SIAM J. COMPUT.
Vol. 50, No. 1, pp. 171–210

QUERY-TO-COMMUNICATION LIFTING USING LOW-DISCREPANCY GADGETS*

ARKADEV CHATTOPADHYAY†, YUVAL FILMUS‡, SAJIN KOROTH§, OR MEIR¶, AND TONIANN PITASSI‖

Abstract. Lifting theorems are theorems that relate the query complexity of a function $f : \{0,1\}^n \to \{0,1\}$ to the communication complexity of the composed function $f \circ g^n$ for some "gadget" $g : \{0,1\}^b \times \{0,1\}^b \to \{0,1\}$. Such theorems allow transferring lower bounds from query complexity to the communication complexity, and have seen numerous applications in recent years. In addition, such theorems can be viewed as a strong generalization of a direct-sum theorem for the gadget g. We prove a new lifting theorem that works for all gadgets g that have logarithmic length and exponentially-small discrepancy, for both deterministic and randomized communication complexity. Thus, we significantly increase the range of gadgets for which such lifting theorems hold. Our result has two main motivations: first, allowing a larger variety of gadgets may support more applications. In particular, our work is the first to prove a randomized lifting theorem for logarithmic-size gadgets, thus improving some applications of the theorem. Second, our result can be seen as a strong generalization of a direct-sum theorem for functions with low discrepancy.

Key words. communication complexity, query complexity, lifting

AMS subject classification. 68Q17

DOI. 10.1137/19M1310153

1. Introduction.

1.1. Background.

In this work, we prove new lifting theorems for a large family of gadgets. Let $f\colon \{0,1\}^n \to \{0,1\}$ 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 inputs $(x_1, y_1), \ldots, (x_n, y_n)$ 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 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

*Received by the editors January 2, 2020; accepted for publication (in revised form) November 10, 2020; published electronically February 18, 2021. This work subsumes an earlier work that appeared in ICALP, LIPIcs 132, Schloss Dagstuhl, Wadern, Germany, 35, 2019 [7].

https://doi.org/10.1137/19M1310153

Funding: The second author was funded by ISF grant 1337/16. The third author was partially supported by the institutional postdoctoral program of the University of Haifa. The third and fourth authors were partially supported by the Israel Science Foundation (grant 1445/16). The fifth author was supported by NSERC and by NSF CCF grant 1900460.

†School of Technology and Computer Science, Tata Institute of Fundamental Research, Mumbai 400005, India (Arkadev.c@tifr.res.in).

‡Computer Science Department, Technion - Israel Institute of Technology, Haifa, 32000, India (filmus.yuval@gmail.com).

§School of Computing Science, Simon Fraser University, Burnaby, B.C., V5A 1S6, Canada (skoroth@sfu.ca).

¶Department of Computer Science, University of Haifa, Haifa, 3498838, India (ormeir2@gmail.com).

‖Department of Computer Science, University of Toronto, Toronto, M5S 3G4, Canada (toni@cs.toronto.edu).

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 ith bit, they compute g on (x_i, y_i) by invoking the best possible communication protocol for g. A *lifting theorem* is a theorem that says that this natural protocol is optimal.

We note that it is often desirable to consider the case where f is a search problem with an arbitrary range rather than a boolean function (see section 2 for the definition of search problems). Most of the known results, as well as the results of this work, apply to this general case. However, for the simplicity of presentation, we focus for now on the case where f is a boolean function.

Applications of lifting theorems. One important reason for why lifting theorems are interesting is that 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 query complexity (which is a relatively simple model) to communication complexity (which is a significantly richer model).

In particular, the first result of this form, due to Raz and McKenzie [35], 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 [22] 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 then lifting it to the setting of communication complexity. This result stimulated a flurry of work on lifting theorems of various kinds, such as lifting for zero-communication protocols [20], round-preserving lifting theorems with applications to time-space trade-offs for proof complexity [10], deterministic lifting theorems with other gadgets [8, 42], lifting theorems from randomized query complexity to randomized communication complexity [23], lifting theorems for directed acyclic graph (DAG)-like protocols [17] with applications to monotone circuit lower bounds, lifting theorems for asymmetric communication problems [9] with applications to data structures, a lifting theorem for the EQUALITY gadget [32], lifting theorems for XOR functions with applications to the log-rank conjecture [24], and lifting theorems for applications to monotone formula complexity, monotone span programs, and proof complexity [21, 37, 33, 34]. There are also lifting theorems which lift more analytic properties of the function like approximate degree due to Sherstov [39] and independently due to Shi and Zhu [40].

In almost all known lifting theorems, the function f can be arbitrary 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? This question is interesting because usually the applications of lifting theorems work by reducing the composed function $f \circ g^n$ to some other problem of interest, and the choice of the gadget g affects the efficiency of such reductions.

In particular, applications of lifting theorems often depend on the size of the gadget, which is the length b of the input to g. Both the deterministic lifting theorem of Raz and McKenzie [35] and the randomized lifting theorem of Göös, Pitassi, and Watson [23] use a gadget of 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 [8] and [42]. Moreover, [8, 29] showed the lifting to

work for a class of gadgets with a certain pseudorandom property rather than just a single specific gadget. A gadget of logarithmic size was also obtained earlier in lifting theorems for more specialized models, such as the work of [20]. However, the randomized lifting theorem of Göös, Pitassi, and Watson [23] seemed to work only with a specific gadget of polynomial size.

In this work, we prove a lifting theorem for a large family of gadgets, namely, all functions g with logarithmic length and exponentially small discrepancy (see section 1.2 for details). Our theorem holds both in the deterministic and the randomized setting. This allows for a considerably larger variety of gadgets; in particular, our theorem is the first lifting theorem in the randomized setting that uses logarithmic-size gadgets; it allows lifting with the inner-product gadget (previously known only in the deterministic setting [8, 42]), and it is also the first lifting theorem that shows that a random function can be used as a gadget.

We would like to point out that, although we reduce the gadget size to logarithmic in this work, it is not enough to obtain the interesting applications a constant sized gadget would have yielded. Nevertheless, our randomized lifting theorem still has some applications. For example, our theorem can be used to simplify the lower bounds of Göös and Jayram [18] 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. In addition, our theorem can be used to derive the separation of randomized separation from partition number (due to [19]) for functions with larger complexity (compared to their input length).

Lifting theorems as a generalization of direct-sum theorems. Lifting theorems can also be motivated from another angle, which is particularly appealing in our case: lifting theorems can be viewed as a generalization of direct-sum theorems. The direct-sum question is a classical question in complexity theory, which asks whether performing a task on n independent inputs is n times harder than performing it on a single input. When specialized to the setting of communication complexity, a direct-sum theorem is a theorem that says that the communication complexity of a computing g on n independent inputs is about n times larger than the communication complexity of g. 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 m times larger than the communication complexity of g.

The direct-sum question for communication complexity has been raised in [26], and has since attracted much attention. While we do not have a general direct-sum theorem for all functions, many works have proved direct-sum theorems and XOR lemmas for large families of functions [13, 38, 3, 31, 28, 2, 6, 4] as well as provided counterexamples [13, 14, 16, 15].

Now, observe that lifting theorems are natural generalizations of direct-sum theorems and XOR lemmas; 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. More generally, a lifting theorem says that the communication complexity of computing any function f of the outputs of g on independent inputs is larger than the complexity of g by a factor that depends on the query complexity of f. This is perhaps the strongest and most natural "direct-sum-like theorem" for g that one could hope for.

From this perspective, it is natural to ask which functions g admit such a strong theorem. The previous works of [35, 23] can be viewed as establishing this theorem only for the index function. The works of [8, 29] have made further progress, establishing this theorem for a class of functions with satisfy certain hitting properties.

However, the latter property is somewhat nonstandard and ad hoc, and their theorem holds only in the deterministic setting. In this work, we establish such theorems for all functions g with low discrepancy, which is a standard and well-studied complexity measure, and we do so in both the deterministic and the randomized setting.

1.2. Our results. In this work, we prove lifting theorems for gadgets of low discrepancy. In what follows, we denote by D^{dt} and D^{cc} the deterministic query complexity and communication complexity of a task, respectively, and by $R^{\mathrm{dt}}_{\varepsilon}$ and $R^{\mathrm{cc}}_{\varepsilon}$ the randomized query complexity and (public-coin) communication complexity with error probability ε, respectively. Given a search problem $\mathcal{S}$ and a gadget $g : \{0,1\}^b \times \{0,1\}^b \to \{0,1\}$, it is easy to see that

$$D^{\mathrm{cc}}(\mathcal{S} \circ g^n) = O\left(D^{\mathrm{dt}}(\mathcal{S}) \cdot b\right),$$
$$R^{\mathrm{cc}}_{\varepsilon}(\mathcal{S} \circ g^n) = O\left(R^{\mathrm{dt}}(\mathcal{S}) \cdot b\right).$$

This upper bound is proved using the simple protocol discussed earlier: the party simulates the optimal decision tree for $\mathcal{S}$, and whenever a query is made, the parties compute g on the corresponding input in order to answer the query (which can be done by communicating at most $b+1$ bits). Our main result says that when g has low discrepancy and b is at least logarithmic, that this upper bound is roughly tight. In order to state this result, we first recall the definition of discrepancy.

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

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

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

Discrepancy is a useful measure for the complexity of g and, in particular, it is well known that for $\varepsilon > 0$,

$$D^{\mathrm{cc}}(g) \geq R^{\mathrm{cc}}_{\varepsilon}(g) \geq \log \frac{1 - 2 \cdot \varepsilon}{\mathrm{disc}(g)}$$

(see, e.g., [30]). We now state our main result.

THEOREM 1.2 (main theorem). *For every $\eta > 0$ there exists $c = O(\frac{1}{\eta^2} \cdot \log \frac{1}{\eta})$ 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* $\mathrm{disc}(g) \leq 2^{-\eta \cdot b}$ *and such that $b \geq c \cdot \log n$. Then*

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

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

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

where $\varepsilon' = \varepsilon + 2^{-\eta \cdot b/8}$.

We note that our results are in fact more general, and preserve the round complexity of $\mathcal{S}$ among other things. See sections 4 and 5 for more details.

Remark 1.3. Note that our main theorem can be applied to a random function $g : \{0,1\}^b \times \{0,1\}^b \to \{0,1\}$, since such a function has a very low discrepancy. As noted above, we believe that our theorem is the first theorem to allow the gadget to be a random function.

Unifying deterministic and randomized lifting theorems. The existing proofs of deterministic lifting theorems and randomized lifting theorems are quite different. While both proofs rely on information-theoretic arguments, they measure information in different ways. In particular, while the randomized lifting theorem of [23] (following [20]) measures information using min-entropy, the deterministic lifting theorems of [35, 22, 8, 42] (following [12]) measure information using a notion known as *thickness* (with [17] being a notable exception). A natural direction of further research is to investigate if these disparate techniques can be unified. Indeed, a related question was raised by [20], who asked if min-entropy based techniques could be used to prove (or simplify the existing proof of) Raz–McKenzie style deterministic lifting theorems.

Our work answers this question affirmatively: we prove both the deterministic and randomized lifting theorems using the same strategy. In particular, both proofs measure information using min-entropy. In doing so, we unify both lifting theorems under the same framework.

1.3. Our techniques. We turn to describe the high-level ideas that underlie the proof of our main theorem. Following the previous works, we use a "simulation argument": we show that given a protocol Π that solves $\mathcal{S} \circ g^n$ with communication complexity C, we can construct a decision tree T that solves $\mathcal{S}$ with query complexity $O(\frac{C}{b})$. The decision tree T works by simulating the action of the protocol Π (hence the name simulation argument). We now describe this simulation in more detail, following the presentation of [23].

The simulation argument. For simplicity of notation, let us denote $\Lambda = \{0,1\}^b$, so g is a function from a "block" in $\Lambda \times \Lambda$ to $\{0,1\}$. Let $G = g^n : \Lambda^n \times \Lambda^n \to \{0,1\}^n$ be the function that takes n disjoint blocks and computes the outputs of g on all of them. We assume that we have a protocol Π that solves $\mathcal{S} \circ G$ with complexity C, and would like to construct a decision tree T that solves $\mathcal{S}$ with complexity $O(\frac{C}{b})$. The basic idea is that given an input $z \in \{0,1\}^n$, the tree T simulates the action of Π on the random inputs (X, Y) that are uniformly distributed over $G^{-1}(z)$. Clearly, it holds that $\mathcal{S} \circ G(X, Y) = \mathcal{S}(z)$, so this simulation, if done right, outputs the correct answer.

The core issue in implementing such a simulation is the following question: how can T simulate the action of Π on $(X, Y) \in G^{-1}(z)$ without knowing z? The answer is that as long as the protocol Π has transmitted less than $\varepsilon \cdot b$ bits of information about every block (X_i, Y_i) (for some specific $\varepsilon > 0$), the distribution of (X, Y) is similar to the uniform distribution in a certain sense (that will be formalized soon). Thus, the tree T can pretend that (X, Y) are distributed uniformly, and simulate the action of Π on such inputs, which can be done without knowing z.

This idea can be implemented as long as the protocol has transmitted less than $\varepsilon \cdot b$ bits of information about every block (X_i, Y_i). However, at some point, the protocol may transmit more than $\varepsilon \cdot b$ bits of information about some blocks. Let $I \subseteq [n]$ denote the set of these blocks. At this point, it is no longer true that the distribution of (X, Y) is similar to the uniform distribution. However, it can be shown

that the distribution of (X, Y) is similar to the uniform distribution *conditioned on* $g^I(X_I, Y_I) = z_I$. Thus, the tree T queries the bits in z_I, and can now continue the simulation of Π on $(X, Y) \in G^{-1}(z)$ by pretending that (X, Y) are distributed uniformly conditioned on $g^I(X_I, Y_I) = z_I$. The tree proceeds in this way, adding blocks to I as necessary, until the protocol Π ends, at which point T halts and outputs the same output as Π.

It remains to show that the query complexity of T is at most $O(\frac{C}{b})$. To this end, observe that the query complexity of T is exactly the size of the set I at the end of the simulation. Moreover, recall that the set I is the set of blocks on which the protocol transmitted at least $\varepsilon \cdot b$ bits of information. Hence, at any given point, the protocol must have transmitted at least $\varepsilon \cdot b \cdot |I|$ bits. On the other hand, we know by assumption that the protocol never transmitted more than C bits. This implies that $\varepsilon \cdot b \cdot |I| \leq C$ and therefore the query complexity of the tree T is at most $|I| \leq \frac{C}{\varepsilon \cdot b} = O(\frac{C}{b})$. This concludes the argument.

Our contribution. In order to implement the foregoing simulation argument, there are two technical issues that need to be addressed and are relevant at this point:

- *The uniform marginals issue:* In the above description, we argued that as long as the protocol has not transmitted too much information, the distribution of (X, Y) is "similar to the uniform distribution." The question is how do we formalize this idea. This issue was dealt with implicitly in several works in the lifting literature since [35], and was made explicit in [23] as the "uniform marginals lemma": if every set of blocks in (X, Y) has sufficient min-entropy, then each of the marginals X, Y on its own is close to the uniform distribution. In [23], they proved this lemma for the case where g is the index function, and in [20] a very similar lemma was proved for the case where g is the inner product function.
- *The conditioning issue:* As we described above, when the protocol transmits too much information about a set of blocks $I \subseteq [n]$, the tree T queries z_I and conditions the distribution of (X, Y) on the event that $g^I(X_I, Y_I) = z_I$. In principle, this conditioning may reveal information on $(X_{[n]-I}, Y_{[n]-I})$, which might reduce their min-entropy and ruin their uniform-marginals property. In order for the simulation argument to work, one needs to show that this cannot happen, and the conditioning will never reveal too much information about $(X_{[n]-I}, Y_{[n]-I})$.

  In the works of [35, 42, 23] this issue was handled by arguments that are tailored to the index and inner product functions. The work of [8] gave this issue a more general treatment, by identifying an abstract property of g that prevents the conditioning from revealing too much information. However, as discussed above, this abstract property is somewhat ad hoc, and only works for deterministic simulation.

Our contribution is dealing with both issues in the general setting where g is an arbitrary low-discrepancy gadget. In order to deal with the first issue, we prove a uniform marginals lemma for such gadgets g: this is relatively easy, since the proof of [20] for the inner product gadget turns out to generalize in a straightforward way to arbitrary low-discrepancy gadgets.

The core of this work is in dealing with the conditioning issue. Our main technical lemma says that as long as every set of blocks in (X, Y) has sufficient min-entropy, there are only a few possible values of X, Y that are "dangerous" (in the sense that they may lead the conditioning to leak too much information). We now modify the simulation such that it discards these dangerous values before performing the

conditioning. Since there are only a few of those dangerous values, discarding them does not reveal too much information on X and Y, and the simulation can proceed as before.

1.4. Open problems. The main question that arises from this work is how much more general can the gadget g be? As was discussed in section 1.1, lifting theorems can be viewed as a generalization of direct-sum theorems. In the setting of randomized communication complexity, it is known that the "ability of g to admit a direct-sum theorem" is characterized exactly by a complexity measure called the *information cost* of g (denoted $\boldsymbol{IC}(g)$). In particular, the complexity of computing a function g on n independent copies is $\approx n \cdot \boldsymbol{IC}(g)$ [2, 5, 4]. This leads to the natural conjecture that a lifting theorem should hold for every gadget g that has sufficiently high information cost.

CONJECTURE 1.4. *There exists a constant $c > 0$ such that the following holds. Let $\mathcal{S}$ be any 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 $\boldsymbol{IC}(g) \geq c \cdot \log n$. Then*

$$R^{\mathrm{cc}}_{\varepsilon}(\mathcal{S} \circ g^n) = \Omega\left(R^{\mathrm{dt}}_{\varepsilon'}(\mathcal{S}) \cdot \boldsymbol{IC}(g)\right).$$

Proving this conjecture would give us a nearly complete understanding of the lifting phenomenon which, in addition to being interesting in its own right, would likely lead to many applications. In particular, this conjecture implies our result, since it is known that $\log \frac{1}{\mathrm{disc}(g)}$ (roughly) lower bounds the information cost of g [27].

Conjecture 1.4 is quite ambitious. As intermediate goals, one could attempt to prove such a lifting theorem for other complexity measures that are stronger than discrepancy and weaker than information cost (see [25, 27] for several measures of this kind). To begin with, one could consider the well-known corruption bound of [43, 1, 36]: could we prove a lifting theorem for an arbitrary gadget g that has a low corruption bound? A particularly interesting example for such a gadget is the disjointness function—indeed, proving a lifting theorem for the disjointness gadget would be interesting in its own right and would likely have applications, in addition to being a step toward Conjecture 1.4.

An even more modest intermediate goal is to gain a better understanding of lifting theorems with respect to discrepancy. For starters, our result only holds[1] for gadgets whose discrepancy is exponentially vanishing in the gadget size. Can we prove a lifting theorem for gadgets g with larger discrepancy? In particular, since the randomized communication complexity of g is lower bounded by $\Omega(\log \frac{1}{\mathrm{disc}(g)})$, the following conjecture comes to mind.

CONJECTURE 1.5. *There exists a constant $c > 0$ such that the following holds. Let $\mathcal{S}$ be any 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 $\log \frac{1}{\mathrm{disc}(g)} \geq c \cdot \log n$. Then*

$$R^{\mathrm{cc}}_{\varepsilon}(\mathcal{S} \circ g^n) = \Omega\left(R^{\mathrm{dt}}_{\varepsilon'}(\mathcal{S}) \cdot \log \frac{1}{\mathrm{disc}(g)}\right).$$

Another interesting direction is to consider discrepancy with respect to other distributions. The definition of discrepancy we gave above (Definition 1.1) is a special case of a more general definition, in which the random variables (U, V) are distributed

[1] More accurately, our result can be applied to gadgets with larger discrepancy, but then the gadget size has to be larger than logarithmic.

according to some fixed distribution μ over $\Lambda \times \Lambda$. Thus, our result works only when μ is the uniform distribution. Can we prove a lifting theorem that holds for an arbitrary choice of μ? While we have not verified it, we believe that our proof can yield a lifting theorem that works whenever μ is a product distribution (after some natural adaptations). However, proving such a lifting theorem for nonproduct distributions seems to require new ideas. We note that direct-sum theorems for discrepancy have been proved by [38, 31], and proving Conjecture 1.4 (and extending it to an arbitrary distribution μ) seems like a natural extension of their results.

Yet another interesting direction is to consider the lifting analogue of strong direct product theorems. Such theorems say that when we compute g on n independent inputs, then not only does the communication complexity increase by a factor of n, but the success probability also drops exponentially in n (see, e.g., [38, 28, 11, 6]). A plausible analogue for lifting theorems is to conjecture that the success probability of computing $\mathcal{S} \circ g^n$ drops exponentially in the query complexity of $\mathcal{S}$. It would be interesting to see a result along these lines.

Finally, there remains the major open problem of the lifting literature to prove a lifting theorem that uses gadgets of constant size.

Organization of the paper. In section 2, we provide the required preliminaries. In section 3, we set up the lifting machinery that is used in both the deterministic and the randomized lifting results, including our uniform marginals lemma and our main technical lemma. We prove the deterministic part of our main theorem in section 4, and the randomized part of our main theorem in section 5.

2. Preliminaries. We assume familiarity with the basic definitions of communication complexity (see, e.g., [30]). For any $n \in \mathbb{N}$, we denote $[n] \stackrel{\text{def}}{=} \{1, \ldots, n\}$. Given a boolean random variable V, we denote the bias of V by

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

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

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

For every $I \subseteq [n]$, we denote by $g^{\oplus I} : \mathcal{X}^I \times \mathcal{Y}^I \to \{0,1\}$ the function that given $x \in \mathcal{X}^I$ and $y \in \mathcal{Y}^I$, outputs the parity of the string $g^I(x,y)$.

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

Intuitively, a search problem $\mathcal{S}$ represents the following task: given an input $z \in \mathcal{I}$, find a solution $o \in \mathcal{S}(z)$. In particular, if $\mathcal{I} = \mathcal{X} \times \mathcal{Y}$ for some finite sets $\mathcal{X}, \mathcal{Y}$, we say that a deterministic protocol Π *solves* $\mathcal{S}$ if for every input $(x,y) \in \mathcal{I}$, the output of Π

is in $\mathcal{S}(x,y)$. We say that a randomized protocol Π *solves* $\mathcal{S}$ *with error* ε if for every input $(x,y) \in \mathcal{I}$, the output of Π is in $\mathcal{S}(x,y)$ with probability at least $1-\varepsilon$.

We denote the deterministic communication complexity of a search problem $\mathcal{S}$ with $D^{\mathrm{cc}}(\mathcal{S})$. Given $\varepsilon > 0$, we denote by $R^{\mathrm{cc}}_{\varepsilon}(\mathcal{S})$ the randomized (public-coin) communication complexity of $\mathcal{S}$ with error ε (i.e., the minimum worst-case complexity of a randomized protocol that solves $\mathcal{S}$ with error ε).

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

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

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

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

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

The *randomized query complexity of* $\mathcal{S}$ *with error* ε, denoted $R^{\mathrm{dt}}_{\varepsilon}$, is the minimal query complexity of a randomized decision tree that solves $\mathcal{S}$ with error ε. Again, when we omit ε, it is assumed to be $\frac{1}{3}$.

2.1.1. Parallel decision trees. Our lifting theorems have the property that they preserve the round complexity of protocols, which is useful for some applications [10]. In order to define this property, we need a notion of a decision tree that has an analogue of "round complexity." Such a notion, due to [41], is called a *parallel decision tree*. Informally, a parallel decision tree is a decision tree that works in "rounds," where in each round multiple queries are issued simultaneously. The round complexity of the tree is the number of rounds, whereas the query complexity is the total number of queries issued.

Formally, a *deterministic parallel decision tree* T from $\{0,1\}^n$ to $\mathcal{O}$ is a rooted tree in which every internal node is labeled with a set $I \subseteq [n]$ (representing the queries issued simultaneously at this round) and has degree $2^{|I|}$. The edges going out of such a node are labeled with all the possible assignments in $\{0,1\}^I$, and every leaf is labeled by some output $o \in \mathcal{O}$. As before, such a tree naturally computes a function that is denoted by T, and it solves a search problem $\mathcal{S} \subseteq \{0,1\}^n \times \mathcal{O}$ if $T(z) \in \mathcal{S}(z)$ for all $z \in \{0,1\}^n$. The depth of such a tree is now the analogue of the number of rounds in a protocol. The query complexity of T is defined as the maximum, over all leaves ℓ, of the sum of the sizes of the sets I that are labeling the vertices on the path from the

root to ℓ. A *randomized parallel decision tree* is defined analogously to the definition of randomized decision trees above.

2.2. Fourier analysis. Given a set $S \subseteq [m]$, the *character* χ_S is the function from $\{0,1\}^m$ to $\mathbb{R}$ that is defined by

$$\chi_S(z) \stackrel{\text{def}}{=} (-1)^{\bigoplus_{i \in S} z_i}.$$

Here, if $S = \emptyset$ then we define $\bigoplus_{i \in S} z_i = 0$. Given a function $f : \{0,1\}^m \to \mathbb{R}$, its *Fourier coefficient* $\hat{f}(S)$ is defined as

$$\hat{f}(S) \stackrel{\text{def}}{=} \frac{1}{2^m} \sum_{z \in \{0,1\}^m} f(z) \cdot \chi_S(z).$$

It is a standard fact of Fourier analysis that f can be written as

$$f(z) = \sum_{S \subseteq [m]} \hat{f}(S) \cdot \chi_S(z). \tag{1}$$

We have the following useful observation.

FACT 2.1. *Let Z be a random variable taking values in $\{0,1\}^m$, and let $\mu : \{0,1\}^m \to \mathbb{R}$ be its density function. Then, for every set $S \subseteq [m]$ it holds that*

$$|\hat{\mu}(S)| = 2^{-m} \cdot \operatorname{bias}\left(\bigoplus_{i \in S} Z_i\right).$$

In particular, $\hat{\mu}(\emptyset) = 2^{-m}$.

Proof. Let $S \subseteq [m]$. It holds that

$$\begin{aligned}
|\hat{\mu}(S)| &= 2^{-m} \cdot \left| \sum_{z \in \{0,1\}^m} \mu(z) \cdot \chi_S(z) \right| \\
&= 2^{-m} \cdot \left| \sum_{z \in \{0,1\}^m} \mu(z) \cdot (-1)^{\bigoplus_{i \in S} z_i} \right| \\
&= 2^{-m} \cdot \left| \sum_{z \in \{0,1\}^m : \bigoplus_{i \in S} z_i = 0} \mu(z) - \sum_{z \in \{0,1\}^m : \bigoplus_{i \in S} z_i = 1} \mu(z) \right| \\
&= 2^{-m} \cdot \left| \Pr\left[\bigoplus_{i \in S} Z_i = 0\right] - \Pr\left[\bigoplus_{i \in S} Z_i = 1\right] \right| \\
&= 2^{-m} \cdot \operatorname{bias}\left(\bigoplus_{i \in S} Z_i\right),
\end{aligned}$$

as required. The "in particular" part follows by noting that in the case of $S = \emptyset$, the character χ_S is the constant function 1, and recalling that the sum of $\mu(z)$ over all z's is 1. ∎

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

$$|\mu_1 - \mu_2| = \max_{\mathcal{E} \subseteq \Omega} \{|\mu_1(\mathcal{E}) - \mu_2(\mathcal{E})|\}.$$

It is not hard to see that the maximum is attained when $\mathcal{E}$ consists of all the values $\omega \in \Omega$ such that $\mu_1(\omega) > \mu_2(\omega)$. We say that μ and μ_2 are ε-*close* if $|\mu - \mu_2| \leq \varepsilon$. The *min-entropy* of a random variable X, denoted $H_\infty(X)$, is the largest number $k \in \mathbb{R}$ such that for every value x it holds that

$$\Pr[X = x] \leq 2^{-k}.$$

Min-entropy has the following easy-to-prove properties.

FACT 2.2. *Let X be a random variable and let $\mathcal{E}$ be an event. Then, $H_\infty(X|\mathcal{E}) \geq H_\infty(X) - \log \frac{1}{\Pr[\mathcal{E}]}$.*

FACT 2.3. *Let X_1, X_2 be random variables taking values from sets $\mathcal{X}_1, \mathcal{X}_2$, respectively. Then, $H_\infty(X_1) \geq H_\infty(X_1, X_2) - \log |\mathcal{X}_2|$.*

We say that a distribution is k-flat if it is uniformly distributed over a subset of the sample space of size at least 2^k. The following standard fact is useful.

FACT 2.4. *If a random variable X has min-entropy k, then its distribution is a convex combination of k-flat distributions.*

2.3.1. Vazirani's lemma. Vazirani's lemma is a useful result which says that a random string is close to being uniformly distributed if the XOR of every set of bits in the string has a small bias. We use the following variant of the lemma due to [20].

LEMMA 2.5 ([20]). *Let $\varepsilon > 0$, and let Z be a random variable taking values in $\{0,1\}^m$. If for every nonempty set $S \subseteq [m]$ it holds that*

$$\text{bias}\left(\bigoplus_{i \in S} Z_i\right) \leq \varepsilon \cdot (2 \cdot m)^{-|S|}, \tag{2}$$

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

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

Proof. Let $\mu : \{0,1\}^m \to \mathbb{R}$ be the density function of Z, and let $z \in \{0,1\}^m$. By (1) it holds that

$$\begin{aligned}
\left|\mu(z) - 2^{-m}\right| &= \left|\sum_{S\subseteq[m]} \hat{\mu}(S)\cdot\chi_S(z) - 2^{-m}\right| \\
(\hat{\mu}(\emptyset) = 2^{-m} \text{ by Fact 2.1}) &= \left|\sum_{S\subseteq[m]:S\neq\emptyset} \hat{\mu}(S)\cdot\chi_S(z)\right| \\
(\text{since } |\chi_S(z)| \text{ is always } 1) &\leq \sum_{S\subseteq[m]:S\neq\emptyset} |\hat{\mu}(S)| \\
(\text{Fact 2.1}) &= 2^{-m}\cdot \sum_{S\subseteq[m]:S\neq\emptyset} \operatorname{bias}\left(\bigoplus_{i\in S} Z_i\right) \\
(\text{inequality (2)}) &\leq 2^{-m}\cdot \sum_{S\subseteq[m]:S\neq\emptyset} \varepsilon\cdot(2\cdot m)^{-|S|} \\
&= \varepsilon\cdot 2^{-m}\cdot \sum_{i=1}^{m}\binom{m}{i}\cdot(2\cdot m)^{-i} \\
&\leq \varepsilon\cdot 2^{-m}\cdot \sum_{i=1}^{m} m^i\cdot(2\cdot m)^{-i} \\
&\leq \varepsilon\cdot 2^{-m}\cdot \sum_{i=1}^{m} 2^{-i} \\
&\leq \varepsilon\cdot 2^{-m},
\end{aligned}$$

as required. ∎

Lemma 2.5 says that if the bias of $\bigoplus_{i\in S} Z_i$ is small for every S, then Z is close to being uniformly distributed. It turns out that if the latter assumption holds only for large sets S, we can still deduce something useful, namely, that the min-entropy of Z is high.

LEMMA 2.6. *Let $t \in \mathbb{N}$ be such that $t \geq 1$, and let Z be a random variable taking values in $\{0,1\}^m$. If for every set $S \subseteq [m]$ such that $|S| \geq t$ it holds that*

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

then $H_\infty(Z) \geq m - t\log m - 1$.

Proof. Observe that if $m = 1$ then the bound holds vacuously, so we may assume that $m \geq 2$. Let $\mu : \{0,1\}^m \to \mathbb{R}$ be the density function of Z, and let $z \in \{0,1\}^m$.

By equality (1) it holds that

$$\begin{aligned}
\mu(z) &= \sum_{S\subseteq[m]} \hat{\mu}(S)\cdot\chi_S(z)\\
(\text{since } |\chi_S(z)| \text{ is always } 1) &\le \sum_{S\subseteq[m]} |\hat{\mu}(S)|\\
(\text{Fact 2.1}) &\le 2^{-m}\cdot \sum_{S\subseteq[m]} \text{bias}(\oplus_{i\in S} Z_i)\\
&\le 2^{-m}\\
&\quad\cdot\left(\sum_{S\subseteq[m]:|S|<t} \text{bias}(\oplus_{i\in S} Z_i) + \sum_{S\subseteq[m]:|S|\ge t} \text{bias}(\oplus_{i\in S} Z_i)\right).
\end{aligned}$$

We now bound each of the two terms separately. The term for sets S whose size is at least t can be upper bounded by 1 using exactly the same calculation as in the proof of Lemma 2.5. In order to upper bound the term for sets whose size is less than t, observe that $\text{bias}(\oplus_{i\in S} Z_i) \le 1$ for every $S \subseteq [m]$ and therefore

$$\begin{aligned}
\sum_{S\subseteq[m]:|S|<t} \text{bias}(\oplus_{i\in S} Z_i) &\le \sum_{i=0}^{t-1} \binom{m}{i}\\
&\le \sum_{i=0}^{t-1} m^i\\
&= \frac{m^t-1}{m-1}\\
(\text{since } m\ge 2) &\le m^t - 1.
\end{aligned}$$

It follows that

$$\begin{aligned}
\mu(z) &\le 2^{-m}\cdot\left[\left(m^t-1\right)+1\right]\\
&= 2^{-(m-t\cdot\log m)}.
\end{aligned}$$

Thus, $H_\infty(Z) \ge m - t\cdot\log m$ as required. Note that this bound is a bit stronger than claimed in the lemma: indeed, we only need the "-1" term in the lemma in order to deal with the case where $m = 1$. ∎

2.3.2. Coupling. Let μ_1, μ_2 be two distributions over sample spaces Ω_1, Ω_2. A *coupling of μ_1 and μ_2* is a distribution ν over the sample space $\Omega_1\times\Omega_2$ whose marginal over the first coordinate is μ_1 and whose marginal over the second coordinate is μ_2. In the case where $\Omega_1 = \Omega_2 = \Omega$, the following standard fact allows us to use couplings to study the statistical distance between μ_1 and μ_2.

FACT 2.7. *Let μ_1, μ_2 be two distributions over a sample space Ω. The statistical distance between μ_1 and μ_2 is equal to the minimum, over all couplings ν of μ_1 and μ_2, of*

$$\Pr_{(X,Y)\leftarrow\nu}[X \ne Y].$$

In particular, we can upper bound the statistical distance between μ_1 and μ_2 by constructing a coupling ν in which the probability that $X \ne Y$ is small.

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

FACT 2.8. *Let $C \subseteq \{0,1\}^*$ be a finite prefix-free code, and let W be a random string taking values from C. Then, there exists a string $w \in C$ such that* $\Pr[W = w] \geq \frac{1}{2^{|w|}}$.

For completeness, we provide the following simple proof of 2.8 that does not rely on Kraft's inequality.

Proof. Let n be the maximal length of a string in C, and let W' be a random string in $\{0,1\}^n$ that is sampled according to the following process: sample a string w from W, choose a uniformly distributed string $z \in \{0,1\}^{n-|w|}$, and set $W' = w \circ z$ (where here $\circ$ denotes string concatenation).

By a simple averaging argument, there exists a string $w' \in \{0,1\}^n$ such that $\Pr[W' = w'] \geq \frac{1}{2^n}$. Since C is a prefix-free code, there exists a unique prefix w of w' that is in C. The definition of W' implies that

$$\Pr[W' = w'] = \Pr[W = w] \cdot \frac{1}{2^{n-|w|}},$$

because the only way the string w' could be sampled is by first sampling w and then sampling z to be the rest of w' (again, since C is a prefix-free code). Hence, it follows that

$$\Pr[W = w] \cdot \frac{1}{2^{n-|w|}} \geq \frac{1}{2^n},$$
$$\Pr[W = w] \geq \frac{1}{2^{|w|}},$$

as required. ∎

2.5. Discrepancy. We start by recalling the definition of discrepancy.

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

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

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

Let $g : \Lambda \times \Lambda \to \{0,1\}$ be a function with discrepancy at most $|\Lambda|^{-\eta}$. Such functions g satisfy the following "extractor-like" property. In what follows, the parameter λ controls bias $(g(X,Y))$.

LEMMA 2.9. *Let X, Y be independent random variables taking values in Λ such that $H_\infty(X) + H_\infty(Y) \geq (2 - \eta + \lambda) \cdot \log|\Lambda|$. Then,*

$$\mathrm{bias}(g(X,Y)) \leq |\Lambda|^{-\lambda}.$$

Proof. By Fact 2.4, it suffices to consider the case where X and Y have flat distributions. Let $A, B \subseteq \Lambda$ be the sets over which X, Y are uniformly distributed,

and denote $R \stackrel{\text{def}}{=} A \times B$. By the assumption on the min-entropies of X and Y, it holds that $|R| \geq |\Lambda|^{2-\eta+\lambda}$.

Let U, V be random variables that are uniformly distributed over Λ. Then, X and Y are distributed like $U|U \in A$ and $V|V \in B$, respectively. It follows that

$$\begin{aligned}
\text{bias}\left(g(X,Y)\right) &= \left|\Pr\left[g(X,Y)=0\right] - \Pr\left[g(X,Y)=1\right]\right| \\
&= \left|\Pr\left[g(U,V)=0|(U,V)\in R\right] - \Pr\left[g(U,V)=1|(U,V)\in R\right]\right| \\
&= \frac{\text{disc}_R(g)}{\Pr\left[(U,V)\in R\right]} \\
&\leq \frac{|\Lambda|^{-\eta}}{\Pr\left[(U,V)\in R\right]} \\
&= \frac{|\Lambda|^{-\eta}}{|R|\,/\,|\Lambda|^2} \\
&\leq \frac{|\Lambda|^{-\eta}}{|\Lambda|^{-\eta+\lambda}} \\
&= |\Lambda|^{-\lambda},
\end{aligned}$$

as required. ∎

Using Lemma 2.9, we can obtain the following sampling property, which says that with high probability X takes a value x for which $\text{bias}\left(g(x,Y)\right)$ is small. In what follows, the parameter λ controls $\text{bias}\left(g(x,Y)\right)$, the parameter γ controls the error probability, and recall that η is the parameter that controls the discrepancy of g (i.e., $\text{disc}(g) \leq 2^{-\eta \cdot b}$).

LEMMA 2.10. *Let $\gamma, \lambda > 0$. Let X, Y be independent random variables taking values in Λ such that*

$$H_\infty(X) + H_\infty(Y) \geq (2 - \eta + \gamma + \lambda) \cdot \log|\Lambda| + 1.$$

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

$$\text{bias}\left(g(x,Y)\right) > |\Lambda|^{-\lambda}$$

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

Proof. For every $x \in \Lambda$, denote

$$p_x \stackrel{\text{def}}{=} \Pr\left[g(x,Y)=1\right].$$

Using this notation, our goal is to prove that

$$\Pr_X\left[\left|p_X - \frac{1}{2}\right| \leq \frac{1}{2}|\Lambda|^{-\lambda}\right] < |\Lambda|^{-\gamma}.$$

We will prove that the probability that $p_X > \frac{1}{2} + \frac{1}{2}|\Lambda|^{-\lambda}$ is less than $\frac{1}{2} \cdot |\Lambda|^{-\gamma}$, and a similar proof can be used to show that the probability that $p_X < \frac{1}{2} - \frac{1}{2}|\Lambda|^{-\lambda}$ is less than $\frac{1}{2} \cdot |\Lambda|^{-\gamma}$. The required result will then follow by the union bound.

Let $\mathcal{E} \subseteq \Lambda$ be the set of values x such that $p_x > \frac{1}{2} + \frac{1}{2}|\Lambda|^{-\lambda}$. Suppose for the sake of contradiction that $\Pr\left[X \in \mathcal{E}\right] \geq \frac{1}{2} \cdot |\Lambda|^{-\gamma}$. It clearly holds that

$$\Pr\left[g(X,Y)=1|X\in\mathcal{E}\right] > \frac{1}{2} + \frac{1}{2}|\Lambda|^{-\lambda}. \tag{3}$$

On the other hand, it holds that

$$\begin{aligned} H_\infty(X|X \in \mathcal{E}) &\geq H_\infty(X) - \log \frac{1}{\Pr[X \in \mathcal{E}]} \\ \text{(by assumption on } \mathcal{E} \text{ and Fact 2.2)} &\geq H_\infty(X) - \gamma \cdot \log|\Lambda| - 1. \end{aligned}$$

This implies that

$$\begin{aligned} H_\infty(X|X \in \mathcal{E}) + H_\infty(Y) &\geq H_\infty(X) + H_\infty(Y) - \gamma \cdot \log|\Lambda| - 1 \\ &\geq (2-\eta) \cdot \log|\Lambda| + \lambda \cdot \log|\Lambda|. \end{aligned}$$

By Lemma 2.9, it follows that

$$\Pr[g(X,Y) = 1 | X \in \mathcal{E}] \leq \frac{1}{2} + \frac{1}{2}|\Lambda|^{-\lambda},$$

which contradicts inequality (3). We reached a contradiction, and therefore the probability that $p_X > \frac{1}{2} + \frac{1}{2}|\Lambda|^{-\lambda}$ is less than $\frac{1}{2} \cdot |\Lambda|^{-\gamma}$, as required. ∎

Recall that $g^{\oplus I}(x,y)$ is the function from $\mathcal{X}^I \times \mathcal{Y}^I$ to $\{0,1\}$ that outputs the parity of the string $g^I(x,y)$. We would like to prove results like Lemmas 2.9 and 2.10 for functions of the form $g^{\oplus I}$. To this end, we use the following XOR lemma for discrepancy due to [31].

THEOREM 2.11 (see [31]). *Let $m \in \mathbb{N}$. Then,*

$$(\mathrm{disc}(g))^m \leq \mathrm{disc}(g^{\oplus m}) \leq (64 \cdot \mathrm{disc}(g))^m.$$

By combining Theorem 2.11 with Lemmas 2.9 and 2.10, we obtain the following results.

COROLLARY 2.12. *Let $\lambda > 0$, $n \in \mathbb{N}$, and $S \subseteq [n]$. Let X, Y be independent random variables taking values in Λ^S such that*

$$H_\infty(X) + H_\infty(Y) \geq \left(2 + \frac{6}{\log|\Lambda|} - \eta + \lambda\right) \cdot |S| \cdot \log|\Lambda|.$$

Then,

$$\mathrm{bias}\left(g^{\oplus S}(X,Y)\right) \leq |\Lambda|^{-\lambda \cdot |S|}.$$

COROLLARY 2.13. *Let $\gamma, \lambda > 0$, $n \in \mathbb{N}$, and $S \subseteq [n]$. Let X, Y be independent random variables taking values in Λ^S such that*

$$H_\infty(X) + H_\infty(Y) \geq \left(2 + \frac{7}{\log|\Lambda|} - \eta + \gamma + \lambda\right) \cdot |S| \cdot \log|\Lambda|.$$

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

$$\mathrm{bias}\left(g^{\oplus S}(x,Y)\right) > |\Lambda|^{-\lambda \cdot |S|}$$

is less than $|\Lambda|^{-\gamma \cdot |I|}$.

3. Lifting machinery. In this section, we set up the machinery we need to prove our main theorem, restated next.

THEOREM 1.2. *For every $\eta > 0$ there exists $c = O(\frac{1}{\eta^2} \cdot \log \frac{1}{\eta})$ 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* $\mathrm{disc}(g) \leq 2^{-\eta \cdot b}$ *and such that $b \geq c \cdot \log n$. Then*

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

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

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

where $\varepsilon' = \varepsilon + 2^{-\eta \cdot b/8}$.

For the rest of this paper, we fix $\eta > 0$ and let $c \in \mathbb{N}$ be some sufficiently large parameter that will be determined later such that $c = O(\frac{1}{\eta^2} \cdot \log \frac{1}{\eta})$. Let $n \in \mathbb{N}$, and let $g : \{0,1\}^b \times \{0,1\}^b \to \{0,1\}$ be a function such that $\mathrm{disc}(g) \leq 2^{-\eta \cdot b}$ and such that $b \geq c \cdot \log n$. Note that when $n = 1$, the theorem holds trivially, so we may assume that $n \geq 2$. For convenience, we denote $\Lambda \stackrel{\mathrm{def}}{=} \{0,1\}^b$ and $G \stackrel{\mathrm{def}}{=} g^n$. Throughout the rest of this section, X and Y will always denote random variables whose domain is Λ^n.

As explained in section 1.3, our simulation argument is based on the idea that as long as the protocol did not transmit too much information about the inputs, their distribution is similar to the uniform distribution. The following definition, due to [20], formalizes the notion that the protocol did not transmit too much information about an input X.

DEFINITION 3.1. *Let $\delta_X > 0$. We say that a random variable X is* δ_X-dense *if for every $I \subseteq [n]$ it holds that $H_\infty(X_I) \geq \delta_X \cdot b \cdot |I|$.*

As explained there, whenever the protocol transmits too much information about a bunch of blocks (X_I, Y_I) (where $I \subseteq [n]$), the simulation queries z_I and conditions the distribution on $g(X_I, Y_I) = z_I$. The following definitions provide a useful way for implementing this argument: restrictions are used to keep track of which bits of z have been queried so far, and the notion of structured variables expresses the desired properties of the distribution of the inputs.

DEFINITION 3.2. *A* restriction *ρ is a string in $\{0,1,*\}^n$. We say that a coordinate $i \in [n]$ is* free *in ρ if $\rho_i = *$ and, otherwise, we say that i is* fixed. *Given a restriction $\rho \in \{0,1,*\}^n$, we denote by* $\mathrm{free}(\rho)$ *and* $\mathrm{fix}(\rho)$ *the set of free and fixed coordinates of ρ, respectively. We say that a string $z \in \{0,1\}^n$ is* consistent *with ρ if $z_{\mathrm{fix}(\rho)} = \rho_{\mathrm{fix}(\rho)}$.*

Intuitively, $\mathrm{fix}(\rho)$ represents the queries that have been made so far, and $\mathrm{free}(\rho)$ represents the coordinates that have not been queried yet.

DEFINITION 3.3 (following [23]). *Let $\rho \in \{0,1,*\}^n$ be a restriction, let $\tau > 0$, and let X, Y be independent random variables. We say that X and Y are* (ρ, τ)-structured *if there exist $\delta_X, \delta_Y > 0$ such that $X_{\mathrm{free}(\rho)}$ and $Y_{\mathrm{free}(\rho)}$ are δ_X-dense and δ_Y-dense, respectively, $\delta_X + \delta_Y \geq \tau$, and*

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

We can now state our version of the uniform marginals lemma of [23], which formalizes the idea that if X and Y are structured then their distribution is similar to

the uniform distribution over $G^{-1}(z)$. In what follows, the parameter γ controls the statistical distance from the uniform distribution, and recall that η is the parameter that controls the discrepancy of g (i.e., $\mathrm{disc}(g) \leq 2^{-\eta \cdot b}$).

LEMMA 3.4 (uniform marginals lemma). *There exists a universal constant h such that the following holds: let $\gamma > 0$, let ρ be a restriction, and let $z \in \{0,1\}^n$ be a string that is consistent with ρ. Let X, Y be independent random variables that are uniformly distributed over sets $\mathcal{X}, \mathcal{Y} \subseteq \Lambda^n$, respectively, and assume that they are (ρ, τ)-structured where*

$$\tau \geq 2 + \frac{h}{c} - \eta + \gamma.$$

Let (X', Y') be uniformly distributed over $G^{-1}(z) \cap (\mathcal{X} \times \mathcal{Y})$. Then, X and Y are $2^{-\gamma \cdot b}$-close to X' and Y', respectively.

Remark 3.5. Here, as well as in other claims in the paper, we denote by h some constant that is large enough to make the proofs go through, and does not depend on any other parameter. The constant h can be calculated explicitly, and we only refrain from doing so in order to streamline the presentation. In all the cases where we use this constant, it can be chosen to be at most 50.

1.1 We defer the proof of Lemma 3.4 to section 3.1, and move to discuss the next issue. Recall that in order for X and Y to be structured, the random variables $X_{\mathrm{free}(\rho)}$ and $Y_{\mathrm{free}(\rho)}$ have to be dense. However, as the simulation progresses and the protocol transmits information, this property may be violated, and $X_{\mathrm{free}(\rho)}$ or $Y_{\mathrm{free}(\rho)}$ may cease to be dense. In order to restore the density, we use the following folklore fact.

PROPOSITION 3.6. *Let X be a random variable, let $\delta_X > 0$, and let $I \subseteq [n]$ be a maximal subset of coordinates such that $H_\infty(X_I) < \delta_X \cdot b \cdot |I|$. Let $x_I \in \Lambda^I$ be a value such that*

$$\Pr[X_I = x_I] > 2^{-\delta_X \cdot b \cdot |I|}.$$

Then, the random variable $X_{[n]-I} | X_I = x_I$ is δ_X-dense.

Proof. Assume for the sake of contradiction that $X_{[n]-I} | X_I = x_I$ is not δ_X-dense. Then, there exists a nonempty set $J \subseteq [n]-I$ such that $H_\infty(X_J | X_I = x_I) < \delta_X \cdot b \cdot |J|$. In particular, there exists a value $x_J \in \Lambda^J$ such that

$$\Pr[X_J = x_J | X_I = x_I] > 2^{-\delta_X \cdot b \cdot |J|}.$$

But this implies that

$$\Pr[X_I = x_I \text{ and } X_J = x_J] > 2^{-\delta_X \cdot b \cdot |I \cup J|},$$

which means that

$$H_\infty(X_{I \cup J}) < \delta_X \cdot b \cdot |I \cup J|.$$

However, this contradicts the maximality of I. ∎

Proposition 3.6 is useful in the deterministic setting, since in this setting the simulation is free to condition the distributions of X, Y in any way that maintains their density. However, in the randomized setting, the simulation is more restricted, and cannot condition the inputs on events such as $X_I = x_I$ which may have very low probability. In [23], this issue was resolved by observing that the probability space can be partitioned to disjoint events of the form $X_I = x_I$, and that the randomized simulation can use such a partition to achieve the same effect as Proposition 3.6. This leads to the following lemma, which we use as well.

LEMMA 3.7 (density-restoring partition [23]). *Let X be a random variable, let $\mathcal{X}$ denote the support of X, and let $\delta_X > 0$. Then, there exists a partition*

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

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

- *$X_{I_j} | X \in \mathcal{X}^j$ is fixed to x_j;*
- *$X_{[n]-I_j} | X \in \mathcal{X}^j$ is δ_X-dense.*

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

$$H_\infty(X_{[n]-I_j} | X \in \mathcal{X}^j) \geq H_\infty(X) - \delta_X \cdot b \cdot |I_j| - \log \frac{1}{p_{\geq j}}.$$

We turn to discuss the "conditioning issue" that was discussed in section 1.3 and its resolution: as mentioned above, the simulation uses Proposition 3.6 and Lemma 3.7 to restore the density of the inputs by conditioning some of the blocks. Specifically, suppose, for example, that $X_{\text{free}(\rho)}$ is no longer dense. Then, the simulation chooses appropriate $I \subseteq \text{free}(\rho)$ and $x_I \in \Lambda^I$, and conditions X on the event $X_I = x_I$. At this point, in order to make X and Y structured again, we need to remove I from $\text{free}(\rho)$, so the simulation queries the bits in z_I, and update the restriction ρ by setting $\rho_I = z_I$. Now, we have to make sure that $g(X_I, Y_I) = z_I$. To this end, the simulation conditions Y on the event $g(x_I, Y_I) = z_I$. However, the latter conditioning reveals information about Y, which may have two harmful effects:

- *Leaking:* As discussed in section 1.3, our analysis of the query complexity assumes that the protocol transmits at most C bits of information. It is important not to reveal more information than that, or otherwise our query complexity may increase arbitrarily. On average, we expect that conditioning on the event $g(x_I, Y_I) = z_I$ would reveal only $|I|$ bits of information, which is sufficiently small for our purposes. However, there could be values of x_I and z_I for which much more information is leaked. In this case, we say the conditioning is *leaking*.
- *Sparsifying:* Even if the conditioning reveals only $|I|$ bits of information on Y, this could still ruin the density of Y if the set I is large. In this case, we say that the conditioning is *sparsifying*.

This is the conditioning issue, and dealing with it is the technical core of the paper. As explained in section 1.3, the simulation deals with this issue by recognizing in advance which values of X are "dangerous," in the sense that they may lead to a bad conditioning, and discarding them before such conditioning may take place. The foregoing discussion leads to the following definition of a dangerous value.

DEFINITION 3.8. *Let Y be a random variable taking values from Λ^n. We say that a value $x \in \Lambda^n$ is* leaking *if there exists a set $I \subseteq [n]$ and an assignment $z_I \in \{0,1\}^I$ such that*

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

Let $\delta_Y, \varepsilon > 0$, and suppose that Y is δ_Y-dense. We say that a value $x \in \Lambda^n$ is ε-sparsifying *if there exists a set $I \subseteq [n]$ and an assignment $z_I \in \{0,1\}^I$ such that the random variable*

$$Y_{[n]-I} | g^I(x_I, Y_I) = z_I$$

is not $(\delta_Y - \varepsilon)$-dense. We say that a value $x \in \Lambda^n$ is ε-dangerous *if it is either leaking or ε-sparsifying.*

We can now state our main technical lemma, which says that X has only a small probability to take a dangerous value. This allows the simulation to discard such values and resolve the conditioning issue. In what follows, the parameter γ controls the error probability, and recall that η is the parameter that controls the discrepancy of g (i.e., $\mathrm{disc}(g) \le 2^{-\eta \cdot b}$).

LEMMA 3.9 (main lemma). *There exists a universal constant*[2] h *such that the following holds: let* $0 < \gamma, \varepsilon, \tau \le 1$ *be such that* $\tau \ge 2 + \frac{h}{c \cdot \varepsilon} - \eta + \gamma$ *and* $\varepsilon \ge \frac{4}{b}$*, and let* X, Y *be* (ρ, τ)*-structured random variables. Then, the probability that* $X_{\mathrm{free}(\rho)}$ *takes a value that is* ε*-dangerous for* $Y_{\mathrm{free}(\rho)}$ *is at most* $2^{-\gamma \cdot b}$.

3.1. Proof of the uniform marginals lemma. Recall that the random variables X and Y are (ρ, τ)*-structured* if there exist $\delta_X, \delta_Y > 0$ such that $X_{\mathrm{free}(\rho)}$ and $Y_{\mathrm{free}(\rho)}$ are δ_X-dense and δ_Y-dense, respectively,

$$\delta_X + \delta_Y \ge \tau, \quad \text{and} \quad g^{\mathrm{fix}(\rho)}\left(X_{\mathrm{fix}(\rho)}, Y_{\mathrm{fix}(\rho)}\right) = \rho_{\mathrm{fix}(\rho)}.$$

In this section we prove the uniform marginals lemma, restated next.

LEMMA 3.4. *There exists a universal constant* h *such that the following holds: let* $\gamma > 0$*, let* ρ *be a restriction, and let* $z \in \{0,1\}^n$ *be a string that is consistent with* ρ*. Let* X, Y *be independent random variables that are uniformly distributed over sets* $\mathcal{X}, \mathcal{Y} \subseteq \Lambda^n$*, respectively, and assume that they are* (ρ, τ)*-structured where*

$$\tau \ge 2 + \frac{h}{c} - \eta + \gamma.$$

Let (X', Y') *be uniformly distributed over* $G^{-1}(z) \cap (\mathcal{X} \times \mathcal{Y})$*. Then,* X *and* Y *are* $2^{-\gamma \cdot b}$*-close to* X' *and* Y'*, respectively.*

In order to prove Lemma 3.4, we first prove the following proposition, which says that the string $g^{\mathrm{free}(\rho)}(X_{\mathrm{free}(\rho)}, Y_{\mathrm{free}(\rho)})$ is close to the uniform distribution in a very strong sense. In what follows, the parameter γ controls the distance from the uniform distribution, and recall that η is the parameter that controls the discrepancy of g (i.e., $\mathrm{disc}(g) \le 2^{-\eta \cdot b}$).

PROPOSITION 3.10 (generalization of [20, Lemma 13]). *There exists a universal constant* h *such that the following holds: let* $\gamma > 0$*. Let* X, Y *be random variables that are* (ρ, τ)*-structured for* $\tau \ge 2 + \frac{h}{c} - \eta + \gamma$*, and let* $I \stackrel{\mathrm{def}}{=} \mathrm{free}(\rho)$*. Then, for every* $z_I \in \{0,1\}^I$ *it holds that*

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

Proof. Let $h \stackrel{\mathrm{def}}{=} 8$. We use Corollary 2.12 to upper bound the biases of $g^I(X_I, Y_I)$, and then apply Vazirani's lemma to show that it is close to the uniform distribution. Let $S \subseteq I$. By assumption, the variables X_I, Y_I are δ_X-dense and δ_Y-dense for some δ_X, δ_Y for which $\delta_X + \delta_Y \ge 2 + \frac{8}{c} - \eta + \gamma$. Therefore, it holds that

$$H_\infty(X_S) + H_\infty(Y_S) \ge \left(2 + \frac{6}{b} - \eta + \gamma + \frac{2}{c}\right) \cdot b \cdot |S|$$

and Corollary 2.12 implies (with $\gamma = \gamma + \frac{2}{c}$) that

$$\mathrm{bias}\left(g^{\oplus S}(X_S, Y_S)\right) \le 2^{-\left(\gamma + \frac{2}{c}\right) \cdot b \cdot |S|} \le 2^{-\gamma \cdot b} \cdot n^{-2 \cdot |S|} \le 2^{-\gamma \cdot b} \cdot (2 \cdot |I|)^{-|S|}.$$

[2] See Remark 3.5 for further explanation on the constant h.

Since the latter inequality holds for every $S \subseteq I$, it follows by Lemma 2.5 that

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

for every $z_I \in \{0,1\}^I$, as required. ∎

We turn to prove the uniform marginals lemma.

Proof of Lemma 3.4. Let h' be the universal constant of Proposition 3.10 and let $h \stackrel{\text{def}}{=} h' + 2$. Let (X', Y') be uniformly distributed over $G^{-1}(z) \cap (\mathcal{X} \times \mathcal{Y})$, and let $I \stackrel{\text{def}}{=} \text{free}(\rho)$. We prove that X is $2^{-\gamma \cdot b}$-close to X', and a similar argument works for Y. Let $\mathcal{E} \subseteq \mathcal{X}$ be any test event. We show that

$$|\Pr[X' \in \mathcal{E}] - \Pr[X \in \mathcal{E}]| \leq 2^{-\gamma \cdot b}.$$

Without loss of generality we may assume that $\Pr[X \in \mathcal{E}] \geq \frac{1}{2}$, since otherwise we can replace $\mathcal{E}$ with its complement. Since X and Y are (ρ, τ)-structured, where

$$\tau \geq 2 + \frac{h}{c} - \eta + \gamma \geq 2 + \frac{h'}{c} - \eta + \gamma + \frac{2}{b},$$

Proposition 3.10 implies that

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

Moreover, since $\Pr[X \in \mathcal{E}] \geq \frac{1}{2}$, conditioning on $\mathcal{E}$ cannot decrease the density of X by more than $\frac{1}{b}$ (since this conditioning increases any probability by a factor of at most 2). Therefore $X|\mathcal{E}$ and Y together are $(\rho, \tau - \frac{1}{b})$-structured, where

$$\tau - \frac{1}{b} \geq 2 + \frac{h'}{c} - \eta + \gamma + \frac{1}{b}.$$

Hence, Proposition 3.10 implies that

$$\Pr\left[g^I(X_I, Y_I) = z_I | X \in \mathcal{E}\right] \in (1 \pm 2^{-\gamma \cdot b - 1}) \cdot 2^{-|I|}.$$

Now, it holds that

$$\begin{aligned}
\Pr[X' \in \mathcal{E}] &= \Pr[X \in \mathcal{E} | G(X,Y) = z] \\
\text{(Bayes' formula)} &= \frac{\Pr[G(X,Y) = z | X \in \mathcal{E}]}{\Pr[G(X,Y) = z]} \cdot \Pr[X \in \mathcal{E}] \\
\text{(since } g^{\text{fix}(\rho)}(X_{\text{fix}(\rho)}, Y_{\text{fix}(\rho)}) = z_{\text{fix}(\rho)} \text{ by assumption)} \\
&= \frac{\Pr\left[g^I(X_I, Y_I) = z_I | X \in \mathcal{E}\right]}{\Pr\left[g^I(X_I, Y_I) = z_I\right]} \\
&\quad \cdot \Pr[X \in \mathcal{E}] \\
&\leq \frac{\left(1 + 2^{-\gamma \cdot b - 2}\right) \cdot 2^{-|I|}}{\left(1 - 2^{-\gamma \cdot b - 1}\right) \cdot 2^{-|I|}} \cdot \Pr[X \in \mathcal{E}] \\
\left(\text{since } \frac{1}{1-\alpha} \leq 1 + 2\alpha \text{ for every } 0 < \alpha \leq \frac{1}{2}\right) &\leq \left(1 + 2^{-\gamma \cdot b - 2}\right) \cdot \left(1 + 2 \cdot 2^{-\gamma \cdot b - 1}\right) \\
&\quad \cdot \Pr[X \in \mathcal{E}] \\
&\leq \Pr[X \in \mathcal{E}] + 2^{-\gamma \cdot b}.
\end{aligned}$$

A similar calculation shows that

$$\begin{aligned}\Pr[X' \in \mathcal{E}] &\geq \frac{\left(1 - 2^{-\gamma \cdot b - 2}\right)}{\left(1 + 2^{-\gamma \cdot b - 1}\right)} \cdot \Pr[X \in \mathcal{E}] \\ &\geq \Pr[X \in \mathcal{E}] - 2^{-\gamma \cdot b}.\end{aligned}$$

It follows that

$$|\Pr[X' \in \mathcal{E}] - \Pr[X \in \mathcal{E}]| \leq 2^{-\gamma \cdot b},$$

as required. ∎

3.2. Proof of the main technical lemma. In this section we prove our main technical lemma, which upper bounds the probability of a variable to take a dangerous value. We first recall the definition of a dangerous value and the lemma.

DEFINITION 3.8. *Let Y be a random variable taking values from Λ^n. We say that a value $x \in \Lambda^n$ is leaking if there exists a set $I \subseteq [n]$ and an assignment $z_I \in \{0,1\}^I$ such that*

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

Let $\delta_Y, \varepsilon > 0$, and suppose that Y is δ_Y-dense. We say that a value $x \in \Lambda^n$ is ε-sparsifying if there exists a set $I \subseteq [n]$ and an assignment $z_I \in \{0,1\}^I$ such that the random variable

$$Y_{[n]-I} | g^I(x_I, Y_I) = z_I$$

is not $(\delta_Y - \varepsilon)$-dense. We say that a value $x \in \Lambda^n$ is ε-dangerous if it is either leaking or ε-sparsifying.

LEMMA 3.9. *There exists a universal constant h such that the following holds: let $0 < \gamma, \varepsilon, \tau \leq 1$ be such that $\tau \geq 2 + \frac{h}{c \cdot \varepsilon} - \eta + \gamma$ and $\varepsilon \geq \frac{4}{b}$, and let X, Y be (ρ, τ)-structured random variables. Then, the probability that $X_{\mathrm{free}(\rho)}$ takes a value that is ε-dangerous for $Y_{\mathrm{free}(\rho)}$ is at most $2^{-\gamma \cdot b}$.*

Let h be a universal constant that will be chosen to be sufficiently large to make the inequalities in the proof hold. Let $\gamma, \varepsilon, \tau, \rho$ be as in the lemma, and assume that X, Y are (ρ, τ)-structured. For simplicity of the presentation, we assume that all the coordinates of ρ are free—this can be assumed without loss of generality since the fixed coordinates of ρ do not play any part in the lemma. Thus, our goal is to prove an upper bound on the probability that X takes a value that is dangerous for Y. By assumption, there exist some parameters $\delta_X, \delta_Y > 0$ such that X and Y are δ_X-dense and δ_Y-dense, respectively, and such that $\delta_X + \delta_Y \geq 2 + \frac{h}{c \cdot \varepsilon} - \eta - \gamma$.

We start by discussing the high-level ideas that underlie the proof. We would like to prove an upper bound on the probability that X takes a value that is either leaking or sparsifying. Proving the upper bound for leaking values is relatively easy and is similar to the proof of Proposition 3.10; basically, since X_I and Y_I are sufficiently dense, the string $g^I(X_I, Y_I)$ is multiplicatively close to uniform, which implies that most values x_I are nonleaking.

The more difficult task is to prove the upper bound for sparsifying values. Basically, a value x is sparsifying if for some disjoint $I, J \subseteq \mathrm{free}(\rho)$, conditioning on the value of $g^I(x_I, Y_I)$ decreases the min-entropy of Y_J by more than $\varepsilon \cdot b \cdot |J|$ bits. Our first step is to apply Bayes' formula to the latter condition, thus obtaining a more convenient condition which we refer to as "skewing": a value x is skewing if conditioning on the value of Y_J decreases the min-entropy of $g^I(x_I, Y_I)$ by more than

$\varepsilon \cdot b \cdot |J|$ bits—in other words, the min-entropy of $g^I(x_I, Y_I)$ conditioned on Y_J should be less than $|I| - \varepsilon \cdot b \cdot |J|$ (roughly).

It remains to prove an upper bound on the probability that X takes a skewing value. This requires proving a lower bound of roughly $|I| - \varepsilon \cdot b \cdot |J|$ of on the min-entropy of $g^I(x_I, Y_I)|Y_J$ for most x's. By the min-entropy version of Vazirani's lemma (Lemma 2.6), in order to prove this lower bound, it suffices to prove an upper bound on the bias of $g^S(x_S, Y_S)|Y_J$ for every set $S \subseteq I$ for which[3] $|S| \gtrapprox \varepsilon \cdot c \cdot |J|$.

To this end, we use the extractor-like property of g^S: recall that by the discrepancy of g (Corollary 2.13), the bias of $g^S(x_S, Y_S)|Y_J$ is small for most x's whenever the min-entropy of X_S and $Y_S|Y_J$ is high. Furthermore, recall that the min-entropy of X_S and Y_S is high since we assumed that X and Y are dense. The key step is to observe that the min-entropy of $Y_S|Y_J$ is still high, since S is large compared to J. Thus, the min-entropy of X_S and $Y_S|Y_J$ is high, so the bias of $g^S(x_S, Y_S)|Y_J$ is small, and this implies the desired lower on the min-entropy of $g^I(x_I, Y_I)|Y_J$.

The argument we explained above almost works, except for a small issue: we said that $H_\infty(Y_S|Y_J)$ is still high, since S is large compared to J. Here, we implicitly assumed that conditioning on the value of Y_J decreases the min-entropy of Y_S by roughly $|J| \cdot b$ bits. This assumption is true for the average value of Y_J, but may fail for values of Y_J that have a very small probability. In order to deal with such values, we define a parameter e_{y_J} which measures the "excess entropy" of y_J, and keep track of it throughout the proof. The key observation is that if we consider a value y_J that has a small probability, then the criterion of skewing actually requires the min-entropy of $g^I(x_I, Y_I)$ to decrease by roughly $\varepsilon \cdot b \cdot |J| + e_{y_J}$. Intuitively, this means that the smaller the probability of y_J, the harder it becomes for x to be skewing. After propagating the additional term of e_{y_J} throughout our proof, we get that the set S can be assumed to satisfy

$$|S| \gtrapprox \varepsilon \cdot c \cdot |J| + \frac{e_{y_J}}{\log n}.$$

This makes the set S sufficiently large compared to e_{y_J} that we can still deduce that $Y_S|Y_J = y_J$ has high min-entropy, which finished the argument. We now turn to provide the formal proof, starting with a formal definition of the parameter e_{y_J} and the criterion of skewing.

DEFINITION 3.11. *Recall that since Y is δ_Y-dense, it holds that $\Pr[Y_J = y_J] \leq 2^{-\delta_Y \cdot b \cdot |J|}$ for every $J \subseteq [n]$ and $y_J \in \Lambda^J$. We denote by $e_{y_J} \in \mathbb{R}$ the (nonnegative) number that satisfies*

$$\Pr[Y_J = y_J] = 2^{-\delta_Y \cdot b \cdot |J| - e_{y_J}}.$$

We say that a value $x \in \Lambda^n$ is ε-skewing if there exist disjoint nonempty sets $I, J \subseteq [n]$ and a value $y_J \in \Lambda^J$ such that

$$H_\infty\left(g^I(x_I, Y_I) \middle| Y_J = y_J\right) < |I| - \varepsilon \cdot b \cdot |J| - e_{y_J} + 1.$$

Next, we show that every dangerous value must be either leaking or skewing by applying Bayes' formula.

CLAIM 3.12. *Let $x \in \Lambda^n$ be an ε-dangerous value that is not leaking for Y. Then x is ε-skewing.*

Proof. Suppose that x is ε-dangerous for Y and that it is not leaking. We prove that x is ε-skewing. By our assumption, x must be ε-sparsifying, so there exists a set

[3] Recall that c is a large constant such that $b \geq c \cdot \log n$.

$I \subseteq [n]$ and an assignment $z_I \in \{0,1\}^I$ such that the random variable

$$Y_{[n]-I} | g^I(x_I, Y_I) = z_I$$

is not $(\delta_Y - \varepsilon)$-dense. Thus, there exists a set $J \subseteq [n] - I$ and a value $y_J \in \Lambda^n$ such that

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

By Bayes' formula, it holds that

$$\begin{aligned}
\Pr\left[Y_J = y_J | g(x_I, Y_I) = z_I\right] &= \frac{\Pr\left[g^I(x_I, Y_I) = z_I | Y_J = y_J\right] \cdot \Pr\left[Y_J = y_J\right]}{\Pr\left[g(x_I, Y_I) = z_I\right]} \\
\text{(definition of } e_{y_J}) &= \frac{\Pr\left[g^I(x_I, Y_I) = z_I | Y_J = y_J\right] \cdot 2^{-\delta_Y \cdot b \cdot |J| - e_{y_J}}.}{\Pr\left[g(x_I, Y_I) = z_I\right]} \\
\text{(since } x \text{ is not leaking for } Y) &\leq \frac{\Pr\left[g^I(x_I, Y_I) = z_I | Y_J = y_J\right] \cdot 2^{-\delta_Y \cdot b \cdot |J| - e_{y_J}}.}{2^{-|I|-1}}.
\end{aligned}$$

Hence, it follows that

$$\frac{\Pr\left[g^I(x_I, Y_I) = z_I | Y_J = y_J\right] \cdot 2^{-\delta_Y \cdot b \cdot |J| - e_{y_J}}}{2^{-|I|-1}} > 2^{-(\delta_Y - \varepsilon)\cdot b \cdot |J|},$$

which implies that

$$\Pr\left[g^I(x_I, Y_I) = z_I | Y_J = y_J\right] > 2^{-|I| + \varepsilon \cdot b \cdot |J| + e_{y_J} - 1}.$$

This means that

$$H_\infty\left(g^I(x_I, Y_I) \middle| Y_J = y_J\right) < |I| - \varepsilon \cdot b \cdot |J| - e_{y_J} + 1.$$

That is, x is ε-skewing, as required. ∎

As explained above, we will upper bound the probability of dangerous values by upper bounding the biases of $g(x_S, Y_S)|Y_J$ for every $S \subseteq I$. To this end, it is convenient to define the notion of a "biasing value," which is a value x for which one of the biases is too large.

DEFINITION 3.13. *We say that a value $x \in \Lambda^n$ is* biasing (for Y) with respect to *disjoint sets $S, J \subseteq [n]$ and an assignment $y_J \in \Lambda^J$ if*

$$\text{bias}\left(g^{\oplus S}(x_S, Y_S) | Y_J = y_J\right) > \frac{1}{2} \cdot (2n)^{-|S|}.$$

We say that x is ε-biasing (for Y) with respect to *a set $S \subseteq [n]$ if there exists a set $J \subseteq [n] - S$ and an assignment $y_J \in \Lambda^J$ that satisfy*

$$|S| \geq c \cdot \varepsilon \cdot |J| + \frac{e_{y_J} + 2}{\log n} \tag{4}$$

such that x is biasing with respect to S, J, and y_J (if J is the empty set, we define $e_{y_J} = 0$). Finally, we say that x is ε-biasing *(for Y) if there exists a nonempty set S with respect to which x is ε-biasing.*

We now apply the min-entropy version of Vazirani's lemma to show that values that are not biasing are not dangerous.

CLAIM 3.14. *If a value $x \in \Lambda^n$ is not ε-biasing for Y then it is not ε-dangerous for Y.*

Proof. Suppose that $x \in \Lambda^n$ is a value that is not ε-biasing for Y. We prove that x is not ε-dangerous for Y. We start by proving that x is not leaking. Let $I \subseteq [n]$ and let $z_I \in \{0,1\}^I$. We wish to prove that

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

Observe that, by the assumption that x is not ε-biasing, it holds for every nonempty set $S \subseteq I$ that

$$\text{bias}\left(g^{\oplus S}(x_S, Y_S)\right) \leq \frac{1}{2} \cdot (2n)^{-|S|}$$

(this follows by substituting $J = \emptyset$ in the definition of ε-biasing and noting that in this case $e_{y_J} = 0$). It now follows from Lemma 2.6 that $\Pr\left[g^I(x_I, Y_I) = z_I\right] \geq 2^{-|I|-1}$, as required.

We turn to prove that x is not ε-skewing. Let $I, J \subseteq [n]$ be disjoint sets and let $y_J \in \Lambda^J$ be an assignment. We wish to prove that

$$H_\infty\left(g^I(x_I, Y_I)\middle| Y_J = y_J\right) \geq |I| - \varepsilon \cdot b \cdot |J| - e_{y_J} + 1.$$

By Lemma 2.6, it suffices to prove that for every set $S \subseteq I$ such that $|S| \geq \frac{\varepsilon \cdot b \cdot |J| + e_{y_J} + 2}{\log n}$ it holds that

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

To this end, observe that every such set S satisfies

$$|S| \geq \varepsilon \cdot c \cdot |J| + \frac{e_{y_J} + 2}{\log n},$$

and since by assumption x is not ε-biasing with respect to S, the required upper bound on the bias must hold. It follows that x is neither leaking nor ε-skewing, and therefore it is not ε-dangerous, as required. ∎

Finally, we prove an upper bound on the probability of X to take an ε-biasing value, which together with Claim 3.14 implies Lemma 3.9. As explained above, the idea is to combine the discrepancy of g with the observation that X_S and Y_S have large min-entropy even conditioned on $Y_J = y_J$ (which holds since X, Y are dense and S is large compared to $|J|$ and e_{y_J}).

PROPOSITION 3.15. *The probability that X takes a value x that is ε-biasing for Y is at most $2^{-\gamma \cdot b}$.*

Proof. We begin with upper bounding the probability of X to take a value that is ε-biasing *with respect to specific choices of S, J, and y_J*, and the rest of the proof will follow by applying union bounds over all possible choices of S, J, and y_J. Let $S, J \subseteq [n]$ be disjoint sets and let $y_J \in \Lambda^J$ be an assignment such that S, J, and y_J together satisfy (4), i.e.,

$$|S| \geq c \cdot \varepsilon \cdot |J| + \frac{e_{y_J} + 2}{\log n}.$$

For simplicity, we assume that J is nonempty (in the case where J is empty, the argument is similar but simpler). Since we assumed that $\varepsilon \geq \frac{4}{b}$ and that J is nonempty, and it holds that $\frac{1}{2} \cdot c \cdot \varepsilon \cdot |J| \geq \frac{2}{\log n}$ and therefore

$$|S| \geq \frac{1}{2} \cdot c \cdot \varepsilon \cdot |J| + \frac{e_{y_J}}{\log n}.$$

In other words, it holds that

$$|S| \cdot \log n \geq \frac{1}{2} \cdot \varepsilon \cdot b \cdot |J| + e_{y_J}. \tag{5}$$

By assumption, Y is δ_Y-dense, so $H_\infty(Y_S) \geq \delta_Y \cdot b \cdot |S|$. By Fact 2.2, it follows that

$$\begin{aligned}
H_\infty(Y_S|Y_J = y_J) &\geq \delta_Y \cdot b \cdot |S| - \log \frac{1}{\Pr[Y_J = y_J]} \\
&= \delta_Y \cdot b \cdot |S| - (\delta_Y \cdot b \cdot |J| + e_{y_J}) \\
&\geq \delta_Y \cdot b \cdot |S| - (b \cdot |J| + e_{y_J}) \\
&\geq \delta_Y \cdot b \cdot |S| - \frac{2}{\varepsilon} \cdot \left(\frac{1}{2} \cdot \varepsilon \cdot b \cdot |J| + e_{y_J} \right) \\
(\text{by (4)}) &\geq \delta_Y \cdot b \cdot |S| - \frac{2}{\varepsilon} \cdot |S| \cdot \log n \\
(\text{since } b \geq c \cdot \log n)) &\geq \left(\delta_Y - \frac{2}{c \cdot \varepsilon} \right) \cdot b \cdot |S|.
\end{aligned}$$

Moreover, X is δ_X-dense and thus

$$H_\infty(X_S) + H_\infty(Y_S|Y_J = y_J) \geq \left(\delta_X + \delta_Y - \frac{2}{c \cdot \varepsilon} \right) \cdot b \cdot |S| \geq \left(2 + \frac{8}{c \cdot \varepsilon} - \eta + \gamma \right) \cdot b \cdot |S|,$$

where the second inequality is made to hold by choosing h to be sufficiently large. It follows by Corollary 2.13 (with $\lambda = \frac{3}{c \cdot \varepsilon}$ and $\gamma = \gamma + \frac{5}{c \cdot \varepsilon}$) that the probability that X_S takes a value $x_S \in \Lambda^S$ for which

$$\text{bias}\left(g^{\oplus S}(\alpha, Y_S) | Y_J = y_J \right) > \frac{1}{2} \cdot (2n)^{-|S|} \geq 2^{-3 \log n \cdot |S|} \geq 2^{-\frac{3}{c \cdot \varepsilon} \cdot b \cdot |S|} \tag{6}$$

is at most $2^{-\left(\gamma + \frac{5}{c \cdot \varepsilon}\right) \cdot b \cdot |S|}$.

We turn to applying the union bounds. First, we show that for every $S \subseteq [n]$, the probability that X takes a value that is ε-biasing with respect to S is at most $2^{-(\gamma + \frac{2}{c \cdot \varepsilon}) \cdot b \cdot |S|}$ by taking the upper bound over all choices of J and y_J. Note that we only need to consider sets $J \subseteq [n]$ for which $|J| \leq \frac{1}{c \cdot \varepsilon} \cdot |S|$. It follows that the probability that X_S takes a value that satisfies (6) for some J and y_J is at most

$$\begin{aligned}
&\sum_{J \subseteq [n] : |J| \leq \frac{1}{c \cdot \varepsilon} \cdot |S|} \sum_{y_J \in \Lambda^J} 2^{-(\gamma + \frac{5}{c \cdot \varepsilon}) \cdot b \cdot |S|} \\
&\leq \sum_{j=1}^{\frac{1}{c \cdot \varepsilon} \cdot |S|} \binom{n}{j} \cdot 2^{b \cdot j} \cdot 2^{-(\gamma + \frac{5}{c \cdot \varepsilon}) \cdot b \cdot |S| + 1} \\
&\leq n \cdot \binom{n}{\frac{1}{c \cdot \varepsilon} \cdot |S|} \cdot 2^{\frac{1}{c \cdot \varepsilon} \cdot |S| \cdot b} \cdot 2^{-(\gamma + \frac{5}{c \cdot \varepsilon}) \cdot b \cdot |S|} \\
&\leq 2^{-\left(\gamma + \frac{5}{c \cdot \varepsilon}\right) \cdot b \cdot |S| + \frac{1}{c \cdot \varepsilon} \cdot |S| \cdot (\log n + b) + \log n} \\
&\leq 2^{-(\gamma + \frac{2}{c \cdot \varepsilon}) \cdot b \cdot |S|},
\end{aligned}$$

where the last inequality follows since

$$\frac{1}{c \cdot \varepsilon} \cdot |S| \cdot (\log n + b) + \log n \leq \frac{1}{c \cdot \varepsilon} \cdot 2b \cdot |S| + b \leq \frac{3}{c \cdot \varepsilon} \cdot b \cdot |S|.$$

The above calculation showed that the probability that X takes a value that is ε-biasing with respect to a fixed set S is at most $2^{-(\gamma+\frac{2}{c\cdot\varepsilon})\cdot b\cdot|S|}$. Taking a union bound over all nonempty sets $S \subseteq [n]$, the probability that X takes a value that is ε-biasing for Y is at most

$$\begin{aligned}
\sum_{\emptyset\neq S\subseteq[n]} 2^{-(\gamma+\frac{2}{c\cdot\varepsilon})\cdot b\cdot|S|} &\leq \sum_{s=1}^{n} \binom{n}{s} \cdot 2^{-(\gamma+\frac{2}{c\cdot\varepsilon})\cdot b\cdot s} \\
&\leq \sum_{s=1}^{n} 2^{-(\gamma+\frac{2}{c\cdot\varepsilon})\cdot b\cdot s - s\cdot\log n} \\
\left(\frac{b}{c\cdot\varepsilon} \geq \log n\right) &\leq \sum_{s=1}^{n} 2^{-(\gamma+\frac{1}{c\cdot\varepsilon})\cdot b\cdot s} \\
&\leq 2^{-(\gamma+\frac{1}{c\cdot\varepsilon})\cdot b+1} \\
&\leq 2^{-\gamma\cdot b}.
\end{aligned}$$

We have thus shown that the probability that X takes a value that is ε-biasing is at most $2^{-\gamma\cdot b}$, as required. ∎

4. The deterministic lifting theorem. In this section, we prove the deterministic part of our main theorem. In fact, we prove the following more general result.

THEOREM 4.1 (deterministic lifting theorem). *For every $\eta > 0$ there exists $c = O(\frac{1}{\eta^2})$ such that the following holds: let $n \in \mathbb{N}$ be such that $n \geq 2$, let $\Lambda \stackrel{\text{def}}{=} \{0,1\}^b$ be such that $b \geq c \cdot \log n$, let $g : \Lambda \times \Lambda \to \{0,1\}$ be a function such that $\mathrm{disc}(g) \leq 2^{-\eta\cdot b}$, and let $G = g^n$. Let Π be a deterministic protocol that takes inputs in $\Lambda^n \times \Lambda^n$ and that has communication complexity C and round complexity r. Then, there exists a deterministic parallel decision tree T that on input $z \in \{0,1\}^n$ outputs a transcript π of Π that is consistent with some pair of inputs $(x, y) \in G^{-1}(z)$, and that has query complexity $O(\frac{C}{b})$ and depth r.*

Observe that this theorem implies the lower bound of the main theorem: given a protocol Π that solves $\mathcal{S} \circ G$ with complexity C, we use the theorem to construct a tree T that on input z outputs the output of Π on some pair of inputs in $G^{-1}(z)$. This tree T clearly solves $\mathcal{S}$, and the query complexity of T is $O(\frac{C}{b})$. This implies that $D^{\mathrm{dt}}(\mathcal{S}) = O\left(D^{\mathrm{cc}}(\mathcal{S} \circ G)/b\right)$ or, in other words, $D^{\mathrm{cc}}(\mathcal{S} \circ G) = \Omega\left(D^{\mathrm{dt}}(\mathcal{S}) \cdot b\right)$, as required.

For the rest of this section, fix Π to be an arbitrary deterministic protocol that takes inputs in $\Lambda^n \times \Lambda^n$, and denote by C and r its communication complexity and round complexity, respectively. The rest of this section is organized as follows: we first describe the construction of the parallel decision tree T in section 4.1. We then prove that the output of T is always correct in section 4.2. Finally, we upper bound the query complexity of T in section 4.3.

4.1. The construction of *T*. Let h' be the maximum among the universal constants of Proposition 3.10 and the main technical lemma (Lemma 3.9), and let h be a universal constant that will be chosen to be sufficiently large to make the inequalities in the proof hold. Let $\varepsilon \stackrel{\text{def}}{=} \frac{h}{c\cdot\eta}$, let $\delta \stackrel{\text{def}}{=} 1 - \frac{\eta}{4} + \frac{\varepsilon}{2}$, and let $\tau \stackrel{\text{def}}{=} 2 \cdot \delta - \varepsilon$. The tree T constructs the transcript π by simulating the protocol Π round-by-round, each time adding a single message to π. Throughout the simulation, the tree maintains

a rectangle $\mathcal{X} \times \mathcal{Y} \subseteq \Lambda^n \times \Lambda^n$ of inputs that are consistent with π (but not necessarily of all such inputs). In what follows, we denote by X and Y random variables that are uniformly distributed over $\mathcal{X}$ and $\mathcal{Y}$, respectively. The tree will maintain the invariant that X and Y are (ρ, τ)-structured, where ρ is a restriction that keeps track of the queries the tree has made so far. In fact, the tree will maintain a more specific invariant: whenever it is Alice's turn to speak, $X_{\text{free}(\rho)}$ is $(\delta - \varepsilon)$-dense and $Y_{\text{free}(\rho)}$ is δ-dense, and whenever it is Bob's turn to speak, the roles of X and Y are reversed.

When the tree T starts the simulation, the tree sets the transcript π to be the empty string, the restriction ρ to $\{*\}^n$, and the sets $\mathcal{X}, \mathcal{Y}$ to Λ^n. At this point the invariant clearly holds. We now explain how T simulates a single round of the protocol while maintaining the invariant. Suppose that the invariant holds at the beginning of the current round, and assume without loss of generality that it is Alice's turn to speak. The tree T performs the following steps:

1. The tree conditions $X_{\text{free}(\rho)}$ on not taking a value that is ε-dangerous for $Y_{\text{free}(\rho)}$ (i.e., the tree removes from $\mathcal{X}$ all the values x for which $x_{\text{free}(\rho)}$ is ε-dangerous for $Y_{\text{free}(\rho)}$).
2. The tree T chooses an arbitrary message M of Alice with the following property: the probability of Alice sending M on input X is at least $2^{-|M|}$ (the existence of M will be justified soon). The tree adds M to the transcript π, and conditions X on the event of sending M (i.e., the tree sets $\mathcal{X}$ to be the subset of inputs that are consistent with M).
3. Let $I \subseteq \text{free}(\rho)$ be a maximal set that violates the δ-density of $X_{\text{free}(\rho)}$ (i.e., $H_\infty(X_I) < \delta \cdot b \cdot |I|$), and let $x_I \in \Lambda^I$ be a value that satisfies $\Pr[X_I = x_I] > 2^{-\delta \cdot b \cdot |I|}$. The tree conditions X on $X_I = x_I$ (i.e., the tree removes from $\mathcal{X}$ all the values that are inconsistent with that event). By Proposition 3.6, $X_{\text{free}(\rho)-I}$ is now δ-dense.
4. The tree queries z_I, and updates ρ accordingly.
5. The tree conditions Y on $g^I(x_I, Y_I) = \rho_I$ (i.e., the tree sets $\mathcal{Y}$ to be the subset of values y for which $g^I(x_I, y_I) = \rho_I$). Due to step 1, the variable $X_{\text{free}(\rho)}$ must take a value that is not ε-dangerous and, therefore, $Y_{\text{free}(\rho)}$ is necessarily $(\delta - \varepsilon)$-dense.

After those steps take place, it becomes Bob's turn to speak and, indeed, $X_{\text{free}(\rho)}$ and $Y_{\text{free}(\rho)}$ are δ-dense and $(\delta - \varepsilon)$-dense, respectively. Thus, the invariant is maintained. When the protocol Π stops, the tree T outputs the transcript π and halts. In order for the foregoing construction to be well-defined, it remains to explain three points:

- First, we should explain why the set $\mathcal{X}$ remains nonempty after step 1 (otherwise, the following steps are not well-defined). To this end, recall that X and Y are (ρ, τ)-structured and observe that τ can be made larger than $2 + \frac{h'}{c \cdot \varepsilon} - \eta$ by choosing h to be sufficiently large (see section 4.3 for a detailed calculation). Hence, by our main lemma (Lemma 3.9), the variable $X_{\text{free}(\rho)}$ has a nonzero probability to take a value that is not ε-dangerous for $Y_{\text{free}(\rho)}$, so $\mathcal{X}$ is nonempty after this step.
- Second, we should explain why the message M in step 2 exists. To see why, observe that the set of possible messages of Alice forms a prefix-free code—otherwise, Bob will not be able to tell when Alice finished speaking and his turn starts. Hence, by Fact 2.8, it follows that there exists a message M with probability at least $2^{-|M|}$.

- Third, we should explain why the set $\mathcal{Y}$ remains nonempty after step 5. To this end, recall that X must take a value that is not ε-dangerous for Y and, in particular, the value of X is necessarily not leaking. This means that in particular that the string $g^I(x_I, Y_I)$ has nonzero probability to be equal to ρ_I, so $\mathcal{Y}$ is nonempty after this step.

The depth of T. We now observe that the depth of T is equal to the round complexity of Π. Note that in each round, the tree T issues a set of queries I simultaneously. Thus, T is a parallel decision tree whose depth equals the maximal number of rounds of Π, as required.

4.2. The correctness of T. We now prove that when the decision tree T halts, the transcript π is consistent with some inputs $(x, y) \in G^{-1}(z)$. Clearly, the transcript π is consistent with all the inputs in the rectangle $\mathcal{X} \times \mathcal{Y}$. Thus, it suffices to show that there exist $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ such that $G(x, y) = z$. To this end, recall that when the tree halts, the random variables X and Y are (ρ, τ)-structured. Since ρ is consistent with z, it holds for every $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ that

$$g^{\mathrm{fix}(\rho)}(x_{\mathrm{fix}(\rho)}, y_{\mathrm{fix}(\rho)}) = z_{\mathrm{fix}(\rho)}. \tag{7}$$

It remains to deal with the free coordinates of ρ. Since τ can be made larger than $2 + \frac{h'}{c} - \eta$ by choosing h to be sufficiently large (see section 4.3 for a detailed calculation), it follows by Proposition 3.10 that

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

In particular, there exist $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ such that

$$g^{\mathrm{free}(\rho)}(x_{\mathrm{free}(\rho)}, y_{\mathrm{free}(\rho)}) = z_{\mathrm{free}(\rho)}. \tag{8}$$

By combining (7) and (8), we get that $G(x, y) = z$, as required.

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

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

We will prove that whenever the protocol transmits a message M, the deficiency increases by $O(|M|)$, and that whenever the tree T makes a query, the deficiency is decreased by $\Omega(b)$. Since the deficiency is always nonnegative, and the protocol transmits at most C bits, it will follow that the tree must make at most $O(\frac{C}{b})$ queries. More specifically, we prove that in every round, the first two steps from section 4.1 increase the deficiency by at most $|M| + 1$ in total, and the rest of the steps decrease the deficiency by at least $\Omega(|I| \cdot b)$, and this will imply the desired result.

Fix a round of the simulation, and assume without loss of generality that the message is sent by Alice. We start by analyzing step 1. At this step, the tree conditions $X_{\mathrm{free}(\rho)}$ on taking dangerous values that are not ε-dangerous for $Y_{\mathrm{free}(\rho)}$. We show that this step increases the deficiency by at most one bit. By applying our main technical lemma (Lemma 3.9) with $\gamma = \frac{1}{b}$, it follows that the probability that $X_{\mathrm{free}(\rho)}$ is ε-dangerous is at most $\frac{1}{2}$. By Fact 2.2, it follows that conditioning on nondangerous values decreases $H_\infty(X_{\mathrm{free}(\rho)})$ by at most one bit and, therefore, it increases the deficiency by at most one bit. To see why we can apply the main lemma with $\gamma = \frac{1}{b}$,

recall that at this point X and Y are (ρ,τ)-structured, where

$$\begin{aligned}\tau &\stackrel{\text{def}}{=} 2\cdot\delta-\varepsilon\\ (\text{by definition of }\delta) &= 2\cdot\left(1-\frac{\eta}{4}+\frac{\varepsilon}{2}\right)-\varepsilon\\ &= 2-\frac{\eta}{2}\\ &= 2+\frac{\eta}{2}-\eta\\ \left(\text{since }\varepsilon\stackrel{\text{def}}{=}\frac{h}{c\cdot\eta}\right) &= 2+\frac{h}{2\cdot c\cdot\varepsilon}-\eta\\ &\geq 2+\frac{h'}{c\cdot\varepsilon}-\eta+\frac{1}{b},\end{aligned}$$

where the last inequality can be made to hold by choosing h to be sufficiently large.

Next, in step 2, the tree conditions X on the event of sending the message M, which has probability at least $2^{-|M|}$. By Fact 2.2, this decreases $H_\infty(X_{\text{free}(\rho)})$ by at most $|M|$ bits, which increases the deficiency by at most $|M|$ bits. All in all, we showed that the first two steps of the simulation increase the deficiency by at most $|M|+1$.

Let I be the set of queries chosen in step 3. We turn to show that the rest of the steps decrease the deficiency by at least $\Omega(b\cdot|I|)$. Without loss of generality, assume that $I\neq\emptyset$ (otherwise the latter bound holds vacuously). The rest of the steps apply the following changes to the deficiency:

- Step 3 conditions X on the event $X_I=x_I$, which has probability greater than $2^{-\delta\cdot b\cdot|I|}$ by the definition of x_I. Hence, this conditioning increases the deficiency by less than $\delta\cdot b\cdot|I|$ (by Fact 2.3).
- Step 4 removes the set I from free(ρ). Looking at the definition of deficiency, this change decreases the term of $2\cdot b\cdot|\text{free}(\rho)|$ by $2\cdot b\cdot|I|$, decreases the term $H_\infty(Y_{\text{free}(\rho)})$ by at most $b\cdot|I|$ (by Fact 2.3), and does not change the term $H_\infty(X_{\text{free}(\rho)})$ (since at this point X_I is fixed to x_I). All in all, the deficiency is decreased by at least $b\cdot|I|$.
- Finally, step 5 conditions Y on the event $g^I(x_I,Y_I)=\rho_I$. This event has probability at least $2^{-|I|-1}$ by the assumption that X is not dangerous (and hence not leaking). Thus, this conditioning increases the deficiency by at most $|I|+1$ (by Fact 2.3).

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

$$b\cdot|I|-\delta\cdot b\cdot|I|-(|I|+1)\geq\left(1-\delta-\frac{2}{b}\right)\cdot b\cdot|I|\,.$$

By choosing c to be sufficiently large, we can make sure that $1-\delta-\frac{2}{b}$ is a positive constant independent of b and n, and therefore the decrease in the deficiency will be at least $\Omega(b\cdot|I|)$, as required. To see it, observe that

$$\begin{aligned}\delta+\frac{2}{b} &= 1-\frac{\eta}{4}+\frac{\varepsilon}{2}+\frac{2}{b}\\ \left(\text{since }\varepsilon\stackrel{\text{def}}{=}\frac{h}{c\cdot\eta}\right) &= 1-\frac{\eta}{4}+\frac{h}{2\cdot c\cdot\eta}+\frac{2}{b}\\ (\text{since } b\geq c) &\leq 1-\frac{\eta}{4}+\frac{h+4}{2\cdot c\cdot\eta}.\end{aligned}$$

Therefore, if we choose $c>\frac{2\cdot(h+4)}{\eta^2}$, the expression on the right-hand side will be a constant that is strictly smaller than 1, as required.

5. The randomized lifting theorem. In this section, we prove the randomized part of our main theorem. In fact, we prove the following more general result.

THEOREM 5.1 (randomized lifting theorem). *For every $\eta > 0$ there exists $c = O(\frac{1}{\eta^2} \cdot \log \frac{1}{\eta})$ such that the following holds: Let $n \in \mathbb{N}$ be such that $n \geq 2$, let $\Lambda \stackrel{\text{def}}{=} \{0,1\}^b$ be such that $b \geq c \cdot \log n$, let $g : \Lambda \times \Lambda \to \{0,1\}$ be a function such that $\mathrm{disc}(g) \leq 2^{-\eta \cdot b}$, and let $G = g^n$. Let Π be a randomized (public-coin) protocol that takes inputs in $\Lambda^n \times \Lambda^n$ that has communication complexity $C \leq 2 \cdot b \cdot n$ and round complexity r. Then, there exists a randomized parallel decision tree T with the following properties:*

- *On input $z \in \{0,1\}^n$, the tree outputs a transcript π of Π, whose distribution is $2^{-\frac{\eta}{8} \cdot b}$-close to the distribution of the transcripts of Π when given inputs that are uniformly distributed in $G^{-1}(z)$.*
- *The tree T has query complexity $O(\frac{C}{b} + 1)$ and depth r.*

We first observe that Theorem 5.1 indeed implies the lower bound of our main theorem.

Proof of Theorem 1.2 *from Theorem* 5.1. Let $\mathcal{S} : \{0,1\}^n \to \mathcal{O}$ be a search problem, and let $\varepsilon > 0$ and $\varepsilon' \stackrel{\text{def}}{=} \varepsilon + 2^{-\frac{\eta}{8} \cdot b}$. We prove that $R^{\mathrm{cc}}_{\varepsilon}(\mathcal{S} \circ G) = \Theta\left(R^{\mathrm{dt}}_{\varepsilon'}(\mathcal{S}) \cdot b\right)$. Let Π be an optimal protocol that solves $\mathcal{S} \circ G$ with complexity $C \stackrel{\text{def}}{=} R^{\mathrm{cc}}(\mathcal{S} \circ G)$, and observe that we can assume without loss of generality that $C \leq 2 \cdot b \cdot n$ (since the players can solve any search problem by sending their whole inputs). By applying the theorem to Π, we construct a tree T that on input z samples a transcript of Π as in the theorem, and outputs the output that is associated with this transcript. It is not hard to see that the output of T will be in $\mathcal{S}(z)$ with probability at least

$$1 - \varepsilon - 2^{-\frac{\eta}{8} \cdot b} \geq 1 - \varepsilon',$$

and that the query complexity of T is $O(\frac{C}{b} + 1)$. This implies that $R^{\mathrm{dt}}_{\varepsilon'}(\mathcal{S}) = O\left(R^{\mathrm{cc}}_{\varepsilon}(\mathcal{S} \circ G)/b + 1\right)$, or in other words, $R^{\mathrm{cc}}_{\varepsilon}(\mathcal{S} \circ G) = \Omega\left(\left(R^{\mathrm{dt}}_{\varepsilon'}(\mathcal{S}) - O(1)\right) \cdot b\right)$ as required. ∎

In the rest of this section we prove Theorem 5.1. We start the proof by observing that it suffices to prove the theorem for the special case in which the protocol Π is deterministic. To see why, recall that a randomized public-coin protocol is a distribution over deterministic protocols. Thus, if we prove the theorem for deterministic protocols, we can extend it to randomized protocols as follows: given a randomized protocol Π, the tree T will start by sampling a deterministic protocol Π_{det} from the distribution Π, and will then apply the theorem to Π_{det}. It is not hard to verify that such a tree T satisfies the requirements of Theorem 5.1. Thus, it suffices to consider the case where Π is deterministic.

For the rest of this section, fix Π to be an arbitrary deterministic protocol that takes inputs in $\Lambda^n \times \Lambda^n$, and denote by C and r its communication complexity and round complexity, respectively. The rest of this section is organized as follows: we first describe the construction of the parallel decision tree T in section 5.1. We then prove that the transcript that T outputs is distributed as required in section 5.2. Finally, we upper bound the query complexity of T in section 5.3.

5.1. The construction of *T*. The construction of the randomized tree T is similar to the construction of the deterministic lifting theorem (section 4.1), but has the following differences in the simulation:

- In the deterministic construction, the tree chose the message M arbitrarily subject to having sufficiently high probability. The reason we could do it is

that it did not matter which transcript the tree would output as long as it was consistent in $G^{-1}(z)$. In the randomized construction, on the other hand, we would like to output a transcript whose distribution is close to the "correct" distribution. Therefore, we change the construction such that the message M is chosen randomly according to the distribution of the inputs.

- Since the messages are now sampled according to the distribution of the inputs, we can no longer guarantee that the message M has sufficiently high probability. Therefore, the tree may choose messages M that have very low probability, and such messages may reveal too much information about the inputs. In order to avoid that, the tree maintains a variable K which keeps track of the amount of information that was revealed by the messages. If at any point K becomes too large, the tree halts and declares failure. This modification is important since if we allow the chosen messages to reveal too much information, then they will lead the tree to make too many queries. In particular, the bound on K is used in section 5.3 to upper bound the query complexity of T.
- In the deterministic construction, the tree restored the density of X by fixing some set of coordinates I to some value x_I (using Proposition 3.6). Again, this was possible since it did not matter which transcript the tree would output. In the randomized construction, we cannot do it, since the transcript has to be distributed in a way that is close to being correct. In order to resolve this issue, we follow [23] and use their "density-restoring partition" (Lemma 3.7). Recall that this lemma says that the probability space of X can be partitioned into dense parts. The tree now samples one of those parts according to their probabilities and conditions X on being in this part. If this conditioning reveals too much information, then the tree halts and declares failure.

We turn to give a formal description of the construction. Let h' be the maximum among the universal constants of the uniform marginals lemma (Lemma 3.4) and the main technical lemma (Lemma 3.9), and let h be a universal constant that will be chosen to be sufficiently large to make the inequalities in the proof hold. Let $\varepsilon \stackrel{\text{def}}{=} \frac{h \cdot \log c}{c \cdot \eta}$ and, as before, $\delta \stackrel{\text{def}}{=} 1 - \frac{\eta}{4} + \frac{\varepsilon}{2}$ and $\tau \stackrel{\text{def}}{=} 2 \cdot \delta - \varepsilon$. As before, the parallel decision tree T constructs the transcript π by simulating the protocol Π round-by-round, each time adding a single message to π. Throughout the simulation, the tree maintains a rectangle $\mathcal{X} \times \mathcal{Y} \subseteq \Lambda^n \times \Lambda^n$ of inputs that are consistent with π (but not necessarily of all such inputs). In what follows, we denote by X and Y random variables that are uniformly distributed over $\mathcal{X}$ and $\mathcal{Y}$, respectively. As before, the tree will maintain the invariant that X and Y are (ρ, τ)-structured, and that moreover, they are $(\delta - \varepsilon)$-dense and δ-dense, respectively, in Alice's rounds and the other way around in Bob's rounds. As mentioned above, the tree will also maintain a variable K from iteration to iteration, which will measure the information revealed so far.

When the tree T starts the simulation, the tree sets the transcript π to be the empty string, the restriction ρ to $\{*\}^n$, the variable K to zero, and the sets $\mathcal{X}, \mathcal{Y}$ to Λ^n. At this point the invariant clearly holds. We now explain how T simulates a single round of the protocol while maintaining the invariant. Suppose that the invariant holds at the beginning of the current round, and assume without loss of generality that it is Alice's turn to speak. The tree T performs the following steps:

1. The tree conditions $X_{\text{free}(\rho)}$ on not taking a value that is ε-dangerous for $Y_{\text{free}(\rho)}$ (i.e., the tree removes from $\mathcal{X}$ all the values x for which $x_{\text{free}(\rho)}$ is ε-dangerous for $Y_{\text{free}(\rho)}$).

2. The tree samples a message M of Alice according to the distribution induced by X. Let p_M be the probability of M. The tree adds M to the transcript, adds $\log \frac{1}{p_M}$ to K, and conditions X on M (i.e., the tree sets $\mathcal{X}$ to be the subset of inputs that are consistent with M).
3. If $K > C + b$, the tree halts and declares error.
4. Let $\mathcal{X}_{\text{free}(\rho)} = \mathcal{X}^1 \cup \ldots \cup \mathcal{X}^\ell$ be the density-restoring partition of Lemma 3.7 with respect to $X_{\text{free}(\rho)}$. The tree chooses a random class in the partition, where the class $\mathcal{X}^j$ is chosen with probability $\Pr\left[X_{\text{free}(\rho)} \in \mathcal{X}^j\right]$. Let $\mathcal{X}^j$ be the chosen class, and let I_j and x_{I_j} be the set and the value associated with $\mathcal{X}^j$. The tree conditions X on the event $X_{\text{free}(\rho)} \in \mathcal{X}^j$ (i.e., the tree sets $\mathcal{X}$ to be the subset of inputs x such that $x_{\text{free}(\rho)} \in \mathcal{X}^j$). The variable $X_{\text{free}(\rho)-I_j}$ is now δ-dense by the properties of the density-restoring partition.
5. Recall that
$$p_{\geq j} \stackrel{\text{def}}{=} \Pr\left[X_{\text{free}(\rho)} \in \mathcal{X}^j \cup \cdots \cup \mathcal{X}^\ell\right]$$
(see Lemma 3.7). If $p_{\geq j} < \frac{1}{8} \cdot 2^{-\frac{\eta}{8}} \cdot \frac{1}{2 \cdot n \cdot b}$, the tree halts and declares error.
6. The tree queries the coordinates in I_j, and updates ρ accordingly.
7. The tree conditions Y on $g^I(x_{I_j}, Y_{I_j}) = \rho_{I_j}$ (i.e., the tree sets $\mathcal{Y}$ to be the subset of values y for which $g^I(x_{I_j}, Y_{I_j}) = \rho_{I_j}$). Due to step 1, the variable $X_{\text{free}(\rho)}$ must take a value that is not ε-dangerous and, therefore, $Y_{\text{free}(\rho)}$ is necessarily $(\delta - \varepsilon)$-dense.

After those steps take place, it becomes Bob's turn to speak, and indeed, $X_{\text{free}(\rho)}$ and $Y_{\text{free}(\rho)}$ are δ-dense and $(\delta - \varepsilon)$-dense respectively. Thus, the invariant is maintained. When the protocol Π stops, the tree T outputs the transcript π and halts. The proof that the above steps are well-defined is similar to the proof for the deterministic construction and is therefore omitted.

The depth of T. As in the proof of the deterministic lifting theorem, it is not hard to see that the depth of T is equal to the round complexity of Π.

5.2. The correctness of T. In this section, we prove the correctness of the construction. For convenience, we first prove the correctness of a modified tree T^*, whose construction is the same as that of T except that step 3 is omitted. Fix an input $z \in \{0,1\}^n$. We define the following (random) transcripts of the protocol Π:

- Let π be a transcript that T outputs when given z.
- Let π^* be a transcript that T^* outputs when given z.
- Let π' be a transcript of Π when given inputs (X', Y') that are uniformly distributed in $G^{-1}(z)$.

Our end goal is to prove that π and π' are $2^{-\frac{\eta}{8} \cdot b}$-close. In order to do so, we will first prove that π^* is $(\frac{1}{2} \cdot 2^{-\frac{\eta}{8} \cdot b})$-close to π'. We will then prove that π is 2^{-b}-close to π^*. Together, the two results imply that π is $2^{-\frac{\eta}{8} \cdot b}$-close to π', as required.

π^* is close to π'. We first prove that π^* is $(\frac{1}{2} \cdot 2^{-\frac{\eta}{8} \cdot b})$-close to π'. To this end, we construct a coupling of π^* and π' such that $\Pr[\pi^* \neq \pi] \leq \frac{1}{2} \cdot 2^{-\frac{\eta}{8} \cdot b}$. Essentially, we construct the coupling by going over the simulation step-by-step and using the uniform marginals lemma to argue that at each step, X and X' are close and can therefore be coupled (and similarly for Y and Y'). We start by setting some notation: for every $i \in [r]$, let us denote by $\mathcal{X}_i \times \mathcal{Y}_i$ the rectangle $\mathcal{X} \times \mathcal{Y}$ from section 5.1 at the end of the ith round of the simulation of T^* (if T^* halts before the ith round ends, set $\mathcal{X}_i \times \mathcal{Y}_i$ to be the rectangle $\mathcal{X} \times \mathcal{Y}$ at the end of the simulation). In our proof, we construct, for every $i \in [r]$,

- a random rectangle $\mathcal{X}'_i \times \mathcal{Y}'_i$ that is jointly distributed with X', Y' with the following property: conditioned on a specific choice of $\mathcal{X}'_i \times \mathcal{Y}'_i$, the pair (X', Y') is uniformly distributed over $(\mathcal{X}'_i \times \mathcal{Y}'_i) \cap G^{-1}(z)$;
- a coupling of $\mathcal{X}_i \times \mathcal{Y}_i$ and $\mathcal{X}'_i \times \mathcal{Y}'_i$ such that $\Pr[\mathcal{X}_i \times \mathcal{Y}_i \neq \mathcal{X}'_i \times \mathcal{Y}'_i] \leq \frac{1}{2} \cdot 2^{-\frac{\eta}{8} \cdot b} \cdot \frac{i}{2 \cdot n \cdot b}$.

Observe that if we can construct such rectangles and couplings, then it follows that π^* and π' are close. To see it, observe that at any given point during the simulation, all the inputs in the rectangle $\mathcal{X} \times \mathcal{Y}$ are consistent with the transcript π. Hence, if $\mathcal{X}_r \times \mathcal{Y}_r = \mathcal{X}'_r \times \mathcal{Y}'_r$, it necessarily means that the inputs (X', Y') are consistent with the transcript π, so $\pi = \pi'$. It follows that

$$\begin{aligned}
\Pr[\pi \neq \pi'] &\leq \Pr[\mathcal{X}_r \times \mathcal{Y}_r \neq \mathcal{X}'_r \times \mathcal{Y}'_r] \\
&\leq \frac{1}{2} \cdot 2^{-\frac{\eta}{8} \cdot b} \cdot \frac{r}{2 \cdot n \cdot b} \\
&\leq \frac{1}{2} \cdot 2^{-\frac{\eta}{8} \cdot b} \cdot \frac{C}{2 \cdot n \cdot b} \\
&\leq \frac{1}{2} \cdot 2^{-\frac{\eta}{8} \cdot b},
\end{aligned}$$

as required.

It remains to construct the rectangles $\mathcal{X}'_i \times \mathcal{Y}'_i$ and the associated couplings. We construct them by induction. Let $i \in [r]$, and suppose we have already constructed $\mathcal{X}'_{i-1} \times \mathcal{Y}'_{i-1}$ and its associated coupling (here, if $i = 1$ we set both $\mathcal{X}_{i-1} \times \mathcal{Y}_{i-1}$ and $\mathcal{X}'_{i-1} \times \mathcal{Y}'_{i-1}$ to $\Lambda^n \times \Lambda^n$). The ith coupling first samples $\mathcal{X}_{i-1} \times \mathcal{Y}_{i-1}$ and $\mathcal{X}'_{i-1} \times \mathcal{Y}'_{i-1}$ from the $(i-1)$th coupling. If they are different, then we set $\mathcal{X}'_i \times \mathcal{Y}'_i$ arbitrarily and assume that the coupling failed (i.e., $\mathcal{X}_i \times \mathcal{Y}_i$ and $\mathcal{X}'_i \times \mathcal{Y}'_i$ are different). Suppose now that $\mathcal{X}_{i-1} \times \mathcal{Y}_{i-1}$ and $\mathcal{X}'_{i-1} \times \mathcal{Y}'_{i-1}$ are equal, and condition on some specific choice of this rectangle. If the tree T^* has already halted by this point, we set $\mathcal{X}'_i \times \mathcal{Y}'_i = \mathcal{X}'_{i-1} \times \mathcal{Y}'_{i-1}$. Otherwise, we proceed as follows.

Let (X, Y) be a random pair that is uniformly distributed over $\mathcal{X}_{i-1} \times \mathcal{Y}_{i-1}$, and recall that due to our conditioning, the pair (X', Y') is uniformly distributed over $(\mathcal{X}'_{i-1} \times \mathcal{Y}'_{i-1}) \cap G^{-1}(z)$. We construct the rest of the coupling by following the simulation step-by-step. For step 1, with probability

$$\Pr\left[X'_{\mathrm{free}(\rho)} \text{ is } \varepsilon\text{-dangerous for } Y_{\mathrm{free}(\rho)}\right],$$

we assume that the coupling failed and set $\mathcal{X}'_i \times \mathcal{Y}'_i$ arbitrarily. Otherwise, we condition both X and X' on not taking a dangerous value. In order to analyze the probability of failure, recall that at the beginning of this step, (X, Y) are (ρ, τ)-structured, where

$$\begin{aligned}
\tau &\stackrel{\text{def}}{=} 2 \cdot \delta - \varepsilon \\
(\text{by definition of } \delta) &= 2 \cdot \left(1 - \frac{\eta}{4} + \frac{\varepsilon}{2}\right) - \varepsilon \\
&= 2 - \frac{\eta}{2} \\
&\geq 2 + \frac{\eta}{4} - \eta + \frac{\eta}{8} \\
\left(\text{since } \varepsilon \stackrel{\text{def}}{=} \frac{h \cdot \log c}{c \cdot \eta}\right) &= 2 + \frac{h \cdot \log c}{4 \cdot c \cdot \varepsilon} - \eta + \frac{\eta}{8} \\
&\geq 2 + \frac{h'}{c \cdot \varepsilon} - \eta + \frac{\eta}{8} + \frac{3 \log c}{c} + \frac{4}{b},
\end{aligned}$$

where the last inequality can be made to hold by choosing h to be sufficiently large. Hence, our main technical lemma (Lemma 3.9) implies that the probability that $X_{\mathrm{free}(\rho)}$ is ε-dangerous for $Y_{\mathrm{free}(\rho)}$ is at most

$$2^{-(\frac{\eta}{8}+\frac{3\log c}{c}+\frac{4}{b})\cdot b} \leq \frac{1}{8}\cdot 2^{-\frac{\eta}{8}\cdot b}\cdot\frac{1}{2\cdot n\cdot b}.$$

Moreover, the uniform marginals lemma (Lemma 3.4) implies that X' is $(\frac{1}{8}\cdot 2^{-\frac{\eta}{8}\cdot b}\cdot\frac{1}{2\cdot n\cdot b})$-close to X and therefore the probability that $X'_{\mathrm{free}(\rho)}$ is ε-dangerous for $Y_{\mathrm{free}(\rho)}$ is at most $2\cdot\frac{1}{8}\cdot 2^{-\frac{\eta}{8}\cdot b}\cdot\frac{1}{2\cdot n\cdot b}$. Hence, the failure probability at this step is at most $\frac{1}{4}\cdot 2^{-\frac{\eta}{8}\cdot b}\cdot\frac{1}{2\cdot n\cdot b}$. Note that if the coupling does not fail, X is conditioned on an event of probability at least $\frac{1}{2}$ and, therefore, after the conditioning X and Y are $(\rho,\tau-\frac{1}{b})$-structured.

For steps 2 and 4, let M and $\mathcal{X}^j$ be the message and partition classes that are distributed according to the input X, respectively. Let M' and $\mathcal{X}^{j'}$ be the corresponding message and class of X'. Since X and Y are $(\rho,\tau-\frac{1}{b})$-structured, it can again be showed by the uniform marginals lemma that X and X' are $(\frac{1}{8}\cdot 2^{-\frac{\eta}{8}\cdot b}\cdot\frac{1}{2\cdot n\cdot b})$-close and, therefore, the pair $(M,\mathcal{X}^j)$ is $(\frac{1}{8}\cdot 2^{-\frac{\eta}{8}\cdot b}\cdot\frac{1}{2\cdot n\cdot b})$-close, to the pair $(M',\mathcal{X}^{j'})$. This implies that there exists a coupling of $(M,\mathcal{X}^j)$ and $(M',\mathcal{X}^{j'})$ such that the probability that they differ is at most $\frac{1}{8}\cdot 2^{-\frac{\eta}{8}\cdot b}\cdot\frac{1}{2\cdot n\cdot b}$. We sample $(M,\mathcal{X}^j)$ and $(M',\mathcal{X}^{j'})$ from this coupling. If they differ, we assume that the coupling failed, and set $\mathcal{X}'_i\times\mathcal{Y}'_i$ arbitrarily. Otherwise, we condition both X and X' on being consistent with the message M and the class $\mathcal{X}^j$, and denote by I_j, x_{I_j} the set and values associated with $\mathcal{X}^j$. Finally, for step 5, if $p_{\geq j}\leq\frac{1}{8}\cdot 2^{-\frac{\eta}{8}\cdot b}\cdot\frac{1}{2\cdot n\cdot b}$, then we assume that the coupling fails and set $\mathcal{X}'_i\times\mathcal{Y}'_i$ arbitrarily (note that this happens with probability at most $\frac{1}{8}\cdot 2^{-\frac{\eta}{8}\cdot b}\cdot\frac{1}{n\cdot b}$).

At this point, we set $\mathcal{X}'_i=\mathcal{X}^j$, and set $\mathcal{Y}'_i$ to be the set of inputs $y\in\mathcal{Y}_{i-1}$ for which $g(x_{I_j},y_{I_j})=z_{I_j}$. It is easy to see that this choice satisfies $\mathcal{X}'_i\times\mathcal{Y}'_i=\mathcal{X}_i\times\mathcal{Y}_i$. To analyze the total failure probability of this coupling, observe that by the induction assumption, the failure probability of the $(i-1)$th coupling is at most $\frac{1}{2}\cdot 2^{-\frac{\eta}{8}\cdot b}\cdot\frac{i-1}{2\cdot n\cdot b}$, and the other failure events discussed above add to that a failure probability of at most

$$\left(\frac{1}{4}+\frac{1}{8}+\frac{1}{8}\right)\cdot 2^{-\frac{\eta}{8}\cdot b}\cdot\frac{1}{2\cdot n\cdot b}=\frac{1}{2}\cdot 2^{-\frac{\eta}{8}\cdot b}\cdot\frac{1}{2\cdot n\cdot b}.$$

Hence, the failure probability of the ith coupling is at most $\frac{3}{4}\cdot 2^{-\frac{\eta}{4}\cdot b}\cdot\frac{i}{n\cdot b}$, as required.

It remains to show that conditioned on any specific choice of $\mathcal{X}'_i\times\mathcal{Y}'_i$, the pair (X',Y') is uniformly distributed over $(\mathcal{X}'_i\times\mathcal{Y}'_i)\cap G^{-1}(z)$. In the cases where the coupling fails, we can ensure this property holds by first sampling (X',Y') and then setting $\mathcal{X}'_i\times\mathcal{Y}'_i=\{(X',Y')\}$. Suppose that the coupling did not fail. Recall that by the induction assumption, it holds that conditioned on the choice of $\mathcal{X}'_{i-1}\times\mathcal{Y}'_{i-1}$, the pair (X',Y') is uniformly distributed over $(\mathcal{X}'_{i-1}\times\mathcal{Y}'_{i-1})\cap G^{-1}(z)$. Observe that all the ith coupling changes in the distribution of (X',Y') is to condition it on being in $\mathcal{X}'_i\times\mathcal{Y}'_i$. Thus, at the end of the ith coupling, the pair (X',Y') is uniformly distributed over $(\mathcal{X}'_i\times\mathcal{Y}'_i)\cap G^{-1}(z)$, as required.

π is close to π^*. We turn to prove that π is 2^{-b}-close to π^*. Let $\mathcal{E}$ denote the event that the tree T halts in step 3. It is not hard to see that the statistical distance between π and π^* is exactly $\Pr[\mathcal{E}]$. We show that $\Pr[\mathcal{E}]<2^{-b}$, and this will conclude the proof of correctness.

Intuitively, the reason that $\Pr[\mathcal{E}]<2^{-b}$ is that the tree halts only if the probability of the transcript up to that point is less than 2^{-C-b}: to see it, observe that the variable

K measures (roughly) the logarithm of the probability of the transcript up to that point, and recall that the tree halts when $K > C + b$. By taking the union bound over all possible transcripts, we get that the halting probability is less than 2^{-b}.

Unfortunately, the formal proof contains a messier calculation: the reason is that the probabilities of the messages as measured by K depend on the choices of the classes $\mathcal{X}^j$ in step 4, so the foregoing intuition only holds for a given choice of these classes. Thus, the formal proof also sums over all the possible choices of classes $\mathcal{X}^j$ and conditions on those choices. However, while the resulting calculation is more complicated, the idea is the same.

In order to facilitate the formal proof, we set up some useful notation. Let $M_1, \ldots, M_r$ be the messages that are chosen in step 2 of the simulation (so $\pi = (M_1, \ldots, M_r)$), and let $J = (j_1, \ldots, j_r)$ be the indices of the classes that are chosen in step 4 (if the tree halts before the ith round, set M_i to the empty string and set $j_i = 1$). Observe that the execution of T is completely determined by π and J and, in particular, π and J determine whether the event $\mathcal{E}$ happens or not. With some abuse of notation, let us denote the fact that a particular choice of (π, J) is consistent with $\mathcal{E}$ by $(\pi, J) \in \mathcal{E}$. For any $i \in [r]$, let us denote $\pi_{\leq i} = (M_1, \ldots, M_{i-1})$ and $J_{\leq i} = (j_1, \ldots, j_i)$. Observe that at the ith round, the probability p_{M_i} in step 2 is determined by $\pi_{<i}$ and $J_{<i}$, and let us denote by $p_{M_i|\pi_{<i},J_{<i}}$ this probability for a given choice of $\pi_{<i}$ and $J_{<i}$. We are now ready to prove the upper bound on $\Pr[\mathcal{E}]$. It holds that

$$\begin{aligned}
\Pr[\mathcal{E}] &= \sum_{(\pi,J)\in\mathcal{E}} \Pr[\pi, J] \\
&= \sum_{(\pi,J)\in\mathcal{E}} \Pr[M_1] \cdot \Pr[j_1|\pi_{\leq 1}] \cdots \Pr[M_r|\pi_{\leq r-1}, J_{\leq r-1}] \cdot \Pr[j_r|\pi_{\leq r}, J_{\leq r-1}] \\
&= \sum_{(\pi,J)\in\mathcal{E}} p_{M_1} \cdot p_{M_2|\pi_{\leq 1},J_{\leq 1}} \cdots p_{M_r|\pi_{\leq r-1},J_{\leq r-1}} \\
&\quad \cdot \Pr[j_1|\pi_{\leq 1}] \cdots \Pr[j_r|\pi_{\leq r}, J_{\leq r-1}].
\end{aligned}$$

Next, observe that for every choice of (π, J), the corresponding value of K at the end of the simulation is

$$\log \frac{1}{p_{M_1}} + \log \frac{1}{p_{M_2|\pi_{<2},J_{<2}}} + \cdots + \log \frac{1}{p_{M_r|\pi_{<r},J_{<r}}}.$$

In particular, if $(\pi, J) \in \mathcal{E}$, then it holds that $K > C + b$ and, therefore,

$$p_{M_1} \cdot p_{M_2|\pi_{<2},J_{<2}} \cdots p_{M_r|\pi_{<r},J_{<r}} < 2^{-C-b}.$$

It follows that

$$\begin{aligned}
\Pr[\mathcal{E}] &= \sum_{(\pi,J)\in\mathcal{E}} p_{M_1} \cdot p_{M_2|\pi_{<2},J_{<2}} \cdots p_{M_r|\pi_{<r},J_{<r}} \\
&\quad \cdot \Pr[j_1|\pi_{\leq 1}] \cdots \Pr[j_r|\pi_{\leq r}, J_{\leq r-1}] \\
&< \sum_{(\pi,J)\in\mathcal{E}} 2^{-C-b} \cdot \Pr[j_1|\pi_{\leq 1}] \cdots \Pr[j_r|\pi_{\leq r}, J_{\leq r-1}] \\
&\leq \sum_{(\pi,J)} 2^{-C-b} \cdot \Pr[j_1|\pi_{\leq 1}] \cdots \Pr[j_r|\pi_{\leq r}, J_{\leq r-1}]
\end{aligned}$$

$$\leq 2^{-C-b} \cdot \sum_{M_1,j_1} \Pr\left[j_1|\pi_{\leq 1}\right] \cdot \sum_{M_2,j_2} \Pr\left[j_2|\pi_{\leq 2}, J_{\leq 1}\right]$$
$$\cdots \sum_{M_r,j_r} \Pr\left[j_2|\pi_{\leq 2}, J_{\leq 1}\right]$$
$$= 2^{-C-b} \cdot \sum_{M_1} 1 \cdot \sum_{M_2} 1 \cdots \sum_{M_r} 1 \tag{9}$$
$$= 2^{-C-b} \cdot \sum_{\pi} 1$$
$$\leq 2^{-b}, \tag{10}$$

as required. In the calculation above, equality (9) follows since each sum goes over all the possible choices of j_i, and inequality (10) follows since Π has at most 2^C distinct transcripts.

5.3. The query complexity of T. The analysis of the query complexity here is similar to the analysis of the deterministic query complexity. The main difference is the following: in the deterministic setting, the increase in the deficiency due to a single message M was upper bounded by $|M|$ and, therefore, the total increase in the deficiency was upper bounded by $|C|$. In the randomized case, the increase in the deficiency due to a single message M is upper bounded by $\log \frac{1}{p_M}$. Thus, we upper bound the total increase in the deficiency by K. Since K is never larger than $C + b$ due to step 3, we conclude that the query complexity is at most $O(\frac{C+b}{b}) = O(\frac{C}{b} + 1)$; details follow.

As before, we define the deficiency of X, Y to be

$$2 \cdot b \cdot |\text{free}(\rho)| - H_\infty(X_{\text{free}(\rho)}) - H_\infty(Y_{\text{free}(\rho)}).$$

We prove that whenever the protocol transmits a message M, the deficiency increases by $O(\log \frac{1}{p_M})$, and that whenever the tree T makes a query, the deficiency is decreased by $\Omega(b)$. Since the deficiency is always nonnegative, and K is never more than $C + b$, it will follow that the tree must make at most $O(\frac{C+b}{b})$ queries. More specifically, we prove that in every round, the first two steps increase the deficiency by $\log \frac{1}{p_M} + 1$, and the rest of the steps decrease the deficiency by $\Omega(|I_j| \cdot b)$, and this will imply the desired result.

Fix a round of the simulation, and assume without loss of generality that the message is sent by Alice. We start by analyzing step 1. At this step, the tree conditions $X_{\text{free}(\rho)}$ on taking dangerous values that are not ε-dangerous for $Y_{\text{free}(\rho)}$. Using the same calculation as in section 5.2, it can be shown that the probability of non-dangerous values is at least $\frac{1}{2}$. Therefore, this step increases the deficiency by at most 1 bit. Next, in step 2, the tree conditions X on an event of choosing the message M, whose probability is p_M by definition. Thus, this step increases the deficiency by at most $\log \frac{1}{p_M}$ bits. All in all, we showed that the first two steps of the simulation increase the deficiency by at most $\log \frac{1}{p_M} + 1$ bits.

Let $\mathcal{X}^j$ be the partition class that is sampled in step 4, and let I_j, x_j be the set and value that are associated with $\mathcal{X}^j$. We turn to show that the rest of the steps decrease the deficiency by $\Omega(b \cdot |I_j|)$. Those steps apply the following changes to the deficiency:

- step 4 conditions X on the event $X_{\text{free}(\rho)} \in \mathcal{X}^j$. By Lemma 3.7, this conditioning increases the deficiency at most $\delta \cdot b \cdot |I| + \log \frac{1}{p_{\geq j}}$. Recall that by

step 5, the probability $p_{\geq j}$ can never be less than $\frac{1}{8} \cdot 2^{-\frac{\eta}{8} \cdot b} \cdot \frac{1}{2 \cdot n \cdot b}$. Thus, this step increases the deficiency by at most

$$\delta \cdot b \cdot |I| + \frac{\eta}{8} \cdot b + \log(2 \cdot n \cdot b) + 3 \leq \left(\delta + \frac{\eta}{4} + \frac{6}{c} \right) \cdot b \cdot |I| .$$

- Step 6 removes the set I from $\mathrm{free}(\rho)$. Looking at the definition of deficiency, this change decreases the term of $2 \cdot b \cdot |\mathrm{free}(\rho)|$ by $2 \cdot b \cdot |I|$, decreases the term of $H_\infty(Y_{\mathrm{free}(\rho)})$ by at most $b \cdot |I|$ (Fact 2.3), and does not change the term $H_\infty(X_{\mathrm{free}(\rho)})$ (since at this point X_I is fixed to x_I). All in all, the deficiency is decreased by at least $b \cdot |I|$.
- Finally, step 7 conditions Y on the event $g^I(x_I, Y_I) = \rho_I$. This event has probability at least $2^{-|I|-1}$ by the assumption that X is not dangerous (and hence not leaking). Thus, this conditioning increases the deficiency by at most $|I| + 1$.

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

$$b \cdot |I| - \left(\delta + \frac{\eta}{8} + \frac{6}{c} \right) \cdot b \cdot |I| - (|I| + 1) \geq \left(1 - \delta - \frac{\eta}{8} - \frac{7}{c} \right) \cdot b \cdot |I| .$$

By choosing c to be sufficiently large, we can make sure that $1 - \delta - \frac{\eta}{8} - \frac{7}{c}$ is a positive constant independent of b and n and, therefore, the decrease in the deficiency will be at least $\Omega(b \cdot |I|)$, as required. To see it, observe that

$$\begin{aligned} \delta + \frac{\eta}{8} + \frac{7}{c} &= 1 - \frac{\eta}{8} + \frac{\varepsilon}{2} + \frac{\eta}{8} + \frac{7}{c} \\ \left(\text{since } \varepsilon \stackrel{\text{def}}{=} \frac{h \cdot \log c}{c \cdot \eta} \right) &= 1 - \frac{\eta}{8} + \frac{h \cdot \log c}{2 \cdot c \cdot \eta} + \frac{7}{c} \\ &\leq 1 - \frac{\eta}{8} + \frac{(h+14) \cdot \log c}{2 \cdot c \cdot \eta}. \end{aligned}$$

Thus, if we choose c such that $\frac{c}{\log c} > \frac{h+14}{2 \cdot \eta^2}$, the expression on the right-hand side will be a constant that is strictly smaller than 1. It is not hard to see that we can choose such a value of c that satisfies $c = O(\frac{1}{\eta^2} \cdot \log \frac{1}{\eta})$.

Acknowledgments. We thank Daniel Kane for some very enlightening conversations and suggestions. The authors would also like to thank anonymous referees for comments that improved the presentation of this work. Part of this work was carried out while the authors were visiting the Simons Institute for the Theory of Computing.

REFERENCES

[1] L. Babai, P. Frankl, and J. Simon, *Complexity classes in communication complexity theory (preliminary version)*, in 27th Annual Symposium on Foundations of Computer Science, Toronto, Canada, IEEE Computer Society, Los Alamitos, CA, 1986, pp. 337–347.

[2] B. Barak, M. Braverman, X. Chen, and A. Rao, *How to compress interactive communication*, in Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, MA, ACM, New York, 2010, pp. 67–76.

[3] P. Beame, T. Pitassi, N. Segerlind, and A. Wigderson, *A strong direct product theorem for corruption and the multiparty communication complexity of disjointness*, Comput. Complexity, 15 (2006), pp. 391–432.

[4] M. Braverman, *Interactive information complexity*, SIAM Rev., 59 (2017), pp. 803–846.

[5] M. Braverman and A. Rao, *Information equals amortized communication*, IEEE Trans. Inform. Theory, 60 (2014), pp. 6058–6069.
[6] M. Braverman, A. Rao, O. Weinstein, and A. Yehudayoff, *Direct products in communication complexity*, in 54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, Berkeley, CA, IEEE Computer Society, Los Alamitos, CA, 2013, pp. 746–755.
[7] A. Chattopadhyay, Y. Filmus, S. Koroth, O. Meir, and T. Pitassi, *Query-to-communication lifting for BPP using inner product*, in 46th International Colloquium on Automata, Languages, and Programming, ICALP 2019, Patras, Greece, C. Baier, I. Chatzigiannakis, P. Flocchini, and S. Leonardi, eds., LIPIcs Leibinz Int. Proc. Inform. 132, Schloss Dagstuhl, Wadern, Germany, 2019, 35, https://doi.org/10.4230/LIPIcs.ICALP.2019.35.
[8] A. Chattopadhyay, M. Koucký, B. Loff, and S. Mukhopadhyay, *Simulation theorems via pseudo-random properties*, Comput. Complexity, 28 (2019), pp. 617–659.
[9] A. Chattopadhyay, M. Koucký, B. Loff, and S. 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, ACM, New York, 2018, pp. 1013–1020.
[10] S. F. de Rezende, J. Nordström, and M. 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, New Brunswick, NJ, IEEE Computer Society, Los Alamitos, CA, 2016, pp. 295–304.
[11] A. Drucker, *Improved direct product theorems for randomized query complexity*, Comput. Complexity, 21 (2012), pp. 197–244.
[12] J. Edmonds, R. Impagliazzo, S. Rudich, and J. Sgall, *Communication complexity towards lower bounds on circuit depth*, Comput. Complexity, 10 (2001), pp. 210–246.
[13] T. Feder, E. Kushilevitz, M. Naor, and N. Nisan, *Amortized communication complexity*, SIAM J. Comput., 24 (1995), pp. 736–750.
[14] A. Ganor, G. Kol, and R. Raz, *Exponential separation of information and communication*, in Proceedings of the 55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, IEEE Computer Society, Los Alamitos, CA, 2014, pp. 176–185.
[15] A. Ganor, G. Kol, and R. Raz, *Exponential separation of communication and external information*, in Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, ACM, New York, 2016, pp. 977–986.
[16] A. Ganor, G. Kol, and R. Raz, *Exponential separation of information and communication for boolean functions*, J. ACM, 63 (2016), 46.
[17] A. Garg, M. Göös, P. Kamath, and D. Sokolov, *Monotone circuit lower bounds from resolution*, in Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, ACM, New York, 2018, pp. 902–911.
[18] M. Göös and T. S. Jayram, *A composition theorem for conical juntas*, in Proceedings of the 31st Conference on Computational Complexity, CCC 2016, Tokyo, 2016, 5, https://doi.org/10.4230/LIPIcs.CCC.2016.5.
[19] M. Göös, T. S. Jayram, T. Pitassi, and T. Watson, *Randomized communication vs. partition number*, Electronic Colloquium on Computational Complexity (ECCC), TR15-169, 2018.
[20] M. Göös, S. Lovett, R. Meka, T. Watson, and D. Zuckerman, *Rectangles are nonnegative juntas*, SIAM J. Comput., 45 (2016), pp. 1835–1869.
[21] M. Göös and T. Pitassi, *Communication lower bounds via critical block sensitivity*, SIAM J. Comput., 47 (2018), pp. 1778–1806, https://doi.org/10.1137/16M1082007.
[22] M. Göös, T. Pitassi, and T. Watson, *Deterministic communication vs. partition number*, in Proceedings of IEEE 56th Annual Symposium on Foundations of Computer Science (FOCS), IEEE Computer Society, Los Alamitos, CA, 2015, pp. 1077–1088.
[23] M. Göös, T. Pitassi, and T. Watson, *Query-to-communication lifting for BPP*, in Proceedings of IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS), IEEE Computer Society, Los Alamitos, CA, 2017, pp. 132–143.
[24] H. Hatami, K. Hosseini, and S. Lovett, *Structure of protocols for XOR functions*, SIAM J. Comput., 47 (2018), pp. 208–217.
[25] R. Jain and H. Klauck, *The partition bound for classical communication complexity and query complexity*, in Proceedings of the 25th Annual IEEE Conference on Computational Complexity, CCC 2010, Cambridge, MA, IEEE, Piscataway, NJ, 2010, pp. 247–258.
[26] M. Karchmer, R. Raz, and A. Wigderson, *Super-logarithmic depth lower bounds via direct sum in communication complexity*, in Proceedings of the Sixth Annual Structure in Complexity Theory Conference, Chicago, IL, IEEE Computer Society, Los Alamitos, CA, 1991, 1991, pp. 299–304.

[27] I. Kerenidis, S. Laplante, V. Lerays, J. Roland, and D. Xiao, *Lower bounds on information complexity via zero-communication protocols and applications*, SIAM J. Comput., 44 (2015), pp. 1550–1572.
[28] H. Klauck, *A strong direct product theorem for disjointness*, in Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, MA, ACM, New York, 2010, pp. 77–86.
[29] A. Kozachinskiy, *From expanders to hitting distributions and simulation theorems*, in Proceedings of the 43rd International Symposium on Mathematical Foundations of Computer Science, MFCS 2018, Liverpool, UK, Schloss Dagstuhl, Wadern, Germany, 2018, 4.
[30] E. Kushilevitz and N. Nisan, *Communication Complexity*, Cambridge University Press, 1997.
[31] T. Lee, A. Shraibman, and R. Spalek, *A direct product theorem for discrepancy*, in Proceedings of the 23rd Annual IEEE Conference on Computational Complexity, CCC 2008, College Park, MD, IEEE, Piscataway, NJ, 2008, pp. 71–80.
[32] B. Loff and S. Mukhopadhyay, *Lifting theorems for equality*, Electronic Colloquium on Computational Complexity (ECCC), TR18-175, 2019.
[33] T. Pitassi and R. Robere, *Strongly exponential lower bounds for monotone computation*, in Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, Canada, ACM, New York, 2017, pp. 1246–1255.
[34] T. Pitassi and R. Robere, *Lifting nullstellensatz to monotone span programs over any field*, in Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, ACM, New York, 2018, pp. 1207–1219.
[35] R. Raz and P. McKenzie, *Separation of the monotone NC hierarchy*, Combinatorica, 19 (1999), pp. 403–435.
[36] A. A. Razborov, *On the distributional complexity of disjointness*, Theoret. Comput. Sci., 106 (1992), pp. 385–390.
[37] R. Robere, T. Pitassi, B. Rossman, and S. A. Cook, *Exponential lower bounds for monotone span programs*, in IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS, New Brunswick, NJ, 2016, IEEE Computer Society, Los Alamitos, CA, pp. 406–415.
[38] R. Shaltiel, *Towards proving strong direct product theorems*, Comput. Complexity, 12 (2003), pp. 1–22.
[39] A. A. Sherstov, *The pattern matrix method*, SIAM J. Comput., 40 (2011), pp. 1969–2000.
[40] Y. Shi and Y. Zhu, *Quantum communication complexity of block-composed functions*, Quantum Inf. Comput., 9 (2009), pp. 444–460.
[41] L. G. Valiant, *Parallelism in comparison problems*, SIAM J. Comput., 4 (1975), pp. 348–355.
[42] X. Wu, P. Yao, and H. S. Yuen, *Raz-McKenzie simulation with the inner product gadget*, Electronic Colloquium on Computational Complexity (ECCC), TR17-010, 2017, https://eccc.weizmann.ac.il/report/2017/010.
[43] A. C. Yao, *Lower bounds by probabilistic arguments (extended abstract)*, in 24th Annual Symposium on Foundations of Computer Science, Tucson, AZ, 1983, IEEE Computer Society, Los Alamitos, CA, 1983, pp. 420–428.