SIAM J. COMPUT.
Vol. 49, No. 4, pp. FOCS17-441–FOCS17-461

QUERY-TO-COMMUNICATION LIFTING FOR BPP[*]

MIKA GÖÖS[†], TONIANN PITASSI[‡], AND THOMAS WATSON[§]

Abstract. For any n-bit boolean function f, we show that the randomized communication complexity of the composed function $f \circ g^n$, where g is an index gadget, is characterized by the randomized decision tree complexity of f. In particular, this means that many query complexity separations involving randomized models (e.g., classical vs. quantum) automatically imply analogous separations in communication complexity.

Key words. query complexity, communication complexity, lifting, BPP

AMS subject classifications. 68Q11, 68Q15, 68Q17

DOI. 10.1137/17M115339X

1. Introduction.

A *query-to-communication lifting theorem* (a.k.a. communication-to-query simulation theorem) translates lower bounds on some type of *query complexity* (a.k.a. decision tree complexity) [42, 10, 24] of a boolean function f into lower bounds on a corresponding type of *communication complexity* [47, 28, 24, 31] of a two-party version of f. See Table 1 for a list of several known results in this vein. In this work, we show a lifting theorem for bounded-error randomized (i.e., BPP-type) query/communication complexity. Such a theorem had been conjectured by [5, 8, 14, 45] and (ad nauseam) by the current authors.

1.1. Our result.

For a function $f\colon \{0,1\}^n \to \{0,1\}$ (called the *outer function*) and a two-party function $g\colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ (called the *gadget*), their composition $f \circ g^n\colon \mathcal{X}^n \times \mathcal{Y}^n \to \{0,1\}$ is defined by

$$(f \circ g^n)(x, y) \coloneqq f(g(x_1, y_1), \ldots, g(x_n, y_n)).$$

Here, Alice holds $x \in \mathcal{X}^n$ and Bob holds $y \in \mathcal{Y}^n$. Our result is proved for the popular *index* gadget $\textsc{Ind}_m\colon [m] \times \{0,1\}^m \to \{0,1\}$ mapping $(x, y) \mapsto y_x$. We use $\mathsf{BPP}^{\mathsf{dt}}$ and $\mathsf{BPP}^{\mathsf{cc}}$ to denote the usual bounded-error randomized query and communication complexities. That is, $\mathsf{BPP}^{\mathsf{dt}}(f)$ is the minimum cost of a randomized decision tree (distribution over deterministic decision trees) which, on each input z, outputs $f(z)$ with probability at least $2/3$, where the cost is the maximum number of queries over all inputs and outcomes of the randomness; $\mathsf{BPP}^{\mathsf{cc}}(F)$ is defined similarly but with communication protocols instead of decision trees.

THEOREM 1.1 (lifting for BPP). *Let $m = m(n) \coloneqq n^{256}$. For every $f\colon \{0,1\}^n \to \{0,1\}$,*

$$\mathsf{BPP}^{\mathsf{cc}}(f \circ \textsc{Ind}_m^n) = \mathsf{BPP}^{\mathsf{dt}}(f) \cdot \Theta(\log n).$$

[*] Received by the editors October 24, 2017; accepted for publication (in revised form) June 29, 2020; published electronically August 12, 2020.
https://doi.org/10.1137/17M115339X
Funding: The work of the third author was supported by NSF grant CCF-1657377.

[†] Computer Science Department, Stanford University, Stanford, CA 94305 (goos@stanford.edu).
[‡] Computer Science Department, University of Toronto, Toronto, ON, M5S 3G4, Canada (toni@cs.toronto.edu).
[§] Computer Science Department, University of Memphis, Memphis, TN 38152 (Thomas.Watson@memphis.edu).

TABLE 1

Query-to-communication lifting theorems. The first five are formulated in the language of boolean functions (as in this paper); the last two are formulated in the language of combinatorial optimization.

Class	Query model	Communication model	References
P	deterministic	deterministic	[33, 22, 15, 23, 45, 14]
NP	nondeterministic	nondeterministic	[21, 18]
many	polynomial degree	rank	[40, 38, 35, 36]
many	conical junta degree	nonnegative rank	[21, 27]
P^{NP}	decision list	rectangle overlay	[20]
	Sherali–Adams	LP extension complexity	[12, 27]
	sum-of-squares	SDP extension complexity	[29]

1.2. What does it mean?

The upshot of our lifting theorem is that it *automates* the task of proving randomized communication lower bounds: we only need to show a problem-specific query lower bound for f (which is often relatively simple), and then invoke the general-purpose lifting theorem to completely characterize the randomized communication complexity of $f \circ \text{IND}_m^n$.

Separation results. The lifting theorem is especially useful for constructing examples of two-party functions that have large randomized communication complexity, but low complexity in some other communication model. For example, one of the main results of Anshu et al. [5] is a nearly 2.5th power separation between randomized and quantum ($\mathsf{BQP}^{\mathsf{cc}}$) communication complexities for a total function F:

$$\mathsf{BPP}^{\mathsf{cc}}(F) \geq \mathsf{BQP}^{\mathsf{cc}}(F)^{2.5-o(1)}. \tag{1.1}$$

Previously, a quadratic separation was known (witnessed by set-disjointness). The construction of F (and its ad hoc analysis) in [5] was closely modeled after an analogous query complexity separation, $\mathsf{BPP}^{\mathsf{dt}}(f) \geq \mathsf{BQP}^{\mathsf{dt}}(f)^{2.5-o(1)}$, shown earlier by [2]. Our lifting theorem can reproduce the separation (1.1) by simply taking $F \coloneqq f \circ \text{IND}_m^n$ and using the query result of [2] as a black box. Here we only note that $\mathsf{BQP}^{\mathsf{cc}}(F)$ is at most a logarithmic factor larger than $\mathsf{BQP}^{\mathsf{dt}}(f)$, since a protocol can always efficiently simulate a decision tree.

In a similar fashion, we can unify (and in some cases simplify) several other existing results in communication complexity [32, 19, 5, 6], including separations between $\mathsf{BPP}^{\mathsf{cc}}$ and the log of the partition number; see section 5 for details.

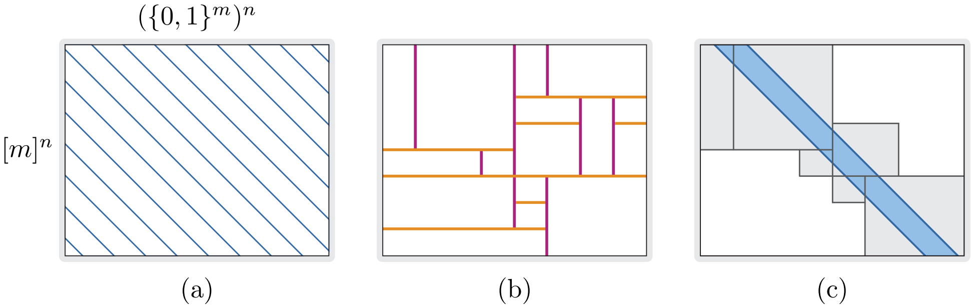
Gadget size. A drawback of our lifting theorem is that it assumes gadget size $m = \text{poly}(n)$, which limits its applicability. For example, we are not able to reproduce tight randomized lower bounds for important functions such as set-disjointness [25, 34, 7] or gap-Hamming [11, 39, 43]. It remains an open problem to prove a lifting theorem for $m = O(1)$ even for the models studied in [21, 27].

Our result has been strengthened to hold for any gadget on $O(\log n)$ bits with small enough discrepancy, such as the inner-product mod 2 gadget [13].

2. Reformulation.

Our lifting theorem holds for all f, even if f is a partial function or a general relation (search problem). Thus the theorem is not *really* about the outer function at all; it is about the obfuscating ability of the index gadget IND_m to hide information about the input bits of f. To focus on what is essential, let us reformulate the lifting theorem in a more abstract way that makes no reference to f.

2.1. Slices. Write $G \coloneqq g^n$ for $g \coloneqq \textsc{Ind}_m$. We view G's input domain $[m]^n \times (\{0,1\}^m)^n$ as being partitioned into *slices* $G^{-1}(z) = \{(x,y) : G(x,y) = z\}$, one for each $z \in \{0,1\}^n$; see (a) below. We will eventually consider *randomized* protocols, but suppose for simplicity that we are given a *deterministic* protocol Π of communication cost $|\Pi|$. The most basic fact about Π is that it induces a partition of the input domain into at most $2^{|\Pi|}$ rectangles (sets of the form $X \times Y$, where $X \subseteq [m]^n$, $Y \subseteq (\{0,1\}^m)^n$); see (b) below. The rectangles are in 1-to-1 correspondence with the leaves of the protocol tree, which are in 1-to-1 correspondence with the protocol's *transcripts* (root-to-leaf paths; each path is a concatenation of messages). Fixing some $z \in \{0,1\}^n$, we are interested in the distribution over transcripts that is generated when Π is run on a uniform random input from the slice $G^{-1}(z)$; see (c) below.



2.2. The reformulation. We devise a *randomized* decision tree that on input z outputs a random transcript distributed close (in total variation distance) to that generated by Π on uniformly random input $(\boldsymbol{x}, \boldsymbol{y}) \sim G^{-1}(z)$. (We always use boldface letters for random variables.)

THEOREM 2.1. *Let Π be a deterministic protocol with inputs from the domain of $G = g^n$. There is a randomized decision tree of cost $O(|\Pi|/\log n)$ that on input $z \in \{0,1\}^n$ samples a random transcript (or outputs $\perp$ for failure) such that the following two distributions are $o(1)$-close:*

$\boldsymbol{t}_z \coloneqq$ *output distribution of the randomized decision tree on input z;*
$\boldsymbol{t}'_z \coloneqq$ *transcript generated by Π when run on a random input $(\boldsymbol{x}, \boldsymbol{y}) \sim G^{-1}(z)$.*

Moreover, the simulation has "one-sided error": $\operatorname{supp}(\boldsymbol{t}_z) \subseteq \operatorname{supp}(\boldsymbol{t}'_z) \cup \{\perp\}$ for every z.

The lifting theorem (Theorem 1.1) follows as a simple consequence of the above reformulation. For the easy direction ("$\leq$"), any randomized decision tree for f making c queries can be converted into a randomized protocol for $f \circ g^n$ communicating $c \cdot O(\log n)$ bits, where the $O(\log n)$ factor is the deterministic communication complexity of the gadget. For the nontrivial direction ("$\geq$"), suppose we have a randomized protocol $\boldsymbol{\Pi}$ (viewed as a probability distribution over deterministic protocols) that computes $f \circ g^n$ (with error $\leq 1/3$, say) and each $\Pi \sim \boldsymbol{\Pi}$ communicates at most $|\Pi| \leq c$ bits. We convert this into a randomized decision tree for f of query cost $O(c/\log n)$ as follows.

On input z:

(1) Pick a deterministic $\Pi \sim \boldsymbol{\Pi}$ (using random coins of the decision tree).
(2) Run the randomized decision tree for Π from Theorem 2.1 that samples a transcript $t \sim \boldsymbol{t}_z(\Pi)$.
(3) Output the value of the leaf reached in t.

The resulting decision tree has bounded error on input z:

$$\begin{aligned}
&\mathbf{Pr}[\,\text{output of decision tree} \neq f(z)]\\
&= \mathbf{E}_{\Pi\sim\boldsymbol{\Pi}}\big[\mathbf{Pr}_{t\sim\boldsymbol{t}_z(\Pi)}[\,\text{value of leaf in } t \neq f(z)]\big]\\
&= \mathbf{E}_{\Pi\sim\boldsymbol{\Pi}}\big[\mathbf{Pr}_{t\sim\boldsymbol{t}'_z(\Pi)}[\,\text{value of leaf in } t \neq f(z)] \pm o(1)\big]\\
&= \mathbf{E}_{\Pi\sim\boldsymbol{\Pi}}\big[\mathbf{Pr}_{(\boldsymbol{x},\boldsymbol{y})\sim G^{-1}(z)}[\Pi(\boldsymbol{x},\boldsymbol{y}) \neq f(z)]\big] \pm o(1)\\
&= \mathbf{E}_{(x,y)\sim G^{-1}(z)}\big[\mathbf{Pr}_{\boldsymbol{\Pi}}[\boldsymbol{\Pi}(x,y)\neq f(z)]\big] \pm o(1)\\
&\le \mathbf{E}_{(x,y)\sim G^{-1}(z)}[1/3] \pm o(1)\\
&\le 1/3 + o(1).
\end{aligned}$$

This simple reformulation is one key conceptual insight that enabled progress on obtaining a BPP lifting theorem.

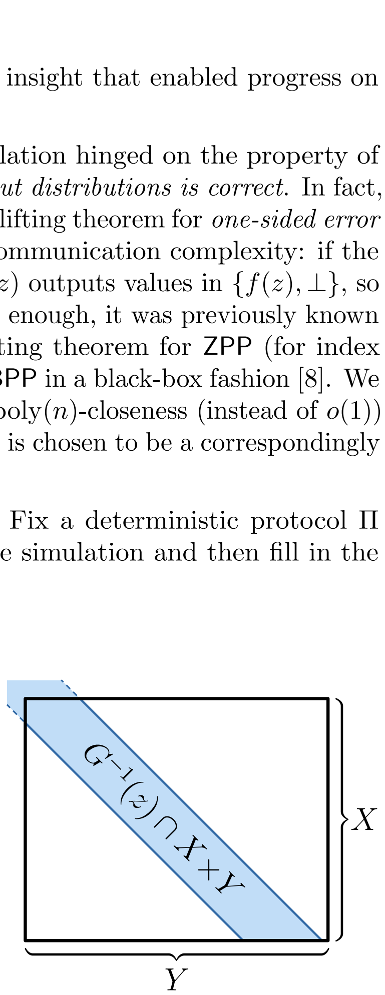
2.3. Extensions. The correctness of our simulation hinged on the property of BPP-type algorithms that the *mixture of correct output distributions is correct.* In fact, the "moreover" part in Theorem 2.1 allows us to get a lifting theorem for *one-sided error* (RP-type) and *zero-sided error* (ZPP-type) query/communication complexity: if the randomized protocol $\boldsymbol{\Pi}$ on every input $(x,y)\in G^{-1}(z)$ outputs values in $\{f(z),\perp\}$, so does our decision tree simulation on input z. Funnily enough, it was previously known that the existence of a query-to-communication lifting theorem for ZPP (for index gadget) implies the existence of a lifting theorem for BPP in a black-box fashion [8]. We also mention that Theorem 2.1 in fact holds with $1/\text{poly}(n)$-closeness (instead of $o(1)$) for an arbitrarily high degree polynomial, provided m is chosen to be a correspondingly high enough degree polynomial in n.

3. Simulation.

We now prove Theorem 2.1. Fix a deterministic protocol Π henceforth. We start with a high-level sketch of the simulation and then fill in the details.

3.1. Executive summary. The randomized decision tree will generate a random transcript of Π by taking a random walk down the protocol tree of Π, guided by occasional queries to the bits of z. The design of our random walk is dictated by one (and only one) property of the slice sets $G^{-1}(z)$, as follows.

> *Uniform marginals lemma (informal):* For every $z\in\{0,1\}^n$ and every rectangle $X\times Y$ where X is "dense" and Y is "large," the uniform distribution on $G^{-1}(z)\cap X\times Y$ has both of its marginal distributions close to uniform on X and Y, respectively.



(The definitions of "dense" and "large" are not needed for this outline of the argument and are given in section 3.2.) This immediately suggests a way to *begin* the randomized simulation. Each node of Π's protocol tree is associated with a rectangle $X\times Y$ of all inputs that reach that node. We start at the root where, initially, $X\times Y = [m]^n\times(\{0,1\}^m)^n$. Suppose Alice communicates the first bit $b\in\{0,1\}$. This induces a partition $X = X^0\cup X^1$ where X^b consists of those inputs where Alice sends b. When Π is run on a random input $(\boldsymbol{x},\boldsymbol{y})\sim G^{-1}(z)$, the above lemma states that $\boldsymbol{x}$ is close to uniform on X, and hence the branch X^b is taken with probability roughly $|X^b|/|X|$. Our idea for a simulation is this: we pretend that $\boldsymbol{x} \sim X$ is perfectly uniform so that our simulation takes the branch X^b with probability exactly $|X^b|/|X|$. It follows that the first bit sent in the two scenarios ($\boldsymbol{t}_z$ and $\boldsymbol{t}'_z$) is distributed close to each other. We can continue the simulation in the same manner, updating $X \leftarrow X^b$ (and similarly $Y \leftarrow Y^b$ when Bob speaks), as long as $X \times Y$ remains "dense × large."

Largeness. A convenient property of the index gadget is that Bob's nm-bit input is much longer than Alice's $n \log m$-bit input. Consequently, the simulation will not need to go out of its way to maintain the "largeness" of Bob's set Y—we will argue that it naturally remains "large" enough with high probability throughout the simulation.

Density. The interesting case is when Alice's set X ceases to be "dense." Our idea is to promptly restore "density" by computing a *density-restoring* partition $X = \bigcup_i X^i$ with the property that each X^i is fixed on some subset of blocks $I_i \subseteq [n]$ (which "caused" a density violation), and such that X^i is again "dense" on the remaining blocks $[n] \setminus I_i$. Moreover, $|I_i|$ will typically be bounded in terms of the number of bits communicated so far.

After Alice has partitioned $X = \bigcup_i X^i$ we will follow the branch X^i (updating $X \leftarrow X^i$) with probability $|X^i|/|X|$; this random choice is justified by the uniform marginals lemma, since it imitates what would happen on a uniform random input from $G^{-1}(z)$. Since we made Alice's pointers $X^i_{I_i}$ fixed, say, to value $\alpha \in [m]^{I_i}$, we need to fix the corresponding pointed-to bits on Bob's side so as to make the output of the gadgets $g^n(X^i, Y)$ consistent with z on the fixed coordinates. At this point, our decision tree queries all the bits $z_{I_i} \in \{0,1\}^{I_i}$ and we argue that we can indeed typically restrict Bob's set to some still-"large" $Y^i \subseteq Y$ to ensure $g^{I_i}(X^i_{I_i} \times Y^i_{I_i}) = \{z_{I_i}\}$. Now that we have recovered "density" on the unfixed blocks, we may continue the simulation as before (relativized to unfixed blocks).

3.2. Tools. Let us make the notions of "dense" and "large" precise. Let $\mathbf{H}_\infty(\boldsymbol{x}) := \min_x \log(1/\mathbf{Pr}[\boldsymbol{x} = x])$ denote the usual min-entropy of a random variable $\boldsymbol{x}$. Supposing $\boldsymbol{x}$ is distributed over a set X, we define the *deficiency* of $\boldsymbol{x}$ as the nonnegative quantity $\mathbf{D}_\infty(\boldsymbol{x}) := \log|X| - \mathbf{H}_\infty(\boldsymbol{x})$. A basic property, which we use freely and repeatedly throughout the proof, is that marginalizing $\boldsymbol{x}$ to some coordinates (assuming X is a product set) cannot increase the deficiency. For a set X we use the boldface $\boldsymbol{X}$ to denote a random variable uniformly distributed on X.

DEFINITION 3.1 (blockwise-density [21]). *A random variable $\boldsymbol{x} \in [m]^J$ (where J is some index set) is called* δ-dense *if for every nonempty $I \subseteq J$ the blocks $\boldsymbol{x}_I$ have min-entropy rate at least δ, that is, $\mathbf{H}_\infty(\boldsymbol{x}_I) \geq \delta \cdot |I| \log m$. (Note that $\boldsymbol{x}_I$ is marginally distributed over $[m]^I$.)*

LEMMA 3.2 (uniform marginals; simple version). *Suppose $\boldsymbol{X}$ is 0.9-dense and $\mathbf{D}_\infty(\boldsymbol{Y}) \leq n^3$. Then for any $z \in \{0,1\}^n$ the uniform distribution on $G^{-1}(z) \cap X \times Y$ (which is nonempty) has both of its marginal distributions $1/n^2$-close to uniform on X and Y, respectively.*

We postpone the proof of the lemma to section 4, and instead concentrate here on the simulation itself—its correctness will mostly rely on this lemma. Actually, we need a slightly more general-looking statement that we can easily apply when some blocks in X have become fixed during the simulation. To this end, we introduce terminology for such rectangles $X \times Y$. Note that Lemma 3.4 below specializes to Lemma 3.2 by taking $\rho = *^n$.

Definition 3.3 (structured rectangles). *For a partial assignment $\rho \in \{0,1,*\}^n$, define its* free *positions as* $\operatorname{free}\rho := \rho^{-1}(*) \subseteq [n]$, *and its* fixed *positions as* $\operatorname{fix}\rho := [n] \smallsetminus \operatorname{free}\rho$. *A rectangle $X \times Y$ is called ρ-*structured *if $\boldsymbol{X}_{\operatorname{free}\rho}$ is 0.9-dense, $\boldsymbol{X}_{\operatorname{fix}\rho}$ is fixed, and each output in $G(X \times Y)$ is consistent with ρ.*

An illustration of a ρ-structured rectangle appears in Figure 1.

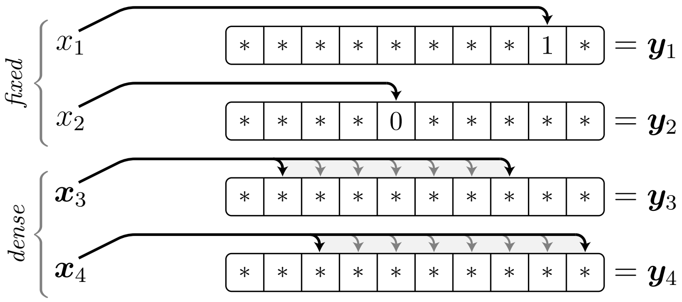


Fig. 1. *Illustration of $\boldsymbol{x} \sim X$ and $\boldsymbol{y} \sim Y$, where $X \times Y$ is ρ-structured for $\rho := 10{*}{*}$*

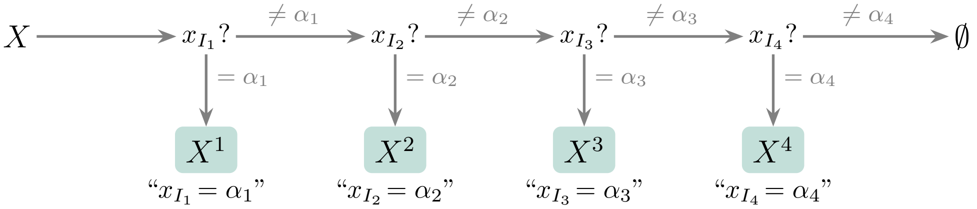
Lemma 3.4 (uniform marginals; general version). *Suppose $X \times Y$ is ρ-structured and $\mathbf{D}_\infty(\boldsymbol{Y}) \le n^3$. Then for any $z \in \{0,1\}^n$ consistent with ρ, the uniform distribution on $G^{-1}(z) \cap X \times Y$ (which is nonempty) has both of its marginal distributions $1/n^2$-close to uniform on X and Y, respectively.*

The uniform marginals lemma is a key technical ingredient that enables us to go beyond the limitations of techniques from previous work on query-to-communication lifting.

3.3. Density-restoring partition. Fix some set $X \subseteq [m]^J$. (In our application, $J \subseteq [n]$ will correspond to the set of free blocks during the simulation.) We describe a procedure that takes X and outputs a *density-restoring* partition $X = \bigcup_i X^i$ such that each $\boldsymbol{X}^i$ is fixed on some subset of blocks $I_i \subseteq J$ and 0.9-dense on $J \smallsetminus I_i$. The procedure associates a *label* of the form "$x_{I_i} = \alpha_i$" with each part X_i, recording which blocks we fixed and to what value. If $\boldsymbol{X}$ is already 0.9-dense, the procedure outputs just one part: X itself.

While X is nonempty:

(1) Let $I \subseteq J$ be a *maximal* subset (possibly $I = \emptyset$) such that $\boldsymbol{X}_I$ has min-entropy rate < 0.9, and let $\alpha \in [m]^I$ be an outcome witnessing this: $\mathbf{Pr}[\boldsymbol{X}_I = \alpha] > m^{-0.9|I|}$.
(2) Output part $X^{(x_I = \alpha)} := \{x \in X : x_I = \alpha\}$ with label "$x_I = \alpha$."
(3) Update $X \leftarrow X \smallsetminus X^{(x_I=\alpha)}$.



We collect below the key properties of the partition $X = \bigcup_i X^i$ output by the procedure. First, the partition indeed restores blockwise-density for the unfixed blocks. Second, the deficiency (relative to unfixed blocks) typically decreases proportional to the number of blocks we fixed.

LEMMA 3.5. *Each X^i (labeled "$x_{I_i} = \alpha_i$") in the density-restoring partition satisfies the following.*

$$\begin{array}{rl} \text{(Density)} & \boldsymbol{X}^i_{J \smallsetminus I_i} \text{ is } 0.9\text{-dense.} \\ \text{(Deficiency)} & \mathbf{D}_\infty(\boldsymbol{X}^i_{J \smallsetminus I_i}) \le \mathbf{D}_\infty(\boldsymbol{X}) - 0.1|I_i| \log m + \delta_i, \\ & \text{where} \quad \delta_i := \log(|X| / |\cup_{j \ge i} X^j|). \end{array}$$

Proof. Write $X^{\geqslant i} := \bigcup_{j \ge i} X^j$ so that $\boldsymbol{X}^i = (\boldsymbol{X}^{\geqslant i} \mid \boldsymbol{X}^{\geqslant i}_{I_i} = \alpha_i)$. Suppose for contradiction that some part $\boldsymbol{X}^i$ was not 0.9-dense on $J \smallsetminus I_i$. Then there is some nonempty $K \subseteq J \smallsetminus I_i$ and an outcome $\beta \in [m]^K$ violating the min-entropy condition: $\mathbf{Pr}[\boldsymbol{X}^i_K = \beta] > m^{-0.9|K|}$. But this contradicts the maximality of I_i since the larger set $I_i \cup K$ now violates the min-entropy condition for $\boldsymbol{X}^{\geqslant i}$:

$$\begin{aligned} \mathbf{Pr}[\boldsymbol{X}^{\geqslant i}_{I_i \cup K} = \alpha_i \beta] &= \mathbf{Pr}[\boldsymbol{X}^{\geqslant i}_{I_i} = \alpha_i] \cdot \mathbf{Pr}[\boldsymbol{X}^i_K = \beta] \\ &> m^{-0.9|I_i|} \cdot m^{-0.9|K|} = m^{-0.9|I_i \cup K|}. \end{aligned}$$

This proves the first part. The second part is a straightforward calculation (intuitively, going from X to $X^{\geqslant i}$ causes a δ_i increase in deficiency, going from $X^{\geqslant i}$ to X^i causes a $\le 0.9|I_i| \log m$ increase, and restricting from J to $J \smallsetminus I_i$ causes a $|I_i| \log m$ decrease):

$$\begin{aligned} \mathbf{D}_\infty(\boldsymbol{X}^i_{J \smallsetminus I_i}) &= |J \smallsetminus I_i| \log m - \log |X^i| \\ &\le \big(|J| \log m - |I_i| \log m\big) - \log\big(|X^{\geqslant i}| \cdot 2^{-0.9|I_i| \log m}\big) \\ &= \big(|J| \log m - \log |X|\big) - 0.1|I_i| \log m + \log\big(|X| / |X^{\geqslant i}|\big) \\ &= \mathbf{D}_\infty(\boldsymbol{X}) - 0.1|I_i| \log m + \delta_i. \qquad \square \end{aligned}$$

3.4. The simulation. To describe our simulation in a convenient language, we modify the deterministic protocol Π into a *refined* deterministic protocol $\overline{\Pi}$; see Figure 2. Namely, we insert two new rounds of communication whose sole purpose is to restore density for Alice's free blocks by fixing some other blocks and Bob's corresponding bits. In short, we maintain the rectangle $X \times Y$ as ρ-structured for some ρ. Each communication round of Π is thus replaced with a whole *iteration* in $\overline{\Pi}$. The new communication rounds do not affect the input/output behavior of the original protocol: any transcript of $\overline{\Pi}$ can be projected back to a transcript of Π (by ignoring messages sent on lines 14 and 16). One way to think about $\overline{\Pi}$ is that it induces a partition of the communication matrix that is a *refinement* of the one Π induces. Therefore, for the purpose of proving Theorem 2.1, we can concentrate on simulating $\overline{\Pi}$ in place of Π. The randomized decision tree becomes simple to describe relative to $\overline{\Pi}$; see Figure 3.

Next, we proceed to show that our randomized decision tree is (1) correct—on input z it samples a transcript distributed close to that of $\overline{\Pi}$ when run on $(\boldsymbol{x}, \boldsymbol{y}) \sim G^{-1}(z)$—and (2) efficient—the number of queries it makes is bounded in terms of $|\Pi|$ (the number of iterations in $\overline{\Pi}$).

3.5. Correctness: Transcript distribution. We show that for every $z \in \{0,1\}^n$ the following distributions are $o(1)$-close:

$\boldsymbol{t} :=$ transcript generated by our simulation of $\overline{\Pi}$ with query access to z;

$\boldsymbol{t}' :=$ transcript generated by $\overline{\Pi}$ when run on a random input from $G^{-1}(z)$.

Refined protocol* $\overline{\Pi}$ *on input* $(\boldsymbol{x}, \boldsymbol{y})$*:

```
1:  initialize:  $v$ = root of $\Pi$,  $X \times Y = [m]^n \times (\{0,1\}^m)^n$,  $\rho = *^n$
2:  while $v$ is not a leaf  [invariant: $X \times Y$ is $\rho$-structured]
3:      let $v_0$, $v_1$ be the children of $v$
4:      if Bob sends a bit at $v$ then
5:          let $Y = Y^0 \cup Y^1$ be the partition according to Bob's function at $v$
6:          let $b$ be such that $y \in Y^b$
7:        ▷ $\underline{\text{Bob sends } b}$ and we update $Y \leftarrow Y^b$, $v \leftarrow v_b$
8:      else Alice sends a bit at $v$
9:          let $X = X^0 \cup X^1$ be the partition according to Alice's function at $v$
10:         let $b$ be such that $x \in X^b$
11:       ▷ $\underline{\text{Alice sends } b}$ and we update $X \leftarrow X^b$, $v \leftarrow v_b$
    - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
12:         let $X = \bigcup_i X^i$ be such that $X_{\text{free}\,\rho} = \bigcup_i X^i_{\text{free}\,\rho}$ is a density-restoring
                partition
13:         let $i$ be such that $x \in X^i$ and suppose $X^i_{\text{free}\,\rho}$ is labeled "$x_I = \alpha$,"
                $I \subseteq \text{free}\,\rho$
14:       ▷ $\underline{\text{Alice sends } i}$ and we update $X \leftarrow X^i$
15:         let $s = g^I(\alpha, y_I) \in \{0,1\}^I$
16:       ▶ $\underline{\text{Bob sends } s}$ and we update $Y \leftarrow \{y' \in Y : g^I(\alpha, y'_I) = s\}$, $\rho_I \leftarrow s$
    - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
17:     end if
18: end while
19: output the value of the leaf $v$
```

FIG. 2. *The refined (deterministic) protocol* $\overline{\Pi}$. *The protocol explicitly keeps track of a rectangle* $X \times Y$ *consisting of all inputs that reach the current node (i.e., produce the same transcript so far). The original protocol* Π *can be recovered by simply ignoring lines* 12–16 *and text in red. The purpose of lines* 12–16 *is to maintain the invariant; they do not affect the input/output behavior. (Color available online.)*

Randomized decision tree on input* z*:

To generate a transcript of $\overline{\Pi}$ we take a random walk down $\overline{\Pi}$'s protocol tree, guided by queries to the bits of z. The following defines the distribution of messages to send at each underlined line.

Lines marked "▷": We simulate an iteration of the protocol $\overline{\Pi}$ pretending that $\boldsymbol{x} \sim X$ and $\boldsymbol{y} \sim Y$ are uniformly distributed over their domains. Namely, in line 7, we send b with probability $|Y^b|/|Y|$; in line 11, we send b with probability $|X^b|/|X|$; in line 14 (after having updated $X \leftarrow X^b$), we send i with probability $|X^i|/|X|$.

Line marked "▶": Here we query z_I and send *deterministically* the message $s = z_I$; except if this message is impossible to send (because $z_I \notin g^I(\alpha, Y_I)$), we output $\perp$ and halt the simulation with failure.

FIG. 3. *The randomized decision tree with query access to* z. *Its goal is to generate a random transcript of* $\overline{\Pi}$ *that is* $o(1)$*-close to the transcript generated by* $\overline{\Pi}$ *on a random input* $(\boldsymbol{x}, \boldsymbol{y}) \sim G^{-1}(z)$.

The following is the heart of the argument.

LEMMA 3.6. *Let $z \in \{0,1\}^n$, and let $\boldsymbol{t}, \boldsymbol{t}'$ be defined as above. Consider a node v at the beginning of an iteration in $\overline{\Pi}$'s protocol tree, such that z is consistent with the associated ρ. Suppose $X \times Y$ is the ρ-structured rectangle at v, and assume that $\mathbf{D}_\infty(\boldsymbol{Y}) \leq n^3$. Let $\boldsymbol{\mu}$ and $\boldsymbol{\mu}'$ denote the messages sent in this iteration under $\boldsymbol{t}$ and $\boldsymbol{t}'$, respectively (conditioned on reaching v). Then*

(i) *$\boldsymbol{\mu}$ and $\boldsymbol{\mu}'$ are $1/n^2$-close;*
(ii) *with probability at least $1 - 4/n^2$ over $\boldsymbol{\mu}$, at least a $2^{-(n \log m + 2)}$ fraction of Y is retained.*

Before proving the lemma, let us use it to show that $\boldsymbol{t}$ and $\boldsymbol{t}'$ are $o(1)$-close. For this, it suffices to exhibit a coupling such that $\mathbf{Pr}[\boldsymbol{t} = \boldsymbol{t}'] \geq 1 - o(1)$. (A coupling of any two random variables $\boldsymbol{a}$ and $\boldsymbol{b}$ is a joint distribution whose marginals are $\boldsymbol{a}$ and $\boldsymbol{b}$; a basic fact is that $\boldsymbol{a}$ and $\boldsymbol{b}$ are ε-close in total variation distance iff there exists a coupling with respect to which $\mathbf{Pr}[\boldsymbol{a} = \boldsymbol{b}] \geq 1 - \varepsilon$.) Our coupling works as follows:

Begin at the root, and for each iteration of $\overline{\Pi}$:

(1) Sample this iteration's messages $\boldsymbol{\mu}$ and $\boldsymbol{\mu}'$ according to an optimal coupling.
(2) If $\boldsymbol{\mu} \neq \boldsymbol{\mu}'$, or if $\boldsymbol{\mu}$ results in $< 2^{-(n \log m + 2)}$ fraction of Y being retained (this includes the simulation's failure case), then proceed to sample the rest of $\boldsymbol{t}$ and $\boldsymbol{t}'$ independently.

It follows by induction on k that after the kth iteration, with probability at least $(1 - 5/n^2)^k$,

(I) $\boldsymbol{t}$ and $\boldsymbol{t}'$ match so far;
(II) $\mathbf{D}_\infty(\boldsymbol{Y}) \leq k \cdot (n \log m + 2) \leq n^3$, where Y is Bob's set under $\boldsymbol{t}$ so far.

This trivially holds for $k = 0$. For $k > 0$, conditioned on (I) and (II) for iteration $k - 1$, the assumptions of Lemma 3.6 are met and hence $\mathbf{Pr}[\boldsymbol{\mu} = \boldsymbol{\mu}'] \geq 1 - 1/n^2$ and

$$\mathbf{Pr}\big[\mathbf{D}_\infty(\boldsymbol{Y}) \leq (k-1) \cdot (n \log m + 2) + (n \log m + 2) = k \cdot (n \log m + 2)\big] \;\geq\; 1 - 4/n^2.$$

By a union bound, with probability $\geq 1 - 5/n^2$, (I) and (II) continue to hold. Thus,

$$\mathbf{Pr}[\text{(I) and (II) hold after the } k\text{th iteration}] \geq (1-5/n^2)^{k-1} \cdot (1-5/n^2) = (1-5/n^2)^k.$$

Since there are at most $n \log m$ iterations, we indeed always have $k \cdot (n \log m + 2) \leq n^3$ (in (II)), and in the end we have $\mathbf{Pr}[\boldsymbol{t} = \boldsymbol{t}'] \geq (1 - 5/n^2)^{n \log m} \geq 1 - (n \log m) \cdot 5/n^2 \geq 1 - o(1)$, and thus $\boldsymbol{t}$ and $\boldsymbol{t}'$ are $o(1)$-close.

Proof of Lemma 3.6. Let $\boldsymbol{x} \coloneqq \boldsymbol{X}$ be uniform over X, let $\boldsymbol{y} \coloneqq \boldsymbol{Y}$ be uniform over Y, and let $(\boldsymbol{x}', \boldsymbol{y}')$ be uniform over $G^{-1}(z) \cap X \times Y$. By Lemma 3.4, $\boldsymbol{x}$ and $\boldsymbol{x}'$ are $1/n^2$-close, and $\boldsymbol{y}$ and $\boldsymbol{y}'$ are $1/n^2$-close.

First assume Bob sends a bit at v. Then $\boldsymbol{\mu}$ is some deterministic function of $\boldsymbol{y}$, and $\boldsymbol{\mu}'$ is the same deterministic function of $\boldsymbol{y}'$ (the bit sent on line 7); thus $\boldsymbol{\mu}$ and $\boldsymbol{\mu}'$ are $1/n^2$-close since $\boldsymbol{y}$ and $\boldsymbol{y}'$ are. Also, the second property in the lemma statement trivially holds.

Henceforth assume Alice sends a bit at v. Write $\boldsymbol{\mu} = \boldsymbol{bis}$ (jointly distributed with $\boldsymbol{x}$) and $\boldsymbol{\mu}' = \boldsymbol{b'i's'}$ (jointly distributed with $(\boldsymbol{x}', \boldsymbol{y}')$) as the concatenation of the three messages sent (on lines 11, 14, and 16). Then $\boldsymbol{bis}$ is some deterministic function of $\boldsymbol{x}$, and $\boldsymbol{b'i's'}$ is the same deterministic function of $\boldsymbol{x}'$ ($\boldsymbol{s}$ and $\boldsymbol{s}'$ depend on z, which is fixed); thus $\boldsymbol{\mu}$ and $\boldsymbol{\mu}'$ are $1/n^2$-close since $\boldsymbol{x}$ and $\boldsymbol{x}'$ are. A subtlety here is that there may be outcomes of $\boldsymbol{bi}$ for which $\boldsymbol{s}$ is not defined (there is no corresponding child in

$\overline{\Pi}$'s protocol tree, since Bob's set would become empty), in which case our randomized decision tree fails and outputs $\perp$. But such outcomes have 0 probability under $\boldsymbol{b'i'}$, so it is still safe to say $\boldsymbol{\mu}$ and $\boldsymbol{\mu}'$ are $1/n^2$-close, treating $\boldsymbol{s}$ as $\perp$ if it is undefined.

We turn to verifying the second property. Define $X^{bi} \times Y^{bi} \subseteq X \times Y$ as the rectangle at the end of the iteration if Alice sends b and i, and note that $\boldsymbol{x} \in X^{\boldsymbol{bi}}$ and $\boldsymbol{x}' \in X^{\boldsymbol{b'i'}}$. We have

$$\begin{aligned}
&\mathbf{Pr}_{bi\sim\boldsymbol{bi}}\big[\mathbf{Pr}[\boldsymbol{y} \in Y^{bi}] < 2^{-(n\log m+2)}\big] \\
&\le \mathbf{Pr}_{bi\sim\boldsymbol{b'i'}}\big[\mathbf{Pr}[\boldsymbol{y} \in Y^{bi}] < 2^{-(n\log m+2)}\big] + 1/n^2 \\
&\le \mathbf{Pr}_{bi\sim\boldsymbol{b'i'}}\big[\mathbf{Pr}[\boldsymbol{y} \in Y^{bi}] < \mathbf{Pr}[\boldsymbol{x} \in X^{bi}]/4\big] + 1/n^2 \\
&\le \mathbf{Pr}_{bi\sim\boldsymbol{b'i'}}\Big[\mathbf{Pr}[\boldsymbol{y} \in Y^{bi}] < \mathbf{Pr}[\boldsymbol{x}' \in X^{bi}]/2 \;\text{ or }\; \mathbf{Pr}[\boldsymbol{x}' \in X^{bi}] < \mathbf{Pr}[\boldsymbol{x} \in X^{bi}]/2\Big] + 1/n^2,
\end{aligned} \tag{3.1}$$

where the second line follows since $\boldsymbol{bi}$ and $\boldsymbol{b'i'}$ are $1/n^2$-close, and the third line follows since $\mathbf{Pr}[\boldsymbol{x} \in X^{bi}] \ge 1/|X| \ge 2^{-n\log m}$. It is straightforward to check that

$$\mathbf{Pr}_{bi\sim\boldsymbol{b'i'}}\big[\mathbf{Pr}[\boldsymbol{x}' \in X^{bi}] < \mathbf{Pr}[\boldsymbol{x} \in X^{bi}]/2\big] \;\le\; 1/n^2 \tag{3.2}$$

since $\boldsymbol{bi}$ and $\boldsymbol{b'i'}$ are $1/n^2$-close. (Note that the inner probabilities "resample" the random variables; e.g., although $\mathbf{Pr}[\boldsymbol{x}' \in X^{\boldsymbol{b'i'}}] = 1$, we cannot say $\mathbf{Pr}[\boldsymbol{x}' \in X^{bi}] = 1$ in (3.2) since the outcome bi is fixed inside the outer probability.) To analyze the other event in (3.1), first note there is a coupling of $\boldsymbol{y}$ and $\boldsymbol{y}'$ such that $\mathbf{Pr}[\boldsymbol{y} \neq \boldsymbol{y}'] \le 1/n^2$, and we may imagine that $\boldsymbol{y}$ is jointly distributed with $(\boldsymbol{x}', \boldsymbol{y}')$: sample $(\boldsymbol{x}', \boldsymbol{y}')$ and then, conditioned on the outcome of $\boldsymbol{y}'$, sample $\boldsymbol{y}$ according to the coupling. For each bi,

$$\begin{aligned}
\mathbf{Pr}[\boldsymbol{y} \in Y^{bi}] &\ge \mathbf{Pr}[\boldsymbol{y} \in Y^{bi} \mid \boldsymbol{x}' \in X^{bi}] \cdot \mathbf{Pr}[\boldsymbol{x}' \in X^{bi}] \\
&\ge \mathbf{Pr}[\boldsymbol{y} = \boldsymbol{y}' \mid \boldsymbol{x}' \in X^{bi}] \cdot \mathbf{Pr}[\boldsymbol{x}' \in X^{bi}]
\end{aligned}$$

(since $\boldsymbol{x}' \in X^{bi}$ implies $\boldsymbol{y}' \in Y^{bi}$), and so

$$\begin{aligned}
\mathbf{Pr}_{bi\sim\boldsymbol{b'i'}}\big[\mathbf{Pr}[\boldsymbol{y} \in Y^{bi}] < \mathbf{Pr}[\boldsymbol{x}' \in X^{bi}]/2\big] &\le \mathbf{Pr}_{bi\sim\boldsymbol{b'i'}}\big[\mathbf{Pr}[\boldsymbol{y} \neq \boldsymbol{y}' \mid \boldsymbol{x}' \in X^{bi}] \ge 1/2\big] \\
&\le 2/n^2.
\end{aligned} \tag{3.3}$$

Combining (3.1), (3.2), and (3.3) using a union bound yields $\mathbf{Pr}_{bi\sim\boldsymbol{bi}}\big[\mathbf{Pr}[\boldsymbol{y} \in Y^{bi}] < 2^{-(n\log m+2)}\big] \le 2/n^2 + 1/n^2 + 1/n^2 = 4/n^2$. ∎

One-sided error. One more detail to iron out is the "moreover" part in the statement of Theorem 2.1. The simulation we described does not quite satisfy this condition, but this is simple to fix: instead of halting with failure only when Y becomes empty, we also halt with failure when $\mathbf{D}_\infty(\boldsymbol{Y}) > n^3$. This does not affect the correctness or efficiency analysis at all, but it ensures that we only output a transcript if $X \times Y$ is ρ-structured and $\mathbf{D}_\infty(\boldsymbol{Y}) \le n^3$ at the end, which by Lemma 3.4 guarantees that the transcript's rectangle intersects the slice $G^{-1}(z)$ and thus $\boldsymbol{t} \in \operatorname{supp}(\boldsymbol{t}')$.

3.6. Efficiency: Number of queries. We show that our randomized decision tree makes $O(|\Pi|/\log n)$ queries with high probability. If we insist on a decision tree that *always* makes this many queries (to match the statement of Theorem 2.1), we may terminate the execution early (with output $\perp$) whenever we exceed the threshold. This would incur only a small additional loss in the closeness of transcript distributions.

LEMMA 3.7. *The simulation makes $O(|\Pi|/\log n)$ queries with probability $\geq 1 - \min(2^{-|\Pi|}, 1/n^{\Omega(1)})$.*

Proof. During the simulation, we view the quantity $\mathbf{D}_\infty(\boldsymbol{X}_{\mathrm{free}\,\rho}) \geq 0$ as a nonnegative potential function. Consider a single iteration where lines 11, 14, and 16 modify the sets X and $\mathrm{free}\,\rho$.

- In line 11, we shrink $X = X^0 \cup X^1$ down to $X^{\boldsymbol{b}}$, where $\mathbf{Pr}[\boldsymbol{b} = b] = |X^b|/|X|$. Hence the increase in the potential function is $\gamma_{\boldsymbol{b}} \coloneqq \log(|X|/|X^{\boldsymbol{b}}|)$.
- In line 14 (after $X \leftarrow X^b$), we shrink $X = \bigcup_i X^i$ down to $X^{\boldsymbol{i}}$, where $\mathbf{Pr}[\boldsymbol{i} = i] = |X^i|/|X|$. Moreover, in line 16, $|\mathrm{free}\,\rho|$ decreases by the number of bits we query. Lemma 3.5 says that the potential changes by $\delta_{\boldsymbol{i}} - \Omega(\log n) \cdot \#(\text{queries in this iteration})$, where $\delta_{\boldsymbol{i}} \coloneqq \log(|X|/|\cup_{j \geq \boldsymbol{i}} X^j|)$.

We will see later that for any iteration, $\mathbf{E}[\gamma_{\boldsymbol{b}}], \mathbf{E}[\delta_{\boldsymbol{i}}] \leq O(1)$.

For $j = 1, \ldots, |\Pi|$, letting $\boldsymbol{\gamma}_j, \boldsymbol{\delta}_j$ be the random variables $\gamma_{\boldsymbol{b}}, \delta_{\boldsymbol{i}}$, respectively, in the jth iteration (and letting $\boldsymbol{\gamma}_j = \boldsymbol{\delta}_j = 0$ for outcomes in which Alice does not communicate in the jth iteration), the potential function at the end of the simulation is $\sum_j (\boldsymbol{\gamma}_j + \boldsymbol{\delta}_j) - \Omega(\log n) \cdot \#(\text{queries in total}) \geq 0$, and hence

$$\mathbf{E}\big[\#(\text{queries in total})\big] \leq O(1/\log n) \cdot \sum_j \big(\mathbf{E}[\boldsymbol{\gamma}_j] + \mathbf{E}[\boldsymbol{\delta}_j]\big) \;\leq\; O(|\Pi|/\log n).$$

By Markov's inequality, this already suffices to show that with probability ≥ 0.9 (say), the simulation uses $O(|\Pi|/\log n)$ queries. To get a better concentration bound, we would like for the $\boldsymbol{\gamma}_j, \boldsymbol{\delta}_j$ variables (over all j) to be mutually independent, which they unfortunately generally are not (e.g., $\boldsymbol{\gamma}_1, \boldsymbol{\delta}_1$ may reveal Alice's message in the first iteration, which in turn affects the set of possible values $\boldsymbol{\gamma}_2, \boldsymbol{\delta}_2$ may take). However, there is a trick to overcome this: we will define mutually independent random variables $\boldsymbol{c}_j, \boldsymbol{d}_j$ (for all j) and couple them with the $\boldsymbol{\gamma}_j, \boldsymbol{\delta}_j$ variables in such a way that each $\boldsymbol{\gamma}_j \leq \boldsymbol{c}_j$ and $\boldsymbol{\delta}_j \leq \boldsymbol{d}_j$ with probability 1, and show that $\sum_j (\boldsymbol{c}_j + \boldsymbol{d}_j)$ is bounded with very high probability, which implies the same for $\sum_j (\boldsymbol{\gamma}_j + \boldsymbol{\delta}_j)$. For each j, do the following:

- Sample a uniform real $\boldsymbol{p}_j \in [0, 1)$ and define $\boldsymbol{c}_j \coloneqq \log(1/\boldsymbol{p}_j) + \log(1/(1 - \boldsymbol{p}_j))$, and let $\boldsymbol{\gamma}_j = \gamma_{\boldsymbol{b}}$, where $\boldsymbol{b} = 0$ if $\boldsymbol{p}_j \in [0, |X^0|/|X|)$ and $\boldsymbol{b} = 1$ if $\boldsymbol{p}_j \in [|X^0|/|X|, 1)$ (where X, X^0, X^1 are the sets that arise in the first half of the jth iteration, conditioned on the outcomes of previous iterations). Note that $\boldsymbol{\gamma}_j$ is correctly distributed, and that $\boldsymbol{\gamma}_j \leq \boldsymbol{c}_j$ with probability 1 (specifically, if $\boldsymbol{b} = 0$, then $\boldsymbol{\gamma}_j = \log(|X|/|X^0|) \leq \log(1/\boldsymbol{p}_j) \leq \boldsymbol{c}_j$, and if $\boldsymbol{b} = 1$, then $\boldsymbol{\gamma}_j = \log(|X|/|X^1|) \leq \log(1/(1 - \boldsymbol{p}_j)) \leq \boldsymbol{c}_j$). Also note that, as claimed earlier, $\mathbf{E}[\boldsymbol{\gamma}_j] \leq \mathbf{E}[\boldsymbol{c}_j] = \int_0^1 \big(\log(1/p) + \log(1/(1 - p))\big)\, \mathrm{d}p = 2/\ln 2 \leq O(1)$. For future use, note that $\mathbf{E}\big[2^{\boldsymbol{c}_j/2}\big] = \int_0^1 (p(1 - p))^{-1/2}\, \mathrm{d}p = \pi \leq O(1)$.
- Sample a uniform real $\boldsymbol{q}_j \in [0, 1)$ and define $\boldsymbol{d}_j \coloneqq \log(1/(1 - \boldsymbol{q}_j))$, and let $\boldsymbol{\delta}_j = \delta_{\boldsymbol{i}}$, where $\boldsymbol{i}$ is such that $\boldsymbol{q}_j$ falls in the $\boldsymbol{i}$th interval, assuming we have partitioned $[0, 1)$ into half-open intervals with lengths $|X^i|/|X|$ in the natural left-to-right order (where $X, X^1, X^2, \ldots$ are the sets that arise in the second half of the jth iteration, conditioned on the outcomes of the first half and previous iterations). Note that $\boldsymbol{\delta}_j$ is correctly distributed, and that $\boldsymbol{\delta}_j \leq \boldsymbol{d}_j$ with probability 1 (specifically, if $\boldsymbol{i} = i$, then $\boldsymbol{\delta}_j = \log(|X|/|\cup_{j \geq i} X^j|) \leq \log(1/(1 - \boldsymbol{q}_j)) = \boldsymbol{d}_j$). Also note that, as claimed earlier, $\mathbf{E}[\boldsymbol{\delta}_j] \leq \mathbf{E}[\boldsymbol{d}_j] \leq \mathbf{E}[\boldsymbol{c}_j] \leq O(1)$. For future use, note that $\mathbf{E}\big[2^{\boldsymbol{d}_j/2}\big] \leq \mathbf{E}\big[2^{\boldsymbol{c}_j/2}\big] \leq O(1)$.

Now for some sufficiently large constants C, C' we have

$$\begin{aligned}\mathbf{Pr}\big[\#(\text{queries in total}) > C'\cdot|\Pi|/\log n\big] &\le \mathbf{Pr}\Big[\sum_j(\boldsymbol{\gamma}_j+\boldsymbol{\delta}_j) > C\cdot|\Pi|\Big]\\ &\le \mathbf{Pr}\Big[\sum_j(\boldsymbol{c}_j+\boldsymbol{d}_j) > C\cdot|\Pi|\Big]\\ &= \mathbf{Pr}\Big[2^{\sum_j(\boldsymbol{c}_j+\boldsymbol{d}_j)/2} > 2^{C\cdot|\Pi|/2}\Big]\\ &\le \mathbf{E}\big[2^{\sum_j(\boldsymbol{c}_j+\boldsymbol{d}_j)/2}\big]/2^{C\cdot|\Pi|/2}\\ &= \Big(\prod_j \mathbf{E}\big[2^{\boldsymbol{c}_j/2}\big]\cdot\mathbf{E}\big[2^{\boldsymbol{d}_j/2}\big]\Big)/2^{C\cdot|\Pi|/2}\\ &\le \big(O(1)/2^{C/2}\big)^{|\Pi|}\\ &\le 2^{-|\Pi|}.\end{aligned}$$

If $|\Pi| \le o(\log n)$, then a similar calculation shows that $\mathbf{Pr}\big[\#(\text{queries in total}) \ge 1\big] \le 1/n^{\Omega(1)}$. ∎

4. Uniform marginals lemma.

LEMMA 4.1 (uniform marginals; general version). *Suppose $X\times Y$ is ρ-structured and $\mathbf{D}_\infty(\boldsymbol{Y}) \le n^3$. Then for any $z\in\{0,1\}^n$ consistent with ρ, the uniform distribution on $G^{-1}(z)\cap X\times Y$ (which is nonempty) has both of its marginal distributions $1/n^2$-close to uniform on X and Y, respectively.*

We prove a slightly stronger statement formulated in Lemma 4.2 below. For terminology, we say a distribution $\mathcal{D}_1$ is *ε-pointwise-close* to a distribution $\mathcal{D}_2$ if for every outcome, the probability under $\mathcal{D}_1$ is within a factor $1\pm\varepsilon$ of the probability under $\mathcal{D}_2$. As a minor technicality (for the purpose of deriving Lemma 3.4 from Lemma 4.2), we say that a random variable $\boldsymbol{x}\in[m]^J$ is *δ-essentially-dense* if for every nonempty $I\subseteq J$, $\mathbf{H}_\infty(\boldsymbol{x}_I) \ge \delta\cdot|I|\log m - 1$ (the difference from Definition 3.1 is the "-1"); we also define *ρ-essentially-structured* in the same way as ρ-structured but requiring $\boldsymbol{X}_{\text{free}\,\rho}$ to be only 0.9-essentially-dense instead of 0.9-dense. The following strengthens a lemma from [20], which implied that $G(\boldsymbol{X},\boldsymbol{Y})$ has full support over the set of all z consistent with ρ.

LEMMA 4.2 (pointwise uniformity). *Suppose $X\times Y$ is ρ-essentially-structured and $\mathbf{D}_\infty(\boldsymbol{Y}) \le n^3+1$. Then $G(\boldsymbol{X},\boldsymbol{Y})$ is $1/n^3$-pointwise-close to the uniform distribution over the set of all z consistent with ρ.*

Proof of Lemma 3.4. Let $(\boldsymbol{x},\boldsymbol{y})$ be uniformly distributed over $G^{-1}(z)\cap X\times Y$. We show that $\boldsymbol{x}$ is $1/n^2$-close to $\boldsymbol{X}$; a completely analogous argument works to show that $\boldsymbol{y}$ is $1/n^2$-close to $\boldsymbol{Y}$. Let $E\subseteq X$ be any test event. Replacing E by $X\smallsetminus E$ if necessary, we may assume $|E|\ge|X|/2$. Since $X\times Y$ is ρ-structured, $E\times Y$ is ρ-essentially-structured. Hence we can apply Lemma 4.2 in both the rectangles $E\times Y$ and $X\times Y$:

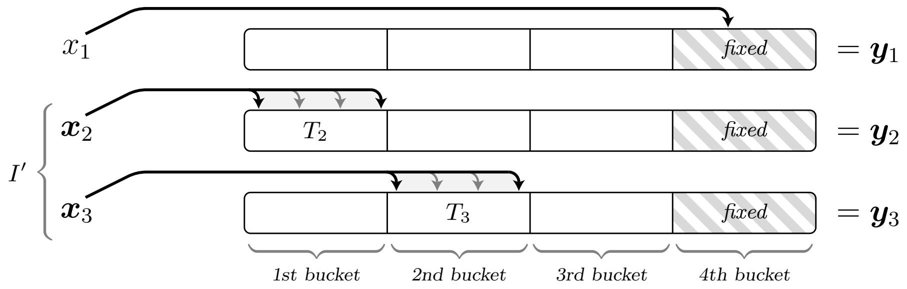
$$\begin{aligned}\mathbf{Pr}[\boldsymbol{x}\in E] = \frac{|G^{-1}(z)\cap E\times Y|}{|G^{-1}(z)\cap X\times Y|} &= \frac{(1\pm 1/n^3)\cdot 2^{-|\text{free}\,\rho|}\cdot|E\times Y|}{(1\pm 1/n^3)\cdot 2^{-|\text{free}\,\rho|}\cdot|X\times Y|}\\ &= (1\pm 3/n^3)\cdot|E|/|X| = |E|/|X| \pm 1/n^2.\end{aligned}$$ ∎

We prove Lemma 4.2 in the rest of this section. An alternative, shorter proof relying on deeper Fourier analysis tools appears in [44].

4.1. Overview for Lemma 4.2. A version of Lemma 4.2 (for the inner-product gadget) was proved in [21, sect. 2.2] under the assumption that $\boldsymbol{X}$ and $\boldsymbol{Y}$ had low deficiencies: $\mathbf{D}_\infty(\boldsymbol{X}_I), \mathbf{D}_\infty(\boldsymbol{Y}_I) \le O(|I| \log n)$ for free blocks I. The key difference is that we only assume $\mathbf{D}_\infty(\boldsymbol{Y}_I) \le n^3 + 1$. We still follow the general plan from [21] but with a new step that allows us to reduce the deficiency of $\boldsymbol{Y}$.

Fourier perspective. We refer the reader to [30] for background on discrete Fourier analysis. The idea in [21] to prove that $\boldsymbol{z} := G(\boldsymbol{X}, \boldsymbol{Y})$ is pointwise-close to uniform is to study $\boldsymbol{z}$ in the Fourier domain, and show that $\boldsymbol{z}$'s Fourier coefficients (corresponding to free blocks) decay exponentially fast. That is, for every nonempty $I \subseteq \operatorname{free}\rho$ we want to show that the bias of $\oplus(\boldsymbol{z}_I)$ (parity of the output bits $\boldsymbol{z}_I$) is exponentially small in $|I|$. Tools tailor-made for this situation exist: various "Xor lemmas" are known to hold for communication complexity (e.g., [37]) that apply as long as $\boldsymbol{X}_I$ and $\boldsymbol{Y}_I$ have low deficiencies. All this is recalled in section 4.2. This suggests that all that remains is to reduce our case of high deficiency (of $\boldsymbol{Y}_I$) to the case of low deficiency.

Reducing deficiency via buckets. For the moment assume $I = [n]$ for simplicity of discussion. Our idea for reducing the deficiency of $\boldsymbol{Y}_I = \boldsymbol{Y}$ is as follows. We partition each m-bit string in $\boldsymbol{Y} \in (\{0,1\}^m)^n$ into $m^{1/2}$ many *buckets* each of length $m^{1/2}$. We argue that $\boldsymbol{Y}$ can be expressed as a mixture of distributions $\boldsymbol{y}$ with the following properties: for each index i, the ith string $\boldsymbol{y}_i$ in $\boldsymbol{y}$ has few of its buckets fixed (where a "fixed bucket" has all the corresponding bits of $\boldsymbol{y}_i$ fixed to constants), and for any way of choosing an unfixed bucket for each $\boldsymbol{y}_i$, the marginal distribution of $\boldsymbol{y}$ on the union T of these buckets has deficiency as low as $\mathbf{D}_\infty(\boldsymbol{y}_T) \le 1$. Correspondingly, we argue that $\boldsymbol{X}$ may be expressed as a mixture of distributions $\boldsymbol{x}$ that have a nice form:



Here each pointer $\boldsymbol{x}_i$ ranges over a single bucket T_i. Moreover, for a large subset $I' \subseteq [n]$ of coordinates, T_i is unfixed in $\boldsymbol{y}_i$ for $i \in I'$, and hence $\boldsymbol{y}$ has deficiency ≤ 1 on the union of these unfixed buckets. The remaining few $i \in [n] \smallsetminus I'$ are associated with fixed pointers $\boldsymbol{x}_i = x_i$ pointing into fixed buckets in $\boldsymbol{y}$. Consequently, we may interpret $(\boldsymbol{x}, \boldsymbol{y})$ as a random input to $\textsc{Ind}^n_{m^{1/2}}$ by identifying each bucket T_i with $[m^{1/2}]$. In this restricted domain, we can show that $(\oplus \circ g^n)(\boldsymbol{x}, \boldsymbol{y})$ is indeed very unbiased: the fixed coordinates do not contribute to the bias of the parity, and $(\boldsymbol{x}_{I'}, \boldsymbol{y}_{I'})$ is a pair of low-deficiency variables for which an Xor lemma–type calculation applies. The heart of the proof will be to find a decomposition of $\boldsymbol{X} \times \boldsymbol{Y}$ into such distributions $\boldsymbol{x} \times \boldsymbol{y}$.

In the remaining subsections, we carry out the formal proof of Lemma 4.2.

4.2. Fourier perspective. Henceforth we abbreviate $J := \operatorname{free}\rho$. We employ the following calculation from [21], whose proof is reproduced in section 4.6 for completeness. Here $\chi(z) := (-1)^{\oplus(z)}$.

Lemma 4.3 (pointwise uniformity from parities). *If a random variable $\boldsymbol{z}_J$ over $\{0,1\}^J$ satisfies $\left|\mathbf{E}\left[\chi(\boldsymbol{z}_I)\right]\right| \le 2^{-5|I| \log n}$ for every nonempty $I \subseteq J$, then $\boldsymbol{z}_J$ is $1/n^3$-*

pointwise-close to uniform.

To prove Lemma 4.2, it suffices to take $\boldsymbol{z}_J = g^J(\boldsymbol{X}_J, \boldsymbol{Y}_J)$ above and show that for every $\emptyset \neq I \subseteq J$,

$$\left|\mathbf{E}\left[\chi(g^I(\boldsymbol{X}_I, \boldsymbol{Y}_I))\right]\right| \leq 2^{-5|I|\log n}. \tag{4.1}$$

In our high-deficiency case, we have

(i) $\mathbf{D}_\infty(\boldsymbol{X}_I) \leq 0.1|I|\log m + 1$,

(ii) $\mathbf{D}_\infty(\boldsymbol{Y}_I) \leq n^3 + 1$.

Low-deficiency case. As a warm-up, let us see how to obtain (4.1) by imagining that we are in the low-deficiency case, i.e., replacing assumption (ii) by

(ii′) $\mathbf{D}_\infty(\boldsymbol{Y}_I) \leq 1$.

We present a calculation that is a very simple special case of, e.g., Shaltiel's [37] Xor lemma for discrepancy (relative to uniform distribution).

We first review a few mathematical concepts. For any real matrix M, its operator 2-norm $\|M\|$ equals its largest singular value and satisfies $\|Mv\| \leq \|M\| \cdot \|v\|$ for any vector v. The k-fold tensor product of M with itself is the matrix $M^{\otimes k}$ defined by $M^{\otimes k}_{x_1\cdots x_k,\, y_1\cdots y_k} \coloneqq \prod_{i=1}^k M_{x_i,y_i}$. A standard fact is that the 2-norm behaves multiplicatively under the tensor product: $\left\|M^{\otimes k}\right\| = \|M\|^k$. We denote the Rényi 2-entropy of a random variable $\boldsymbol{a}$ by $\mathbf{H}_2(\boldsymbol{a}) \coloneqq -\log \sum_a \mathbf{Pr}[\boldsymbol{a} = a]^2$. A standard fact is that $\mathbf{H}_2(\boldsymbol{a}) \geq \mathbf{H}_\infty(\boldsymbol{a})$.

Now let M be the communication matrix of $g \coloneqq \textsc{Ind}_m$ but with $\{+1, -1\}$ instead of $\{0,1\}$ entries. The operator 2-norm of M is $\|M\| = 2^{m/2}$ since the rows are orthogonal and each has 2-norm $2^{m/2}$. The $|I|$-fold tensor product of M then satisfies $\left\|M^{\otimes|I|}\right\| = 2^{|I|m/2}$. Here $M^{\otimes|I|}$ is the communication matrix of the 2-party function $\chi \circ g^I$. We think of the distribution of $\boldsymbol{X}_I$ as an $m^{|I|}$-dimensional vector $\mathcal{D}_{\boldsymbol{X}_I}$, and of the distribution of $\boldsymbol{Y}_I$ as a $(2^m)^{|I|}$-dimensional vector $\mathcal{D}_{\boldsymbol{Y}_I}$. By (i) we have

$$\begin{aligned}\left\|\mathcal{D}_{\boldsymbol{X}_I}\right\| = 2^{-\mathbf{H}_2(\boldsymbol{X}_I)/2} \;&\leq\; 2^{-\mathbf{H}_\infty(\boldsymbol{X}_I)/2}\\ &\leq 2^{-(|I|\log m - 0.1|I|\log m - 1)/2} = 2^{-0.45|I|\log m + 1/2}.\end{aligned}$$

Similarly, by (ii′) we would have

$$\left\|\mathcal{D}_{\boldsymbol{Y}_I}\right\| \leq 2^{-(|I|m-1)/2} = 2^{-|I|m/2+1/2}.$$

The left side of (4.1) is now

$$\begin{aligned}\left|\mathcal{D}_{\boldsymbol{X}_I}^\top\, M^{\otimes|I|}\, \mathcal{D}_{\boldsymbol{Y}_I}\right| &\leq \left\|\mathcal{D}_{\boldsymbol{X}_I}\right\| \cdot \left\|M^{\otimes|I|}\right\| \cdot \left\|\mathcal{D}_{\boldsymbol{Y}_I}\right\|\\ &\leq 2^{-0.45|I|\log m+1/2} \cdot 2^{|I|m/2} \cdot 2^{-|I|m/2+1/2}\\ &= 2^{-0.45|I|\log m+1} \leq 2^{-5|I|\log n}. \qquad (4.2)\end{aligned}$$

Therefore our goal becomes to reduce (via buckets) from case (ii) to case (ii′).

4.3. Buckets.

We introduce some bucket terminology for random $(\boldsymbol{x}, \boldsymbol{y}) \in [m]^I \times (\{0,1\}^m)^I$.

- Each string $\boldsymbol{y}_i$ is partitioned into $m^{1/2}$ buckets, each of length $m^{1/2}$.
- We think of $\boldsymbol{x}_i$ as a pair $\boldsymbol{\ell}_i\boldsymbol{r}_i$, where $\boldsymbol{\ell}_i$ specifies which bucket and $\boldsymbol{r}_i$ specifies which element of the bucket. (Or, viewing $\boldsymbol{x}_i \in \{0,1\}^{\log m}$, $\boldsymbol{\ell}_i \in \{0,1\}^{(\log m)/2}$ would be the left half and $\boldsymbol{r}_i \in \{0,1\}^{(\log m)/2}$ would be the right half.) Thus $\boldsymbol{x} = \boldsymbol{\ell}\boldsymbol{r}$, where the random variable $\boldsymbol{\ell} \in [m^{1/2}]^I$ picks a bucket for each

coordinate, and the random variable $\boldsymbol{r} \in [m^{1/2}]^I$ picks an element from each of the buckets specified by $\boldsymbol{\ell}$. Every outcome ℓ of $\boldsymbol{\ell}$ has an associated *bucket union* (one bucket for each string) given by $T_\ell \coloneqq \bigcup_{i \in I}(\{i\} \times T_{\ell_i})$, where $T_{\ell_i} \subseteq [m]$ is the bucket specified by ℓ_i. Here a bit index $(i,j) \in I \times [m]$ refers to the jth bit of the string $\boldsymbol{y}_i$.

4.4. Focused decompositions. Our goal is to express the product distribution $\boldsymbol{X}_I \times \boldsymbol{Y}_I$ as a convex combination of product distributions $\boldsymbol{x} \times \boldsymbol{y}$ that are *focused*, which informally means that many pointers in $\boldsymbol{x}$ point into buckets that collectively have low deficiency in $\boldsymbol{y}$, and the remaining pointers produce constant gadget outputs. A formal definition follows.

DEFINITION 4.4. *A product distribution $\boldsymbol{x} \times \boldsymbol{y}$ over $[m]^I \times (\{0,1\}^m)^I$ is called* focused *if there is a partial assignment $\sigma \in \{0,1,*\}^I$ such that, letting $I' \coloneqq \operatorname{free}\sigma$, we have $|I'| \geq |I|/2$, and $g^I(\boldsymbol{x},\boldsymbol{y})$ is always consistent with σ, and for each $i \in I'$, $\boldsymbol{x}_i = \ell_i \boldsymbol{r}_i$ is always in a specific bucket $T_{\ell_i} \subseteq [m]$, and*

(i*) $\mathbf{D}_\infty(\boldsymbol{x}_{I'}) \leq 0.6|I'|\log(m^{1/2})$ *with respect to* $\bigtimes_{i \in I'} T_{\ell_i}$*;*

(ii*) $\mathbf{D}_\infty(\boldsymbol{y}_T) \leq 1$*, where* $T \coloneqq \bigcup_{i \in I'}(\{i\} \times T_{\ell_i})$*.*

We elaborate on this definition. Since $g^I(\boldsymbol{x},\boldsymbol{y})$ is always consistent with σ, the coordinates $\operatorname{fix}\sigma = I \smallsetminus I'$ are irrelevant to the bias of the parity of $g^I(\boldsymbol{x},\boldsymbol{y})$. For each $i \in I'$, we might as well think of the domain of $\boldsymbol{x}_i$ as T_{ℓ_i} instead of $[m]$, and of the domain of $\boldsymbol{y}_i$ as $\{0,1\}^{T_{\ell_i}}$ instead of $\{0,1\}^m$. Hence, out of the $|I'|m$ bits of $\boldsymbol{y}_{I'}$, the only relevant ones are the $|I'|m^{1/2}$ bits indexed by T. We may thus interpret $(\boldsymbol{x}_{I'},\boldsymbol{y}_T)$ as a random input to $\textsc{Ind}^{I'}_{m^{1/2}}$. In summary,

$$\big|\mathbf{E}\big[\chi(g^I(\boldsymbol{x},\boldsymbol{y}))\big]\big| = \big|\mathbf{E}\big[\chi(g^{I'}(\boldsymbol{x}_{I'},\boldsymbol{y}_{I'}))\big]\big| = \big|\mathbf{E}\big[\chi(\textsc{Ind}^{I'}_{m^{1/2}}(\boldsymbol{x}_{I'},\boldsymbol{y}_T))\big]\big|. \tag{4.3}$$

If $\boldsymbol{x} \times \boldsymbol{y}$ is focused, then the calculation leading to (4.2) can be applied to $\boldsymbol{x}_{I'} \times \boldsymbol{y}_T$, with m replaced by $m^{1/2}$, $|I|$ replaced by $|I'| \geq |I|/2$, and min-entropy rate 0.9 replaced by 0.4, to show that

$$\text{value of (4.3)} \leq 2^{-0.2|I'|\log(m^{1/2})+1} \leq 2^{-(0.2/4)|I|\log m+1} \leq 2^{-5|I|\log n-1}$$

using $m = n^{256}$.

LEMMA 4.5. *The product distribution $\boldsymbol{X}_I \times \boldsymbol{Y}_I$ can be decomposed into a mixture of product distributions $\mathbf{E}_{d\sim\boldsymbol{d}}[\boldsymbol{x}^d \times \boldsymbol{y}^d]$ over $[m]^I \times (\{0,1\}^m)^I$ (d stands for "data") such that $\boldsymbol{x}^d \times \boldsymbol{y}^d$ is focused with probability at least $1 - 2^{-5|I|\log n-1}$ over $d \sim \boldsymbol{d}$.*

Using Lemma 4.5, which we prove in the following subsection, we can derive (4.1):

$$\begin{aligned}
\big|\mathbf{E}\big[\chi(g^I(\boldsymbol{X}_I,\boldsymbol{Y}_I))\big]\big| &\leq \mathbf{E}_{d\sim\boldsymbol{d}}\big|\mathbf{E}\big[\chi(g^I(\boldsymbol{x}^d,\boldsymbol{y}^d))\big]\big| \\
&\leq \mathbf{Pr}[\boldsymbol{d} \text{ is not focused}] + \max_{\text{focused } d}\big|\mathbf{E}\big[\chi(g^I(\boldsymbol{x}^d,\boldsymbol{y}^d))\big]\big| \\
&\leq 2^{-5|I|\log n-1} + 2^{-5|I|\log n-1} = 2^{-5|I|\log n}.
\end{aligned}$$

4.5. Finding a focused decomposition. We now prove Lemma 4.5. By assumption, $\boldsymbol{X}_I = \boldsymbol{\ell r}$ is 0.9-essentially-dense (since $\boldsymbol{X}_J$ is) and $\mathbf{D}_\infty(\boldsymbol{Y}_I) \leq \mathbf{D}_\infty(\boldsymbol{Y}) \leq n^3 + 1$. We carry out the decomposition in the following three steps. Define $\varepsilon \coloneqq 2^{-5|I|\log n-1}$.

CLAIM 4.6. *$\boldsymbol{Y}_I$ can be decomposed into a mixture of distributions $\mathbf{E}_{c\sim\boldsymbol{c}}[\boldsymbol{y}^c]$ over $(\{0,1\}^m)^I$ such that, with probability at least $1 - \varepsilon/3$ over $c \sim \boldsymbol{c}$,*

(P1) *each string in* $\boldsymbol{y}^c$ *has at most* $2n^3$ *fixed buckets;*
(P2) *each bucket union* T_ℓ *not containing fixed buckets has* $\mathbf{D}_\infty(\boldsymbol{y}^c_{T_\ell}) \leq 1$.

Claim 4.7. *For any* c *satisfying* (P1), *with probability at least* $1-\varepsilon/3$ *over* $\ell \sim \boldsymbol{\ell}$,
(Q1) *the bucket union* T_ℓ *contains at most* $|I|/2$ *fixed buckets of* $\boldsymbol{y}^c$*;*
(Q2) $\mathbf{D}_\infty(\boldsymbol{r} \mid \boldsymbol{\ell} = \ell) \leq 0.25|I|\log(m^{1/2})$.

Claim 4.8. *For any* c *and* ℓ *satisfying* (Q1), (Q2), *letting*

$$I^* \coloneqq \big\{i \in I \,:\, \textit{the } \ell_i \textit{ bucket of } \boldsymbol{y}^c_i \textit{ is fixed}\big\} \qquad \textit{and} \qquad I' \coloneqq I \smallsetminus I^*,$$

with probability at least $1-\varepsilon/3$ *over* $r_{I^*} \sim (\boldsymbol{r}_{I^*} \mid \boldsymbol{\ell} = \ell)$, *we have*

$$\mathbf{D}_\infty(\boldsymbol{r}_{I'} \mid \boldsymbol{\ell} = \ell,\, \boldsymbol{r}_{I^*} = r_{I^*}) \leq 0.6|I'|\log(m^{1/2}).$$

We now finish the proof of Lemma 4.5 assuming these three claims. Take $\boldsymbol{d} \coloneqq (\boldsymbol{c}, \boldsymbol{\ell}, \boldsymbol{r}_{\boldsymbol{I}^*})$; that is, the data $d \sim \boldsymbol{d}$ is sampled by first sampling $c \sim \boldsymbol{c}$, then $\ell \sim \boldsymbol{\ell}$, then $r_{I^*} \sim (\boldsymbol{r}_{I^*} \mid \boldsymbol{\ell} = \ell)$, where I^* implicitly depends on c and ℓ. Take $\boldsymbol{y}^d \coloneqq \boldsymbol{y}^c$ and $\boldsymbol{x}^d \coloneqq (\boldsymbol{X}_I \mid \boldsymbol{\ell} = \ell,\, \boldsymbol{r}_{I^*} = r_{I^*})$, and note that $\mathbf{E}_{d\sim\boldsymbol{d}}[\boldsymbol{x}^d \times \boldsymbol{y}^d]$ indeed forms a decomposition of $\boldsymbol{X}_I \times \boldsymbol{Y}_I$. By a union bound, with probability at least $1-\varepsilon$ over $d \sim \boldsymbol{d}$, the properties of all three claims hold, in which case we just need to check that $\boldsymbol{x}^d \times \boldsymbol{y}^d$ is focused.

Since for each $i \in I^*$, $\boldsymbol{x}^d_i \in T_{\ell_i}$ and $\boldsymbol{y}^d_{i,T_{\ell_i}}$ are both fixed, we have that $g^{I^*}(\boldsymbol{x}^d_{I^*}, \boldsymbol{y}^d_{I^*})$ is fixed, and hence $g^I(\boldsymbol{x}^d, \boldsymbol{y}^d)$ is always consistent with some partial assignment σ with $\operatorname{fix}\sigma = I^*$ and $\operatorname{free}\sigma = I'$. We have $|I'| \geq |I|/2$ by (Q1). For each $i \in I'$, note that $\boldsymbol{x}^d_i$ is always in T_{ℓ_i} since we conditioned on $\boldsymbol{\ell} = \ell$. Note that (i*) for $\boldsymbol{x}^d$ holds by Claim 4.8. To see that (ii*) for $\boldsymbol{y}^d$ holds, pick any ℓ' that agrees with ℓ on I' and such that for every $i \in I^*$ the ℓ'_i bucket of $\boldsymbol{y}^d_i$ is not fixed—this is possible since each $\boldsymbol{y}^d_i$ has $m^{1/2}$ buckets but at most $2n^3 < m^{1/2}$ fixed buckets by (P1), hence at least one unfixed bucket. Since the bucket union $T_{\ell'}$ contains no fixed buckets of $\boldsymbol{y}^d$, we have $\mathbf{D}_\infty(\boldsymbol{y}^d_T) \leq \mathbf{D}_\infty(\boldsymbol{y}^d_{T_{\ell'}}) \leq 1$ by (P2).

Proof of Claim 4.6. We use a process highly reminiscent of the "density-restoring partition" process described in section 3.3. We maintain an event E which is initially all of $(\{0,1\}^m)^I$.

```
While $\mathbf{Pr}[\boldsymbol{Y}_I \in E] > \varepsilon/3$:
(1) Choose a maximal set of pairwise disjoint bucket unions $\mathcal{T} = \{T_{\ell^1}, \ldots, T_{\ell^k}\}$
    with the property that $\mathbf{D}_\infty(\boldsymbol{Y}_{\cup\mathcal{T}} \mid E) > k$ (possibly $\mathcal{T} = \emptyset$) and let $\beta \in$
    $\{0,1\}^{\cup\mathcal{T}}$ be an outcome witnessing this: $\mathbf{Pr}[\boldsymbol{Y}_{\cup\mathcal{T}} = \beta \mid E] > 2^{-(k|I|m^{1/2}-k)}$.
(2) Output the distribution $(\boldsymbol{Y}_I \mid \boldsymbol{Y}_{\cup\mathcal{T}} = \beta, E)$ with associated probability
    $\mathbf{Pr}[\boldsymbol{Y}_{\cup\mathcal{T}} = \beta, E] > 0$.
(3) Update $E \leftarrow \big\{y_I \in E \,:\, y_{\cup\mathcal{T}} \neq \beta\big\}$.
Output the distribution $(\boldsymbol{Y}_I \mid E)$ with associated probability $\mathbf{Pr}[\boldsymbol{Y}_I \in E]$ if the latter
is nonzero.
```

The distributions output throughout the process are the $\boldsymbol{y}^c$'s; note that with the associated probabilities, they indeed form a decomposition of $\boldsymbol{Y}_I$. Each time (1) is executed, we have

$$k < \mathbf{D}_\infty(\boldsymbol{Y}_{\cup\mathcal{T}} \mid E) \leq \mathbf{D}_\infty(\boldsymbol{Y}_I) + \log(1/\mathbf{Pr}[\boldsymbol{Y}_I \in E]) \leq n^3 + 1 + \log(3/\varepsilon) \leq 2n^3.$$

Also, any $\boldsymbol{y}^c = (\boldsymbol{Y}_I \mid \boldsymbol{Y}_{\cup\mathcal{T}} = \beta,\, E)$ output in (2) has the property that for any bucket union T_ℓ not containing fixed buckets, $\mathbf{D}_\infty(\boldsymbol{y}^c_{T_\ell}) \leq 1$. To see this, first note that T_ℓ is disjoint from $\cup\mathcal{T}$ since the latter buckets are fixed to β. If $\mathbf{D}_\infty(\boldsymbol{y}^c_{T_\ell}) > 1$ were witnessed by some $\gamma \in \{0,1\}^{T_\ell}$, then

$$\begin{aligned}\mathbf{Pr}[\boldsymbol{Y}_{(\cup\mathcal{T})\cup T_\ell} = \beta\gamma \mid E] &= \mathbf{Pr}[\boldsymbol{Y}_{\cup\mathcal{T}} = \beta \mid E] \cdot \mathbf{Pr}[\boldsymbol{Y}_{T_\ell} = \gamma \mid \boldsymbol{Y}_{\cup\mathcal{T}} = \beta,\, E] \\ &> 2^{-(k|I|m^{1/2}-k)} \cdot 2^{-(|I|m^{1/2}-1)} = 2^{-((k+1)|I|m^{1/2}-(k+1))},\end{aligned}$$

and so $\mathbf{D}_\infty(\boldsymbol{Y}_{(\cup\mathcal{T})\cup T_\ell} \mid E) > k+1$, which would contradict the maximality of k since $\{T_{\ell^1}, \ldots, T_{\ell^k}, T_\ell\}$ is a set of pairwise disjoint bucket unions. □

In the proofs of both Claim 4.7 and Claim 4.8, we use the chain rule for min-entropy [41, Lem. 6.30], which states that if $\boldsymbol{a}$ and $\boldsymbol{b}$ are any joint random variables, then for any $\delta > 0$, with probability at least $1-\delta$ over $a \sim \boldsymbol{a}$ we have $\mathbf{D}_\infty(\boldsymbol{b} \mid \boldsymbol{a} = a) \leq \mathbf{D}_\infty(\boldsymbol{ab}) + \log(1/\delta)$.

Proof of Claim 4.7. Assume that for each coordinate $i \in I$, $\boldsymbol{y}^c_i$ has at most $2n^3$ fixed buckets. Since $\boldsymbol{X}_I$ is 0.9-essentially-dense, $\boldsymbol{\ell}$ is 0.8-essentially-dense (for each nonempty $H \subseteq I$,

$$\mathbf{D}_\infty(\boldsymbol{\ell}_H) \leq \mathbf{D}_\infty(\boldsymbol{X}_H) \leq 0.1|H|\log m + 1 = 0.2|H|\log(m^{1/2}) + 1$$

holds). Thus, the probability that $T_{\boldsymbol{\ell}}$ hits fixed buckets in all coordinates in some set $H \subseteq I$ is at most the number of ways of choosing a fixed bucket from each of those coordinates ($\leq (2n^3)^{|H|}$) times the maximum probability that $T_{\boldsymbol{\ell}}$ hits all the chosen buckets ($\leq 2^{-(0.8|H|\log(m^{1/2})-1)}$ since $\boldsymbol{\ell}$ is 0.8-essentially-dense). We can now calculate

$$\begin{aligned}&\mathbf{Pr}[T_{\boldsymbol{\ell}} \text{ hits} \geq |I|/2 \text{ fixed buckets}] \\ &\leq \sum_{H\subseteq I,\,|H|=|I|/2} \mathbf{Pr}[T_{\boldsymbol{\ell}} \text{ hits fixed buckets in coordinates } H] \\ &\leq \binom{|I|}{|I|/2} \cdot (2n^3)^{|I|/2} \cdot 2^{-(0.8(|I|/2)\log(m^{1/2})-1)} \\ &\leq 2^{|I|} \cdot 2^{1.5|I|\log n+1} \cdot 2^{-(51.2|I|\log n-1)} \qquad (\text{using } m = n^{256}) \\ &\leq 2^{|I|-49.7|I|\log n+2} \\ &\leq \varepsilon/6.\end{aligned}$$

For convenience, we assumed above that $|I|$ is even; if $|I|$ is odd (including the case $|I|=1$), the same calculation works with $\lceil |I|/2 \rceil$ instead of $|I|/2$.

(Q2) follows by a direct application of the chain rule for min-entropy: with probability at least $1-\varepsilon/6$ over $\ell \sim \boldsymbol{\ell}$, we have

$$\begin{aligned}\mathbf{D}_\infty(\boldsymbol{r} \mid \boldsymbol{\ell} = \ell) &\leq \mathbf{D}_\infty(\boldsymbol{X}_I) + \log(6/\varepsilon) \\ &\leq \big(0.1|I|\log m + 1\big) + \big(5|I|\log n + 4\big) \\ &\leq 0.25|I|\log(m^{1/2}).\end{aligned}$$

By a union bound, with probability at least $1-\varepsilon/3$ over $\boldsymbol{\ell}$, (Q1) and (Q2) hold simultaneously. □

Proof of Claim 4.8. This is again a direct application of the chain rule for min-entropy: with probability at least $1-\varepsilon/3$ over $r_{I^*} \sim (\boldsymbol{r}_{I^*} \mid \boldsymbol{\ell} = \ell)$, we have

$$\begin{aligned}\mathbf{D}_\infty(\boldsymbol{r}_{I'} \mid \boldsymbol{\ell} = \ell,\, \boldsymbol{r}_{I^*} = r_{I^*}) &\le \mathbf{D}_\infty(\boldsymbol{r} \mid \boldsymbol{\ell} = \ell) + \log(3/\varepsilon)\\ &\le \big(0.25|I|\log(m^{1/2})\big) + \big(5|I|\log n + 3\big)\\ &\le 0.6|I'|\log(m^{1/2}),\end{aligned}$$

where the middle inequality uses (Q2), and the last inequality uses (Q1) ($|I'| \ge |I|/2$) and $m = n^{256}$. ∎

4.6. Pointwise uniformity from parities.

Lemma 4.9 (pointwise uniformity from parities). *If a random variable $\boldsymbol{z}_J$ over $\{0,1\}^J$ satisfies $\big|\mathbf{E}\big[\chi(\boldsymbol{z}_I)\big]\big| \le 2^{-5|I|\log n}$ for every nonempty $I \subseteq J$, then $\boldsymbol{z}_J$ is $1/n^3$-pointwise-close to uniform.*

Proof (from [21, sect. 2.2]*).* We let $\varepsilon := 1/n^3$ and write $\boldsymbol{z}_J$ as $\boldsymbol{z}$ throughout the proof. We think of the distribution of $\boldsymbol{z}$ as a function $\mathcal{D}\colon \{0,1\}^J \to [0,1]$ and write it in the Fourier basis as

$$\mathcal{D}(z) = \sum_{I \subseteq J} \widehat{\mathcal{D}}(I)\chi_I(z),$$

where $\chi_I(z) := (-1)^{\oplus(z_I)}$ and $\widehat{\mathcal{D}}(I) := 2^{-|J|}\sum_z \mathcal{D}(z)\chi_I(z) = 2^{-|J|}\cdot\mathbf{E}[\chi_I(\boldsymbol{z})]$. Note that $\widehat{\mathcal{D}}(\emptyset) = 2^{-|J|}$ because $\mathcal{D}$ is a distribution. Our assumption says that for all nonempty $I \subseteq J$, $2^{|J|}\cdot|\widehat{\mathcal{D}}(I)| \le 2^{-5|I|\log n}$, which is at most $\varepsilon 2^{-2|I|\log|J|}$. Hence,

$$\begin{aligned}2^{|J|}\textstyle\sum_{I\neq\emptyset}|\widehat{\mathcal{D}}(I)| &\le \varepsilon\sum_{I\neq\emptyset}2^{-2|I|\log|J|} = \varepsilon\sum_{k=1}^{|J|}\binom{|J|}{k}2^{-2k\log|J|}\\ &\le \varepsilon\sum_{k=1}^{|J|}2^{-k\log|J|} \le \varepsilon.\end{aligned}$$

We use this to show that $\big|\mathcal{D}(z) - 2^{-|J|}\big| \le \varepsilon 2^{-|J|}$ for all $z \in \{0,1\}^J$, which proves the lemma. To this end, let $\mathcal{U}$ denote the uniform distribution (note that $\widehat{\mathcal{U}}(I) = 0$ for all nonempty $I \subseteq J$) and let $\mathbb{1}_z$ denote the indicator for z defined by $\mathbb{1}_z(z) = 1$ and $\mathbb{1}_z(z') = 0$ for $z' \neq z$ (note that $|\widehat{\mathbb{1}}_z(I)| = 2^{-|J|}$ for all I). We can now calculate

$$\begin{aligned}\big|\mathcal{D}(z) - 2^{-|J|}\big| &= \big|\langle\mathbb{1}_z,\mathcal{D}\rangle - \langle\mathbb{1}_z,\mathcal{U}\rangle\big| = |\langle\mathbb{1}_z,\mathcal{D}-\mathcal{U}\rangle| = 2^{|J|}\cdot|\langle\widehat{\mathbb{1}}_z,\widehat{\mathcal{D}}-\widehat{\mathcal{U}}\rangle|\\ &\le 2^{|J|}\cdot\sum_{I\neq\emptyset}|\widehat{\mathbb{1}}_z(I)|\cdot|\widehat{\mathcal{D}}(I)| = \sum_{I\neq\emptyset}|\widehat{\mathcal{D}}(I)| \le \varepsilon 2^{-|J|}.\end{aligned}$$ ∎

5. Applications.

In this section, we collect some recent results in communication complexity, which we can derive (often with simplifications) from our lifting theorem.

Classical vs. quantum. Anshu et al. [5] gave a nearly 2.5th power total function separation between quantum and classical randomized protocols. Our lifting theorem can reproduce this separation by lifting an analogous separation in query complexity due to Aaronson, Ben-David, and Kothari [2]. Let us also mention that Aaronson and Ambainis [1] conjectured that a slight generalization of Forrelation witnesses an $O(\log n)$-vs.-$\tilde{\Omega}(n)$ quantum/classical query separation. If true, our lifting theorem implies that "2.5" can be improved to "3" above; see [2] for a discussion. (Such an improvement is not black-box implied by the techniques of Anshu et al. [5].)

Raz [32] gave an exponential partial function separation between quantum and classical randomized protocols. Our lifting theorem can reproduce this separation by lifting, say, the FORRELATION partial function [1], which witnesses a 1-vs.-$\tilde{\Omega}(\sqrt{n})$ separation for quantum/classical query complexity. However, qualitatively stronger separations are known [26, 17] where the quantum protocol can be taken to be *one-way* or even *simultaneous*.

Partition numbers. Anshu et al. [5] gave a nearly quadratic separation between (the log of) the *two-sided partition number* (number of monochromatic rectangles needed to partition the domain of F) and randomized communication complexity. This result now follows by lifting an analogous separation in query complexity due to Ambainis, Kokainis, and Kothari [4].

In [19], a nearly quadratic separation was shown between (the log of) the *one-sided partition number* (number of rectangles needed to partition $F^{-1}(1)$) and randomized communication complexity. This separation question can be equivalently phrased as proving randomized lower bounds for the *Clique vs. Independent Set* game [46]. This result now follows by lifting an analogous separation in query complexity, obtained in several papers [19, 3, 2]; it was previously shown using the lifting theorem of [21], which requires a query lower bound in a model stronger than $\mathsf{BPP}^{\mathsf{dt}}$.

Approximate Nash equilibria. Babichenko and Rubinstein [6] showed a randomized communication lower bound for finding an approximate Nash equilibrium in a two-player game. Their approach was to show a lower bound for a certain query version of the PPAD-complete END-OF-LINE problem, and then lift this lower bound into communication complexity using [21]. However, as in the above Clique vs. Independent Set result, the application of [21] here requires that the query lower bound be established for a model stronger than $\mathsf{BPP}^{\mathsf{dt}}$, which required some additional busywork. Our lifting theorem can be used to streamline their proof.

Direct sum. In [9], our lifting theorem has been applied to show that there exists a total two-party function F such that $\mathsf{BPP}^{\mathsf{cc}}(F^k) = \Theta(k \log k \cdot \mathsf{BPP}^{\mathsf{cc}}(F))$ holds for all $k \leq 2^{n^{O(1)}}$, answering a question of [16].

Acknowledgments. Thanks to Shalev Ben-David and Robin Kothari for quantum references. Thanks to Anurag Anshu, Rahul Jain, Raghu Meka, Aviad Rubinstein, and Henry Yuen for discussions.

REFERENCES

[1] S. AARONSON AND A. AMBAINIS, *Forrelation: A problem that optimally separates quantum from classical computing*, SIAM J. Comput., 47 (2018), pp. 982–1038, https://doi.org/10.1137/15M1050902.

[2] S. AARONSON, S. BEN-DAVID, AND R. KOTHARI, *Separations in query complexity using cheat sheets*, in Proceedings of the 48th Symposium on Theory of Computing (STOC), ACM, 2016, pp. 863–876, https://doi.org/10.1145/2897518.2897644.

[3] A. AMBAINIS, K. BALODIS, A. BELOVS, T. LEE, M. SANTHA, AND J. SMOTROVS, *Separations in query complexity based on pointer functions*, J. ACM, 64 (2017), pp. 32:1–32:24, https://doi.org/10.1145/3106234.

[4] A. AMBAINIS, M. KOKAINIS, AND R. KOTHARI, *Nearly optimal separations between communication (or query) complexity and partitions*, in Proceedings of the 31st Computational Complexity Conference (CCC), Schloss Dagstuhl, 2016, pp. 4:1–4:14, https://doi.org/10.4230/LIPIcs.CCC.2016.4.

[5] A. Anshu, A. Belovs, S. Ben-David, M. Göös, R. Jain, R. Kothari, T. Lee, and M. Santha, *Separations in communication complexity using cheat sheets and information complexity*, in Proceedings of the 57th Symposium on Foundations of Computer Science (FOCS), IEEE, 2016, pp. 555–564, https://doi.org/10.1109/FOCS.2016.66.

[6] Y. Babichenko and A. Rubinstein, *Communication complexity of approximate Nash equilibria*, in Proceedings of the 49th Symposium on Theory of Computing (STOC), ACM, 2017, pp. 878–889, https://doi.org/10.1145/3055399.3055407.

[7] Z. Bar-Yossef, T. Jayram, R. Kumar, and D. Sivakumar, *An information statistics approach to data stream and communication complexity*, J. Comput. System Sci., 68 (2004), pp. 702–732, https://doi.org/10.1016/j.jcss.2003.11.006.

[8] S. Ben-David and R. Kothari, *Randomized query complexity of sabotaged and composed functions*, Theory Comput., 14 (2018), pp. 1–27, https://doi.org/10.4086/toc.2018.v014a005.

[9] E. Blais and J. Brody, *Optimal separation and strong direct sum for randomized query complexity*, in Proceedings of the 34th Computational Complexity Conference (CCC), Schloss Dagstuhl, 2019, pp. 29:1–29:17, https://doi.org/10.4230/LIPIcs.CCC.2019.29.

[10] H. Buhrman and R. de Wolf, *Complexity measures and decision tree complexity: A survey*, Theoret. Comput. Sci., 288 (2002), pp. 21–43, https://doi.org/10.1016/S0304-3975(01)00144-X.

[11] A. Chakrabarti and O. Regev, *An optimal lower bound on the communication complexity of gap-Hamming-distance*, SIAM J. Comput., 41 (2012), pp. 1299–1317, https://doi.org/10.1137/120861072.

[12] S. O. Chan, J. Lee, P. Raghavendra, and D. Steurer, *Approximate constraint satisfaction requires large LP relaxations*, J. ACM, 63 (2016), pp. 34:1–34:22, https://doi.org/10.1145/2811255.

[13] A. Chattopadhyay, Y. Filmus, S. Koroth, O. Meir, and T. Pitassi, *Query-to-Communication Lifting Using Low-Discrepancy Gadgets*, Tech. Report TR19-103, Electronic Colloquium on Computational Complexity (ECCC), 2019, https://eccc.weizmann.ac.il/report/2019/103/.

[14] A. Chattopadhyay, M. Koucký, B. Loff, and S. Mukhopadhyay, *Simulation theorems via pseudo-random properties*, Comput. Complexity, 28 (2019), pp. 617–659, https://doi.org/10.1007/s00037-019-00190-7,

[15] S. de Rezende, J. Nordström, and M. Vinyals, *How limited interaction hinders real communication (and what it means for proof and circuit complexity)*, in Proceedings of the 57th Symposium on Foundations of Computer Science (FOCS), IEEE, 2016, pp. 295–304, https://doi.org/10.1109/FOCS.2016.40.

[16] T. Feder, E. Kushilevitz, M. Naor, and N. Nisan, *Amortized communication complexity*, SIAM J. Comput., 24 (1995), pp. 736–750, https://doi.org/10.1137/S0097539792235864.

[17] D. Gavinsky, *Entangled simultaneity versus classical interactivity in communication complexity*, in Proceedings of the 48th Symposium on Theory of Computing (STOC), ACM, 2016, pp. 877–884, https://doi.org/10.1145/2897518.2897545.

[18] M. Göös, *Lower bounds for clique vs. independent set*, in Proceedings of the 56th Symposium on Foundations of Computer Science (FOCS), IEEE, 2015, pp. 1066–1076, https://doi.org/10.1109/FOCS.2015.69.

[19] M. Göös, T. Jayram, T. Pitassi, and T. Watson, *Randomized communication vs. partition number*, ACM Trans. Comput. Theory, 10 (2018), pp. 4:1–4:20, https://doi.org/10.1145/3170711.

[20] M. Göös, P. Kamath, T. Pitassi, and T. Watson, *Query-to-communication lifting for P^{NP}*, Comput. Complexity, 28 (2019), pp. 113–144, https://doi.org/10.1007/s00037-018-0175-5.

[21] M. Göös, S. Lovett, R. Meka, T. Watson, and D. Zuckerman, *Rectangles are nonnegative juntas*, SIAM J. Comput., 45 (2016), pp. 1835–1869, https://doi.org/10.1137/15M103145X.

[22] M. Göös, T. Pitassi, and T. Watson, *Deterministic communication vs. partition number*, SIAM J. Comput., 47 (2018), pp. 2435–2450, https://doi.org/10.1137/16M1059369.

[23] H. Hatami, K. Hosseini, and S. Lovett, *Structure of protocols for XOR functions*, SIAM J. Comput., 47 (2018), pp. 208–217, https://doi.org/10.1137/17M1136869.

[24] S. Jukna, *Boolean Function Complexity: Advances and Frontiers*, Algorithms Combin. 27, Springer, 2012.

[25] B. Kalyanasundaram and G. Schnitger, *The probabilistic communication complexity of set intersection*, SIAM J. Discrete Math., 5 (1992), pp. 545–557, https://doi.org/10.1137/0405044.

[26] B. Klartag and O. Regev, *Quantum one-way communication can be exponentially stronger than classical communication*, in Proceedings of the 43rd Symposium on Theory of Computing (STOC), ACM, 2011, pp. 31–40, https://doi.org/10.1145/1993636.1993642.

[27] P. Kothari, R. Meka, and P. Raghavendra, *Approximating rectangles by juntas and weakly-exponential lower bounds for LP relaxations of CSPs*, in Proceedings of the 49th Symposium on Theory of Computing (STOC), ACM, 2017, pp. 590–603, https://doi.org/10.1145/3055399.3055438.

[28] E. Kushilevitz and N. Nisan, *Communication Complexity*, Cambridge University Press, 1997.

[29] J. Lee, P. Raghavendra, and D. Steurer, *Lower bounds on the size of semidefinite programming relaxations*, in Proceedings of the 47th Symposium on Theory of Computing (STOC), ACM, 2015, pp. 567–576, https://doi.org/10.1145/2746539.2746599.

[30] R. O'Donnell, *Analysis of Boolean Functions*, Cambridge University Press, 2014.

[31] A. Rao and A. Yehudayoff, *Communication Complexity*, in preparation, 2017.

[32] R. Raz, *Exponential separation of quantum and classical communication complexity*, in Proceedings of the 31st Symposium on Theory of Computing (STOC), ACM, 1999, pp. 358–367, https://doi.org/10.1145/301250.301343.

[33] R. Raz and P. McKenzie, *Separation of the monotone NC hierarchy*, Combinatorica, 19 (1999), pp. 403–435, https://doi.org/10.1007/s004930050062.

[34] A. Razborov, *On the distributional complexity of disjointness*, Theoret. Comput. Sci., 106 (1992), pp. 385–390, https://doi.org/10.1016/0304-3975(92)90260-M.

[35] A. Razborov and A. Sherstov, *The sign-rank of AC^0*, SIAM J. Comput., 39 (2010), pp. 1833–1855, https://doi.org/10.1137/080744037.

[36] R. Robere, T. Pitassi, B. Rossman, and S. Cook, *Exponential lower bounds for monotone span programs*, in Proceedings of the 57th Symposium on Foundations of Computer Science (FOCS), IEEE, 2016, pp. 406–415, https://doi.org/10.1109/FOCS.2016.51.

[37] R. Shaltiel, *Towards proving strong direct product theorems*, Comput. Complexity, 12 (2003), pp. 1–22, https://doi.org/10.1007/s00037-003-0175-x.

[38] A. Sherstov, *The pattern matrix method*, SIAM J. Comput., 40 (2011), pp. 1969–2000, https://doi.org/10.1137/080733644.

[39] A. Sherstov, *The communication complexity of gap Hamming distance*, Theory Comput., 8 (2012), pp. 197–208, https://doi.org/10.4086/toc.2012.v008a008.

[40] Y. Shi and Y. Zhu, *Quantum communication complexity of block-composed functions*, Quantum Inf. Comput., 9 (2009), pp. 444–460.

[41] S. Vadhan, *Pseudorandomness*, Found. Trends Theoret. Comput. Sci., 7 (2012), pp. 1–336, https://doi.org/10.1561/0400000010.

[42] N. Vereshchagin, *Relativizability in complexity theory*, in Provability, Complexity, Grammars, Amer. Math. Soc. Transl. Ser. 2, 192, American Mathematical Society, 1999, pp. 87–172.

[43] T. Vidick, *A concentration inequality for the overlap of a vector on a large set, with application to the communication complexity of the gap-Hamming-distance problem*, Chicago J. Theoret. Comput. Sci., 2013 (2013), pp. 1–12, https://doi.org/10.4086/cjtcs.2012.001.

[44] X. Wu, *The Uniform Marginals Lemma in "Query-to-Communication Lifting for BPP"*, 2018, https://www.andrew.cmu.edu/user/xinyuw1/papers/uniform-marginals-lemma.pdf.

[45] X. Wu, P. Yao, and H. Yuen, *Raz–McKenzie Simulation with the Inner Product Gadget*, Tech. Report TR17-010, Electronic Colloquium on Computational Complexity (ECCC), 2017, https://eccc.weizmann.ac.il/report/2017/010/.

[46] M. Yannakakis, *Expressing combinatorial optimization problems by linear programs*, J. Comput. System Sci., 43 (1991), pp. 441–466, https://doi.org/10.1016/0022-0000(91)90024-Y.

[47] A. Yao, *Some complexity questions related to distributive computing*, in Proceedings of the 11th Symposium on Theory of Computing (STOC), ACM, 1979, pp. 209–213, https://doi.org/10.1145/800135.804414.