Query-to-communication lifting for BPP using inner product

Arkadev Chattopadhyay

School of Technology and Computer Science, Tata Institute of Fundamental Research, Mumbai, India
http://www.tcs.tifr.res.in/~arkadev/
arkadev@tifr.res.in

Yuval Filmus

Department of Computer Science, Technion Israel Institute of Technology, Haifa, Israel
https://filmus.net.technion.ac.il/
yuvalfi@cs.technion.ac.il

Sajin Koroth

Department of Computer Science, University of Haifa, Haifa, Israel
https://sites.google.com/csweb.haifa.ac.il/sajin
sajin@csweb.haifa.ac.il

Or Meir

Department of Computer Science, University of Haifa, Haifa, Israel
http://cs.haifa.ac.il/~ormeir/
ormeir@cs.haifa.ac.il

Toniann Pitassi

Department of Computer Science, University of Toronto, Canada
https://www.cs.toronto.edu/~toni/
toni@cs.toronto.edu

Abstract

We prove a new query-to-communication lifting for randomized protocols, with inner product as gadget. This allows us to use a much smaller gadget, leading to a more efficient lifting. Prior to this work, such a theorem was known only for deterministic protocols, due to Chattopadhyay et al. [3] and Wu et al. [20]. The only query-to-communication lifting result for randomized protocols, due to Göös, Pitassi and Watson [11], used the much larger indexing gadget.

Our proof also provides a unified treatment of randomized and deterministic lifting. Most existing proofs of deterministic lifting theorems use a measure of information known as *thickness*. In contrast, Göös, Pitassi and Watson [11] used blockwise min-entropy as a measure of information. Our proof uses the blockwise min-entropy framework to prove lifting theorems in both settings in a unified way.

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

Keywords and phrases Lifting theorems, Innner product, BPP Lifting, Deterministic Lifting

Digital Object Identifier 10.4230/LIPIcs...

Funding *Yuval Filmus*: Taub Fellow — supported by the Taub Foundations. The research was funded by ISF grant 1337/16.
Sajin Koroth: Supported by the Israel Science Foundation (grant No. 1445/16)
Or Meir: Partially supported by ISF grant by the Israel Science Foundation (grant No. 1445/16).

Acknowledgements We thank Daniel Kane for some very enlightening conversations and suggestions. This work was done (in part) while the authors were visiting the Simons Institute for the Theory of Computing.

1 Introduction

In this work, we prove new lifting theorems that use the inner-product function as a gadget. Let $f\colon \{0,1\}^n \to \{0,1\}^m$ and $g\colon \{0,1\}^b \times \{0,1\}^b \to \{0,1\}$ be functions (where g is referred to as a *gadget*). The block-composed function $f \circ g^n$ is the function that takes n instances $(x_1, y_1), \ldots, (x_n, y_n)$ of inputs for g and computes $f \circ g^n$ as,

$$f \circ g^n((x_1, y_1), \ldots, (x_n, y_n)) = f(g(x_1, y_1), g(x_2, y_2), \ldots, g(x_n, y_n)).$$

Lifting theorems are theorems that relate the communication complexity of $f \circ g^n$ to the query complexity[1] of f and the communication complexity of g.

More specifically, consider the following communication problem: Alice gets $x_1, \ldots, x_n$, Bob gets $y_1, \ldots, y_n$, and they wish to compute the output of $f \circ g^n$ on their inputs. The natural protocol for doing so is the following: 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 the i-th instance by invoking the best possible communication protocol for g. A *lifting theorem* is a theorem that says that this natural protocol is optimal.

Lifting theorems are interesting because they create a connection between query complexity and communication complexity. This connection, besides being interesting in its own right, allows us to transfer lower bounds and separations from the from query complexity (which is a relatively simple model) to a communication complexity (which is a significantly richer model).

In particular, the first result of this form, due to Raz and McKenzie [17], proved a lifting theorem from *deterministic* query complexity to *deterministic* communication complexity when g is the index function. They then used it to prove new lower bounds on communication complexity by lifting query-complexity lower-bounds. More recently, Göös, Pitassi and Watson [10] applied that theorem to separate the logarithm of the partition number and the deterministic communication complexity of a function, resolving a long-standing open problem. This too was done by proving such a separation in the setting of query complexity and lifting it to the setting of communication complexity. This result stimulated a flurry of work on lifting theorems of various kinds, such as: round-preserving lifting theorems with applications to time-space trade-offs for proof complexity [5], deterministic lifting theorems with other gadgets [3, 20], lifting theorems from randomized query complexity to randomized communication complexity [11], lifting theorems for DAG-like protocols [7] with applications to monotone circuit lower bounds, lifting theorems for asymmetric communication problems [4] with applications to data-structures, and a lifting theorem [16] for the EQUALITY gadget.

Viewed from another angle, lifting theorems are natural generalizations of classic theorems such as direct-sum theorems and XOR lemmas [21, 13, 6, 14, 1, 2]: in particular, if we set f to be the identity function or the parity function, we get a direct sum theorem or an XOR lemma for g, respectively. This point of view motivates the work of Hatami et al. [12] that made progress towards proving a lifting theorem with a constant-size gadget.

In almost all known lifting theorems, the function f can be arbitrary (and may also be a general search problem) while g is usually a specific function (e.g., the index function). This raises the following natural question: for which choices of g can we prove lifting theorems?

[1] Here, we limit ourselves mostly to theorems lifting precisely the query complexity of f to the communication complexity. Consequently, we do not discuss lifting-like theorems due to Sherstov [18] and independently due to Shi and Zhu [19], that enabled several important later developments. Moreover, it is not clear how to make this line of work for relations f that are not necessarily Boolean functions.

This question is interesting both because many applications depend on the choice of g, and because if we view lifting theorems as generalizations of direct-sum theorems, we would like them to work for as many choices of g as possible.

In particular, applications of lifting theorems often depend on the size of the gadget, which is the length of the input to g. Both the deterministic lifting theorem of Raz and McKenzie [17] and the randomized lifting theorem of Göös et al. [11] use the indexing function INDEX, which has very large size (polynomial in n). Reducing the gadget size to a constant would have many interesting applications.

In the deterministic setting, the gadget size was recently improved to logarithmic by the independent works of [3] and [20], who chose the gadget g to be the inner product function. Moreover, [3, 15] showed the lifting to work for a large class of gadgets. However, the randomized lifting theorem of Göös et al. [11], until our work, seemed to work only with INDEX as gadget.

In this work, we prove a randomized lifting theorem using an inner product gadget of logarithmic size. This has the immediate application that any lower bound on the outer function f can now be lifted to a much stronger lower bound on the composed function $f \circ g^n$, since hardness is measured as a function of the input length. This allows us, for example, to simplify the lower bounds of Göös, and Jayram [8] on AND-OR trees and MAJORITY trees, since we can now obtain them directly from the randomized query complexity lower bounds rather than going through conical juntas.

We now turn to state our main result more formally. Let $n \in \mathbb{N}$ be such that $n \geq 2$ and let $b \stackrel{\text{def}}{=} 10,000 \cdot \log n$. Let $\Lambda \stackrel{\text{def}}{=} \{0,1\}^b$, and let $g\colon \Lambda \times \Lambda \to \{0,1\}$ denote the inner product (mod 2) gadget. We prove lifting theorems for various lifted versions of $G \stackrel{\text{def}}{=} g^n$. That is, $G\colon \Lambda^n \times \Lambda^n \to \{0,1\}^n$ is the function that takes n independent instances of g and computes g on all of them. Here is our main result:

▶ **Theorem 1** (Randomized lifting). *Let $S\colon \{0,1\}^n \to \Sigma$ be any search problem and let Π be a bounded-error randomized communication protocol that solves $S \circ G$ with complexity c and error probability ε. Then, there exists a randomized decision tree T that solves S with complexity $O(\frac{c}{b})$ and bounded error probability.*

Using essentially the same proof method, we also prove a similar result in the deterministic setting:

▶ **Theorem 2** (Deterministic lifting). *Let S be any search problem that takes inputs from $\{0,1\}^n$, and let Π be a deterministic communication protocol that solves $S \circ G$ with complexity c. Then, there exists a deterministic decision tree T that solves S with complexity $O(\frac{c}{b})$.*

Most existing proofs of deterministic lifting theorems employ an information measure known as *thickness*, borrowed from earlier work on the KRW conjecture. The one deviation from this is the recent beautiful work of Garg et al. [7] who prove a deterministic lifting theorem in the dag-like setting. Curiously, their result does not use the thickness measure of information, but rather uses the blockwise min-entropy measure of information that was used by Göös, Pitassi and Watson [11] in order to prove a randomized lifting theorem. A natural direction of further research is to investigate if these disparate techniques can be unified. Indeed, a related question was asked in the first work to employ the measures of min-entropy for lifting by Göös et al. [9]: they asked if min-entropy and density based techniques could be used to prove (or simplify the existing proof of) Raz–McKenzie style deterministic lifting theorems.

Our unified proof answers this question by showing that the same information measure (blockwise min entropy) can in fact be used in both the deterministic and randomized settings.

The main difference between the two proofs is the way in which we decide the next bit of the communication protocol: in the deterministic setting, we make a greedy choice, and in the randomized setting, we make a (non-uniform) random choice. Whereas in the randomized setting, our information measure guarantees that we are able to estimate the distribution of the next bit of the protocol, in the deterministic setting it guarantees *richness*, that is, when the protocol ends, there is some input consistent with answers of all queries made by the decision tree.

Organization of the paper In Section 2 we set up the machinery that is used in both the deterministic and the randomized lifting theorems. We prove the deterministic lifting theorem in Section 3, and the randomized lifting theorem in Section 4. Both proofs use a Fourier-theoretic lemma, proved in Section 5.

2 Common Machinery

In this paper we consider lifting theorems for the most general case of search problems. A search problem $\mathcal{S}$ is defined by a relation $\mathcal{I} \times \mathcal{O}$ where $\mathcal{I}$ is a finite set of inputs and $\mathcal{O}$ is a finite set of outputs. The goal of the search problem, given an input $x \in \mathcal{I}$ is to find at least one output $o \in \mathcal{O}$ such that $(x, o) \in \mathcal{S}$. Like in the statement of the main theorem, let S be any search problem that takes inputs from $\{0,1\}^n$, and let Π be a bounded-error randomized communication protocol that solves $S \circ G$ with complexity c and error probability ε. We prove the randomized and deterministic lifting theorems, by building deterministic and randomized decision trees of cost $O(c/b)$ based on respective protocols of cost c. Intuitively, in both theorems, on input $z \in \{0,1\}^n$, the tree T will simulate the action of the protocol Π on inputs $(x, y) \in G^{-1}(z)$. More specifically, the tree will simulate the protocol bit by bit, and maintain a rectangle $\mathcal{X} \times \mathcal{Y}$ that is consistent with the protocol so far such that all the strings in $G(\mathcal{X} \times \mathcal{Y})$ are consistent with the queries made so far. To this end, we consider random variables X and Y that are distributed uniformly over $\mathcal{X}$ and $\mathcal{Y}$ respectively. We now state a few useful definitions and results about such random variables

The first such definition ensures that the random variables we consider have enough blockwise min-entropy.

▶ **Definition 3.** *Let X be a random variable taking values in Λ^n. We say that X is δ-*dense *if for every $I \subseteq [n]$ it holds that $H_\infty(X_I) \geq \delta \cdot b \cdot |I|$.*

We would like these random variables to be consistent with the query answers obtained by the decision tree thus far in the simulation. To this end, we also define the following notion of restrictions.

▶ **Definition 4.** *Given a restriction $\rho \in \{0,1,*\}^n$, we denote by $\mathrm{fix}(\rho)$ and $\mathrm{free}(\rho)$ the set of fixed and free coordinates of ρ respectively.*

Intuitively, $\mathrm{fix}(\rho)$ represents the query answers obtained thus far, and $\mathrm{free}(\rho)$ represents the yet unqueried coordinates. With these definitions, we define the property that we would like to maintain for X and Y during the simulation.

▶ **Definition 5** (following [11]). *Let X, Y be random variables taking values in Λ^n, and let $\rho \in \{0,1,*\}^n$ be a restriction. We say that X and Y are ρ-structured if $X_{\mathrm{free}(\rho)}$ and $Y_{\mathrm{free}(\rho)}$ are 0.9-dense, and*

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

In both lifting theorems, the decision tree T starts by setting X and Y to be uniform over Λ^n, and maintains throughout the simulation the invariant that, if ρ is the restriction that represents the current "state of knowledge" regarding the input z, then X and Y are ρ-structured. In order to maintain this invariant, we use the following Fourier-analytic result, which is proved in Section 5.

▶ **Definition 6.** *Let $\alpha \in \Lambda^n$ and let Y be a random variable taking values in Λ^n. We say that α is* η-bad *for Y if there exists a set $I \subset [n]$ and a string $\sigma \in \{0,1\}^I$ such that the random variable*

$$Y_{[n]-I} \left| g^I(\alpha_I, Y_I) = \sigma_I \right.$$

is not η-dense or

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

▶ **Theorem 7** (Main Technical Tool). *Let $n \in \mathbb{N}$ and let $b \in \mathbb{N}$ such that $b \geq 10000 \cdot \log(n)$. Let X and Y be random variables taking values in Λ^n that are δ_X-dense and δ_Y-dense respectively. Suppose that $\delta_X + \delta_Y \geq 1.3$ and $\delta_Y \geq 0.1$. Then, the probability that X takes a value that is $\frac{\delta_Y}{2.01}$-bad for Y is at most $2^{-0.01 \cdot b}$.*

We also use the following analogue of the "uniform marginals lemma" of [11] for the inner product gadget.

▶ **Lemma 8** (Uniform marginals lemma). *Let X, Y be random variables uniformly distributed over sets $\mathcal{X}, \mathcal{Y} \subseteq \Lambda^n$, and suppose they are ρ-structured. Then, for any $z \in \{0,1\}^n$ that is consistent with ρ, the uniform distribution over $G^{-1}(z) \cap (\mathcal{X} \times \mathcal{Y})$ has its marginal distributions $\frac{1}{n^3}$-close to X and Y respectively.*

In order to prove Lemma 8, we use the following definition and lemma from Göös et al. [9].

▶ **Definition 9.** *Let $\varepsilon > 0$ and let V be a random variable taking values from a set $\mathcal{V}$. We say that V is* ε-pointwise close to uniform *if for every $v \in \mathcal{V}$ it holds that $\Pr\left[V = v\right] \in (1 \pm \varepsilon) \cdot \frac{1}{|\mathcal{V}|}$.*

▶ **Lemma 10.** *Let A, B be 0.6-dense random variables taking values from Λ^m. Then $g^m(A, B)$ is $2^{-\frac{b}{20}}$-uniform.*

The proof of this lemma, which is similar to the proof of the uniform marginals lemma in [11], appears in Appendix A.

We use the following simple folklore fact about density.

▶ **Proposition 11.** *Let X be a random variable over Λ^J, and let $I \subseteq J$ be maximal subset of coordinates such that $H_\infty(X_I) < \delta \cdot b \cdot |I|$. Let $\alpha \in \Lambda^I$ be a value such that*

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

Then, the random variable $X_{J-I} | X_I = \alpha$ is δ-dense.

We also use the following decomposition result from Göös et al. [11], which extends the last proposition.

▶ **Lemma 12** (Density-restoring partition). *Let X be a random variable over $\mathcal{X} \subseteq \Lambda^J$. Then, there exists a partition*

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

such that every $\mathcal{X}^i$ is associated with a set $I_i \subseteq J$, a value $\alpha_i \in \Lambda^{I_i}$, and a probability $p_{\geq i} \stackrel{\text{def}}{=} \Pr\left[X \in \mathcal{X}^i \cup \ldots \cup \mathcal{X}^r\right]$ that satisfy the following properties: Denote by X^i the random variable X conditioned on $X \in \mathcal{X}^i$.

- *$X^i_{I_i}$ is fixed to α_i.*
- *$X^i_{J-I_i}$ is 0.9-dense.*
- *$H_\infty(X^i) \geq H_\infty(X) - 0.9 \cdot b \cdot |I_i| - \log \frac{1}{p_{\geq i}}$.*

3 The deterministic lifting theorem

In this section, we prove the deterministic lifting theorem, restated from the Introduction.

▶ **Theorem 13** (Restatement of Theorem 2)**.** *Let S be any search problem that takes inputs from $\{0,1\}^n$, and let Π be a deterministic communication protocol that solves $S \circ G$ with complexity c. Then, there exists a decision tree T that solves S with complexity $O(\frac{c}{b})$.*

As noted earlier, the decision tree T we construct would simulate the protocol Π. Throughout the simulation, the tree keeps track of random variables X, Y, which represent the inputs to the protocol, and maintains the invariant that they are ρ-structured. When the protocol Π ends, the decision tree T ends as well and outputs the output of Π. In order to complete the proof of Theorem 2, we need to show three things:

- How to simulate a single bit of the protocol while maintaining the above invariant.
- After the decision tree ends, its output is a correct output of S on z.
- The total number of queries made by the decision tree T during the lifting is $O(\frac{c}{b})$.

Due to space constraints, we will only briefly describe the simulation, relegating its analysis to Appendix B.

Consider a given step in the simulation where the tree is at a particular node of the protocol Π. Let $\mathcal{X}, \mathcal{Y}$ be the current set of inputs that are being maintained which are consistent with this node, and let X, Y be random variables uniformly distributed over $\mathcal{X}, \mathcal{Y}$. Let $\rho \in \{0, 1, *\}^n$ denote the restriction that represents the queries that have been made so far and their answers, i.e., coordinates that were queried are fixed to the answers that were received, and coordinates that were not queried are free. By the invariant we maintain, the variables X, Y are ρ-structured.

We would like to simulate the next bit of the protocol. Suppose without loss of generality that it is Alice's turn to speak. The tree T chooses the next bit to be the bit that has the highest probability of being sent by Alice, if the inputs are chosen according to X. The tree then updates the set $\mathcal{X}$ to be consistent with the new bit, and updates the random variable X accordingly. Now, if the ρ-structure property of X, Y has been violated, then it must be because $X_{\mathrm{free}(\rho)}$ is no longer 0.9-dense, since the new bit did not affect Y. The tree now modifies the sets $\mathcal{X}, \mathcal{Y}$ and the restriction ρ to restore the structuredness of X, Y. In order to do so, the tree T repeats the following steps iteratively until X and Y are ρ-structured:

1. Condition $X_{\mathrm{free}(\rho)}$ on not taking a value that is 0.4-bad for $Y_{\mathrm{free}(\rho)}$, and update $\mathcal{X}$ accordingly.
2. If $X_{\mathrm{free}(\rho)}$ is now 0.9-dense, then we are done — the structuredness has been restored. Otherwise continue.
3. Let $I \subseteq \mathrm{free}(\rho)$ be a maximal set that violates the density of $X_{\mathrm{free}(\rho)}$ (i.e., $H_\infty(X_I) < 0.9 \cdot b \cdot |I|$), and let $\alpha_I \in \Lambda^I$ be a "heavy" value that satisfies $\Pr[X_I = \alpha_I] > 2^{-0.9 \cdot b \cdot |I|}$.
4. Condition X on $X_I = \alpha_I$, and update $\mathcal{X}$ accordingly. Proposition 11 implies that $X_{\mathrm{free}(\rho)-I}$ is now 0.9-dense.
5. Query the coordinates in I, and update ρ accordingly.
6. Condition Y on $g^I(\alpha_I, Y_I) = \rho_I$, and update $\mathcal{Y}$ accordingly.
7. If $Y_{\mathrm{free}(\rho)}$ is now 0.9-dense then we are done — the structuredness has been restored. Otherwise go back to Step 1 but replace the roles of X and Y.

In order for the steps of the above process to always be well-defined, we need to show that we never condition on events with probability 0. If this is always satisfied, it follows that the algorithm terminates and at termination the random variables X, Y are ρ-structured. To see this, note that the process only stops if $X_{\mathrm{free}(\rho)}$ and $Y_{\mathrm{free}(\rho)}$ are 0.9-dense, and the process clearly maintains the invariant that

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

Moreover, the process always stops, since in every iteration the size of the set $\mathrm{free}(\rho)$ decreases, and it cannot decrease below 0.

We turn to show that we never condition on a zero probability event. To this end, we will show that the process preserves the following property: At the beginning of every iteration, one of the variables $X_{\mathrm{free}(\rho)}$ and $Y_{\mathrm{free}(\rho)}$ is 0.9-dense, and the other is at least 0.4-dense. Observe that this property indeed holds at the beginning of the first iteration: at this point, Y is 0.9-dense, and X must be at least 0.4-dense — since we chose the next bit of Alice to be the one with the highest probability, and therefore the min-entropy of any set of coordinates could have dropped by at most 1.

Suppose that the property holds at the beginning of a given iteration. The first conditioning takes place at Step 1. When Step 1 is performed, we know by Theorem 7 that the event that $X_{\mathrm{free}(\rho)}$ does not take values that are 0.4-bad for $Y_{\mathrm{free}(\rho)}$ has non-zero probability: to see it, note that by assumption $\delta_X \geq 0.4$ and $\delta_Y \geq 0.9$, so it holds that $\delta_X + \delta_Y \geq 1.3$ and $\frac{\delta_Y}{2.01} \geq 0.4$, so the requirements of the theorem are satisfied.

The next conditioning takes place at Step 4, but here the event has non-zero probability by definition. The last conditioning takes place at Step 6, and here the event has non-zero probability due to the assumption that $X_{\mathrm{free}(\rho)}$ does not take values that are bad for $Y_{\mathrm{free}(\rho)}$ — and in particular

$$\Pr\left[g^I(\alpha_I, Y_I) = \rho_I\right] \geq 2^{-|I|-1}.$$

Finally, we need to show that the above property is maintained for the next iteration. As stated in Step 4, at this point X is 0.9-dense. Moreover, since we know that $X_{\mathrm{free}(\rho)}$ does not take values that are 0.4-bad for $Y_{\mathrm{free}(\rho)}$, it follows in particular that

$$Y_{\mathrm{free}(\rho)} \left| g^I(\alpha_I, Y_I) = \rho_I \right.$$

is 0.4-dense. This concludes the proof. The rest of the analysis can be found in Appendix B.

4 The randomized lifting theorem

In this section, we prove the randomized lifting theorem, restated next.

▶ **Theorem 14** (Restatement of Theorem 1). *Let S be any search problem that takes inputs from $\{0,1\}^n$, and let Π be a randomized communication protocol that solves $S \circ G$ with complexity c and error probability ε. Then, there exists a decision tree T that solves S with complexity $O(\frac{c}{b})$ and error probability $\varepsilon + \frac{1}{10}$.*

As noted earlier, the decision tree T we construct simulates the protocol Π. The simulation is similar to the deterministic one, with two main differences:

- Instead of choosing the next bit of the protocol to be the most likely bit, we choose it randomly according to the distribution of the next bit (except that we abort the simulation on bits of very small probability).

- Instead of choosing I and α_I arbitrarily, we choose them from the density-restoring partition of Lemma 12, according to the distribution induced by this partition (except that we truncate parts of the partition that have very small probability).

In the following sections, we describe the simulation, analyze its error probability, and analyze its query complexity, respectively. For simplicity, we describe a simulation that has a better error probability of $\varepsilon + o(1)$ but query complexity that is efficient *only in expectation*. This simulation can be transformed into one with error probability $\varepsilon + \frac{1}{10}$, and efficient query complexity in the worst case, using standard arguments.

4.1 The simulation

As before, the decision tree T simulates the protocol Π while maintaining a rectangle $\mathcal{X} \times \mathcal{Y}$ that is contained in the rectangle of the current node of Π. When the simulation ends, T outputs the output of Π. Throughout the simulation, the decision tree T considers random variables X, Y that are uniformly distributed over $\mathcal{X} \times \mathcal{Y}$ and maintains the invariant that they are ρ-structured (for a restriction ρ that records the queries made so far). For the purpose of the simulation, we may assume without loss of generality that Π is deterministic (since T can use its randomness to choose the randomness of Π, and then pretend that Π is deterministic for the rest of the simulation).

We turn to explain how to simulate a single bit of the protocol. Suppose that at a given point it is Alice's turn to speak. The protocol partitions $\mathcal{X}$ into $\mathcal{X}_0 \cup \mathcal{X}_1$. The tree now chooses the next bit to be 0 with probability $\frac{|\mathcal{X}_0|}{|\mathcal{X}|}$ and to be 1 otherwise. If the bit that was chosen had probability less than $\frac{1}{n^2}$, the tree halts and declares error. Otherwise, the tree updates $\mathcal{X}$ to the corresponding set among $\mathcal{X}_0, \mathcal{X}_1$ and updates the random variable X accordingly.

Now, if the ρ-structure property of X, Y has been violated, then it must be because $X_{\text{free}(\rho)}$ is no longer 0.9-dense, since the new bit did not affect Y. The tree now modifies the sets $\mathcal{X}, \mathcal{Y}$ and the restriction ρ to restore the structuredness of X, Y. In order to do so, the tree T repeats the following steps iteratively until X, Y are ρ-structured:

1. Condition $X_{\text{free}(\rho)}$ on not taking a value that is 0.4-bad for $Y_{\text{free}(\rho)}$, and update $\mathcal{X}$ accordingly.
2. If X is now 0.9-dense, then we are done — the structuredness has been restored. Otherwise continue.
3. Let $\mathcal{X}_{\text{free}(\rho)} = \mathcal{X}^1 \cup \ldots \cup \mathcal{X}^r$ be the density-restoring partition of Lemma 12 with respect to $X_{\text{free}(\rho)}$. Choose a random class in the partition, where the class $\mathcal{X}^i$ is chosen with probability $\Pr\left[X_{\text{free}(\rho)} \in \mathcal{X}^i\right]$.
4. Recall that we defined the probability

$$p_{\geq i} \stackrel{\text{def}}{=} \Pr\left[X_{\text{free}(\rho)} \in \mathcal{X}^i \cup \ldots \cup \mathcal{X}^r\right].$$

   If $p_{\geq i} < \frac{1}{n^3}$, the tree T halts and declares error.
5. Let I_i and α_i be the set and the value associated with the class $\mathcal{X}^i$. The tree conditions X on the event $X_{\text{free}(\rho)} \in \mathcal{X}^i$ and updates $\mathcal{X}$ accordingly. The variable $X_{\text{free}(\rho)-I_i}$ is now 0.9-dense by the properties of the density-restoring partition.
6. Query the coordinates in I_i, and update ρ based on the query answers.
7. Condition Y on $g^I(\alpha_i, Y_{I_i}) = \rho_{I_i}$, and update $\mathcal{Y}$ accordingly.
8. If $Y_{\text{free}(\rho)}$ is now 0.9-dense then we are done — the structuredness has been restored. Otherwise go back to Step 1 but replace the roles of X and Y.

The proof that the process is well-defined and always halts, and that the ρ-structuredness invariant is maintained, is the same as in the deterministic simulation. The only difference here is that choosing the next bit of the protocol decreases the min-entropy of the blocks by at most $2 \log n$ bits rather than by at most 1 bit. Nevertheless, since the random variable X started as 0.9-dense and $b > 20 \log n$, the variable X is still 0.4-dense after choosing the next bit.

4.2 Correctness

We prove that the decision tree errs with probability at most $\varepsilon + o(1)$ (recall that ε is the error probability of the protocol Π). Fix an input $z \in \{0,1\}^n$. Let π be the (random) transcript generated by the simulation of T on z (if we the simulation declares error, we set $\pi = \perp$). Let π' denote the (random) transcript of Π on random inputs (X', Y') that are distributed uniformly over $G^{-1}(z)$ (again, we assume that Π' is deterministic and that the only randomness comes from the choice of (X', Y')). We will prove that the distributions of π and π' are $o(1)$-close. Since π' outputs the correct answer on z with probability at least $1-\varepsilon$, it will follow that π outputs the correct answer on z with probability at least $1 - \varepsilon - o(1)$.

To prove that π and π' are $o(1)$-close, we describe a coupling of π with π' that satisfies that $\pi = \pi'$ with probability at least $1 - o(1)$. To this end, we show that there exists a coupling of the random choices of the simulation with X', Y' such that, up to some bad event $\mathcal{E}$ of small probability, it holds that the pair (X', Y') is uniformly distributed in $G^{-1}(z) \cap (\mathcal{X} \times \mathcal{Y})$. Since $\mathcal{X} \times \mathcal{Y}$ determines the transcript π of the simulation (as $\mathcal{X} \times \mathcal{Y}$ is contained the rectangle of the current node in the protocol), whenever $(X', Y') \in (\mathcal{X}, \mathcal{Y})$ it holds that $\pi = \pi'$.

More specifically, we prove that there exists a coupling and an event $\mathcal{E}$ with probability at most $\frac{6 \cdot b}{n} = o(1)$ such that, when the simulation ends, conditioned on $\neg\mathcal{E}$ it holds that the pair (X', Y') is uniformly distributed in $G^{-1}(z) \cap (\mathcal{X} \times \mathcal{Y})$. To this end, we define a sequence of events $\mathcal{E}_1, \mathcal{E}_2, \ldots$ such that $\Pr[\mathcal{E}_t] \leq \frac{6}{n^2} \cdot (t-1)$ and at the begining of the t-th iteration, conditioned on $\neg\mathcal{E}_t$ it holds that the pair (X', Y') is uniformly distributed in $G^{-1}(z) \cap (\mathcal{X} \times \mathcal{Y})$. We then set $\mathcal{E}$ to be the event at the end of the last iteration. Since the number of iterations is at most $c \leq n \cdot b$ (as each iteration transmits 1-bit), it follows that the probability of $\mathcal{E}$ is at most $\frac{6}{n^2} \cdot c \leq \frac{6b}{n}$. In order to construct the coupling and the events $\mathcal{E}_1, \mathcal{E}_2, \ldots$, we prove the following auxiliary result.

▶ **Lemma 15.** *Suppose that we constructed the coupling until the beginning of the t-th iteration, and there is an event $\mathcal{E}_t$ such that conditioned on $\neg\mathcal{E}_t$ it holds that the pair (X', Y') is uniformly distributed in $G^{-1}(z) \cap (\mathcal{X} \times \mathcal{Y})$. Then, there exists a way to extend the coupling until the end of the t-th iteration, and there exists an event $\mathcal{E}_{t+1}$, such that $\Pr[\mathcal{E}_{t+1}] \leq \Pr[\mathcal{E}_t] + \frac{6}{n^2}$ and at the end of the t-th iteration, conditioned on $\neg\mathcal{E}_{t+1}$ it holds that the pair (X', Y') is uniformly distributed in $G^{-1}(z) \cap (\mathcal{X} \times \mathcal{Y})$.*

Given Lemma 15, we design the coupling and the events $\mathcal{E}_1, \mathcal{E}_2, \ldots$ by setting $\mathcal{E}_1$ to be the empty event and then applying Lemma 15 repeatedly until we reach the last iteration.

Proof. Suppose that the simulation ran until the beginning of the t-th iteration according to our coupling. If the event $\mathcal{E}_t$ happened, then the coupling behaves arbitrarily until the end of the simulation, and we assume that the simulation failed. Let us now condition on the event $\mathcal{E}_t$ not having happened, so we may assume that at the beginning of the t-th iteration, the pair (X', Y') is uniformly distributed in $G^{-1}(z) \cap (\mathcal{X} \times \mathcal{Y})$. We start by setting $\mathcal{E}_{t+1}$ to be the event $\mathcal{E}_t$, and we will add more events to it as the simulation progresses.

The simulation starts by choosing the next bit of the protocol, and suppose that it is Alice's turn to speak. The simulation has probability $\frac{|\mathcal{X}_0|}{|\mathcal{X}|}$ to choose 0, and by the uniform marginals lemma (Lemma 8), the random variable X' has probability $\frac{|\mathcal{X}_0|}{|\mathcal{X}|} \pm \frac{1}{n^3}$ to be in $\mathcal{X}_0$. In other words, the distribution of the class that the simulation chooses among $\mathcal{X}_0, \mathcal{X}_1$, and the distribution of the class that X' chooses, are $\frac{1}{n^3}$-close, and therefore there exists a coupling of those choices such that the same class is chosen in both with probability at least $1 - \frac{1}{n^3}$, so we use it to extend our coupling. We add to $\mathcal{E}_{t+1}$ the event in which the simulation and X' choose a different class among $\mathcal{X}_0, \mathcal{X}_1$, and for the rest of the proof we assume that it did not happen. We also add to $\mathcal{E}_{t+1}$ the event in which the simulation declared failure since it chose a bit with probability less than $\frac{1}{n^2}$ (clearly, this event has probability less than $\frac{1}{n^2}$), and for the rest of the proof we assume that it did not happen. We may thus assume that after this step, the pair (X', Y') is uniformly distributed in $G^{-1}(z) \cap (\mathcal{X} \times \mathcal{Y})$.

Next, the simulation removes from $\mathcal{X}$ the values that are 0.4-bad for Y. The probability that X takes such a value is at most $2^{-0.01 \cdot b} \le \frac{1}{n^3}$, and therefore the probability that X' takes such a value is at most $\frac{2}{n^3}$ by the uniform marginals lemma. We add the event that X' takes a bad value to $\mathcal{E}_{t+1}$ and assume for the rest of the proof that it did not happen. Hence, we may again assume that after this step, X' belongs to $\mathcal{X}$, and that the pair (X', Y') is uniformly distributed in $G^{-1}(z) \cap (\mathcal{X} \times \mathcal{Y})$.

In the following step, a class $\mathcal{X}^i$ is chosen according to the distribution induced by $X_{\mathrm{free}(\rho)}$. Let us now choose the class $\mathcal{X}^{i'}$ to which $X'_{\mathrm{free}(\rho)}$ belongs. By the uniform marginals lemma, the distributions of $\mathcal{X}^i$ and $\mathcal{X}^{i'}$ are $\frac{1}{n^3}$-close, and therefore there is a coupling of those classes such that they are equal with probability at least $1 - \frac{1}{n^3}$, so we use it to extend our coupling. We add to $\mathcal{E}_{t+1}$ the event in $\mathcal{X}^i \neq \mathcal{X}^{i'}$, and for the rest of the proof we assume that it did not happen. We also add to $\mathcal{E}_{t+1}$ the event in which the simulation declared error since $p_{\le i} < \frac{1}{n^3}$ (clearly, this event has probability less than $\frac{1}{n^3}$), and for the rest of the proof we assume that it did not happen. We therefore assume again that after this step, X' belongs to $\mathcal{X}$, and that the pair (X', Y') is uniformly distributed in $G^{-1}(z) \cap (\mathcal{X} \times \mathcal{Y})$.

Finally, the simulation conditions Y on $g^I(\alpha_i, Y_{I_i}) = \rho_{I_i}$. This conditioning trivially holds for Y' (since by assumption $(X', Y') \in G^{-1}(z)$ and by this point we chose $X'_{I_i} = \alpha_{I_i}$), and no further coupling needs to be done.

We conclude the proof by upper bounding the probability of the event $\mathcal{E}_{t+1}$. At the beginning, we set $\mathcal{E}_{t+1}$ to be $\mathcal{E}_t$, and therefore at this point its probability is $\Pr[\mathcal{E}_t]$. The step of choosing the next bit of the protocol contribute to $\mathcal{E}_{t+1}$ events whose total probability is at most $\frac{1}{n^3} + \frac{1}{n^2}$. Steps 1 to 7 above add to $\mathcal{E}_{t+1}$ events of total probability at most $\frac{4}{n^3}$. Those latter steps are now repeated until (X, Y) are ρ-structured. However, they may be repeated at most n times, since each time they are repeated, the tree makes at least one query, and it cannot make more than n queries. Hence, in all of those repetitions together, those steps in the simulation contribute to $\mathcal{E}_{t+1}$ events whose total probability is at most $\frac{4}{n^2}$. It follows that

$$\Pr[\mathcal{E}_{t+1}] \le \Pr[\mathcal{E}_t] + \frac{1}{n^3} + \frac{1}{n^2} + \frac{4}{n^2} \le \Pr[\mathcal{E}_t] + \frac{6}{n^2},$$

as required. ◀

4.3 The query complexity

We show that the *expected* query complexity of this simulation is $O(\frac{c}{b})$. Again, we define the deficiency of X, Y to be

$$\Delta \stackrel{\mathrm{def}}{=} 2 \cdot b \cdot |\mathrm{free}(\rho)| - H_\infty(X_{\mathrm{free}(\rho)}) - H_\infty(Y_{\mathrm{free}(\rho)}).$$

We will show that whenever the simulation sends one bit in the protocol, the deficiency is increased by $O(1)$ *in expectation*. On the other hand, we will show that whenever a query is made, the deficiency is always decreased by at least $\Omega(b)$. Thus, the expected deficiency at any point is at most

$$O(\#\text{bits communicated}) - \Omega(b \cdot \#\text{queries}).$$

Since the deficiency is always at least 0 and the number of bits communicated is at most c, it follows that the expected number of queries is upper bounded by $O(\frac{c}{b})$.

Whenever we choose the next bit for Alice, the deficiency increases by $\log \frac{|\mathcal{X}|}{|\mathcal{X}_0|}$ (if the next bit is 0) or by $\log \frac{|\mathcal{X}|}{|\mathcal{X}_1|}$ (if the next bit is 1). Thus, the expected increase in deficiency is

$$\frac{|\mathcal{X}_0|}{|\mathcal{X}|} \cdot \log \frac{|\mathcal{X}|}{|\mathcal{X}_0|} + \frac{|\mathcal{X}_1|}{|\mathcal{X}|} \cdot \log \frac{|\mathcal{X}|}{|\mathcal{X}_1|}.$$

This is the value of the binary entropy function on $\frac{|\mathcal{X}_0|}{|\mathcal{X}|}$, and hence it is upper bounded by 1. Conditioning on X not taking a value that is 0.4-bad for Y increases the deficiency by at most 1 bit since its probability is at least $\frac{1}{2}$. All in all, the expected increase in the deficiency is at most 2.

We turn to show that when a query is being made, the deficiency decreases by $\Omega(b)$. Suppose that the decision tree queried a set $I_i \subseteq \text{free}(\rho)$. This brings about the following changes to the deficiency:

- The variable X was conditioned on the event $X_{\text{free}(\rho)} \in \mathcal{X}^i$. By Lemma 12, this decreases the min-entropy of X by at most $0.9 \cdot b \cdot |I_i| + \log \frac{1}{p_{\geq i}}$. Now, Step 4 guarantees that $p_i \geq \frac{1}{n^3}$, and therefore $\log \frac{1}{p_i} \leq 3 \log n < 0.01 \cdot b$. All in all, this step increases the deficiency by at most $0.91 \cdot |I_i|$
- The variable Y is conditioned on the event $g^{I_i}(\alpha_{I_i}, Y_{I_i}) = \rho_{I_i}$, which has probability at least $2^{-|I_i|-1}$ by the assumption that X does not take bad values. This increases the deficiency by at most $|I_i| + 1$.
- The set I_i is removed from the set $\text{free}(\rho)$. By definition of deficiency, this dereases the term of $2 \cdot b \cdot |\text{free}(\rho)|$ by $2 \cdot b \cdot |I_i|$, decreases $H_\infty(Y_{\text{free}(\rho)})$ by at most $b \cdot |I_i|$, and does not change $H_\infty(X_{\text{free}(\rho)})$ (since at this point X_{I_i} is fixed to α_{I_i}). All in all, the deficiency is decreased by at least $b \cdot |I_i|$.
- Finally, the queries may make the process repeat for another iteration, so Step 1 may be performed again, increasing the deficiency by another 2 bits.

Summing all those effects together, we get that the deficiency was decreased by at least

$$b \cdot |I_i| - 0.91 \cdot b \cdot |I_i| - (|I_i| + 1) - 2 \geq 0.05 \cdot b \cdot |I_i|,$$

as required. This concludes the proof.

5 Fourier-theoretic result

We recall our notation, some definitions and the result. Let $n \in \mathbb{N}$ and let $b \in \mathbb{N}$ be such that $b \geq 10,000 \cdot \log n$. We denote the domain of the inner product gadget by $\Lambda = \{0,1\}^b$ (so the inner product is over $\Lambda \times \Lambda$), and denote $q = |\Lambda| = 2^b$. Given a string $\gamma \in \Lambda$, we denote the corresponding Fourier character by $\chi_\gamma(x) \stackrel{\text{def}}{=} (-1)^{\langle \gamma, x \rangle}$. When considering a set $I \subseteq [n]$ and the space of functions $f \colon \Lambda^I \to \mathbb{R}$, we index the corresponding Fourier characters by tuples from Λ^I, such that for every $\gamma \in \Lambda^I$ it holds that $\chi_\gamma = \prod_{i \in I} \chi_{\gamma_i}$.

▶ **Definition 16.** *Let $\alpha \in \Lambda^n$ and let Y be a random variable taking values in Λ^n. We say that α is η-*bad *for Y if there exists a set $I \subset [n]$ and a string $\sigma \in \{0,1\}^I$ such that the random variable*

$$Y_{[n]-I} \,|\forall_{i\in I} \langle \alpha_i, Y_i\rangle = \sigma_i$$

is not η-dense or

$$\Pr\left[\forall_{i\in I} \langle \alpha_i, Y_i\rangle = \sigma_i\right] < 2^{-|I|-1}.$$

In this section we prove the following result.

▶ **Theorem 17** (Restatement of Theorem 7)**.** *Let X and Y be random variables taking values in Λ^n that are δ_X-dense and δ_Y-dense respectively. Suppose that $\delta_X + \delta_Y \geq 1.3$ and $\delta_Y \geq 0.1$. Then, the probability that X takes a value that is $\frac{\delta_Y}{2.01}$-bad for Y is at most $q^{-0.01}$.*

For the rest of this section, fix the random variables X and Y, and suppose that they are δ_X-dense and δ_Y-dense respectively where $\delta_X + \delta_Y \geq 1.3$ and $\delta_Y \geq 0.1$. We use the following definition, which essentially isolates "badness" to a particular set of coordinates.

▶ **Definition 18.** *Let $\varepsilon > 0$. We say that $\alpha \in \Lambda^n$ is ε-*bad *for Y on $J \subseteq [n]$ if there exist a string $\beta_J \in \Lambda^J$, a non-empty set $I \subset [n] - J$ and a string $\sigma \in \{0,1\}^I$ such that*

$$\Pr\left[Y_J = \beta_J \text{ and } \forall_{i\in I} \langle \alpha_i, Y_i\rangle = \sigma_i\right] \notin 2^{-|I|} \cdot \left(\Pr\left[Y_J = \beta_J\right] \pm \varepsilon\right).$$

In particular, if $J = \emptyset$, we view Y_J, β_J as the empty string and the event $Y_J = \beta_J$ as an event that occurs with probability 1 vacuously.

Morally, a value is not bad if it is not bad on any J. Theorem 17 will follow as a corollary from the following result (see that last part of Appendix C).

▶ **Lemma 19.** *For every $J \subseteq [n]$, the probability that X takes a value that is ε-bad for Y on J is at most $q^{-\delta_Y \cdot |J| - 0.05}/\varepsilon^2$.*

In order to analyze the probability of bad values, it is more convenient to consider "unbiased" values, i.e., values α for which the event $Y_J = \beta_J$ is not correlated with inner products of the form $\forall_{i\in I} \langle \alpha_i, Y_i\rangle = \sigma_i$. This bias is naturally measured using Fourier coefficients. We denote by $D\colon \Lambda^n \to [0,1]$ the distribution of Y, i.e., the function that for every $\beta \in \Lambda^n$ outputs $\Pr\left[Y = \beta\right]$. For a set of indices $K \subseteq [n]$, we denote by D_K the function corresponding to the marginal distribution over K. Moreover, given disjoint sets $J, K \subseteq [n]$ and a string $\beta_J \in \Lambda^J$ we denote by $D_{K,\beta_J}\colon \Lambda^K \to [0,1]$ the function that maps each $\beta_K \in \Lambda^K$ to $\Pr\left[Y_K = \beta_K \text{ and } Y_J = \beta_J\right]$.

▶ **Definition 20.** *We say that a value $\alpha \in \Lambda^n$ is ε-*biased *for Y with respect to $J \subseteq [n]$ if for every non-empty $I \subseteq [n] - J$ and for every $\beta_J \in \Lambda^J$ it holds that $\left|\hat{D}_{I,\beta_J}(\alpha_I)\right| \leq \varepsilon \cdot q^{-1.1\cdot|I|}$.*

Lemma 19 follows immediately from the next two propositions. The first proposition is a "Vazirani lemma" type of result that shows that small bias implies small distortion of probabilities.

▶ **Proposition 21.** *If a value $\alpha \in \Lambda^n$ is ε-biased for Y with respect to $J \subseteq [n]$, then it is not ε-bad with respect to J.*

The second proposition upper bounds the probability of X taking a value with large bias using the fact that X and Y are δ_X-dense and δ_Y-dense respectively.

▶ **Proposition 22.** *For every $J \subseteq [n]$, the probability that X takes a value that is not ε-biased for Y with respect to J is at most $q^{-\delta_Y \cdot |J| - 0.05}/\varepsilon^2$.*

The rest of the proof can be found in Appendix C.

References

1 Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. In *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, Massachusetts, USA, 5-8 June 2010*, pages 67–76, 2010.

2 Mark Braverman, Anup Rao, Omri Weinstein, and Amir Yehudayoff. Direct products in communication complexity. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, 26-29 October, 2013, Berkeley, CA, USA*, pages 746–755, 2013.

3 Arkadev Chattopadhyay, Michal Koucký, Bruno Loff, and Sagnik Mukhopadhyay. Simulation theorems via pseudorandom properties. *CoRR*, abs/1704.06807, 2017. URL: `http://arxiv.org/abs/1704.06807`.

4 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.

5 Susanna F. de Rezende, Jakob Nordström, and Marc Vinyals. How limited interaction hinders real communication (and what it means for proof and circuit complexity). In *IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS 2016, 9-11 October 2016, Hyatt Regency, New Brunswick, New Jersey, USA*, pages 295–304, 2016.

6 Tomás Feder, Eyal Kushilevitz, Moni Naor, and Noam Nisan. Amortized communication complexity. *SIAM J. Comput.*, 24(4):736–750, 1995. URL: `https://doi.org/10.1137/S0097539792235864`, `doi:10.1137/S0097539792235864`.

7 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.

8 Mika Göös and T. S. Jayram. A composition theorem for conical juntas. In *31st Conference on Computational Complexity, CCC 2016, May 29 to June 1, 2016, Tokyo, Japan*, pages 5:1–5:16, 2016. URL: `https://doi.org/10.4230/LIPIcs.CCC.2016.5`, `doi:10.4230/LIPIcs.CCC.2016.5`.

9 Mika Göös, Shachar Lovett, Raghu Meka, Thomas Watson, and David Zuckerman. Rectangles are nonnegative juntas. *SIAM J. Comput.*, 45(5):1835–1869, 2016.

10 Mika Göös, Toniann Pitassi, and Thomas Watson. Deterministic communication vs. partition number. In *Proceedings of IEEE 56th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1077–1088, 2015.

11 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.

12 Hamed Hatami, Kaave Hosseini, and Shachar Lovett. Structure of protocols for XOR functions. *SIAM J. Comput.*, 47(1):208–217, 2018.

13 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, 1991.

14 Hartmut Klauck. A strong direct product theorem for disjointness. In *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, Massachusetts, USA, 5-8 June 2010*, pages 77–86, 2010.

15 Alexander Kozachinskiy. From expanders to hitting distributions and simulation theorems. In *43rd International Symposium on Mathematical Foundations of Computer Science, MFCS 2018, August 27-31, 2018, Liverpool, UK*, pages 4:1–4:15, 2018.

16 Bruno Loff and Sagnik Mukhopadhyay. Lifting theorems for equality. *Electronic Colloquium on Computational Complexity (ECCC)*, 25:175, 2018.

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

18 Alexander A. Sherstov. The pattern matrix method. *SIAM J. Comput.*, 40(6):1969–2000, 2011.

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

20 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. URL: https://eccc.weizmann.ac.il/report/2017/010.

21 Andrew C. Yao. Theory and application of trapdoor functions. In *Proceedings of IEEE 23rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 80–91, 1982.

A Missing proofs from Section 2

Proof of Lemma 8. Let (X', Y') be uniformly distributed over $G^{-1}(z) \cap (\mathcal{X} \times \mathcal{Y})$. We prove that X is $\frac{1}{n^3}$-close to X', and a similar argument works for Y. Let $\mathcal{E} \subseteq \mathcal{X}$ be any test event, and without loss of generality assume that $\Pr[X \in \mathcal{E}] \geq \frac{1}{2}$ (otherwise replace E with its complement). Let us denote by $X^{\mathcal{E}}$ the random variable that is uniformly distributed over $\mathcal{E}$, i.e., it distributed like $X|\mathcal{E}$. Since X, Y are ρ-structured, it holds that $X_{\text{free}(\rho)}$, $X^{\mathcal{E}}_{\text{free}(\rho)}$, and $Y_{\text{free}(\rho)}$ are 0.6-dense and therefore by Lemma 10 and our choice of b it holds that $g^{\text{free}(\rho)}(X_{\text{free}(\rho)}, Y_{\text{free}(\rho)})$ and $g^{\text{free}(\rho)}(X^{\mathcal{E}}_{\text{free}(\rho)}, Y_{\text{free}(\rho)})$ are $\frac{1}{n^4}$-pointwise close to uniform. It follows that

$$\begin{aligned}
\Pr[X' \in \mathcal{E}] &= \frac{\left|G^{-1}(z) \cap (\mathcal{E} \times \mathcal{Y})\right|}{\left|G^{-1}(z) \cap (\mathcal{X} \times \mathcal{Y})\right|} \\
&= \frac{\left|G^{-1}(z) \cap (\mathcal{E} \times \mathcal{Y})\right| / |\mathcal{X} \times \mathcal{Y}|}{\left|G^{-1}(z) \cap (\mathcal{X} \times \mathcal{Y})\right| / |\mathcal{X} \times \mathcal{Y}|} \\
&= \frac{\Pr[G(X,Y) = z \text{ and } X \in \mathcal{E}]}{\Pr[G(X,Y) = z]} \\
&= \frac{\Pr[G(X,Y) = z | X \in \mathcal{E}]}{\Pr[G(X,Y) = z]} \cdot \Pr[X \in \mathcal{E}] \\
&= \frac{\Pr\left[G(X^{\mathcal{E}}, Y) = z\right]}{\Pr[G(X,Y) = z]} \cdot \Pr[X \in \mathcal{E}] \\
&= \frac{\Pr\left[g^{\text{free}(\rho)}(X^{\mathcal{E}}_{\text{free}(\rho)}, Y_{\text{free}(\rho)}) = z_{\text{free}(\rho)}\right]}{\Pr\left[g^{\text{free}(\rho)}(X_{\text{free}(\rho)}, Y_{\text{free}(\rho)}) = z_{\text{free}(\rho)}\right]} \cdot \Pr[X \in \mathcal{E}] \\
&\in \frac{(1 \pm \frac{1}{n^4})}{(1 \pm \frac{1}{n^4})} \cdot \Pr[X \in \mathcal{E}] \\
&\in \left(1 \pm \frac{1}{n^3}\right) \cdot \Pr[X \in \mathcal{E}]
\end{aligned}$$

as required. ◀

B Missing proofs from Section 3

B.1 Concluding the simulation

In this section, we prove that when the simulation ends, the protocol Π outputs an answer in $S(z)$. To this end, all we need to prove is that when the simulation ends, we can find $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ such that $G(x, y) = z$: To see why, observe that the output of the protocol at this point must be its output on (x, y), since the rectangle $\mathcal{X} \times \mathcal{Y}$ is contained in the rectangle of the leaf to which the protocol arrived. Now, since we assumed that Π computes $S \circ G$, it follows that its output must be $(S \circ G)(x, y) = S(z)$.

We thus turn to show that there exist $x, y \in \mathcal{X} \times \mathcal{Y}$ such that $G(x, y) = z$. Recall that when the protocol ends, it holds that X, Y are ρ-structured (by the invariant that we maintained). This means that $g^{\text{fix}(\rho)}(X_{\text{fix}(\rho)}, Y_{\text{fix}(\rho)}) = z_{\text{fix}(\rho)}$, and that $X_{\text{free}(\rho)}, Y_{\text{free}(\rho)}$ are 0.9-dense. By Theorem 7, it follows that $X_{\text{free}(\rho)}$ takes a value that is not 0.4-bad for $Y_{\text{free}(\rho)}$ with non-zero probability. This means that there exists some $x \in \mathcal{X}$ such that $x_{\text{free}(\rho)}$ is not 0.4-bad for $Y_{\text{free}(\rho)}$. By the definition of badness, it follows that

$$\Pr\left[g^{\text{free}(\rho)}(x_{\text{free}(\rho)}, Y_{\text{free}(\rho)}) = z_{\text{free}(\rho)}\right] \geq 2^{-|\text{free}(\rho)|-1} > 0$$

and therefore there exists some $y \in \mathcal{Y}$ such that $g^{\mathrm{free}(\rho)}(x_{\mathrm{free}(\rho)}, y_{\mathrm{free}(\rho)}) = z_{\mathrm{free}(\rho)}$. It follows that x and y satisfy

$$g^{\mathrm{fix}(\rho)}(x_{\mathrm{fix}(\rho)}, y_{\mathrm{fix}(\rho)}) = z_{\mathrm{fix}(\rho)}$$
$$g^{\mathrm{free}(\rho)}(x_{\mathrm{free}(\rho)}, y_{\mathrm{free}(\rho)}) = z_{\mathrm{free}(\rho)}$$

and therefore $G(x, y) = z$, as required.

B.2 The query complexity

We conclude by showing that the total number of queries the tree T makes is $O(\frac{c}{b})$. To this end, we define the deficiency of X, Y to be

$$\Delta \stackrel{\mathrm{def}}{=} 2 \cdot b \cdot |\mathrm{free}(\rho)| - H_\infty(X_{\mathrm{free}(\rho)}) - H_\infty(Y_{\mathrm{free}(\rho)}).$$

We prove that whenever the protocol Π transmits a bit in the simulation, the deficiency increases by $O(1)$, and that whenever the tree T makes a query, the deficiency is decreased by $\Omega(b)$. Since the deficiency is always non-negative, and the protocol transmits at most c bits, it follows that the tree must make at most $O(\frac{c}{b})$ bits.

We start by showing that when the protocol Π transmits a bit in the simulation, the deficiency increases by $O(1)$. When a bit is transmitted, either X or Y is conditioned on an event of probability at least $\frac{1}{2}$, depending on which player spoke, and the other variable remains unchanged. This means that the sum $H_\infty(X_{\mathrm{free}(\rho)}) + H_\infty(Y_{\mathrm{free}(\rho)})$ decreases by at most 1, and therefore the deficiency increases by at most 1. Next, the simulation might perform Step 1 in the process above, i.e., condition X or Y on taking a value that is not bad. This event has probability $1 - 2^{-0.01 \cdot b} \geq \frac{1}{2}$, so conditioning on it increases the deficiency by at most 1. All in all, we increased the deficiency by at most 2. All the other steps that might be taken are only taken if a query is being made, so we account their deficiency increases to the following "query part" of the analysis.

We turn to show that when a query is being made, the deficiency decreases by $\Omega(b)$. Suppose that the decision tree queried a set $I \subseteq \mathrm{free}(\rho)$. This applies the following changes to the deficiency:

- The variable X is conditioned on the event $X_I = \alpha_I$, which has probability greater than $2^{-0.9 \cdot b \cdot |I|}$ by the definition of α_I. Hence, this conditioning increases the deficiency by at most $0.9 \cdot b \cdot |I|$.
- The variable Y is conditioned on the event $g^I(\alpha_I, Y_I) = \rho_I$, which has probability at least $2^{-|I|-1}$ by the assumption that X does not take bad values. This increases the deficiency by at most $|I| + 1$.
- The set I is removed from the set $\mathrm{free}(\rho)$. Looking at the definition of deficiency, this decreases the first term, $2 \cdot b \cdot |\mathrm{free}(\rho)|$, by at most $2 \cdot b \cdot |I|$, decreases $H_\infty(Y_{\mathrm{free}(\rho)})$ by at most $b \cdot |I|$, and does not change $H_\infty(X_{\mathrm{free}(\rho)})$ (since at this point X_I is fixed to α_I). All in all, the deficiency is decreased by $b \cdot |I|$.
- Finally, the queries may make the process repeat for another iteration, so Step 1 may be performed again, increasing the deficiency by another 2 bits.

Summing all those effects together, we get that the deficiency was decreased by at least

$$b \cdot |I| - 0.9 \cdot b \cdot |I| - (|I| + 1) - 2 \geq 0.05 \cdot b \cdot |I|$$

in each iteration, as required. This concludes the proof.

C Missing proofs from Section 5

Proof of Proposition 21

Let $\alpha \in \Lambda^n$ be an ε-biased value for Y with respect to a set $J \subseteq [n]$, let $\beta_J \in \Lambda^J$ be a string, $I \subseteq [n] - J$ be a non-empty set, and $\sigma \in \{0,1\}^I$ be a string. Let E denote the event that $\langle \alpha_i, Y_i \rangle = \sigma_i$ for all $i \in I$, and for every $K \subseteq I$, let $\sigma_K = \sum_{i \in K} \sigma_i$. It holds that

$$
\begin{aligned}
\Pr[Y_J = \beta_J \text{ and } E_\alpha] &= \sum_{\beta_I \in \Lambda^I} \Pr[Y_J = \beta_J \text{ and } Y_I = \beta_I] \cdot \mathbf{1}_{\forall_{i \in I} \langle \alpha_i, \beta_i \rangle = \sigma_i} \\
&= \sum_{\beta_I \in \Lambda^I} \Pr[Y_J = \beta_J \text{ and } Y_I = \beta_I] \cdot \prod_{i \in I} \left(\frac{1 + (-1)^{\sigma_i} \cdot \chi_{\alpha_i}(\beta_i)}{2} \right) \\
\text{(Expanding the product)} &= \sum_{\beta_I \in \Lambda^I} D_{I,\beta_J}(\beta_I) \cdot 2^{-|I|} \cdot \sum_{K \subseteq I} (-1)^{\sigma_K} \cdot \chi_{\alpha_K}(\beta_K) \\
&= 2^{-|I|} \cdot \sum_{K \subseteq I} (-1)^{\sigma_K} \cdot \sum_{\beta_I \in \Lambda^I} D_{I,\beta_J}(\beta_I) \cdot \chi_{\alpha_K}(\beta_K) \\
&= 2^{-|I|} \cdot \sum_{K \subseteq I} (-1)^{\sigma_K} \cdot \sum_{\beta_K \in \Lambda^K} D_{K,\beta_J}(\beta_K) \cdot \chi_{\alpha_K}(\beta_K) \\
&= 2^{-|I|} \cdot D_{\emptyset,\beta_J} + 2^{-|I|} \cdot \sum_{\emptyset \neq K \subseteq I} (-1)^{\sigma_K} \cdot \sum_{\beta_K \in \Lambda^K} D_{K,\beta_J}(\beta_K) \cdot \chi_{\alpha_K}(\beta_K)
\end{aligned}
$$

Next, observe that $D_{\emptyset,\beta_j} = \Pr[Y_J = \beta_J]$ by definition, and therefore

$$
\Pr[Y_J = \beta_J \text{ and } E_\alpha] = 2^{-|I|} \cdot \left(\Pr[Y_J = \beta_J] + \sum_{\emptyset \neq K \subseteq I} (-1)^{\sigma_K} \cdot \sum_{\beta_K \in \Lambda^K} D_{K,\beta_J}(\beta_K) \cdot \chi_{\alpha_K}(\beta_K) \right).
$$

Now,

$$
\begin{aligned}
&\left| \sum_{\emptyset \neq K \subseteq I} (-1)^{\sigma_K} \cdot \sum_{\beta_K \in \Lambda^K} D_{K,\beta_J}(\beta_K) \cdot \chi_{\alpha_K}(\beta_K) \right| \\
\text{(Formula for Fourier coefficients)} &\leq \left| \sum_{\emptyset \neq K \subseteq I} (-1)^{\sigma_K} \cdot q^{|K|} \cdot \hat{D}_{K,\beta_J}(\alpha_K) \right| \\
\text{(Triangle inequality)} &\leq \sum_{\emptyset \neq K \subseteq I} q^{|K|} \cdot \left| \hat{D}_{K,\beta_J}(\alpha_K) \right| \\
(\alpha \text{ is } \varepsilon\text{-biased}) &\leq \sum_{\emptyset \neq K \subseteq I} q^{|K|} \cdot \varepsilon \cdot q^{-1.1 \cdot |K|} \\
&= \varepsilon \cdot \sum_{k=1}^{|I|} \binom{|I|}{k} \cdot q^{-0.1 \cdot k} \\
&\leq \varepsilon \cdot \sum_{k=1}^{n} n^k \cdot q^{-0.1 \cdot k} \\
(q \stackrel{\text{def}}{=} n^{10000}) &= \varepsilon \cdot \sum_{k=1}^{n} n^k \cdot n^{-1000 \cdot k} \\
&\leq \varepsilon
\end{aligned}
$$

The required result follows.

Proof of Proposition 22

Fix $J \subseteq [n]$. We first upper bound the probability that X takes a value α that violates the ε-biased property for a specific subset $I \subseteq [n] - J$, and then take a union bound over all subsets I. Let $I \subseteq [n] - J$ be a non-empty set. For every value α that is not ε-biased due to I, there exists a value $\beta_J \in \Lambda^J$ such that $\left|\hat{D}_{I,\beta_J}(\alpha_I)\right| > \varepsilon \cdot q^{-1.1\cdot|I|}$. We upper bound the number of large coefficients of the form $\left|\hat{D}_{I,\beta_J}(\alpha_I)\right|$ by showing that the sum of their squares is not too large, which follows from the high min-entropy of $H_\infty(Y_{I\cup J})$. For simplicity of notation, denote $K = I \cup J$. It holds that

$$\begin{aligned}
\sum_{\alpha_I\in\Lambda^I}\sum_{\beta_J\in\Lambda^J}\hat{D}_{I,\beta_J}(\alpha_I)^2 &= \sum_{\beta_J\in\Lambda^J}\sum_{\alpha_I\in\Lambda^I}\hat{D}_{I,\beta_J}(\alpha_I)^2\\
\text{(Parseval's inequality)} &= \sum_{\beta_J\in\Lambda^J} q^{-|I|}\cdot\sum_{\beta_I\in\Lambda^I} D_{I,\beta_J}(\beta_I)^2\\
&= q^{-|I|}\cdot\sum_{\beta_J\in\Lambda^J}\sum_{\beta_I\in\Lambda^I} D_{I\cup J}(\beta_I,\beta_J)^2\\
&= q^{-|I|}\cdot\sum_{\beta_K\in\Lambda^K} D_K(\beta_K)^2\\
&= q^{-|I|}\cdot\sum_{\beta_K\in\Lambda^K}\Pr\left[Y_K=\beta_K\right]^2\\
&\le q^{-|I|}\cdot\max\left\{\Pr\left[Y_K=\beta_K\right]\right\}\cdot\sum_{\beta_K\in\Lambda^K}\Pr\left[Y_K=\beta_K\right]\\
&= q^{-|I|}\cdot\max\left\{\Pr\left[Y_K=\beta_K\right]\right\}\\
&\le q^{-|I|}\cdot q^{-\delta_Y\cdot|K|}\\
&= q^{-(1+\delta_Y)\cdot|I|-\delta_Y\cdot|J|}.
\end{aligned}$$

We wish to upper bound the number of strings $\alpha_I \in \Lambda^I$ for which there is some β_J such that $\left|\hat{D}_{I,\beta_J}(\alpha_I)\right| > \varepsilon \cdot q^{-1.1\cdot|I|}$. For every such string α_I, it holds in particular that

$$\sum_{\beta_J\in\Lambda^J}\hat{D}_{I,\beta_J}(\alpha_I)^2 > \varepsilon^2\cdot q^{-2.2\cdot|I|}.$$

Therefore, the number such strings α_I is at most

$$\frac{q^{-(1+\delta_Y)\cdot|I|-\delta_Y\cdot|J|}}{\varepsilon^2\cdot q^{-2.2\cdot|I|}} \le \frac{q^{(1.2-\delta_Y)\cdot|I|-\delta_Y\cdot|J|}}{\varepsilon^2}.$$

Since X is δ_X-dense, the probability that $X_I = \alpha_I$ for any α_I is at most $q^{-\delta_X\cdot|I|}$ and therefore the total probability of the bad α_I's is at most

$$\begin{aligned}
\frac{q^{(1.2-\delta_Y)\cdot|I|-\delta_Y\cdot|J|}}{\varepsilon^2}\cdot q^{-\delta_X\cdot|I|} &= \frac{q^{(1.2-\delta_X-\delta_Y)\cdot|I|-\delta_Y\cdot|J|}}{\varepsilon^2}\\
\text{(By the assumption that } \delta_X+\delta_Y\ge 1.3) &\le \frac{q^{-0.1\cdot|I|-\delta_Y\cdot|J|}}{\varepsilon^2}.
\end{aligned}$$

Finally, by taking union bound over all bad I's, we get that the probability that X takes a bad value is at most

$$\begin{aligned}
\sum_{\emptyset\neq I\subseteq[n]} \frac{q^{-0.1\cdot|I|-\delta_Y\cdot|J|}}{\varepsilon^2} &= \frac{q^{-\delta_Y\cdot|J|}}{\varepsilon^2}\cdot\sum_{\emptyset\neq I\subseteq[n]} q^{-0.1\cdot|I|}\\
&= \frac{q^{-\delta_Y\cdot|J|}}{\varepsilon^2}\cdot\sum_{i=1}^{n}\binom{n}{i}q^{-0.1\cdot i}\\
&\leq \frac{q^{-\delta_Y\cdot|J|}}{\varepsilon^2}\cdot\sum_{i=1}^{n} n^i\cdot q^{-0.1\cdot i}\\
(q \overset{\text{def}}{=} n^{10000}) &\leq \frac{q^{-\delta_Y\cdot|J|}}{\varepsilon^2}\cdot\sum_{i=1}^{q^{0.01}} q^{0.01\cdot i}\cdot q^{-0.1\cdot i}\\
&\leq \frac{q^{-\delta_Y\cdot|J|}}{\varepsilon^2}\cdot q^{0.01}\cdot q^{0.01}\cdot q^{-0.1}\\
&\leq \frac{q^{-\delta_Y\cdot|J|-0.05}}{\varepsilon^2}
\end{aligned}$$

Proof of Theorem 17 from Lemma 19

We consider two "bad events" that might happen, and upper bound the probability of both events using Lemma 19:

- X takes a value that $\frac{1}{2}$-bad for the empty set (i.e., $J=\emptyset$). By Lemma 19, the probability of this event is at most $4\cdot q^{-0.05}$.
- For any non-empty set $J\subseteq[n]$, the variable X takes a value that is ε-bad for J with $\varepsilon = q^{-\frac{\delta_Y}{2.02}\cdot|J|}$. By applying Lemma 19 and the union bound, the probability of this event is at most

$$\begin{aligned}
\sum_{\emptyset\neq J\subseteq[n]} \frac{q^{-\delta_Y\cdot|J|-0.05}}{q^{-\frac{2}{2.02}\cdot\delta_Y\cdot|J|}} &= q^{-0.05}\cdot\sum_{\emptyset\neq J\subseteq[n]} q^{\delta_Y\cdot|J|\cdot\left(\frac{1}{1.01}-1\right)}\\
&= q^{-0.05}\cdot\sum_{\emptyset\neq J\subseteq[n]} q^{-0.001\cdot\delta_Y\cdot|J|}\\
&= q^{-0.05}\cdot\sum_{j=1}^{n}\binom{n}{j}\cdot q^{-0.001\cdot\delta_Y\cdot j}\\
&\leq q^{-0.05}\cdot\sum_{j=1}^{n} n^j\cdot q^{-0.001\cdot\delta_Y\cdot j}\\
(q \overset{\text{def}}{=} n^{40000}, \delta_Y\geq 0.1) &\leq q^{-0.05}\cdot\sum_{j=1}^{n} n^j\cdot n^{-4\cdot j}\\
&\leq q^{-0.05}\cdot\sum_{j=1}^{n} n^{-3\cdot j}\\
&\leq q^{-0.05}
\end{aligned}$$

Hence, with probability at least $1-5\cdot q^{-0.05}\geq 1-q^{-0.01}$, none of these bad events happen. We now prove that whenever these events do not happen, the variable X takes a value that is not $\frac{\delta_Y}{2.01}$-bad for Y.

Let $\alpha \in \Lambda^n$ be a value for X that does not give rise to the foregoing bad events. Let $I \subset [n]$ and $\sigma \in \{0,1\}^I$, and let E_α denote the event $\forall_{i \in I} \langle \alpha_i, Y_i \rangle = \sigma_i$. We want to show that for every $I \subset [n]$ and for every $\sigma \in \{0,1\}^I$, the random variable

$$Y_{[n]-I} | \forall_{i \in I} \langle Y_i, \alpha_i \rangle = \sigma_i$$

is $\frac{\delta_Y}{2.01}$-dense, and that $\Pr[E_\alpha] \geq 2^{-|I|-1}$. We start with the latter condition. Since we know that α is not $\frac{1}{2}$-bad for the empty set, it holds that

$$\Pr[E_\alpha] \geq 2^{-|I|} \cdot \left(1 - \frac{1}{2}\right) = 2^{-|I|-1},$$

as required. Next, let $J \subseteq [n] - I$ and $\beta_J \in \Lambda^J$. We prove that

$$\Pr[Y_J = \beta_J | E_\alpha] \leq 2^{-\frac{\delta_Y}{2.01} \cdot b \cdot |J|} = q^{-\frac{\delta_Y}{2.01} \cdot |J|}.$$

Since we know that α is not ε-bad for J with $\varepsilon = q^{-\frac{\delta_Y}{2.02} \cdot |J|}$, it holds that

$$\begin{aligned}
\Pr[Y_J = \beta_J \text{ and } E_\alpha] &\leq 2^{-|I|} \cdot \left(\Pr[Y_J = \beta_J] + q^{-\frac{\delta_Y}{2.02} \cdot |J|}\right) \\
&\leq 2^{-|I|} \cdot \left(q^{-\delta_Y} + q^{-\frac{\delta_Y}{2.02} \cdot |J|}\right) \\
&\leq 2^{-|I|+1} \cdot q^{-\frac{\delta_Y}{2.02} \cdot |J|}.
\end{aligned}$$

It follows that

$$\begin{aligned}
\Pr[Y_J = \beta_J | E_\alpha] &= \frac{\Pr[Y_J = \beta_J \text{ and } E_\alpha]}{\Pr[E_\alpha]} \\
&\leq \frac{2^{-|I|+1} \cdot q^{-\frac{\delta_Y}{2.02} \cdot |J|}}{2^{-|I|-1}} \\
&= 4 \cdot q^{-\frac{\delta_Y}{2.02} \cdot |J|} \\
&\leq q^{-\frac{\delta_Y}{2.01} \cdot |J|},
\end{aligned}$$

as required.