in \Lambda^J : \Pr\left[Y_J = y_J \mid g^I\left(x_I, Y_I\right) = z_I\right] > 2^{(\sigma_Y + \varepsilon - \frac{1}{\Delta}) \cdot \Delta \cdot |J| - b \cdot |J|} \right\}$$

*the set of "heavy" values y_J. We say that a value x is (ε, α)-*light *if for every disjoint I, J and $z_I \in \{0,1\}^I$ it holds that*

$$\Pr\left[Y_J \in \mathcal{H}_{x,J,I,z_I} \mid g^I\left(x_I, Y_I\right) = z_I\right] \leq 2^{-\alpha\Delta} \cdot \left(\frac{1}{2n}\right)^{|J|}.$$

▶ **Proposition 19.** *Let $\alpha \geq \frac{1}{\Delta}$. If $x \in \Lambda^n$ is (ε, α)-light with respect to every $J \subseteq [n]$ then it is (ε, α)-recoverable.*

Proof. Let $\alpha \geq \frac{1}{\Delta}$ and let $x \in \Lambda^n$ be (ε, α)-light with respect to every $J \subseteq [n]$. We show that x is (ε, 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 random variable

$$Y_{[n]-I} \mid \mathcal{E} \text{ and } g^I\left(x_I, Y_I\right) = z_I$$

is $(\sigma_Y + \varepsilon)$-sparse. We choose $\mathcal{E}$ to be the event that $Y_J \notin \mathcal{H}_{x,J,I,z_I}$ for any non-empty set $J \subseteq [n] - I$. We first prove that $\Pr\left[\neg\mathcal{E} \mid g\left(x_I, Y_I\right) = z_I\right] < 2^{-\alpha\Delta}$. By the union bound, it holds that

$$\begin{aligned}
&\Pr\left[\neg\mathcal{E} \mid g\left(x_I, Y_I\right) = z_I\right] \\
=&\Pr\left[\bigvee_{\emptyset \neq J \subseteq [n]-I} Y_J \in \mathcal{H}_{x,J,I,z_I} \mid g\left(x_I, Y_I\right) = z_I\right] \\
\leq&\sum_{\emptyset \neq J \subseteq [n]-I} \Pr\left[Y_J \in \mathcal{H}_{x,J,I,z_I}\right] \\
\leq&\sum_{\emptyset \neq J \subseteq [n]-I} 2^{-\alpha\Delta} \cdot \left(\frac{1}{2n}\right)^{|J|} \\
\leq&2^{-\alpha\Delta} \qquad \text{(binomial like bound)}
\end{aligned}$$

It remains to prove that the random variable

$$Y_{[n]-I} \mid \mathcal{E} \text{ and } g^I(x_I, Y_I) = z_I$$

is $(\sigma_Y + \varepsilon)$-sparse. For every $J \subseteq [n] - I$, it holds that

$$\begin{aligned}
&\Pr\left[Y_J = y_J \mid \mathcal{E} \text{ and } g^I(x_I, Y_I) = z_I\right] \\
=&\frac{\Pr\left[Y_J = y_J \text{ and } \mathcal{E} \mid g(x_I, Y_I) = z_I\right]}{\Pr\left[\mathcal{E} \mid g(x_I, Y_I) = z_I\right]} \\
\leq&\frac{\Pr\left[Y_J = y_J \mid g(x_I, Y_I) = z_I\right]}{\Pr\left[\mathcal{E} \mid g(x_I, Y_I) = z_I\right]} \\
\leq&\frac{2^{(\sigma_Y+\varepsilon-\frac{1}{\Delta})\cdot\Delta\cdot|J|-b\cdot|J|}}{1-2^{-\alpha\Delta}} \\
\leq&\frac{2^{(\sigma_Y+\varepsilon)\cdot\Delta\cdot|J|-b\cdot|J|-1}}{1-\frac{1}{2}} && (\text{since } \alpha \geq \frac{1}{\Delta}) \\
\leq&2^{(\sigma_Y+\varepsilon)\cdot\Delta\cdot|J|-b\cdot|J|},
\end{aligned}$$

and therefore the above random variable is $(\sigma_Y + \varepsilon)$-sparse, as required. ◀

▶ **Definition 20.** *Let $x \in \Lambda^n$, $J \subseteq [n]$, $I \subseteq [n] - J$ and $z_I \in \{0,1\}^I$, recall that*

$$\mathcal{H}_{x,J,I,z_I} \stackrel{\text{def}}{=} \left\{y_J \in \Lambda^J : \Pr\left[Y_J = y_J \mid g^I(x_I, Y_I) = z_I\right] > 2^{(\sigma_Y+\varepsilon-\frac{1}{\Delta})\cdot\Delta\cdot|J|-b\cdot|J|}\right\}.$$

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

$$\Pr\left[Y_J \in \mathcal{H}_{x,J,I,z_I} \mid g^I(x_I, Y_I) = z_I\right] \leq 2^{-\alpha\Delta} \cdot \left(\frac{1}{2n}\right)^{|J|}.$$

It is easy to see that by definition a value x is (ε, α)-light if it is (ε, α)-light with respect to every $J \subseteq [n]$. We now use this notion to bound the probability of x been not light for every J and later get bound on the probability that X is (ε, α)-light by binomial like bound.

▶ **Proposition 21.** *Assume that $\sigma_X + 2 \cdot \sigma_Y \leq 1 - \frac{19}{c} - \gamma - \alpha$. For every $J \subseteq [n]$, the probability that X takes an almost uniform value x that is not (ε, α)-light for J is at most $2^{-\gamma\cdot\Delta\cdot|J|}$.*

Proof. Let $J \subseteq [n]$. 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} denote the maximal value t such that x is $(\varepsilon - \frac{1}{\Delta}, t)$-sparsifying for y_J. Next, consider a two dimensional 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

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

Now we use bucketing argument to bound the probabilities of high values of $\text{ent}(x, y_J)$ to occur. Let $\gamma' = \gamma + \sigma_Y + \frac{2}{c} + \alpha + \frac{3}{c}$. By applying Proposition 12 with $\gamma = \gamma'$, we get that for every y_J and every $t \in \mathbb{Z}_{>0}$ it holds that

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

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

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

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

$$\begin{aligned}
&\mathbb{E}\left[\sum_{y_J\in\mathcal{Y}_J} \mathrm{ent}(X, y_J)\right] \\
=&\sum_{y_J\in\mathcal{Y}_J} \mathbb{E}\left[\mathrm{ent}(X, y_J)\right] \\
\leq&\sum_{y_J\in\mathcal{Y}_J} 2^{(\sigma_Y+\varepsilon-\frac{1}{\Delta}-\gamma'\cdot\frac{c\cdot\varepsilon}{2}+\frac{2}{c})\cdot\Delta\cdot|J|-b\cdot|J|} \\
=&2^{(\sigma_Y+\varepsilon-\frac{1}{\Delta}-\gamma'\cdot\frac{c\cdot\varepsilon}{2}+\frac{2}{c})\cdot\Delta\cdot|J|} \\
=&2^{(\sigma_Y+\varepsilon-\frac{1}{\Delta}-\gamma-\sigma_Y-\frac{2}{c}\cdot\frac{c\cdot\varepsilon}{2}-\alpha-\frac{3}{c}+\frac{2}{c})\cdot\Delta\cdot|J|} && \left(\text{definition of } \gamma', \cdot\, \frac{c\cdot\varepsilon}{2} \geq 1\right) \\
\leq&2^{-(\gamma+\alpha)\cdot\Delta\cdot|J|}. && (\Delta \geq c)
\end{aligned}$$

By Markov's inequality, the probability that X is almost uniform and the sum of the X-th row is more than $2^{-\alpha\cdot\Delta|J|}$ 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 $2^{-\alpha\cdot\Delta|J|}$, then x is (ε, 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 (ε, 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

$$\Pr\left[Y_J \in \mathcal{H}_{x,J,I,z_I} \mid g^I\left(x_I, Y_I\right) = z_I\right] \leq 2^{-\alpha\Delta} \cdot \left(\frac{1}{2n}\right)^{|J|}.$$

Observe that for every value $y_J \in \mathcal{H}_{x,J,I,z_I}$, it holds that x is $(\varepsilon - \frac{1}{\Delta}, t')$-sparsifying for y_J with some $t' \geq 0$. Therefore, for every such y_J it holds that $t_{x,y_J} > 0$ and in particular $\mathrm{ent}(x, y_J) = 2^{(\sigma_Y+\varepsilon-\frac{1}{\Delta})\cdot\Delta\cdot|J|-b\cdot|J|+t_{x,y_J}}$. Furthermore, recall that by the definition of t_{x,y_J} it holds that

$$\Pr\left[Y_J = y_J \mid g^I\left(x_I, Y_I\right) = z_I\right] \leq 2^{(\sigma_Y+\varepsilon-\frac{1}{\Delta})\cdot\Delta\cdot|J|-b\cdot|J|+t_{x,y_J}}.$$

It follows that

$$\begin{aligned}
&\Pr\left[Y_J \in \mathcal{H}_{x,J,I,z_I} \mid g^I\left(x_I, Y_I\right) = z_I\right] \\
\leq&\sum_{y_J\in\mathcal{H}_{x,J,I,z_I}} \Pr\left[Y_J = y_J \mid g^I\left(x_I, Y_I\right) = z_I\right] \\
\leq&\sum_{y_J\in\mathcal{H}_{x,J,I,z_I}} \mathrm{ent}(x, y_J). \\
\leq&\, 2^{-\alpha\Delta\cdot|J|}
\end{aligned}$$

as required. ◀

▶ Remark. In the preceded proof we bound the probability that

$$\Pr\left[\lceil t_{X,y_J} \rceil = t \text{ and } X \text{ is almost uniform}\right]$$

. At a first glance the usage of ceiling may be unclear. The ceiling in this argument is merely the implantation of the bucketing argument that used in the proof. Furthermore, the bucketing argument is needed as our tools such as Proposition 12 bound the probability of t to pass some threshold, if we not additionally give upper bound on t then the increment of the contribution for the exception become unlimited and those we need some for of bucketing. On the other hand one suggest creating "zero sized" buckets around every value of t, and thus removing the need for ceiling but that way the sum can be infinite.

We conclude the following bound by taking union bound of the probability for x to be not light over all J, yielding a bound for the probability that x is not light.

▶ **Corollary 22.** *Assume that $\sigma_X + 2 \cdot \sigma_Y \leq 1 - \frac{21}{c} - \gamma - \alpha$. Then, the probability that X takes an almost uniform value x that is not (ε, α)-recoverable is at most $2^{-\gamma \cdot \Delta}$.*

Proof. By applying Proposition 21 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 (ε, α)-light for J is at most $2^{-\gamma \cdot \Delta} \cdot \frac{1}{(2n)^{|J|}}$. By binomial like bound, we obtain that with probability at least $1 - 2^{-\gamma \cdot \Delta}$, the random variable X takes a value x that is (ε, α)-light for $J \subseteq [n]$. Such a value x is (ε, α)-recoverable by Proposition 19, so the required result follows. ◀

We finally complete the proof of Proposition 9, restated next.

▶ **Proposition 9** (Main Lemma). *Let $\varepsilon \geq \frac{5}{c}$, $\alpha > \frac{1}{c}$, $\gamma > 0$, and let X and Y be independent (ρ, τ)-structured random variables. Let $\sigma_X, \sigma_Y > 0$ be such that $X_{\mathrm{free}(\rho)}$ is σ_X-sparse, and $Y_{\mathrm{free}(\rho)}$ is σ_Y-sparse. If $\sigma_X + 2\sigma_Y \leq \frac{9}{10} - \frac{22}{c} - \gamma - \alpha$. Then*

$$\Pr_{x \sim X}\left[x \text{ is not } (\varepsilon, \alpha)\text{-safe}\right] \leq 2^{-\gamma \cdot \Delta}.$$

Proof. Any value that is not (ε, α)-safe must be not almost uniform or almost uniform but not (ε, α)-recoverable. By applying Proposition 10 with $\gamma = \gamma + \frac{1}{c}$, it follows that the probability that X takes a non almost uniform value is at most $2^{-(\gamma + \frac{1}{c}) \cdot \Delta} \leq 2^{-\gamma\Delta - 1}$, By applying Corollary 22 with $\gamma = \gamma + \frac{1}{c}$, $\alpha = \alpha$, and $\varepsilon = \varepsilon$, it follows that the probability that X takes an almost uniform and not (ε, α)-recoverable value is at most $2^{-(\gamma + \frac{1}{c}) \cdot \Delta} \leq 2^{-\gamma\Delta - 1}$. Therefore, the probability that X takes that is not (ε, α)-safe value is at most $2^{-\gamma\Delta - 1} + 2^{-\gamma\Delta - 1} = 2^{-\gamma\Delta}$. ◀

References

1. 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.
2. Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. In *STOC*, pages 67–76, 2010.
3. 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.
4. Mark Braverman. Interactive information complexity. In *STOC*, pages 505–524, 2012.

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

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

7 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.

8 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.

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

10 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.

11 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.

12 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.

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

14 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.

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.

16 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.

17 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.

18 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.

19 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.

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

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

22 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.

23 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.

24 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.

25 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.

26 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.

27 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.

28 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.

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

30 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.

31 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.

32 Alexander A. Sherstov. The pattern matrix method (journal version). *CoRR*, abs/0906.4291, 2009. `arXiv:0906.4291`.

33 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.

34 Yaoyun Shi and Yufan Zhu. Quantum communication complexity of block-composed functions. *Quantum Information & Computation*, 9(5):444–460, 2009. URL: `http://www.rintonpress.com/xxqic9/qic-9-56/0444-0460.pdf`.

35 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.