SIAM J. COMPUT.
Vol. 45, No. 5, pp. 1835–1869
© 2016 Society for Industrial and Applied Mathematics

RECTANGLES ARE NONNEGATIVE JUNTAS*

MIKA GÖÖS†, SHACHAR LOVETT‡, RAGHU MEKA§, THOMAS WATSON¶, AND DAVID ZUCKERMAN‖

Abstract. We develop a new method to prove communication lower bounds for composed functions of the form $f \circ g^n$, where f is any boolean function on n inputs and g is a sufficiently "hard" two-party gadget. Our main structure theorem states that each rectangle in the communication matrix of $f \circ g^n$ can be simulated by a *nonnegative combination of juntas*. This is a new formalization for the intuition that each low-communication randomized protocol can only "query" a few inputs of f as encoded by the gadget g. Consequently, we characterize the communication complexity of $f \circ g^n$ in all known one-sided (i.e., not closed under complement) zero-communication models by a corresponding query complexity measure of f. These models in turn capture important lower bound techniques such as corruption, smooth rectangle bound, relaxed partition bound, and extended discrepancy. As applications, we resolve several open problems from prior work. We show that $\mathsf{SBP}^{\mathsf{cc}}$ (a class characterized by corruption) is not closed under intersection. An immediate corollary is that $\mathsf{MA}^{\mathsf{cc}} \neq \mathsf{SBP}^{\mathsf{cc}}$. These results answer questions of Klauck [*Proceedings of the* 18*th Conference on Computational Complexity (CCC)*, IEEE Computer Society, Los Alamitos, CA, 2003, pp. 118–134] and Böhler, Glasser, and Meister [*J. Comput. System Sci.*, 72 (2006), pp. 1043–1076]. We also show that the approximate nonnegative rank of partial boolean matrices does not admit efficient error reduction. This answers a question of Kol et al. [*Proceedings of the* 41*st International Colloquium on Automata, Languages, and Programming (ICALP)*, Springer, Berlin, 2014, pp. 701–712] for partial matrices. In subsequent work, our structure theorem has been applied to resolve the communication complexity of the clique versus independent set problem.

Key words. rectangles, nonnegative, juntas

AMS subject classifications. 68Q10, 68Q15, 68Q17, 68Q87

DOI. 10.1137/15M103145X

1. Introduction.

Many functions studied in communication complexity (e.g., equality, set disjointness, inner product, gap Hamming; see [38, 31]) are *composed functions* of the form $f \circ g^n$, where $f\colon \{0,1\}^n \to \{0,1,*\}$ is a partial function and $g\colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ is some small two-party function, often called a *gadget*. Here Alice and Bob are given inputs $x \in \mathcal{X}^n$ and $y \in \mathcal{Y}^n$, respectively; we think of the inputs as being partitioned into *blocks* $x_i \in \mathcal{X}$ and $y_i \in \mathcal{Y}$ for $i \in [n]$. Their goal is to compute

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

*Received by the editors July 20, 2015; accepted for publication (in revised form) May 17, 2016; published electronically October 19, 2016. Most of this work was done while the authors were visiting Microsoft Research, Silicon Valley Lab at various times. An extended abstract of this work was published by Göös et al., in *Proceedings of the* 47*th Symposium on Theory of Computing (STOC)*, ACM, New York, 2015, pp. 257–266 [21].

http://www.siam.org/journals/sicomp/45-5/M103145.html

Funding: The second author was supported in part by NSF CAREER award 1350481. The fifth author was supported in part by NSF grant CCF-1218723.

†Department of Computer Science, University of Toronto, Toronto, ON, Canada (mika.goos@mail.utoronto.ca, http://www.cs.utoronto.ca/~mgoos/).

‡Computer Science and Engineering, University of California, San Diego, La Jolla, CA 92093 (shachar.lovett@gmail.com, http://cseweb.ucsd.edu/~slovett/).

§Department of Computer Science, University of California, Los Angeles, CA 90095 (raghuvardhan@gmail.com, http://www.raghumeka.org/).

¶Department of Computer Science, University of Memphis, Memphis, TN 38152 (Thomas.Watson@memphis.edu, http://umdrive.memphis.edu/twwtson1/public/).

‖Department of Computer Science, University of Texas at Austin, Austin, TX 78705 (diz@cs.utexas.edu, https://www.cs.utexas.edu/~diz/).

Intuitively, the difficulty in computing $f \circ g^n$ stems from the fact that for any i, the ith input $z_i := g(x_i, y_i)$ to f remains unknown to either party until they decide to communicate enough information about x_i and y_i. Indeed, an educated guess is that—assuming g is chosen carefully—the communication complexity of $f \circ g^n$ should be explained by some *query* measure of f.

This work is about formalizing the above intuition. Our main result is the following.

Simulation theorem (Theorem 2, informally). *Many types of randomized protocols for $f \circ g^n$ can be simulated by a corresponding type of randomized decision tree for f.*

This result makes it easy to prove strong lower bounds for $f \circ g^n$ in all known one-sided (and some two-sided) *zero-communication* models. Here a zero-communication protocol is understood in the sense of [32] as a probability distribution over (labeled) rectangles $R = X \times Y$ (where $X \subseteq \mathcal{X}^n$ and $Y \subseteq \mathcal{Y}^n$) together with some acceptance criterion (and hence no communication is needed for Alice and Bob to select a rectangle, since it can be sampled with public randomness). Such models can be used to capture all known rectangle-based lower bound techniques used in communication complexity. This includes widely studied measures such as corruption [67, 6, 49, 33, 7, 57, 25], smooth rectangle bound [29, 35, 10, 30, 28, 37], relaxed partition bound [32], and extended discrepancy [33, 16]; see [29] for an extensive catalog. The simulation theorem applies to all these measures: it reduces the task of understanding a specific communication complexity measure of $f \circ g^n$ to the task of understanding a corresponding query complexity measure of f, which is typically a far easier task.

1.1. Main structural result: Junta theorem. In order to motivate our approach (and to introduce notation), we start by reviewing some previous influential work in communication complexity.

Prior work: Approximation by polynomials. A long line of prior work has developed a framework of *polynomial approximation* to analyze the communication complexity of composed functions. Building on the work of Razborov [50], a general framework was introduced by Sherstov [54, 55] (called the pattern matrix method) and independently by Shi and Zhu [60] (called the block-composition method). See also the survey [53]. Both methods have since been studied in the two-party setting [42, 51, 56] and also the multiparty setting [40, 3, 13, 58, 59, 46].

One way to phrase the approach taken in these works (a "primal" point of view championed in [58]) is as follows. Let Π be a randomized protocol and let $\mathrm{acc}_\Pi(x, y)$ denote the probability that Π accepts an input (x, y). For example, if Π computes a two-party function F with error at most $1/4$, then $\mathrm{acc}_\Pi(x, y) \in [3/4, 1]$ for every 1-input $(x, y) \in F^{-1}(1)$ and $\mathrm{acc}_\Pi(x, y) \in [0, 1/4]$ for every 0-input $(x, y) \in F^{-1}(0)$. When $F := f \circ g^n$ is a composed function, we can define $\mathrm{acc}_\Pi(z)$ for $z \in \mathrm{dom}\, f$ (domain of f) meaningfully as the probability that Π accepts a *random two-party encoding* of z. More specifically, letting $\mathbf{E}$ denote expectation and $\mathcal{U}_z$ the uniform distribution over $(g^n)^{-1}(z)$ we define

$$\mathrm{acc}_\Pi(z) := \mathop{\mathbf{E}}_{(\boldsymbol{x},\boldsymbol{y})\sim\mathcal{U}_z} \mathrm{acc}_\Pi(\boldsymbol{x}, \boldsymbol{y}).$$

The centerpiece in the framework is the following type of structure theorem: assuming g is chosen carefully, for any cost-c protocol Π there is a degree-$O(c)$ multivariate polynomial $p(z)$ such that $\mathrm{acc}_\Pi(z) \approx p(z)$. Here the approximation error is typically measured pointwise. Consequently, if f cannot be approximated pointwise with a low-degree polynomial, one obtains lower bounds against any bounded-error protocol computing $f \circ g^n$.

A technical convenience that will be useful for us is that since randomized protocols are essentially linear combinations of 0/1-labeled rectangles R, it suffices to study the acceptance probability of each individual rectangle R. More formally, it suffices to understand $\mathrm{acc}_R(z)$, defined as the probability that $(\boldsymbol{x}, \boldsymbol{y}) \in R$ for a random encoding $(\boldsymbol{x}, \boldsymbol{y}) \sim \mathcal{U}_z$ of z. Put succinctly,

$$\mathrm{acc}_R(z) \coloneqq \mathcal{U}_z(R).$$

An important feature of the polynomial framework is that it often yields tight lower bounds for *two-sided* (i.e., closed under complement) randomized models. However, polynomials are not always the most precise modeling choice when it comes to understanding *one-sided* (i.e., not closed under complement) randomized models, such as randomized generalizations of NP and measures like nonnegative rank.

This work: Approximation by conical juntas. In this work, we show that randomized protocols for composed functions can be simulated by *conical juntas*, a nonnegative analog of polynomials. Let $h\colon \{0,1\}^n \to \mathbb{R}_{\geq 0}$ be a function. We say that h is a *d-junta* if it only depends on at most d of its input bits—we stress that all juntas in this work are nonnegative by definition. More generally, we call h a *conical d-junta* if it lies in the nonnegative cone generated by d-juntas, i.e., if we can write $h = \sum_i a_i h_i$, where $a_i \geq 0$ are nonnegative coefficients and h_i are d-juntas. Equivalently, a conical d-junta can be viewed as a nonnegative combination of width-d conjunctions (i.e., functions of the form $(\ell_1 \wedge \cdots \wedge \ell_w)$, where $w \leq d$ and each ℓ_i is an input variable or its negation). Note that a conical d-junta is, in particular, a polynomial of degree at most d.

For concreteness, we state and prove our results for logarithmic-size inner-product gadgets. That is, throughout this work, we restrict our attention to the following setting of parameters:

(†)
- The gadget is given by $g(x, y) \coloneqq \langle x, y\rangle \bmod 2$, where $x, y \in \{0,1\}^b$.
- The block length $b = b(n)$ satisfies $b(n) \geq 100 \log n$.

(However, our results hold more generally whenever g is a sufficiently strong two-source extractor; see Remark 14. Further, lower bounds for the inner-product gadget as above can be used to get lower bounds for other gadgets with worse parameters. See section 1.4 for more discussion.)

We are now ready to state our key structural result. The result essentially characterizes the computational power of a single rectangle in the communication matrix of $f \circ g^n$. Note that the theorem makes no reference to f.

THEOREM 1 (junta theorem). *Assume (†). For any $d \geq 0$ and any rectangle R in the domain of g^n there exists a conical d-junta h such that, for all $z \in \{0,1\}^n$,*

$$\mathrm{acc}_R(z) \in (1 \pm 2^{-\Theta(b)}) \cdot h(z) \pm 2^{-\Theta(db)}. \tag{1}$$

Discussion. Theorem 1 is similar in spirit to the approach taken by Chan et al. [12]. They gave a black-box method for converting Sherali–Adams lower bounds into size lower bounds for extended formulations. A key step in their proof is to approximate a single nonnegative rank-1 matrix with a single junta. In our approach, we approximate a single rectangle with a whole nonnegative combination of juntas. This allows us to achieve better error bounds that yield tight characterizations for many communication models (as discussed in section 1.2 below). In the language of

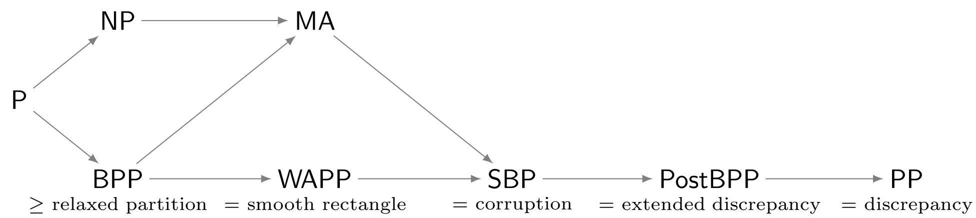

FIG. 1. *Models and lower bound methods at a glance. Arrows denote class inclusions.*

communication complexity, the lower bounds of [12] went up to about $\Omega(\log^2 n)$. See [12, section 3.1] for more discussion.

The additive error $2^{-\Theta(db)}$ in Theorem 1 is essentially optimal, and the same additive error appears in the polynomial approximation framework. The multiplicative error $(1 \pm 2^{-\Theta(b)})$ is new: this is the cost we end up incurring for using juntas instead of polynomials. Such multiplicative error does not appear in the polynomial approximation framework. Whether one can achieve better multiplicative accuracy in Theorem 1 is left as an open problem (see section 1.4).

Maybe the biggest drawback with Theorem 1 is that our proof assumes block length $b = \Omega(\log n)$ (cf. the pattern matrix method works even when $b = \Theta(1)$). Whether Theorem 1 (or some relaxed form of it) is true for $b = \Theta(1)$ is left as an open problem.

1.2. Communication versus query: Simulation theorem. The most intuitive way to formalize our simulation theorem is in terms of different randomized models of computation rather than in terms of different lower bound measures. Indeed, we consider several models originally introduced in the context of Turing machine complexity theory: for any such model $\mathcal{C}$ one can often associate, in a canonical fashion, a communication model $\mathcal{C}^{\mathsf{cc}}$ and a decision tree model $\mathcal{C}^{\mathsf{dt}}$. We follow the convention of using names of models as complexity measures so that $\mathcal{C}^{\mathsf{cc}}(F)$ denotes the communication complexity of F in model $\mathcal{C}^{\mathsf{cc}}$, and $\mathcal{C}^{\mathsf{dt}}(f)$ denotes the query complexity of f in model $\mathcal{C}^{\mathsf{dt}}$. In this work, we further identify $\mathcal{C}^{\mathsf{cc}}$ with the class of partial functions F with $\mathcal{C}^{\mathsf{cc}}(F) \leq \text{poly}(\log n)$. We stress that our complexity classes consist of partial functions (i.e., promise problems)—for total functions many surprising collapses are possible (e.g., $\mathsf{NP}^{\mathsf{cc}} \cap \mathsf{coNP}^{\mathsf{cc}} = \mathsf{P}^{\mathsf{cc}}$ for total functions [38, section 2.3]).

Our methods allow us to accurately analyze the models listed below (see also Figure 1). Our discussion in this introduction is somewhat informal; see section 3 for precise definitions.

- NP: *Nondeterminism.* We view an NP computation as a randomized computation where 1-inputs are accepted with nonzero probability and 0-inputs are accepted with zero probability. The communication analog $\mathsf{NP}^{\mathsf{cc}}$ was formalized in the work of Babai, Frankl, and Simon [6] that introduced communication complexity analogs of classical complexity classes.
- WAPP: *Weak almost-wide* PP [8]. A WAPP computation is a randomized computation such that 1-inputs are accepted with probability in $[(1-\epsilon)\alpha, \alpha]$, and 0-inputs are accepted with probability in $[0, \epsilon\alpha]$ where $\alpha = \alpha(n) > 0$ is arbitrary and $\epsilon < 1/2$ is a constant. The communication analog $\mathsf{WAPP}^{\mathsf{cc}}$ is equivalent to the (one-sided) *smooth rectangle bound* of Jain and Klauck [29] and also to *approximate nonnegative rank* by a result of Kol et al. [37]. We also study a two-sided model $\mathsf{WAPP} \cap \mathsf{coWAPP}$ whose communication analog

corresponds to the two-sided smooth rectangle bound, which was called the *relaxed partition bound* by [32].

- SBP: *Small bounded-error probability* [8]. An SBP computation is a randomized computation such that 1-inputs are accepted with probability in $[\alpha, 1]$ and 0-inputs are accepted with probability in $[0, \alpha/2]$, where $\alpha = \alpha(n) > 0$ is arbitrary. The communication analog $\mathsf{SBP}^{\mathsf{cc}}$ is equivalent to the (one-sided) *corruption bound* originally defined in [67] (see [25]).
- $\mathsf{PostBPP}$: *Postselected* BPP [1] (equivalent to $\mathsf{BPP}_{\mathsf{path}}$ [27]). A $\mathsf{PostBPP}$ computation is a randomized computation that may sometimes output $\perp$ (representing "abort" or "don't know"), but conditioned on not outputting $\perp$ the output is correct with probability at least 3/4. The communication analog $\mathsf{PostBPP}^{\mathsf{cc}}$ was first studied in [33] (under the name "approximate majority covers") and subsequently in [16] (under the generic name "zero-communication protocols"), where the term *extended discrepancy* was coined for the dual characterization of $\mathsf{PostBPP}^{\mathsf{cc}}$.

We apply the junta theorem to show that when $\mathcal{C}$ is one of the above models, any $\mathcal{C}^{\mathsf{cc}}$ protocol for $f \circ g^n$ can be converted into a corresponding $\mathcal{C}^{\mathsf{dt}}$ decision tree for f. Roughly speaking, this is because such a protocol can be formulated as a distribution over (labeled) rectangles, and each rectangle can be converted (via the junta theorem) into a distribution over conjunctions. Hence lower bounds on $\mathcal{C}^{\mathsf{cc}}(f \circ g^n)$ follow in a black-box way from lower bounds on $\mathcal{C}^{\mathsf{dt}}(f)$.

THEOREM 2 (simulation theorem). *Assume* (†). *For any partial* $f\colon \{0,1\}^n \to \{0,1,*\}$ *we have*

$$\mathcal{C}^{\mathsf{cc}}(f \circ g^n) = \Theta(\mathcal{C}^{\mathsf{dt}}(f) \cdot b) \qquad \text{for } \mathcal{C} \in \{\mathsf{NP}, \mathsf{WAPP}, \mathsf{SBP}\},$$
$$\mathcal{C}^{\mathsf{cc}}(f \circ g^n) \geq \Omega(\mathcal{C}^{\mathsf{dt}}(f) \cdot b) \qquad \text{for } \mathcal{C} = \mathsf{PostBPP}.$$

(Here we crucially ignore constant factors in the error parameter ϵ for $\mathcal{C} = \mathsf{WAPP}$.)

Naturally, the upper bounds in Theorem 2 follow from the fact that a communication protocol for $f \circ g^n$ can simulate the corresponding decision tree for f: when the decision tree queries the ith input of f, the protocol spends $b+1$ bits of communication to figure out $z_i = g(x_i, y_i)$ in a brute-force manner. (There is one subtlety concerning the two-sided model $\mathsf{PostBPP}$; see Remark 36.)

We also mention that the result for the simplest model $\mathcal{C} = \mathsf{NP}$ does not require the full power of the junta theorem: it is possible to prove it using only a proper subset of the ideas that we present for the other randomized models (see [18]).

1.3. Applications. Using the simulation theorem we can resolve several questions from prior work.

SBP *and corruption.* Our first application is the following.

THEOREM 3. $\mathsf{SBP}^{\mathsf{cc}}$ *is not closed under intersection.*

We prove this theorem by first giving an analogous lower bound for query complexity: there exists a partial f such that $\mathsf{SBP}^{\mathsf{dt}}(f) \leq O(1)$, but $\mathsf{SBP}^{\mathsf{dt}}(f_\wedge) \geq n^{\Omega(1)}$, where $f_\wedge\colon \{0,1\}^{2n} \to \{0,1,*\}$ is defined by $f_\wedge(z, z') \coloneqq f(z) \wedge f(z')$. This query separation alone yields via standard diagonalization (e.g., [2, section 5]) an oracle relative to which the classical complexity class SBP is not closed under intersection, solving an open problem posed by [8]. Applying the simulation theorem to $f \circ g^n$ and $f_\wedge \circ g^{2n} = (f \circ g^n)_\wedge$ we then obtain Theorem 3.

Theorem 3 has consequences for Arthur–Merlin communication ($\mathsf{MA}^{\mathsf{cc}}$, $\mathsf{AM}^{\mathsf{cc}}$) which has been studied in [33, 48, 2, 17, 36, 26, 24]. Namely, Klauck [33] asked (using the language of uniform threshold covers) whether the known inclusion $\mathsf{MA}^{\mathsf{cc}} \subseteq \mathsf{SBP}^{\mathsf{cc}}$ is strict. (This was also reasked in [25].) Put diffferently, is corruption a complete lower bound method for $\mathsf{MA}^{\mathsf{cc}}$ up to polynomial factors? Since $\mathsf{MA}^{\mathsf{cc}}$ is closed under intersection, we conclude that the answer is "no."

COROLLARY 4. $\mathsf{SBP}^{\mathsf{cc}} \not\subseteq \mathsf{MA}^{\mathsf{cc}}$.

Proving explicit lower bounds for $\mathsf{AM}^{\mathsf{cc}}$ remains one of the central challenges in communication complexity. Motivated by this [24] studied a certain unambiguous restriction of $\mathsf{AM}^{\mathsf{cc}}$, denoted $\mathsf{UAM}^{\mathsf{cc}}$, as a stepping stone towards $\mathsf{AM}^{\mathsf{cc}}$. They asked whether $\mathsf{UAM}^{\mathsf{cc}} \subseteq \mathsf{SBP}^{\mathsf{cc}}$. In other words, does corruption give lower bounds against $\mathsf{UAM}^{\mathsf{cc}}$ in a black-box fashion? They showed that the answer is "no" for query complexity. Using the simulation theorem it is now straightforward to convert this result into an analogous communication separation.

COROLLARY 5. $\mathsf{UAM}^{\mathsf{cc}} \not\subseteq \mathsf{SBP}^{\mathsf{cc}}$.

Intriguingly, we still lack $\mathsf{UAM}^{\mathsf{cc}}$ lower bounds for set disjointness. Corollary 5 implies that such lower bounds cannot be blindly derived from Razborov's corruption lemma [49].

WAPP *and nonnegative rank.* Kol et al. [37] asked whether the error in the definition of WAPP can be efficiently *amplified*, i.e., whether the parameter ϵ can be *reduced* without blowing up the cost too much. It is known that such amplification is possible for the closely related two-sided model AWPP, *almost-wide* PP (related to smooth discrepancy and approximate rank), using "amplification polynomials"; see [15, section 3] (or [39, section 3.2] and [4] for approximate rank). In [37] it was shown that no one-sided analog of amplification polynomials exists, ruling out one particular approach to amplification.

We show unconditionally that $\mathsf{WAPP}^{\mathsf{cc}}$ (and hence rank^+_ϵ, approximate nonnegative rank) does not admit efficient error amplification in the case of partial functions. For total functions, this at least shows that no "pointwise" method can be used to amplify ϵ, since such methods would also work for partial functions. We write $\mathsf{WAPP}^{\mathsf{cc}}_\epsilon$ for the measure corresponding to error ϵ.

THEOREM 6. *For all constants $0 < \epsilon < \delta < 1/2$ there exists a two-party partial function F such that $\mathsf{WAPP}^{\mathsf{cc}}_\delta(F) \le O(\log n)$ but $\mathsf{WAPP}^{\mathsf{cc}}_\epsilon(F) \ge \Omega(n)$.*

COROLLARY 7. *For all constants $0 < \epsilon < \delta < 1/2$ there exists a partial boolean matrix F such that $\mathrm{rank}^+_\delta(F) \le n^{O(1)}$ but $\mathrm{rank}^+_\epsilon(F) \ge 2^{\Omega(n)}$.*

In order to conclude Corollary 7 from Theorem 6 we actually need a stronger equivalence of $\mathsf{WAPP}^{\mathsf{cc}}$ and approximate nonnegative rank than the one proved by Kol et al. [37]: they showed the equivalence for total functions while we need the equivalence for partial functions. The extension to partial functions is nontrivial, and is related to the issue of "unrestricted" versus "restricted" models of communication.

Unrestricted versus restricted models. So far we have discussed restricted communication models. We can also define their unrestricted counterparts in analogy to the well-studied pair of classes $\mathsf{PP}^{\mathsf{cc}}$ (also known as discrepancy [34, section 8]) and $\mathsf{UPP}^{\mathsf{cc}}$ (also known as sign rank [45]). Recall that a PP computation is a randomized computation such that 1-inputs are accepted with probability in $[1/2 + \alpha, 1]$, and 0-inputs are accepted with probability in $[0, 1/2 - \alpha]$, where $\alpha = \alpha(n) > 0$ is arbitrary. In the unrestricted model $\mathsf{UPP}^{\mathsf{cc}}$ the parameter $\alpha > 0$ can be arbitrarily small (consequently, the model is defined using private randomness), whereas in the restricted model $\mathsf{PP}^{\mathsf{cc}}$

the cost of a protocol with parameter α is defined as the usual communication cost plus $\log(1/\alpha)$. It is known that $\mathsf{PP}^{\mathsf{cc}} \subsetneq \mathsf{UPP}^{\mathsf{cc}}$ where the separation is exponential [9].

One can analogously ask whether the unrestricted–restricted distinction is relevant for the models considered in this work. (The question was raised and left unresolved for SBP in [25].) In fact, the separation of [9] already witnesses $\mathsf{PostBPP}^{\mathsf{cc}} \subsetneq \mathsf{UPostBPP}^{\mathsf{cc}}$, where the latter is the unrestricted version of the former. By contrast, we prove that the distinction is immaterial for WAPP and SBP, even for partial functions: the unrestricted models $\mathsf{UWAPP}^{\mathsf{cc}}$ and $\mathsf{USBP}^{\mathsf{cc}}$ (see section 3 for definitions) are essentially no more powerful than their restricted counterparts. Consequently, the simulation theorem can be applied to analyze these unrestricted models, too—but the equivalences are also interesting in their own right.

THEOREM 8. $\mathsf{SBP}^{\mathsf{cc}}(F) \leq O(\mathsf{USBP}^{\mathsf{cc}}(F) + \log n)$ *for all* F*.*

THEOREM 9. $\mathsf{WAPP}^{\mathsf{cc}}_{\delta}(F) \leq O(\mathsf{UWAPP}^{\mathsf{cc}}_{\epsilon}(F) + \log(n/(\delta - \epsilon)))$ *for all* F *and all* $0 < \epsilon < \delta < 1/2$*.*

The seemingly more powerful models $\mathsf{USBP}^{\mathsf{cc}}$ and $\mathsf{UWAPP}^{\mathsf{cc}}$ admit characterizations in terms of the nonnegative rank of matrices: instead of rectangles, the protocols compute using nonnegative rank-1 matrices. In particular, $\mathsf{UWAPP}^{\mathsf{cc}}$ turns out to capture $\mathrm{rank}^{+}_{\epsilon}$; it is Theorem 9 that will be used in the proof of Corollary 7 above.

1.4. Open problems and subsequent developments. Our main open question is whether Theorem 1 continues to hold for $b = O(1)$. If true, such a result would be very useful as the inner product on b bits can be simulated by most other gadgets on blocks of length roughly 2^b (which would be $O(1)$ again). This in turn would give new and more unified proofs of important communication complexity lower bounds such as Razborov's corruption lower bound for set disjointness [49] and the lower bound for gap Hamming [11, 57, 64]. A first hurdle in understanding the case $b = O(1)$ seems to be Lemma 13—does some version of it hold for $b = O(1)$? In particular, using notions from section 2.2, we can ask the following concrete question as a starting point: for b a sufficiently big constant, g the inner-product gadget, and two independent 0.9-dense sources $\boldsymbol{X}, \boldsymbol{Y}$ over $(\{0,1\}^b)^n$, does $g^n(\boldsymbol{X}, \boldsymbol{Y})$ have full support over $\{0,1\}^n$?

The following are some other relevant open problems.

- Can the multiplicative accuracy in Theorem 1 be improved? This issue seems to be what is preventing us from quantitatively improving on the lower bounds obtained by [12] for the linear programming extension complexity of approximating max-cut.
- Raz and McKenzie [47] (see also [22]) obtained a simulation theorem that converts deterministic communication protocols for $f \circ g^n$ into deterministic decision trees for f, where g is a certain polynomial-size gadget. Can our methods be used to simplify their proof, or to extend their result to other g's?
- Our focus in this work has been on partial functions. It remains open whether $\mathsf{SBP}^{\mathsf{cc}} = \mathsf{MA}^{\mathsf{cc}}$ for total functions, or whether efficient error amplification exists for $\mathsf{WAPP}^{\mathsf{cc}}$ for total functions.

Since this paper first appeared, our main results have found several further applications.

- In [18], Theorem 2 (specialized to NP) has been applied to obtain the first superlogarithmic communication lower bound for the clique versus independent set problem.

- In [23], Theorem 2 (more precisely, the key technical component of the proof, Theorem 17) has been applied to obtain a result exploring the question of whether rank-1 matrices are inherently more powerful than rectangles in communication complexity. This is motivated by the open question of whether $\mathsf{PP}^{\mathsf{cc}} \not\subseteq \mathsf{UPostBPP}^{\mathsf{cc}}$.
- In [20], Theorem 2 has been applied to obtain an essentially tight randomized communication lower bound for the clique versus independent set problem, as well as to prove that there exist boolean matrices for which the randomized communication complexity can be superlogarithmic in the number of monochromatic rectangles needed to partition the matrix.
- In [19], Theorem 2 has been applied to obtain strong randomized communication lower bounds for the recursive NAND function and the recursive majority-of-3 function.
- In [66], Theorem 17 has been applied to obtain an exponential separation between nonnegative rank and binary rank for partial boolean matrices.
- In [5], Theorem 2 has been applied to obtain a superquadratic separation between randomized and quantum communication complexities of a total function.

1.5. Notational conventions. We always write random variables in bold (e.g., $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$). Capital letters X, Y are reserved for subsets of inputs to $G = g^n$ (so all rectangles R are of the form $X \times Y$). We identify such sets with flat distributions: we denote by $\boldsymbol{X}$ the random variable that is uniformly distributed on X. Given a distribution $\mathcal{D}$ and an event E we denote by $(\mathcal{D} \mid E)$ the conditional distribution of $\mathcal{D}$ given E, specifically, $(\mathcal{D} \mid E)(\,\cdot\,) \coloneqq \mathcal{D}(\,\cdot \cap E)/\mathcal{D}(E)$. We also use the shorthand $\mathcal{D}(\,\cdot \mid E) \coloneqq (\mathcal{D} \mid E)(\,\cdot\,)$.

2. Proof of the junta theorem.

In this section we prove Theorem 1, restated here for convenience.

THEOREM 1 (junta theorem). *Assume (†). For any $d \geq 0$ and any rectangle R in the domain of g^n there exists a conical d-junta h such that, for all $z \in \{0,1\}^n$,*

$$\operatorname{acc}_R(z) \in (1 \pm 2^{-\Theta(b)}) \cdot h(z) \pm 2^{-\Theta(db)}. \tag{1}$$

2.1. Proof overview. We write $G \coloneqq g^n$ for short. Fix $d \geq 0$ and a rectangle $L \subseteq \operatorname{dom} G$. Our goal is to approximate $\operatorname{acc}_L(z)$ by some conical d-junta $h(z)$. (We are going to use the symbol L for the "main" rectangle so as to keep the symbol R free for later use as a more generic rectangle.) The high-level idea in our proof is extremely direct: to find a suitable h we partition—or at least almost partition—the rectangle L into subrectangles $R \subseteq L$ that behave like width-d conjunctions.

DEFINITION 10 (conjunction rectangles). *A rectangle R is a* (d, ϵ)-conjunction *if there exists a width-d conjunction $h_R\colon \{0,1\}^n \to \{0,1\}$ (i.e., h_R can be written as $(\ell_1 \wedge \cdots \wedge \ell_w)$, where $w \leq d$ and each ℓ_i is an input variable or its negation) such that $\operatorname{acc}_R(z) \in (1 \pm \epsilon) \cdot a_R h_R(z)$ for some $a_R \geq 0$ and all $z \in \{0,1\}^n$.*

The proof is split into three subsections.

(section 2.2) *Blockwise density*: We start by discussing a key property that is a sufficient condition for a subrectangle $R \subseteq L$ to be a conjunction rectangle.

(section 2.3) *Reduction to a packing problem*: Instead of partitioning L into conjunctions, we show that it suffices to find a packing (disjoint collection) of conjunction subrectangles of L that cover most of L relative to a given distribution over inputs. This will formalize our main technical task: solving a type of packing-with-conjunctions problem.

(section 2.4) *Solving the packing problem*: This is the technical heart of the proof: we describe an algorithm to find a good packing for L.

2.2. Blockwise density. In this subsection we introduce a central notion that will allow us to extract close to uniform output from sufficiently random inputs to $G = g^n\colon \{0,1\}^{bn} \times \{0,1\}^{bn} \to \{0,1\}^n$. Recall that in the setting of two-source extractors (e.g., [62]), one considers a pair of independent random inputs $\boldsymbol{x}$ and $\boldsymbol{y}$ that have high *min-entropy*, defined by $\mathbf{H}_\infty(\boldsymbol{x}) := \min_x \log(1/\mathbf{Pr}[\boldsymbol{x} = x])$. In our setting we think of $G = g^n$ as a *local* two-source extractor: each of the n output bits depends only on few of the input bits. Hence we need a stronger property than high min-entropy on $\boldsymbol{x}$ and $\boldsymbol{y}$ to guarantee that $\boldsymbol{z} := G(\boldsymbol{x}, \boldsymbol{y})$ will be close to uniform. This property we call *blockwise density*. Below, for $I \subseteq [n]$, we write $\boldsymbol{x}_I$ for the restriction of $\boldsymbol{x}$ to the *blocks* determined by I.

DEFINITION 11 (blockwise density). *A random variable $\boldsymbol{x} \in \{0,1\}^{bn}$ is δ-*dense *if for all $I \subseteq [n]$ the blocks $\boldsymbol{x}_I$ have* min-entropy rate *at least δ, that is, $\mathbf{H}_\infty(\boldsymbol{x}_I) \geq \delta b|I|$.*

DEFINITION 12 (multiplicative uniformity). *A distribution $\mathcal{D}$ on $\{0,1\}^m$ is ϵ-*uniform *if $\mathcal{D}(z) \in (1 \pm \epsilon) \cdot 2^{-m}$ for all outcomes z.*

LEMMA 13. *Assume (†). If $\boldsymbol{x}$ and $\boldsymbol{y}$ are independent and* 0.6*-dense, then $G(\boldsymbol{x}, \boldsymbol{y})$ is $2^{-b/20}$-uniform.*

Proof. Let $\boldsymbol{z} := G(\boldsymbol{x}, \boldsymbol{y})$. First observe that for any $I \subseteq [n]$ the parity of the output bits $\boldsymbol{z}_I$ is simply $\langle \boldsymbol{x}_I, \boldsymbol{y}_I \rangle \bmod 2$. We use the fact that the inner product is a good two-source extractor to argue that this parity is close to an unbiased random bit. Indeed, by 0.6-density we have $\mathbf{H}_\infty(\boldsymbol{x}_I) + \mathbf{H}_\infty(\boldsymbol{y}_I) \geq 1.2 \cdot b|I|$ and this implies by a basic theorem of Chor and Goldreich [14, Theorem 9] that for $I \neq \emptyset$,

$$\big|\, \mathbf{Pr}[\, \langle \boldsymbol{x}_I, \boldsymbol{y}_I \rangle \bmod 2 \,=0\,] - 1/2 \,\big| \leq 2^{-0.1 \cdot b|I|+1}. \tag{2}$$

This bound is enough to yield ϵ-uniformity for $\epsilon := 2^{-b/20}$, as we next verify using standard Fourier analysis (see, e.g., [44]).[1] Let $\mathcal{D}$ be the distribution of $\boldsymbol{z}$. We think of $\mathcal{D}$ as a function $\{0,1\}^n \to [0,1]$ and write it in the Fourier basis as

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

where $\chi_I(z) := (-1)^{\sum_{i\in I} z_i}$ and $\widehat{\mathcal{D}}(I) := 2^{-n} \sum_z \mathcal{D}(z)\chi_I(z) = 2^{-n} \cdot \mathbf{E}_{\boldsymbol{z}\sim\mathcal{D}}[\,\chi_I(\boldsymbol{z})\,]$. Note that $\widehat{\mathcal{D}}(\emptyset) = 2^{-n}$ because $\mathcal{D}$ is a distribution. In this language, property (2) says that, for all $I \neq \emptyset$, $2^n \cdot |\widehat{\mathcal{D}}(I)| = |\,\mathbf{E}[\,(-1)^{\langle \boldsymbol{x}_I, \boldsymbol{y}_I\rangle}\,]| \leq 2^{-0.1\cdot b|I|+2}$, which is at most $\epsilon 2^{-2|I|\log n}$ by our definition of b and ϵ. Hence,

$$2^n \sum_{I\neq\emptyset} |\widehat{\mathcal{D}}(I)| \leq \epsilon \sum_{I\neq\emptyset} 2^{-2|I|\log n} = \epsilon \sum_{k=1}^{n} \binom{n}{k} 2^{-2k\log n} \leq \epsilon \sum_{k=1}^{n} 2^{-k\log n} \leq \epsilon.$$

We use this to show that $|\mathcal{D}(z) - 2^{-n}| \leq \epsilon 2^{-n}$ for all $z \in \{0,1\}^n$, which proves the lemma. To this end, let $\mathcal{U}$ denote the uniform distribution (note that $\widehat{\mathcal{U}}(I) = 0$ for all

[1]This fact resembles the classic "Vazirani XOR lemma" [63], except that the latter only guarantees the distribution is close to uniform in statistical distance, and it assumes a single bound on the bias of all parities (whereas we assume a bound that depends on the size of the parity).

$I \neq \emptyset$) 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 $|\widehat{\mathbb{1}}_z(I)| = 2^{-n}$ for all I). We can now calculate

$$\begin{aligned}|\mathcal{D}(z) - 2^{-n}| &= |\langle \mathbb{1}_z, \mathcal{D}\rangle - \langle \mathbb{1}_z, \mathcal{U}\rangle| = |\langle \mathbb{1}_z, \mathcal{D} - \mathcal{U}\rangle| = 2^n \cdot |\langle \widehat{\mathbb{1}}_z, \widehat{\mathcal{D}} - \widehat{\mathcal{U}}\rangle| \\ &\leq 2^n \cdot \textstyle\sum_{I\neq\emptyset} |\widehat{\mathbb{1}}_z(I)| \cdot |\widehat{\mathcal{D}}(I)| = \sum_{I\neq\emptyset} |\widehat{\mathcal{D}}(I)| \leq \epsilon 2^{-n}. \qquad \square\end{aligned}$$

Remark 14. The only properties of the inner product we needed in the above proof were that it is a strong two-source extractor and that it satisfies an XOR lemma. However, all sufficiently strong two-source extractors have the latter property automatically [52], so we could have fixed g to be any such extractor in Theorem 1. It is known [41] that an XOR lemma holds even under the weaker assumption of g having low discrepancy (not necessarily under the uniform distribution over $\operatorname{dom} g$). Hence it is plausible that Theorem 1 could be extended to handle such g, as well.

We have the following corollary; here we write $\bar{I} \coloneqq [n] \setminus I$ for short.

COROLLARY 15. *Assume* (†). *Let $R = X \times Y$ and suppose there is an $I \subseteq [n]$ such that $\boldsymbol{X}_I$ and $\boldsymbol{Y}_I$ are fixed while $\boldsymbol{X}_{\bar{I}}$ and $\boldsymbol{Y}_{\bar{I}}$ are* 0.6*-dense. Then R is an* $(|I|, O(2^{-b/20}))$*-conjunction.*

Proof. Let $\boldsymbol{z} \coloneqq G(\boldsymbol{X}, \boldsymbol{Y})$ and note that $\boldsymbol{z}_I$ is fixed. Write $\epsilon \coloneqq 2^{-b/20}$ for short. Applying Lemma 13 to $\boldsymbol{x} = \boldsymbol{X}_{\bar{I}}$ and $\boldsymbol{y} = \boldsymbol{Y}_{\bar{I}}$ ($\boldsymbol{x}$ and $\boldsymbol{y}$ are 0.6-dense) shows that $|G^{-1}(z) \cap R|/|R| \in (1 \pm \epsilon) \cdot 2^{-|\bar{I}|}$ whenever $z_I = \boldsymbol{z}_I$ (and 0 otherwise). If g were perfectly balanced, then we would have $|G^{-1}(z)|/2^{2bn} = 2^{-n}$ for all $z \in \{0,1\}^n$; instead, since g is only approximately balanced ($|g^{-1}(1)|, |g^{-1}(0)| \in 2^{2b-1} \pm 2^{b-1}$), it can be seen by direct calculation that $|G^{-1}(z)|/2^{2bn} \in (1 \pm \epsilon) \cdot 2^{-n}$ for all $z \in \{0,1\}^n$ (though this can also be seen by another application of Lemma 13—to uniform $\boldsymbol{x}, \boldsymbol{y} \in \{0,1\}^{bn}$, which are 1-dense). Therefore $\operatorname{acc}_R(z) = |G^{-1}(z) \cap R|/|G^{-1}(z)| \in (1 \pm O(\epsilon)) \cdot 2^{|I| - 2bn}|R|$ if $z_I = \boldsymbol{z}_I$ and $\operatorname{acc}_R(z) = 0$ if $z_I \neq \boldsymbol{z}_I$. This is of the form $(1 \pm O(\epsilon)) \cdot a_R h_R(z)$ (where $h_R(z) = 1$ iff $z_I = \boldsymbol{z}_I$), as required. $\square$

2.3. Reduction to a packing problem. The purpose of this subsection is to massage the statement of the junta theorem into an alternative form in order to uncover its main technical content. We will end up with a certain type of packing problem, formalized in Theorem 17 at the end of this subsection.

Fix some "multiplicative" error bound $\epsilon \coloneqq 2^{-\Theta(b)}$ for the purposes of the following discussion. Whenever $\mathscr{C}$ is a packing (disjoint collection) of (d, ϵ)-conjunction subrectangles of L we let

$$h_{\mathscr{C}} \coloneqq \sum_{R \in \mathscr{C}} a_R h_R.$$

Write $\cup\mathscr{C} \coloneqq \cup_{R\in\mathscr{C}} R$ for short. Then $\operatorname{acc}_{\cup\mathscr{C}} \coloneqq \sum_{R\in\mathscr{C}} \operatorname{acc}_R$ is multiplicatively approximated by the conical d-junta $h_{\mathscr{C}}$ in the sense that $\operatorname{acc}_{\cup\mathscr{C}}(z) \in (1 \pm \epsilon) \cdot h_{\mathscr{C}}(z)$. Hence if we could find a $\mathscr{C}$ that partitioned $L = \cup\mathscr{C}$, we would have proved the theorem—without incurring any additive error.

Unfortunately, there are a few obstacles standing in the way of finding a perfect partition $\mathscr{C}$. One unavoidable issue is that we cannot multiplicatively approximate a tiny rectangle L with a low-degree conical junta. This is why we allow a small additive error and only multiplicatively approximate the acceptance probabilities of those z that have large enough $\operatorname{acc}_L(z)$. Indeed, we set

$$Z \coloneqq \{\, z \in \{0,1\}^n : \operatorname{acc}_L(z) \geq 2^{-db/20} \,\},$$

and look for an $\mathscr{C}$ that covers most of each of the sets $G^{-1}(z) \cap L$ for $z \in Z$. More precisely, suppose for a moment that we had a packing $\mathscr{C}$ such that for each $z \in Z$,

$$\mathcal{U}_z(\cup\mathscr{C} \mid L) \geq 1 - \epsilon, \tag{3}$$

where $\mathcal{U}_z(\cup\mathscr{C} \mid L) = \mathrm{acc}_{\cup\mathscr{C}}(z)/\,\mathrm{acc}_L(z)$ by definition. Indeed, assuming (3) we claim that

$$(1-\epsilon)\cdot h_{\mathscr{C}}(z) \leq \mathrm{acc}_L(z) \leq (1+O(\epsilon))\cdot h_{\mathscr{C}}(z) + 2^{-\Theta(db)}. \tag{4}$$

In particular, $h_{\mathscr{C}}$ achieves the desired approximation (1). For the first inequality, since $\cup\mathscr{C} \subseteq L$ we never multiplicatively overestimate acc_L, that is, we have $\mathrm{acc}_L \geq \mathrm{acc}_{\cup\mathscr{C}} \geq (1-\epsilon)\cdot h_{\mathscr{C}}$. For the second inequality, for $z \in Z$ we have $\mathrm{acc}_L(z) \leq (1-\epsilon)^{-1}\cdot \mathrm{acc}_{\cup\mathscr{C}}(z) \leq (1-\epsilon)^{-1}\cdot(1+\epsilon)\cdot h_{\mathscr{C}}(z) \leq (1+O(\epsilon))\cdot h_{\mathscr{C}}(z)$, and for $z \notin Z$ we have simply $\mathrm{acc}_L(z) < 2^{-\Theta(db)}$ by the definition of Z.

Unfortunately, we do not know how to construct a packing $\mathscr{C}$ satisfying (3) either. Instead, we show how to find a *randomized* packing $\boldsymbol{\mathscr{C}}$ that guarantees (3) *in expectation.* More precisely, our construction goes through the following primal/dual pair of statements that are equivalent by the minimax theorem.

Primal: $\exists$ distribution $\mathcal{C}$ over $\mathscr{C}$'s $\quad \forall\, z \in Z \quad \mathbf{E}_{\boldsymbol{\mathscr{C}}\sim\mathcal{C}}\,\mathcal{U}_z(\cup\boldsymbol{\mathscr{C}} \mid L) \geq 1-\epsilon,$

Dual: $\forall$ distribution μ over $Z \quad \exists\, \mathscr{C} \quad \mathbf{E}_{\boldsymbol{z}\sim\mu}\,\mathcal{U}_{\boldsymbol{z}}(\cup\mathscr{C} \mid L) \geq 1-\epsilon.$

Suppose the primal statement holds for some $\mathcal{C}$. Then we claim that the convex combination $h := \mathbf{E}_{\boldsymbol{\mathscr{C}}\sim\mathcal{C}}\, h_{\boldsymbol{\mathscr{C}}}$ achieves the desired approximation. The right side of (4) can be reformulated as

$$h_{\boldsymbol{\mathscr{C}}}(z) \geq (1 - O(\epsilon + \boldsymbol{\epsilon}_z))\cdot(\mathrm{acc}_L(z) - 2^{-\Theta(db)}), \tag{5}$$

where $\boldsymbol{\epsilon}_z := 1 - \mathcal{U}_z(\cup\boldsymbol{\mathscr{C}} \mid L)$ is a random variable depending on $\boldsymbol{\mathscr{C}}$ (so $\mathbf{E}_{\boldsymbol{\mathscr{C}}\sim\mathcal{C}}[\,\boldsymbol{\epsilon}_z\,] \leq \epsilon$). Applying linearity of expectation to (5) shows (along with the left side of (4)) that h satisfies (1).

Therefore, to prove Theorem 1 it remains to prove the dual statement. This will preoccupy us for the whole of section 2.4 where, for convenience, we will prove a slightly more general claim formalized below.

Definition 16 (lifted distributions). *A distribution $\mathcal{D}$ on the domain of G is said to be a* lift *of a distribution μ on the codomain of G if $\mathcal{D}(x,y) = \mu(z)/|G^{-1}(z)|$, where $z := G(x,y)$. Note that a lifted distribution is a convex combination of distributions of the form $\mathcal{U}_z$.*

Theorem 17 (packing with conjunctions). *Assume (†). Let $d \geq 0$ and let L be a rectangle. There is an $\epsilon := 2^{-\Theta(b)}$ such that for any lifted distribution $\mathcal{D}$ with $\mathcal{D}(L) \geq 2^{-db/20}$ there exists a packing $\mathscr{C}$ consisting of (d,ϵ)-conjunction subrectangles of L such that $\mathcal{D}(\cup\mathscr{C} \mid L) \geq 1-\epsilon$.*

The dual statement can be derived from Theorem 17 as follows. We need to check that for any distribution μ on Z there is some lifted distribution $\mathcal{D}$ such that $\mathcal{D}(L) \geq 2^{-db/20}$ and $\mathcal{D}(\,\cdot \mid L) = \mathcal{E}(\,\cdot\,)$, where $\mathcal{E}(\,\cdot\,) := \mathbf{E}_{\boldsymbol{z}\sim\mu}\mathcal{U}_{\boldsymbol{z}}(\,\cdot \mid L)$ is the probability measure relevant to the dual statement. For intuition, a seemingly natural candidate would be to choose $\mathcal{D} = \mathbf{E}_{\boldsymbol{z}\sim\mu}\,\mathcal{U}_{\boldsymbol{z}}$; however, this does not ensure $\mathcal{D}(\,\cdot \mid L) =$

$\mathcal{E}(\,\cdot\,)$ as conditioning on L may not commute with taking convex combinations of the $\mathcal{U}_z$'s. This is why we instead define a slightly different distribution $\mu'(z) \coloneqq \gamma\mu(z)/\mathcal{U}_z(L)$, where $\gamma \coloneqq (\mathbf{E}_{\boldsymbol{z}\sim\mu}\, 1/\mathcal{U}_{\boldsymbol{z}}(L))^{-1}$ is a normalizing constant. If we now choose $\mathcal{D} \coloneqq \mathbf{E}_{\boldsymbol{z}\sim\mu'}\,\mathcal{U}_{\boldsymbol{z}}$ the conditioning on L works out. Indeed, noting that $\gamma = \mathcal{D}(L)$ we have $\mathcal{D}(\,\cdot \mid L) = \mathcal{D}(L)^{-1}\mathcal{D}(\,\cdot\, \cap L) = \gamma^{-1}\sum_z \mu'(z)\mathcal{U}_z(\,\cdot\, \cap L) = \sum_z \mu(z)\mathcal{U}_z(\,\cdot\, \cap L)/\mathcal{U}_z(L) = \mathbf{E}_{\boldsymbol{z}\sim\mu}\mathcal{U}_{\boldsymbol{z}}(\,\cdot \mid L) = \mathcal{E}(\,\cdot\,)$, as desired. Also note that $\mathcal{D}(L) = \mathbf{E}_{\boldsymbol{z}\sim\mu'}\,\mathcal{U}_{\boldsymbol{z}}(L) \geq \mathbf{E}_{\boldsymbol{z}\sim\mu'}\, 2^{-db/20} = 2^{-db/20}$ since μ' is supported on Z.

2.4. Solving the packing problem. In this section we prove Theorem 17. Fix an error parameter $\epsilon \coloneqq 2^{-b/100}$.

Notation. In the course of the argument, for any rectangle $R = X \times Y$, we are going to associate a bipartition of $[n]$ into *free* blocks, denoted $\mathrm{free}\, R$, and *fixed* blocks, denoted $\mathrm{fix}\, R \coloneqq [n] \smallsetminus \mathrm{free}\, R$. We will always ensure that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are fixed on the blocks in $\mathrm{fix}\, R$. However, if $\boldsymbol{X}$ and $\boldsymbol{Y}$ are fixed on some block i, we may or may not put i into $\mathrm{fix}\, R$; thus the sets $\mathrm{fix}\, R$ and $\mathrm{free}\, R$ are not predefined functions of R, but rather will be chosen during the proof of Theorem 17. We say that *the free marginals of R are $(\delta, \mathcal{D})$-dense* if for $\boldsymbol{xy} \sim (\mathcal{D} \mid R)$ we have that $\boldsymbol{x}_{\mathrm{free}\, R}$ and $\boldsymbol{y}_{\mathrm{free}\, R}$ are δ-dense. Note that if $\mathcal{D} = \mathcal{U}$ is the uniform distribution, then the definition states that $\boldsymbol{X}_{\mathrm{free}\, R}$ and $\boldsymbol{Y}_{\mathrm{free}\, R}$ are δ-dense. The following is a rephrasing of Corollary 15.

PROPOSITION 18. *If the free marginals of R are $(0.6, \mathcal{U})$-dense then R is a $(|\mathrm{fix}\, R|, \epsilon)$-conjunction.*

We also use the following notation: if C is a *condition* (e.g., of the form $(x_I = \alpha)$ or $(x_I \neq \alpha)$) we write X_C for the set of $x \in X$ that satisfy C. For example, $X_{(x_I=\alpha)} \coloneqq \{x \in X : x_I = \alpha\}$.

Roadmap. The proof is in two steps. In the first step we find a packing with subrectangles whose free marginals are $(0.8, \mathcal{D})$-dense. In the second step we "prune" these subrectangles so that their free marginals become $(0.6, \mathcal{U})$-dense. These two steps are encapsulated in the following two lemmas.

LEMMA 19 (core packing step). *There is a packing $\mathscr{C}'$ of subrectangles of L such that $\mathcal{D}(\cup\mathscr{C}' \mid L) \geq 1-\epsilon$ and for each $R \in \mathscr{C}'$ we have $|\mathrm{fix}\, R| \leq d$ and the free marginals of R are $(0.8, \mathcal{D})$-dense (for some choice of the sets $\mathrm{fix}\, R$ and $\mathrm{free}\, R$).*

LEMMA 20 (pruning step). *For each $R \in \mathscr{C}'$ there is a subrectangle $R' \subseteq R$ with $\mathrm{fix}\, R' = \mathrm{fix}\, R$ such that $\mathcal{D}(R' \mid R) \geq 1 - \epsilon$ and the free marginals of R' are $(0.6, \mathcal{U})$-dense.*

Theorem 17 follows immediately by stringing together Lemmas 19 and 20 and Proposition 18. In particular, the final packing $\mathscr{C}$ will consist of the pruned rectangles R' (which are (d, ϵ)-conjunctions by Proposition 18) and we have $\mathcal{D}(\cup\mathscr{C} \mid L) \geq (1-\epsilon)^2 \geq 1 - 2\epsilon$. (We proved the theorem with error parameter 2ϵ instead of ϵ.)

2.4.1. Core packing step. We will now prove Lemma 19. The desired packing $\mathscr{C}'$ of subrectangles of L will be found via a packing algorithm given in Figure 2.

Informal overview. The principal goal in the algorithm is to find subrectangles $R \subseteq L$ whose free marginals are $(0.8, \mathcal{D})$-dense while keeping $|\mathrm{fix}\, R|$ small. To do this, we proceed in rounds. The main loop of the algorithm maintains a *pool* $\mathscr{P}$ of disjoint subrectangles of L and in each round we inspect each $R \in \mathscr{P}$ in the subroutine PARTITION. If we find that R does not have dense free marginals, we partition R further. The output of PARTITION(R) is a partition of R into subrectangles each labeled as either *dense*, *live*, or *error*. We are simply going to ignore the *error* rectangles, i.e., they do not reenter the pool $\mathscr{P}$. For the *live* subrectangles $R' \subseteq R$

Packing Algorithm for L:

Initialize $\mathscr{P} \coloneqq \{L\}$ where $\operatorname{fix} L \coloneqq \emptyset$ and L is labeled *live*
Repeat for $n+1$ rounds
  Replace each $R \in \mathscr{P}$ by all the non*error* subrectangles output by PARTITION(R)
Output $\mathscr{C}' \coloneqq \mathscr{P}$

Subroutine PARTITION (with error parameter $\delta \coloneqq \epsilon/2n$)

Input: A rectangle R_{in}
Output: A partition of R_{in} into *dense*/*live*/*error* subrectangles

Initialize $R \coloneqq R_{\mathsf{in}}$ with $\operatorname{fix} R \coloneqq \operatorname{fix} R_{\mathsf{in}}$
While the following two conditions hold
  (C1): $\mathcal{D}(R \mid R_{\mathsf{in}}) > \delta$
  (C2): The free marginals of R are not both $(0.8, \mathcal{D})$-dense
  Let $\boldsymbol{xy} \sim (\mathcal{D} \mid R)$ and let X and Y be such that $R = X \times Y$
  We may assume $\boldsymbol{x}_{\operatorname{free} R}$ is not 0.8-dense (otherwise consider $\boldsymbol{y}_{\operatorname{free} R}$)
  Let $I \subseteq \operatorname{free} R$ and α be such that $\mathbf{Pr}[\,\boldsymbol{x}_I = \alpha\,] > 2^{-0.8 \cdot b|I|}$
  Let $\mathcal{B} \coloneqq \left\{\beta : \mathbf{Pr}[\,\boldsymbol{y}_I = \beta \mid \boldsymbol{x}_I = \alpha\,] > \delta \cdot 2^{-b|I|}\right\}$
  For each $\beta \in \mathcal{B}$
    Let $R_{\mathsf{out}} \coloneqq X_{(x_I=\alpha)} \times Y_{(y_I=\beta)}$ with $\operatorname{fix} R_{\mathsf{out}} \coloneqq \operatorname{fix} R \cup I$
    Output R_{out} labeled as *live*
  End for
  Output $X_{(x_I=\alpha)} \times Y_{(y_I \notin \mathcal{B})}$ labeled as *error*
  Update $R \coloneqq X_{(x_I \neq \alpha)} \times Y$ (with the same $\operatorname{fix} R$)
End while
Output R labeled as *dense* if **(C2)** failed, or as *error* if **(C1)** failed

FIG. 2. *Packing algorithm.*

we will have made progress: the subroutine will ensure that the free marginals of R' will become more dense as compared to the free marginals of R.

The subroutine PARTITION works as follows. If the input rectangle R_{in} satisfies the density condition on its free marginals, we simply output R_{in} labeled as *dense*. Otherwise we find some subset I of free blocks that violates the density condition on one of the marginals. Then we consider the subrectangle $R_{\mathsf{out}} \subseteq R_{\mathsf{in}}$ that is obtained from R_{in} by fixing the nondense marginal to its overly likely value on I and the other marginal to each of its typical values on I. Intuitively, these fixings have the effect of increasing the "relative density" in the remaining free blocks, and so we have found a single subrectangle where we have made progress. We then continue iteratively on the rest of R_{in} until only a $\delta \coloneqq \epsilon/2n$ fraction of R_{in} remains, which we deem as *error*.

Note that, at the end of $n+1$ rounds, each $R \in \mathscr{C}'$ must be labeled *dense* because once a rectangle R reaches $\operatorname{fix} R = [n]$, the density condition on the free marginals is satisfied vacuously. It remains to argue that the other two properties in Lemma 19 hold for $\mathscr{C}'$.

Error analysis. We claim that in each run of PARTITION at most a fraction 2δ of the distribution $(\mathcal{D} \mid R_{\mathsf{in}})$ gets classified as *error*. This claim implies that $\cup\mathscr{C}'$ covers all but an ϵ fraction of $(\mathcal{D} \mid L)$ since the total error relative to $(\mathcal{D} \mid L)$ can be easily bounded by the number of rounds (excluding the last round, which only labels the remaining *live* rectangles as *dense*) times the error in PARTITION, which is $n \cdot 2\delta = \epsilon$ under our claim.

To prove our claim, we first note that the *error* rectangle output on line 18 contributes a fraction $\leq \delta$ of error relative to $(\mathcal{D} \mid R_{\mathsf{in}})$ by **(C1)**. Consider then *error* rectangles output on line 15. Here we have (using notation from the algorithm) $\mathbf{Pr}[\,\boldsymbol{y}_I \notin \mathcal{B} \mid \boldsymbol{x}_I = \alpha\,] \leq \delta$ by the definition of $\mathcal{B}$ so we only incur $\leq \delta$ fraction of error relative to $(\mathcal{D} \mid R')$, where $R' \coloneqq X_{(x_I=\alpha)} \times Y$. In the subsequent line we redefine $R \coloneqq R \smallsetminus R'$, which ensures that the errors on line 15 do not add up over the different iterations. Hence, altogether, line 15 contributes a fraction $\leq \delta$ of error relative to $(\mathcal{D} \mid R_{\mathsf{in}})$. The total error in PARTITION is then at most $\delta + \delta = 2\delta$, which was our claim.

Number of fixed blocks. Let $R \in \mathscr{C}'$. We need to show that $|\mathrm{fix}\, R| \leq d$. Let R_i, $i \in [n+1]$, be the unique rectangle in the pool at the start of the ith round such that $R \subseteq R_i$. Let ℓ be the largest number such that R_ℓ is labeled *live*. Hence $|\mathrm{fix}\, R| = |\mathrm{fix}\, R_\ell|$. Let $Q \supseteq R_\ell$ consist of all the inputs that agree with R_ℓ on the fixed coordinates $\mathrm{fix}\, R$. We claim that

$$\mathcal{D}(Q) \leq 2^{-(2b-2)|\mathrm{fix}\, R|}, \tag{6}$$

$$\mathcal{D}(R_\ell) \geq 2^{-1.9\cdot b|\mathrm{fix}\, R| - db/20}. \tag{7}$$

Let us first see how to conclude the proof of Lemma 19 assuming the above inequalities. Since $\mathcal{D}(Q) \geq \mathcal{D}(R_\ell)$ we can require that (6) $\geq$ (7) and (taking logarithms) obtain the inequality $-(2b-2)|\mathrm{fix}\, R| \geq -1.9 \cdot b|\mathrm{fix}\, R| - db/20$. But this implies $|\mathrm{fix}\, R| \leq d$, as desired.

To prove (6), write $\mathcal{D}(Q) = \mathbf{E}_{\boldsymbol{z}\sim\mu}\, \mathcal{U}_{\boldsymbol{z}}(Q)$ for some μ since $\mathcal{D}$ is a lifted distribution. Here for each fixed z we either have $\mathcal{U}_z(Q) = 0$ in case the fixings of Q are inconsistent with z, or otherwise $\mathcal{U}_z(Q) = \prod_{j \in \mathrm{fix}\, R} 1/|g^{-1}(z_j)| \leq 2^{-(2b-2)|\mathrm{fix}\, R|}$ (where we used the fact that the gadget g is approximately balanced: $|g^{-1}(1)|, |g^{-1}(0)| \geq 2^{2b}/4$). Hence $\mathcal{D}(Q)$ is a convex combination of values that satisfy (6).

To prove (7), note that $\mathcal{D}(R_\ell) = \mathcal{D}(R_\ell \mid L) \cdot \mathcal{D}(L) \geq \mathcal{D}(R_\ell \mid L) \cdot 2^{-db/20}$. Hence it suffices to show that $\mathcal{D}(R_\ell \mid L) \geq 2^{-1.9\cdot b|\mathrm{fix}\, R|}$. To this end, write $|\mathrm{fix}\, R| = \sum_{i=1}^{\ell-1} |I_i|$, where I_i is the set of blocks that were fixed to obtain $R_{i+1} = R_{\mathsf{out}}$ from $R_i = R_{\mathsf{in}}$ and use the following claim inductively.

CLAIM 21. *Each R_{out} output labeled as live (on line 13) satisfies $\mathcal{D}(R_{\mathsf{out}} \mid R_{\mathsf{in}}) \geq 2^{-1.9\cdot b|I|}$.*

Proof. Using notation from the algorithm,

$$\begin{aligned}
\mathcal{D}(R_{\mathsf{out}} \mid R_{\mathsf{in}}) &= \mathcal{D}(R_{\mathsf{out}} \mid R) \cdot \mathcal{D}(R \mid R_{\mathsf{in}}) \\
&\geq \mathcal{D}(R_{\mathsf{out}} \mid R) \cdot \delta && \text{(by } \mathbf{(C1)}\text{)} \\
&= \mathbf{Pr}[\,\boldsymbol{x}_I = \alpha \text{ and } \boldsymbol{y}_I = \beta\,] \cdot \delta \\
&\geq 2^{-0.8\cdot b|I|} \cdot \delta \cdot 2^{-b|I|} \cdot \delta \\
&= 2^{-1.8\cdot b|I| - b/50 - 2\log n - 2} && \text{(definition of } \epsilon, \delta\text{)} \\
&\geq 2^{-1.9\cdot b|I|}.
\end{aligned}$$

∎

2.4.2. Pruning step. We will now prove Lemma 20. Let $R = X \times Y \in \mathscr{C}'$ and $\boldsymbol{xy} \sim (\mathcal{D} \mid R)$. For notational convenience, we assume that $\operatorname{fix} R = \emptyset$, i.e., we forget about the fixed blocks and think of $\boldsymbol{x}$ and $\boldsymbol{y}$ as 0.8-dense. As will be clear from the proof, if $\operatorname{fix} R$ was nonempty, it would only help us in the ensuing calculations.

We want to find a "pruned" subrectangle $R' \coloneqq X' \times Y' \subseteq R$ such that

(i) $\mathbf{Pr}[\,\boldsymbol{xy} \in X' \times Y'\,] \geq 1 - \epsilon$,

(ii) $\boldsymbol{X'}$ and $\boldsymbol{Y'}$ are 0.6-dense.

In fact, it is enough to show how to find an $X' \subseteq X$ such that

(i′) $\mathbf{Pr}[\,\boldsymbol{x} \in X'\,] \geq 1 - \epsilon/2$,

(ii′) $\boldsymbol{X'}$ is 0.6-dense.

Indeed, we can run the argument for (i′, ii′) twice, once for X and once for Y in place of X. The property (i) then follows by a union bound.

We will obtain X' by forbidding some outcomes of $\boldsymbol{X}_I$ that are too likely. We build up a set $\mathcal{C}$ of conditions via the following algorithm. We use the notation $X_{\mathcal{C}} = \cap_{C \in \mathcal{C}} X_C$ below.

Initialize $\mathcal{C} \coloneqq \emptyset$
Repeat
    If $X_{\mathcal{C}} = \emptyset$, then halt with a *failure*
    If $\boldsymbol{X_{\mathcal{C}}}$ is 0.6-dense, then halt with a *success*
    Otherwise let I and α be such that $\mathbf{Pr}[\,(\boldsymbol{X_{\mathcal{C}}})_I = \alpha\,] > 2^{-0.6 \cdot b|I|}$
    Add the condition $(x_I \neq \alpha)$ to $\mathcal{C}$
End repeat

This process eventually halts since $|X_{\mathcal{C}}|$ decreases every time we add a new condition to $\mathcal{C}$. Let $\mathcal{F}$ denote the set of final conditions when the process halts. We show that $X' \coloneqq X_{\mathcal{F}}$ satisfies (i′, ii′). Write $\mathcal{F} = \cup_{s \in [n]} \mathcal{F}_s$, where $\mathcal{F}_s$ denotes conditions of the form $(x_I \neq \alpha)$, $|I| = s$ in $\mathcal{F}$.

Claim 22. $|\mathcal{F}_s| \leq 2^{0.7 \cdot bs}$.

Proof of claim. The effect of adding a new condition $(x_I \neq \alpha)$, $|I| = s$, to $\mathcal{C}$ is to shrink the size of $X_{\mathcal{C}}$ by a factor of $\mathbf{Pr}[\,(\boldsymbol{X_{\mathcal{C}}})_I \neq \alpha\,] < 1 - \delta$, where $\delta \coloneqq 2^{-0.6 \cdot bs}$. Our initial set has size $|X| \leq 2^{bn}$ and hence we cannot shrink it by such a condition more than $k \geq |\mathcal{F}_s|$ times, where k is the smallest number satisfying $|X|(1-\delta)^k < 1$. Solving for k gives $k \leq O(bn/\delta) = O(bn \cdot 2^{0.6 \cdot bs})$, which is at most $2^{0.7 \cdot bs}$ given our definition of b. ∎

We can now verify (i′) by a direct calculation:

$$\begin{aligned}
&& \mathbf{Pr}[\,\boldsymbol{x} \notin X'\,] &= \mathbf{Pr}[\,\boldsymbol{x} \notin X_{\mathcal{F}}\,] \\
&& &\leq \textstyle\sum_s \mathbf{Pr}[\,\boldsymbol{x} \notin X_{\mathcal{F}_s}\,] \\
&& &\leq \textstyle\sum_s \sum_{(x_I \neq \alpha) \in \mathcal{F}_s} \mathbf{Pr}[\,\boldsymbol{x}_I = \alpha\,] \\
&(\mathbf{H}_\infty(\boldsymbol{x}_I) \geq 0.8 \cdot b|I|) & &\leq \textstyle\sum_s |\mathcal{F}_s| \cdot 2^{-0.8 \cdot bs} \\
&(\text{Claim 22}) & &\leq \textstyle\sum_s 2^{-0.1 \cdot bs} \\
&& &\leq \epsilon/2.
\end{aligned}$$

This also proves (ii′) because the calculation implies that $X' \neq \emptyset$ which means that our process halted with a *success*. This concludes the proof of Lemma 20.

3. Definitions of models. In section 3.1 we introduce our restricted-by-default communication models, justify why they can be viewed as "zero-communication" models, and explain their relationships to known lower bound techniques. In section 3.2 we define their corresponding unrestricted versions. In section 3.3 we describe the query complexity counterparts of our communication models.

3.1. Restricted communication models. We define NP protocols in a slightly nonstandard way as randomized protocols, just for stylistic consistency with the other models. The acronyms WAPP and SBP were introduced in [8] (their communication versions turn out to be equivalent to the smooth rectangle bound and the corruption bound, as argued below). We introduce the acronym $\mathsf{2WAPP}$ (for lack of existing notation) to correspond to a two-sided version of WAPP (which is equivalent to the zero communication with abort model of [32]). We use the notation $\mathsf{PostBPP}$ [1] instead of the more traditional $\mathsf{BPP}_{\mathsf{path}}$ [27] as it is more natural for communication protocols.

A protocol outputs 0 or 1, and in some of these models it may also output $\perp$ representing "abort" or "don't know." In the following definition, α can be arbitrarily small and should be thought of as a function of the input size n for a family of protocols.

Definition 23. *For $\mathcal{C} \in \{\mathsf{NP}, \mathsf{2WAPP}_\epsilon, \mathsf{WAPP}_\epsilon, \mathsf{SBP}, \mathsf{PostBPP}\}$ and $F\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1,*\}$ a partial function, define $\mathcal{C}^{\mathsf{cc}}(F)$ as the minimum over all $\alpha > 0$ and all "α-correct" public-randomness protocols for F of the communication cost plus* $\log(1/\alpha)$ *(this sum is considered to be the cost), where α-correctness is defined as follows.*

- NP: *If $F(x,y) = 1$ then $\mathbf{Pr}[\,\Pi(x,y) = 1\,] \geq \alpha$, and if $F(x,y) = 0$ then $\mathbf{Pr}[\,\Pi(x,y) = 1\,] = 0$.*
- $\mathsf{2WAPP}_\epsilon$: *The protocol may output $\perp$, and for all $(x,y) \in \operatorname{dom} F$, $\mathbf{Pr}[\Pi(x,y) = F(x,y)] \geq (1-\epsilon)\alpha$ and $\mathbf{Pr}[\,\Pi(x,y) \neq \perp\,] \leq \alpha$.*
- WAPP_ϵ: *If $F(x,y) = 1$ then $\mathbf{Pr}[\,\Pi(x,y) = 1\,] \in [(1-\epsilon)\alpha, \alpha]$, and if $F(x,y) = 0$ then $\mathbf{Pr}[\,\Pi(x,y) = 1\,] \in [0, \epsilon\alpha]$.*[2]
- SBP: *If $F(x,y) = 1$ then $\mathbf{Pr}[\,\Pi(x,y) = 1\,] \geq \alpha$, and if $F(x,y) = 0$ then $\mathbf{Pr}[\,\Pi(x,y) = 1\,] \leq \alpha/2$.*
- $\mathsf{PostBPP}$: *The protocol may output $\perp$, and for all $(x,y) \in \operatorname{dom} F$, $\mathbf{Pr}[\,\Pi(x,y) \neq \perp\,] \geq \alpha$ and $\mathbf{Pr}[\,\Pi(x,y) = F(x,y) \mid \Pi(x,y) \neq \perp\,] \geq 3/4$.*

The "syntactic relationships" among the four models $\mathsf{2WAPP}$, WAPP, SBP, $\mathsf{PostBPP}$ is summarized in Table 1. The meaning of the column and row labels is as follows. For the columns, "two-sided" means that the protocol outputs values in $\{0,1,\perp\}$ and conditioned on not outputting $\perp$, the output is correct with high probability. A "one-sided" protocol outputs values in $\{0,1\}$, and we measure its probability of outputting 1 and compare it against the correctness parameter $\alpha > 0$. For the rows, "bounded" means that the *nonabort* probability—that is, the probability of not outputting $\perp$ for two-sided models, or the probability of outputting 1 for one-sided models—is uniformly upper bounded by α, whereas "unbounded" means that the nonabort probability need not be upper bounded by α.

It is straightforward to see that the relative computational power ("semantic relationships") of the models is as follows (recall Figure 1): for all F and all constants

[2] The definition of WAPP in [8] uses ϵ in a different way: $\frac{1}{2}+\epsilon$ and $\frac{1}{2}-\epsilon$ instead of $1-\epsilon$ and ϵ.

TABLE 1

	Two sided	One sided
Bounded nonabort	2WAPP	WAPP
Unbounded nonabort	PostBPP	SBP

$0 < \epsilon < 1/2$, we have $\mathsf{2WAPP}^{\mathsf{cc}}_{\epsilon}(F) \geq \mathsf{WAPP}^{\mathsf{cc}}_{\epsilon}(F) \geq \Omega(\mathsf{SBP}^{\mathsf{cc}}(F)) \geq \Omega(\mathsf{PostBPP}^{\mathsf{cc}}(F))$ and $\mathsf{NP}^{\mathsf{cc}}(F) \geq \mathsf{SBP}^{\mathsf{cc}}(F)$. Furthermore, exponential separations are known for all these relationships: unique set intersection is easy for $\mathsf{WAPP}^{\mathsf{cc}}_0$ but hard for $\mathsf{2WAPP}^{\mathsf{cc}}_{\epsilon}$ (indeed, for $\mathsf{coSBP}^{\mathsf{cc}}$ [49, 25]); set intersection is easy for $\mathsf{SBP}^{\mathsf{cc}}$ (indeed, for $\mathsf{NP}^{\mathsf{cc}}$) but hard for $\mathsf{WAPP}^{\mathsf{cc}}_{\epsilon}$ [35]; set disjointness is easy for $\mathsf{PostBPP}^{\mathsf{cc}}$ (indeed, for $\mathsf{coNP}^{\mathsf{cc}}$) but hard for $\mathsf{SBP}^{\mathsf{cc}}$ [49, 25]; equality is easy for $\mathsf{SBP}^{\mathsf{cc}}$ (indeed, for $\mathsf{coRP}^{\mathsf{cc}}$) but hard for $\mathsf{NP}^{\mathsf{cc}}$. Moreover, $\mathsf{WAPP}^{\mathsf{cc}}$ is a one-sided version of $\mathsf{2WAPP}^{\mathsf{cc}}$ in the sense that $\mathsf{2WAPP}^{\mathsf{cc}}_{\epsilon}(F) \leq O(\mathsf{WAPP}^{\mathsf{cc}}_{\epsilon/2}(F) + \mathsf{coWAPP}^{\mathsf{cc}}_{\epsilon/2}(F))$ (so the classes would satisfy $\mathsf{2WAPP}^{\mathsf{cc}} = \mathsf{WAPP}^{\mathsf{cc}} \cap \mathsf{coWAPP}^{\mathsf{cc}}$ if we ignore the precise value of the constant ϵ).

The reason we do not include an ϵ parameter in the $\mathsf{SBP}^{\mathsf{cc}}$ and $\mathsf{PostBPP}^{\mathsf{cc}}$ models is because standard amplification techniques could be used to efficiently decrease ϵ in these models (rendering the exact value immaterial up to constant factors). Another subtlety concerns the behavior of correct protocols on the *undefined* inputs $\{0,1\}^n \times \{0,1\}^n \smallsetminus \operatorname{dom} F$. For example, for $\mathsf{2WAPP}^{\mathsf{cc}}_{\epsilon}$, the corresponding definitions in [32] also require that for every undefined input (x, y), $\mathbf{Pr}[\,\Pi(x, y) \neq \perp\,] \in [(1-\epsilon)\alpha, \alpha]$. We allow arbitrary behavior on the undefined inputs for stylistic consistency, but our results also hold for the other version. As a final remark, we mention that our definition of $\mathsf{NP}^{\mathsf{cc}}$ is only equivalent to the usual definition within an additive logarithmic term; see Remark 27 below.

Relation to zero-communication models. The following fact shows that protocols in our models can be expressed simply as distributions over (labeled) rectangles; thus these models can be considered zero-communication since Alice and Bob can each produce an output with no communication, and then have the output of the protocol be a simple function of their individual outputs.[3]

FACT 24. *Without loss of generality (w.l.o.g.), in each of the five models from Definition* 23, *for each outcome of the public randomness the associated deterministic protocol is of the following form.*

NP, WAPP_{ϵ}, SBP : *There exists a rectangle R such that the output is* 1 *iff the input is in R.*

$\mathsf{2WAPP}_{\epsilon}$, $\mathsf{PostBPP}$: *There exists a rectangle R and a bit b such that the output is b if the input is in R and is $\perp$ otherwise.*

Proof. Consider a protocol Π in one of the models from Definition 23, and suppose it has communication cost c and associated $\alpha > 0$, so the cost is $c + \log(1/\alpha)$. We may assume that each deterministic protocol has exactly 2^c possible transcripts. Transform Π into a new protocol Π' that operates as follows on input (x, y): sample an outcome of the public randomness of Π, then sample a uniformly random transcript with associated rectangle R and output value b, then execute the following.

$$\text{If } (x,y) \in R \text{ then output } b, \text{ otherwise output } \begin{cases} 0 & \text{if } \mathsf{NP},\ \mathsf{WAPP}_{\epsilon},\ \mathsf{SBP}, \\ \perp & \text{if } \mathsf{2WAPP}_{\epsilon},\ \mathsf{PostBPP}. \end{cases}$$

[3]Admittedly, for Alice and Bob themselves to know the output of this simple function, they would need to use a constant amount of communication.

We have $\mathbf{Pr}[\,\Pi'(x,y)=1\,]=2^{-c}\,\mathbf{Pr}[\,\Pi(x,y)=1\,]$, and for $\mathsf{2WAPP}_\epsilon$, $\mathsf{PostBPP}$ we also have $\mathbf{Pr}[\,\Pi'(x,y)=0\,]=2^{-c}\,\mathbf{Pr}[\,\Pi(x,y)=0\,]$. Thus in all cases Π' is $(2^{-c}\alpha)$-correct. Formally, it takes two bits of communication to check whether $(x,y)\in R$, so the cost of Π' is $2+\log(1/2^{-c}\alpha)$, which is the cost of Π plus 2. ∎

Relation to lower bound measures. Using Fact 24 it is straightforward to see that, ignoring the $+2$ cost of checking whether the input is in a rectangle, $\mathsf{2WAPP}^{\mathsf{cc}}_\epsilon$ is exactly equivalent to the relaxed partition bound of [32] (with the aforementioned caveat about undefined inputs) and $\mathsf{WAPP}^{\mathsf{cc}}_\epsilon$ is exactly equivalent to the (one-sided) smooth rectangle bound,[4] denoted srec^1 [29]. For completeness, the definition of srec^1 and the proof of the following fact appear in Appendix A.1.

FACT 25. $\mathsf{srec}^1_\epsilon(F)\le\mathsf{WAPP}^{\mathsf{cc}}_\epsilon(F)\le\mathsf{srec}^1_\epsilon(F)+2$ *for all F and all $0<\epsilon<1/2$.*

It was shown in [25] that $\mathsf{SBP}^{\mathsf{cc}}$ is equivalent (within constant factors) to the (one-sided) corruption bound. We remark that by a simple application of the minimax theorem, $\mathsf{PostBPP}^{\mathsf{cc}}$ also has a dual characterization analogous to the corruption bound.[5]

3.2. Unrestricted communication models. For all the models described above, we can define their unrestricted versions, denoted by prepending U to the acronym (not to be confused with complexity classes where U stands for "unambiguous"). The distinction is that the restricted versions charge $+\log(1/\alpha)$ in the cost, whereas the unrestricted versions do not charge anything for α in the cost (and hence they are defined using private randomness; otherwise every function would be computable with constant cost.)

DEFINITION 26. *For $\mathcal{C}\in\{\mathsf{NP},\mathsf{2WAPP}_\epsilon,\mathsf{WAPP}_\epsilon,\mathsf{SBP},\mathsf{PostBPP}\}$ and $F\colon\{0,1\}^n\times\{0,1\}^n\to\{0,1,*\}$ a partial function, define $\mathsf{U}\mathcal{C}^{\mathsf{cc}}(F)$ as the minimum over all $\alpha>0$ and all α-correct private-randomness protocols for F of the communication cost, where the α-correctness criteria are as in Definition* 23.

Standard sparsification of randomness (à la Newman's theorem [43], [38, Theorem 3.14]) can be used to show that the unrestricted models are essentially at least as powerful as their restricted versions for all F: for $\mathcal{C}\in\{\mathsf{NP},\mathsf{SBP},\mathsf{PostBPP}\}$ we have $\mathsf{U}\mathcal{C}^{\mathsf{cc}}(F)\le O(\mathcal{C}^{\mathsf{cc}}(F)+\log n)$, and for $\mathcal{C}\in\{\mathsf{2WAPP},\mathsf{WAPP}\}$ we have $\mathsf{U}\mathcal{C}^{\mathsf{cc}}_\delta(F)\le O(\mathcal{C}^{\mathsf{cc}}_\epsilon(F)+\log(n/(\delta-\epsilon)))$, where $0<\epsilon<\delta$. (The additive logarithmic terms come from converting public randomness to private.)

Remark 27. We note that $\mathsf{UNP}^{\mathsf{cc}}$ is actually equivalent to the standard definition of nondeterministic communication complexity, while our $\mathsf{NP}^{\mathsf{cc}}$ from Definition 23 is only equivalent within an additive logarithmic term. It is fair to call this an abuse of notation, but it does not affect our communication query equivalence for NP since we consider block length $b=\Omega(\log n)$ anyway.

$\mathsf{UWAPP}^{\mathsf{cc}}$ *and nonnegative rank.* Of particular interest to us will be $\mathsf{UWAPP}^{\mathsf{cc}}$ which turns out to be equivalent to approximate nonnegative rank. Recall that for M a nonnegative matrix, the *nonnegative rank* $\mathrm{rank}^+(M)$ is defined as the minimum r such that M can be written as the sum of r nonnegative rank-1 matrices or, equivalently,

[4]The paper that introduced this bound [29] defined it as the optimum value of a certain linear program, but following [37] we define it as the log of the optimum value.

[5]$\mathsf{PostBPP}^{\mathsf{cc}}(F)$ is big-Θ of the maximum over all distributions μ over $\{0,1\}^n\times\{0,1\}^n$ of the minimum $\log(1/\mu(R))$ over all rectangles R that are unbalanced in the sense that $\mu(R\cap F^{-1}(1))$ and $\mu(R\cap F^{-1}(0))$ are not within a factor of 2 of each other. In the corruption bound, the maximum is only over balanced μ, and R is considered unbalanced if $\mu(R\cap F^{-1}(1))$ is more than some constant factor greater than $\mu(R\cap F^{-1}(0))$.

$M = UV$ for nonnegative matrices U, V with inner dimension r for the multiplication. Below, we view a partial function $F\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1,*\}$ as a $2^n \times 2^n$ partial boolean matrix.

DEFINITION 28 (approximate nonnegative rank). *For partial F, $\operatorname{rank}^+_\epsilon(F)$ is defined as the minimum $\operatorname{rank}^+(M)$ over all nonnegative matrices M such that $M_{x,y} \in F(x,y) \pm \epsilon$ for all $(x,y) \in \operatorname{dom} F$ (in other words, $\|F - M\|_\infty \le \epsilon$ on $\operatorname{dom} F$).*

For completeness, the straightforward proof of the following fact appears in Appendix A.2.

FACT 29. $\log \operatorname{rank}^+_\epsilon(F) \le \mathsf{UWAPP}^{\mathsf{cc}}_\epsilon(F) \le \lceil \log \operatorname{rank}^+_{\epsilon/2}(F) \rceil + 2$ *for all F and all $0 < \epsilon < 1/2$.*

3.3. Query models. A randomized decision tree $\mathcal{T}$ is a probability distribution over deterministic decision trees, and the query cost is the maximum height of a decision tree in the support.

DEFINITION 30. *For $\mathcal{C} \in \{\mathsf{NP}, \mathsf{2WAPP}_\epsilon, \mathsf{WAPP}_\epsilon, \mathsf{SBP}, \mathsf{PostBPP}\}$ and $f\colon \{0,1\}^n \to \{0,1,*\}$ a partial function, define $\mathcal{C}^{\mathsf{dt}}(f)$ as the minimum over all $\alpha > 0$ and all α-correct randomized decision trees for f of the query cost, where the α-correctness criteria are as in Definition 23 (but where protocols $\Pi(x,y)$ are replaced with randomized decision trees $\mathcal{T}(z)$).*

Completely analogously to how the zero-communication models can be viewed w.l.o.g. as distributions over (labeled) rectangles (Fact 24), their query counterparts can be viewed w.l.o.g. as distributions over (labeled) conjunctions.

FACT 31. *W.l.o.g., in each of the five models from Definition 30, for each outcome of the randomness the associated deterministic decision tree is of the following form.*

NP, WAPP_ϵ, SBP: *There exists a conjunction h such that the output is 1 iff the input is in $h^{-1}(1)$.*

$\mathsf{2WAPP}_\epsilon$, $\mathsf{PostBPP}$: *There exists a conjunction h and a bit b such that the output is b if the input is in $h^{-1}(1)$ and is $\perp$ otherwise.*

Proof. Consider a randomized decision tree $\mathcal{T}$ in one of the models from Definition 30, and suppose it has query cost d and associated $\alpha > 0$. We may assume that each deterministic decision tree has a full set of 2^d leaves and the queries along each root-to-leaf path are distinct. Hence each leaf is associated with a width-d conjunction that checks whether the input is consistent with the queries made in its root-to-leaf path. Transform $\mathcal{T}$ into a new randomized decision tree $\mathcal{T}'$ that operates as follows on input z: sample an outcome of the randomness of $\mathcal{T}$, then sample a uniformly random leaf with associated conjunction h and output value b, then execute the following.

$$\text{If } h(z) = 1 \text{ then output } b \text{, otherwise output } \begin{cases} 0 & \text{if } \mathsf{NP},\ \mathsf{WAPP}_\epsilon,\ \mathsf{SBP}, \\ \perp & \text{if } \mathsf{2WAPP}_\epsilon,\ \mathsf{PostBPP}. \end{cases}$$

We have $\mathbf{Pr}[\,\mathcal{T}'(z) = 1\,] = 2^{-d}\,\mathbf{Pr}[\,\mathcal{T}(z) = 1\,]$, and for $\mathsf{2WAPP}_\epsilon$, $\mathsf{PostBPP}$ we also have $\mathbf{Pr}[\,\mathcal{T}'(z) = 0\,] = 2^{-d}\,\mathbf{Pr}[\,\mathcal{T}(z) = 0\,]$. Thus in all cases $\mathcal{T}'$ is $(2^{-d}\alpha)$-correct, and $\mathcal{T}'$ also has query cost d. ∎

We defined our query models without charging anything for α, i.e., α is unrestricted. This means that deriving communication upper bounds for $f \circ g^n$ in restricted models from corresponding query upper bounds for f is nontrivial; this is discussed in section 4.2. Nevertheless, we contend that Definitions 23 and 30 are the "right" definitions that correspond to one another. The main reason is because in the "normal forms" (Facts 24 and 31), all the cost in the communication version comes from α, and all the cost in the query version comes from the width of the conjunctions—and when we apply the junta theorem in section 4.1, the communication α directly determines the conjunction width.

4. Proof of the simulation theorem. In this section we derive the simulation theorem (Theorem 2) from the junta theorem (Theorem 1). The proof is in two parts: section 4.1 for lower bounds and section 4.2 for upper bounds.

4.1. Communication lower bounds. The junta theorem implies that for functions lifted with our hard gadget g, every distribution $\mathcal{R}$ over rectangles can be transformed into a distribution $\mathcal{H}$ over conjunctions such that for every $z \in \{0,1\}^n$, the acceptance probability under $\mathcal{H}$ is related in a simple way to the acceptance probability under $\mathcal{R}$ averaged over all two-party encodings of z. This allows us to convert zero-communication protocols (which are distributions over (labeled) rectangles by Fact 24) into corresponding decision trees (which are distributions over (labeled) conjunctions by Fact 31).

More precisely, let $\mathcal{R}$ be a distribution over rectangles in the domain of $G = g^n$. First, apply the junta theorem to each R in the support of $\mathcal{R}$ to get an approximating conical d-junta h_R. Now we can approximate the convex combination

$$\begin{aligned} \mathrm{acc}_{\mathcal{R}}(z) &= \mathop{\mathbf{E}}_{\boldsymbol{R}\sim\mathcal{R}} \mathrm{acc}_{\boldsymbol{R}}(z) \\ &\in \mathop{\mathbf{E}}_{\boldsymbol{R}\sim\mathcal{R}} \Big((1 \pm o(1)) \cdot h_{\boldsymbol{R}}(z) \pm 2^{-\Theta(db)} \Big) \\ &\subseteq (1 \pm o(1)) \cdot \Big(\mathop{\mathbf{E}}_{\boldsymbol{R}\sim\mathcal{R}} h_{\boldsymbol{R}}(z) \Big) \pm 2^{-\Theta(db)} \end{aligned}$$

by the conical d-junta $\mathbf{E}_{\boldsymbol{R}\sim\mathcal{R}}\, h_{\boldsymbol{R}}$ with the same parameters as in the junta theorem (we settle for multiplicative error $(1 \pm o(1))$ since it suffices for the applications). But conical d-juntas are—up to scaling—convex combinations of width-d conjunctions. Specifically, we may write any conical d-junta as $\mathrm{acc}_{\mathcal{H}}(z)/a$, where $a > 0$ is some constant of proportionality and $\mathrm{acc}_{\mathcal{H}}(z) := \mathbf{E}_{\boldsymbol{h}\sim\mathcal{H}}\, \boldsymbol{h}(z)$, where $\mathcal{H}$ is a distribution over width-d conjunctions. Finally, we rearrange the approximation so the roles of $\mathrm{acc}_{\mathcal{H}}(z)$ and $\mathrm{acc}_{\mathcal{R}}(z)$ are swapped, since it is more convenient for the applications. Hence we arrive at the following reformulation of the junta theorem.

COROLLARY 32 (junta theorem—reformulation). *Assume (†). For any $d \geq 0$ and any distribution $\mathcal{R}$ over rectangles in the domain of g^n there exist a distribution $\mathcal{H}$ over width-d conjunctions and a constant of proportionality $a > 0$ such that, for all $z \in \{0,1\}^n$,*

$$\mathrm{acc}_{\mathcal{H}}(z) \in a \cdot \big((1 \pm o(1)) \cdot \mathrm{acc}_{\mathcal{R}}(z) \pm 2^{-\Theta(db)} \big). \tag{8}$$

We will now prove the lower bounds in Theorem 2. Here the error parameters for WAPP are made more explicit.

THEOREM 33. *Assume (†). For any partial $f\colon \{0,1\}^n \to \{0,1,*\}$ and constants $0 < \epsilon < \delta < 1/2$,*

$$\begin{aligned} &\mathcal{C}^{\mathsf{cc}}(f \circ g^n) \geq \Omega(\mathcal{C}^{\mathsf{dt}}(f) \cdot b) \qquad \textit{for} \quad \mathcal{C} \in \{\mathsf{NP}, \mathsf{SBP}, \mathsf{PostBPP}\}, \\ &\mathcal{C}^{\mathsf{cc}}_{\epsilon}(f \circ g^n) \geq \Omega(\mathcal{C}^{\mathsf{dt}}_{\delta}(f) \cdot b) \qquad \textit{for} \quad \mathcal{C} \in \{\mathsf{2WAPP}, \mathsf{WAPP}\}. \end{aligned}$$

Proof. For convenience of notation we let $\mathcal{C}^{\mathsf{cc}} := \mathcal{C}^{\mathsf{cc}}_{\epsilon}$ and $\mathcal{C}^{\mathsf{dt}} := \mathcal{C}^{\mathsf{dt}}_{\delta}$ in the $\mathcal{C} \in \{\mathsf{2WAPP}, \mathsf{WAPP}\}$ cases. Given an α-correct cost-c $\mathcal{C}^{\mathsf{cc}}$ protocol Π for $f \circ g^n$ assumed to be in the normal form given by Fact 24, we convert it into a cost-$O(c/b)$ $\mathcal{C}^{\mathsf{dt}}$ decision tree $\mathcal{T}$ for f.

For $\mathcal{C} \in \{\mathsf{NP}, \mathsf{WAPP}, \mathsf{SBP}\}$, Π is a distribution over rectangles, so applying Corollary 32 with $d \coloneqq O(c/b)$ so that $2^{-\Theta(db)} \leq o(2^{-c}) = o(\alpha)$, there exists a distribution $\mathcal{T}$ over width-d conjunctions and an $a > 0$ such that for all $z \in \{0,1\}^n$, $\mathrm{acc}_{\mathcal{T}}(z) \in a \cdot ((1 \pm o(1)) \cdot \mathrm{acc}_{\Pi}(z) \pm o(\alpha))$. Note that $\mathrm{acc}_{\Pi}(z)$ obeys the α-correctness criterion of f since it obeys the α-correctness criterion of $f \circ g^n$ for each encoding of z. Hence $\mathrm{acc}_{\mathcal{T}}(z)$ obeys the $(a\alpha')$-correctness criterion for some $\alpha' \in \alpha \cdot (1 \pm o(1))$. (For $\mathcal{C} = \mathsf{SBP}$, slight amplification may be needed. Also, for $\mathcal{C} = \mathsf{NP}$ we need to ensure that $\mathrm{acc}_{\mathcal{T}}(z) = 0$ whenever $\mathrm{acc}_{\Pi}(z) = 0$, but this is implicit in the proof of the junta theorem; see the left side of (4).) In conclusion, $\mathcal{T}$ is a cost-d $\mathcal{C}^{\mathsf{dt}}$ decision tree for f.

For $\mathcal{C} \in \{\mathsf{2WAPP}, \mathsf{PostBPP}\}$, Π can be viewed as a convex combination $\pi_0\Pi_0 + \pi_1\Pi_1$, where Π_0 is a distribution over 0-labeled rectangles and Π_1 is a distribution over 1-labeled rectangles. Applying the above argument to Π_0 and Π_1 separately, we may assume the scaling factor a is the same for both, by assigning some probability to a special "contradictory" conjunction that accepts nothing. We get a distribution over labeled width-d conjunctions $\mathcal{T} \coloneqq \pi_0\mathcal{T}_0 + \pi_1\mathcal{T}_1$ such that $\mathbf{Pr}[\,\mathcal{T}(z) = 0\,] = \pi_0\,\mathrm{acc}_{\mathcal{T}_0}(z) \in \pi_0 a \cdot \big((1 \pm o(1)) \cdot \mathrm{acc}_{\Pi_0}(z) \pm o(\alpha)\big) \subseteq a \cdot \big((1 \pm o(1)) \cdot \mathbf{Pr}[\,\Pi(z) = 0\,] \pm o(\alpha)\big)$, where we use the shorthand $\mathbf{Pr}[\,\Pi(z) = 0\,] \coloneqq \mathbf{E}_{\boldsymbol{xy} \sim \mathcal{U}_z}\, \mathbf{Pr}[\,\Pi(\boldsymbol{x}, \boldsymbol{y}) = 0\,]$. An analogous property holds for outputting 1 instead of 0. Note that $\mathbf{Pr}[\,\Pi(z) = 0\,]$ and $\mathbf{Pr}[\,\Pi(z) = 1\,]$ obey the α-correctness criterion since they do for each encoding of z. Hence $\mathbf{Pr}[\,\mathcal{T}(z) = 0\,]$ and $\mathbf{Pr}[\,\mathcal{T}(z) = 1\,]$ obey the $(a\alpha')$-correctness criterion for some $\alpha' \in \alpha \cdot (1 \pm o(1))$. (For $\mathcal{C} = \mathsf{PostBPP}$, slight amplification may be needed.) In conclusion, $\mathcal{T}$ is a cost-d $\mathcal{C}^{\mathsf{dt}}$ decision tree for f. ∎

4.2. Communication upper bounds.

THEOREM 34. *Let* $\mathcal{C} \in \{\mathsf{NP}, \mathsf{2WAPP}_\epsilon, \mathsf{WAPP}_\epsilon, \mathsf{SBP}\}$. *For any partial* $f\colon \{0,1\}^n \to \{0,1,*\}$ *and any gadget* $g\colon \{0,1\}^b \times \{0,1\}^b \to \{0,1\}$, *we have* $\mathcal{C}^{\mathsf{cc}}(f \circ g^n) \leq O(\mathcal{C}^{\mathsf{dt}}(f) \cdot (b + \log n))$.

Proof. On input (x, y) the communication protocol just simulates the randomized decision tree on input $z \coloneqq g^n(x, y)$, and when the decision tree queries the ith bit of z, the communication protocol evaluates $z_i \coloneqq g(x_i, y_i)$ by brute force. This has communication cost $\mathcal{C}^{\mathsf{dt}}(f) \cdot (b + 1)$, and it inherits the α parameter from the randomized decision tree. The nontrivial part is that the query models allow arbitrarily small α, which could give arbitrarily large $+\log(1/\alpha)$ cost to the communication protocol. For these particular query models, it turns out that we can assume w.l.o.g. that $\log(1/\alpha) \leq O(\mathcal{C}^{\mathsf{dt}}(f) \cdot \log n)$. We state and prove this for $\mathsf{SBP}^{\mathsf{dt}}$ below. (The other three models are no more difficult to handle.) ∎

PROPOSITION 35. *Every partial function* f *admits an* α*-correct* $\mathsf{SBP}^{\mathsf{dt}}$ *decision tree of query cost* $d \coloneqq \mathsf{SBP}^{\mathsf{dt}}(f)$, *where* $\alpha \geq 2^{-d}\binom{n}{d}^{-1} \geq 2^{-O(d \cdot \log n)}$.

Proof. Consider an α'-correct cost-d $\mathsf{SBP}^{\mathsf{dt}}$ decision tree for f in the normal form given by Fact 31. We may assume each deterministic decision tree in the support is a conjunction with exactly d literals (and there are $2^d\binom{n}{d}$ many such conjunctions). The crucial observation is that it never helps to assign a probability larger than α' to any conjunction: if some conjunction appears with probability $p > \alpha'$, we may replace its probability with α' and assign the leftover probability $p - \alpha'$ to a special contradictory conjunction that accepts nothing. This modified randomized decision tree is still α'-correct for f. Finally, remove all probability from the contradictory conjunction and scale the remaining probabilities (along with α') to sum up to 1.

Let α be the scaled version of α'. Now we have that α is greater than or equal to each of $2^d\binom{n}{d}$ many probabilities, and hence α must be at least the reciprocal of this number. ∎

Remark 36. In the case of $\mathsf{PostBPP}^{\mathsf{dt}}$ we cannot assume w.l.o.g. that $\log(1/\alpha) \leq \mathrm{poly}(d, \log n)$. The canonical counterexample is a *decision list* function $f\colon \{0,1\}^n \to \{0,1\}$ defined relative to a binary vector $(a_1, \dots, a_n) \in \{0,1\}^n$ so that $f(x) \coloneqq a_i$, where $i \in [n]$ is the smallest number such that $x_i = 1$, or $f(x) \coloneqq 0$ if no such i exists. Each decision list admits a cost-1 $\mathsf{PostBPP}^{\mathsf{dt}}$ decision tree, but for some decision lists the associated α must be exponentially small in n; see, e.g., [9] for more details. Indeed, two-party lifts of decision lists have been used in separating unrestricted communication models from restricted ones as we will discuss in section 6.

5. Applications of the simulation theorem. In this section we use the simulation theorem to derive our applications. We prove Theorem 3 and Theorem 6 in sections 5.1 and 5.2, respectively. Throughout this section we use $o(1)$ to denote a quantity that is upper bounded by some sufficiently small constant, which may be different for the different instances of $o(1)$. (For example, $a \leq o(b)$ formally means there exists a constant $\epsilon > 0$ such that $a \leq \epsilon \cdot b$.)

5.1. Nonclosure under intersection. Recall that $f_\wedge(z, z') \coloneqq f(z) \wedge f(z')$. Here $f_\wedge$ is *not* to be thought of as a two-party function; we study the query complexity of $f_\wedge$, whose input we happen to divide into two halves called z and z'. We start with the following lemma.

LEMMA 37. *There exists a partial f such that $\mathsf{SBP}^{\mathsf{dt}}(f) \leq O(1)$, but $\mathsf{SBP}^{\mathsf{dt}}(f_\wedge) \geq \Omega(n^{1/4})$.*

Let $k \coloneqq o(\sqrt{n})$ and define a partial function $f\colon \{0,1\}^n \to \{0,1,*\}$ by

$$f(z) \coloneqq \begin{cases} 1 & \text{if } |z| \geq k, \\ 0 & \text{if } |z| \leq k/2, \\ * & \text{otherwise,} \end{cases}$$

where $|z|$ denotes the Hamming weight of z.

In proving the lower bound in Lemma 37 we make use of the following duality principle for $\mathsf{SBP}^{\mathsf{dt}}$, which we phrase abstractly in terms of a collection $\mathscr{H}$ of "basic functions" over some finite set of inputs Z. In our concrete case $\mathscr{H}$ consists of decision trees of height d or, equivalently width-d conjunctions by Fact 31, and $Z \subseteq \{0,1\}^n$ is the domain of the partial function f. We state the duality principle for acceptance gap $[0, \alpha/2)$ versus $(\alpha, 1]$ rather than $[0, \alpha/2]$ versus $[\alpha, 1]$ as this implicitly ensures $\alpha > 0$. The slight difference in the multiplicative gap, (> 2) versus (≥ 2), is immaterial as the gap can be efficiently amplified for SBP affecting only constant factors.

FACT 38. *For all $\mathscr{H} \subseteq \{0,1\}^Z$ and nonconstant $f\colon Z \to \{0,1\}$, the following are equivalent.*

(i) *There exists a distribution $\mathcal{H}$ over $\mathscr{H}$ such that for all $(z_1, z_0) \in f^{-1}(1) \times f^{-1}(0)$,*

$$\Pr_{\boldsymbol{h}\sim\mathcal{H}}[\,\boldsymbol{h}(z_1) = 1\,] > 2 \cdot \Pr_{\boldsymbol{h}\sim\mathcal{H}}[\,\boldsymbol{h}(z_0) = 1\,]. \tag{9}$$

(ii) *For each pair of distributions (μ_1, μ_0) over $f^{-1}(1)$ and $f^{-1}(0)$ there is an $h \in \mathscr{H}$ with*

$$\Pr_{\boldsymbol{z_1}\sim\mu_1}[\,h(\boldsymbol{z_1}) = 1\,] > 2 \cdot \Pr_{\boldsymbol{z_0}\sim\mu_0}[\,h(\boldsymbol{z_0}) = 1\,]. \tag{10}$$

The direction (i) $\Rightarrow$ (ii) is trivial and is all we need for our proof, but it is interesting that the converse direction (ii) $\Rightarrow$ (i) also holds, by a slightly nonstandard argument. We include a full proof in Appendix A.4.

We also use the following basic calculation (given in Appendix A.3 for completeness).

FACT 39. *Let $h\colon \{0,1\}^n \to \{0,1\}$ be a width-d conjunction with i positive literals. Then h accepts a uniformly random string of Hamming weight w with probability $\in (w/n)^i \cdot (1 \pm o(1))$ provided $w \le o(\sqrt{n})$ and $d \le o(\sqrt{w})$.*

Proof of Lemma 37. Let f and $f_\wedge$ be as above. We have $\mathsf{SBP}^{\mathsf{dt}}(f) = 1$ via the decision tree $\mathcal{T}$ that picks a random coordinate and accepts iff the coordinate is 1. For the lower bound on $\mathsf{SBP}^{\mathsf{dt}}(f_\wedge)$, we use the contrapositive of (i) $\Rightarrow$ (ii). Let $\mathscr{H}$ consist of all conjunctions of width $o(n^{1/4})$. Let $\mathcal{Z}_w$ denote the uniform distribution over n-bit strings of weight w, intended to be used as either the first input z or the second input z' to $f_\wedge$. We construct a hard pair of distributions (μ_1, μ_0) over $f_\wedge^{-1}(1)$ and $f_\wedge^{-1}(0)$, respectively, by

$$\mu_1 := \mathcal{Z}_k \times \mathcal{Z}_k, \qquad \mu_0 := \frac{1}{2}(\mathcal{Z}_{k/2} \times \mathcal{Z}_{2k}) + \frac{1}{2}(\mathcal{Z}_{2k} \times \mathcal{Z}_{k/2}).$$

Here $\times$ denotes concatenation of strings, e.g., $(\boldsymbol{z}, \boldsymbol{z'}) \sim \mu_1$ is such that $\boldsymbol{z}, \boldsymbol{z'} \sim \mathcal{Z}_k$ and $\boldsymbol{z}$ and $\boldsymbol{z'}$ are independent. For intuition why the pair (μ_1, μ_0) is hard, consider the natural decision tree $\mathcal{T}_\wedge$ attempting to compute $f_\wedge$ that runs $\mathcal{T}$ (defined above) twice, once for z and once for z', accepting iff both runs accept. Since $\mathcal{T}$ accepts $\mathcal{Z}_k$ with probability k/n, we have that $\mathcal{T}_\wedge$ accepts μ_1 with probability k^2/n^2. Similarly, $\mathcal{T}_\wedge$ accepts μ_0 with probability $\frac{1}{2}(k/2n) \cdot (2k/n) + \frac{1}{2}(2k/n) \cdot (k/2n) = k^2/n^2$. Hence $\mathcal{T}_\wedge$ fails to distinguish between μ_1 and μ_0. More generally, we make a similar calculation for any width-$o(n^{1/4})$ conjunction. Indeed, let $h\colon \{0,1\}^{2n} \to \{0,1\}$ be an arbitrary conjunction in $\mathscr{H}$, and suppose h has i positive literals in z and j positive literals in z'. Then by Fact 39 we have

$$\begin{aligned}
\frac{\mathbf{Pr}_{(\boldsymbol{z},\boldsymbol{z'})\sim\mu_1}[\,h(\boldsymbol{z},\boldsymbol{z'}) = 1\,]}{\mathbf{Pr}_{(\boldsymbol{z},\boldsymbol{z'})\sim\mu_0}[\,h(\boldsymbol{z},\boldsymbol{z'}) = 1\,]} &\in \frac{(k/n)^i \cdot (k/n)^j}{\frac{1}{2} \cdot (k/2n)^i \cdot (2k/n)^j + \frac{1}{2} \cdot (2k/n)^i \cdot (k/2n)^j} \cdot (1 \pm o(1)) \\
&= \frac{1}{\frac{1}{2} \cdot 2^{j-i} + \frac{1}{2} \cdot 2^{i-j}} \cdot (1 \pm o(1)) \\
&\le 1 \cdot (1 \pm o(1)) \\
&\le 2.
\end{aligned}$$

This means that $\neg$(ii) and hence $\neg$(i). Therefore $f_\wedge$ has no cost-$o(n^{1/4})$ $\mathsf{SBP}^{\mathsf{dt}}$ decision tree. ∎

We can now prove Theorem 3, restated here from the introduction.

THEOREM 3. $\mathsf{SBP}^{\mathsf{cc}}$ *is not closed under intersection.*

Proof. Let f and $f_\wedge$ be as above. Define $F := f \circ g^n$ and $F_\wedge := f_\wedge \circ g^{2n} = (f \circ g^n)_\wedge$, where $g\colon \{0,1\}^b \times \{0,1\}^b \to \{0,1\}$, $b = \Theta(\log n)$, is our hard gadget from (†). Then by the simulation theorem (Theorem 2), we have $\mathsf{SBP}^{\mathsf{cc}}(F_\wedge) \ge \Omega(\mathsf{SBP}^{\mathsf{dt}}(f_\wedge) \cdot b) \ge \Omega(n^{1/4} \cdot b)$ which is not polylogarithmic in the input length so that $F_\wedge \notin \mathsf{SBP}^{\mathsf{cc}}$. Furthermore, we have $\mathsf{SBP}^{\mathsf{cc}}(F) \le O(\mathsf{SBP}^{\mathsf{dt}}(f) \cdot b) \le O(b)$ which is logarithmic in the input length. Thus $F \in \mathsf{SBP}^{\mathsf{cc}}$, which implies that $F_\wedge$ is the intersection of two functions in $\mathsf{SBP}^{\mathsf{cc}}$ (one that evaluates F on the first half of the input, and one that evaluates F on the second half). ∎

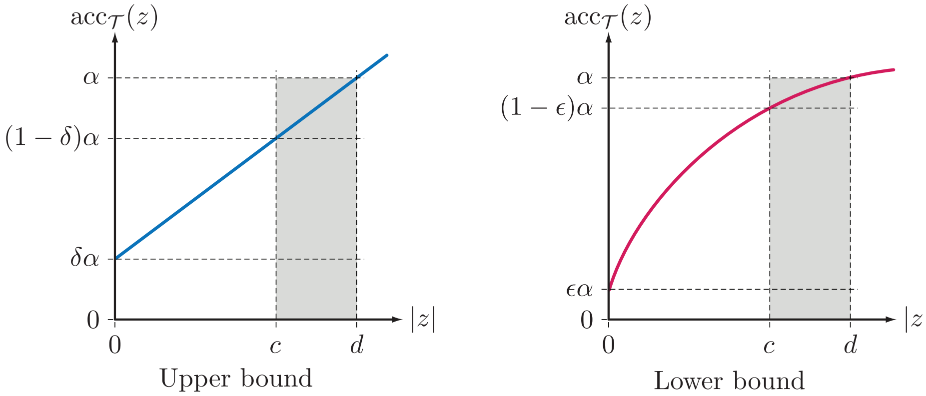

FIG. 3. *Illustration for the proof of Theorem* 6.

5.2. Unamplifiability of error. Our next application of the simulation theorem shows that the error parameter ϵ for $\mathsf{WAPP}^{\mathsf{cc}}$ cannot be efficiently amplified. Combining this with the results illustrated in Figure 4 (in particular, the fact that the equivalence holds for partial functions) shows that also for approximate nonnegative rank, ϵ cannot be efficiently amplified.

THEOREM 6. *For all constants $0 < \epsilon < \delta < 1/2$ there exists a two-party partial function F such that $\mathsf{WAPP}^{\mathsf{cc}}_\delta(F) \le O(\log n)$ but $\mathsf{WAPP}^{\mathsf{cc}}_\epsilon(F) \ge \Omega(n)$.*

Proof. Let c/d be a rational (expressed in lowest terms) such that $(1-2\delta)/(1-\delta) \le c/d < (1-2\epsilon)/(1-\epsilon)$. Note that such c, d exist (since $\epsilon < \delta$) and that they are constants depending only on ϵ and δ. Define a partial function $f\colon \{0,1\}^n \to \{0,1,*\}$ by

$$f(z) \coloneqq \begin{cases} 1 & \text{if } |z| \in \{c,d\}, \\ 0 & \text{if } |z| = 0, \\ * & \text{otherwise,} \end{cases}$$

where $|z|$ denotes the Hamming weight of z. By the simulation theorem (Theorems 33 and 34), it suffices to prove that $\mathsf{WAPP}^{\mathsf{dt}}_\delta(f) \le O(1)$ and $\mathsf{WAPP}^{\mathsf{dt}}_\epsilon(f) \ge \Omega(n)$.

Upper bound. Consider a cost-1 decision tree $\mathcal{T}'$ that picks a random coordinate and accepts iff the coordinate is 1. Then $\mathrm{acc}_{\mathcal{T}'}(z) = |z|/n$. Let $\alpha \coloneqq d/n$ and define $\mathcal{T}$ as follows: on input z accepts with probability $\delta\alpha$, rejects with probability $\delta(1-\alpha)$, and runs $\mathcal{T}'(z)$ with the remaining probability $(1-\delta)$. Now $\mathrm{acc}_\mathcal{T}(z)$ behaves as plotted on the left side of Figure 3: if $|z| = 0$ then $\mathrm{acc}_\mathcal{T}(z) = \delta\alpha$, if $|z| = d$ then $\mathrm{acc}_\mathcal{T}(z) = \delta\alpha + (1-\delta)d/n = \alpha$, and if $|z| = c$ then $\mathrm{acc}_\mathcal{T}(z) = \delta\alpha + (1-\delta)c/n$ which is at most α and at least $\delta\alpha + (1-\delta)d(1-2\delta)/((1-\delta)n) = \delta\alpha + (1-2\delta)\alpha = (1-\delta)\alpha$. In particular, $\mathcal{T}$ is an α-correct $\mathsf{WAPP}^{\mathsf{dt}}_\delta$ decision tree for f.

Lower bound. The $\mathsf{WAPP}^{\mathsf{dt}}_\delta$ decision tree designed above is "tight" for f in the following sense: if we decrease the error parameter from δ to ϵ, there is no longer any convex function of $|z|$ that would correspond to the acceptance probability of an α-correct $\mathsf{WAPP}^{\mathsf{dt}}_\epsilon$ decision tree for f. This is suggested on the right side of Figure 3: only a nonconvex function of $|z|$ can satisfy the α-correctness requirements for f. We show that the acceptance probability of any low-cost $\mathsf{WAPP}^{\mathsf{dt}}_\epsilon$ decision tree can indeed be accurately approximated by a convex function, which then yields a contradiction.

We now give the details. Suppose for contradiction that $\mathcal{T}$ is a distribution over width-$o(n)$ conjunctions (by Fact 31) forming an α-correct $\mathsf{WAPP}^{\mathsf{dt}}_\epsilon$ decision tree for

f for some arbitrary $\alpha > 0$. Consider the function $Q\colon \{0, c, d\} \to [0, 1]$ defined by $Q(w) \coloneqq \mathbf{E}_{\boldsymbol{z}\,:\,|\boldsymbol{z}|=w}\, \mathrm{acc}_{\mathcal{T}}(\boldsymbol{z})$, where the expectation is over a uniformly random string of Hamming weight w. Note that $Q(0) \in [0, \epsilon\alpha]$ and $Q(w) \in [(1-\epsilon)\alpha, \alpha]$ for $w \in \{c, d\}$ by the correctness of $\mathcal{T}$. A function $R\colon \{0, c, d\} \to \mathbb{R}$ is convex iff $(R(c) - R(0))/c \leq (R(d)-R(0))/d$. Note that Q is nonconvex since $((1-\epsilon)\alpha-\epsilon\alpha)/c > (\alpha-\epsilon\alpha)/d$. In fact, this shows that there cannot exist a convex function R that pointwise multiplicatively approximates Q within $1 \pm o(1)$. However, we claim that there exists such an R, which provides the desired contradiction.

We now argue the claim. For a width-$o(n)$ conjunction h, let $Q_h\colon \{0, c, d\} \to [0, 1]$ be defined by $Q_h(w) \coloneqq \mathbf{Pr}_{\boldsymbol{z}\,:\,|\boldsymbol{z}|=w}[\, h(\boldsymbol{z}) = 1\,]$, and note that $Q = \mathbf{E}_{\boldsymbol{h}\sim\mathcal{T}}\, Q_{\boldsymbol{h}}$. We show that for each such h, Q_h can be multiplicatively approximated by a convex function R_h. Hence Q is multiplicatively approximated by the convex function $R \coloneqq \mathbf{E}_{\boldsymbol{h}\sim\mathcal{T}}\, R_{\boldsymbol{h}}$.

Let $\ell \leq o(n)$ denote the number of literals in h, and let i denote the number of positive literals in h. If $i > c$, we have $Q_h(0) = Q_h(c) = 0$ and thus Q_h is convex and we can take $R_h \coloneqq Q_h$. Henceforth suppose $i \leq c$. Using the notation $(t)_m$ for the falling factorial $t(t-1)\cdots(t-m+1)$, for $w \in \{c, d\}$ we have $Q_h(w) = \binom{n-\ell}{w-i}/\binom{n}{w} = (w)_i(n-\ell)_{w-i}/(n)_w$.

Suppose $i = 0$. Then $Q_h(0) = 1$, and for $w \in \{c, d\}$ we have $Q_h(w) = (n-\ell)_w/(n)_w \geq (1-o(1))^w \geq 1 - o(1)$ (since $\ell \leq o(n)$). Thus we can let R_h be the constant 1 function. Now suppose $1 \leq i \leq c$. Then $Q_h(0) = 0$, and for $w \in \{c, d\}$ we denote the "0 to w slope" as $s_w \coloneqq (Q_h(w)-Q_h(0))/w = (w-1)_{i-1}(n-\ell)_{w-i}/(n)_w$. We have

$$\frac{s_c}{s_d} = \frac{(c-1)_{i-1}}{(d-1)_{i-1}} \cdot \frac{(n-\ell)_{c-i}}{(n-\ell)_{d-i}} \cdot \frac{(n)_d}{(n)_c} = \frac{(c-1)_{i-1}}{(d-1)_{i-1}} \cdot \frac{(n-c)_{d-c}}{(n-\ell-c+i)_{d-c}}.$$

The second multiplicand on the right side is at least 1 and at most $(1 + o(1))^{d-c} \leq 1 + o(1)$ since $\ell \leq o(n)$. Now we consider two subcases. If $2 \leq i \leq c$ then the first multiplicand on the right side is at most $1-\Omega(1)$ since $c < d$; hence $s_c/s_d \leq 1$ and thus Q_h is convex and we can take $R_h \coloneqq Q_h$. Suppose $i = 1$. Then the first multiplicand on the right side is 1, and hence $s_c/s_d \in 1 \pm o(1)$. This means Q_h is approximately linear. More precisely, defining $R_h(w) \coloneqq s_c \cdot w$, we have $R_h(0) = Q_h(0)$, $R_h(c) = Q_h(c)$, and $R_h(d) = Q_h(d) \cdot s_c/s_d \in Q_h(d) \cdot (1 \pm o(1))$. ∎

COROLLARY 40. *For all constants* $0 < \epsilon < \delta < 1/2$ *there exists a partial boolean matrix* F *such that* $\mathrm{rank}^+_\delta(F) \leq n^{O(1)}$ *but* $\mathrm{rank}^+_\epsilon(F) \geq 2^{\Omega(n)}$.

Proof sketch. Theorem 6 together with Theorem 9 (proved in the next section) imply that for all $0 < \epsilon < \delta < 1/2$ there is a partial F such that $\mathsf{UWAPP}^{\mathsf{cc}}_\delta(F) \leq O(\log n)$ and $\mathsf{UWAPP}^{\mathsf{cc}}_\epsilon(F) \geq \Omega(n)$. Unfortunately, there is a slight problem with applying Fact 29 to conclude a similar separation for rank^+_ϵ as this direct simulation loses a factor of 2 in the error parameter ϵ. This loss results from the following asymmetry between the measures $\mathsf{UWAPP}^{\mathsf{cc}}_\epsilon$ and rank^+_ϵ: the acceptance probabilities of 1-inputs are in $[(1-\epsilon)\alpha, \alpha]$ in the former, whereas 1-entries can be approximated with values in $[1-\epsilon, 1+\epsilon]$ in the latter. However, this annoyance is easily overcome by considering modified versions of $\mathsf{WAPP}^{\mathsf{cc}}_\epsilon$ and $\mathsf{UWAPP}^{\mathsf{cc}}_\epsilon$ where the acceptance probability on 1-inputs is allowed to lie in $[(1-\epsilon)\alpha, (1+\epsilon)\alpha]$. It can be verified that under such a definition Theorems 6 and 9 and Fact 29 continue to hold, and the "new" Fact 29 does not lose the factor 2 in the error. ∎

6. Unrestricted–restricted equivalences. In this section we prove our unrestricted–restricted equivalence results, Theorems 8 and 9, restated below. In sec-

$$\begin{array}{ccccc} & & \text{Fact 25} & & \\ & \mathsf{WAPP}^{\mathsf{cc}} & \equiv & \mathsf{srec}^1 & \\ \text{Theorem 9, all } F & ||| & & ||| & \text{[37], total } F \\ & \mathsf{UWAPP}^{\mathsf{cc}} & \equiv & \log \operatorname{rank}_\epsilon^+ & \\ & & \text{Fact 29} & & \end{array}$$

FIG. 4. *Summary of equivalences.*

tion 6.1 we prove a key "truncation lemma," and in section 6.2 we use the lemma to prove the equivalences.

As already alluded to in the introduction, Buhrman, Vereshchagin, and de Wolf [9] exhibited a function F with $\mathsf{UPostBPP}^{\mathsf{cc}}(F) \leq O(\log n)$ and $\mathsf{PP}^{\mathsf{cc}}(F) \geq \Omega(n^{1/3})$. This simultaneously gives an exponential separation between $\mathsf{PostBPP}^{\mathsf{cc}}$ and $\mathsf{UPostBPP}^{\mathsf{cc}}$ and between $\mathsf{PP}^{\mathsf{cc}}$ and $\mathsf{UPP}^{\mathsf{cc}}$. For our other models, we will show that the unrestricted and restricted versions are essentially equivalent. We state and prove this result only for $\mathsf{SBP}^{\mathsf{cc}}$ and $\mathsf{WAPP}^{\mathsf{cc}}$ as the result for $\mathsf{2WAPP}^{\mathsf{cc}}$ is very similar.

THEOREM 8. $\mathsf{SBP}^{\mathsf{cc}}(F) \leq O(\mathsf{USBP}^{\mathsf{cc}}(F) + \log n)$ *for all* F*.*

THEOREM 9. $\mathsf{WAPP}^{\mathsf{cc}}_\delta(F) \leq O(\mathsf{UWAPP}^{\mathsf{cc}}_\epsilon(F) + \log(n/(\delta - \epsilon)))$ *for all* F *and all* $0 < \epsilon < \delta < 1/2$*.*

Hence, roughly speaking, $\mathsf{SBP}^{\mathsf{cc}}$ and $\mathsf{USBP}^{\mathsf{cc}}$ are equivalent and $\mathsf{WAPP}^{\mathsf{cc}}$ and $\mathsf{UWAPP}^{\mathsf{cc}}$ are equivalent. Here "equivalence" is ignoring constant factors and additive logarithmic terms in the cost, but much more significantly it is ignoring constant factors in ϵ (for $\mathsf{WAPP}^{\mathsf{cc}}$), which is important as we know that ϵ cannot be efficiently amplified (Theorem 6).

Discussion of Theorem 8*.* The equivalence of $\mathsf{SBP}^{\mathsf{cc}}$ and $\mathsf{USBP}^{\mathsf{cc}}$ implies an alternative proof of the lower bound $\mathsf{USBP}^{\mathsf{cc}}(\textsc{Disj}) \geq \Omega(n)$ for set disjointness from [25] without using information complexity. Indeed, that paper showed that $\mathsf{SBP}^{\mathsf{cc}}(\textsc{Disj}) \geq \Omega(n)$ follows from Razborov's corruption lemma [49]. It was also noted in [25] that the greater-than function $\textsc{Gt}$ (defined by $\textsc{Gt}(x, y) \coloneqq 1$ iff $x > y$ as n-bit numbers) satisfies $\mathsf{USBP}^{\mathsf{cc}}(\textsc{Gt}) = \Theta(1)$ and $\mathsf{SBP}^{\mathsf{cc}}(\textsc{Gt}) = \Theta(\log n)$, and thus the $+\log n$ gap in Theorem 8 is tight. Our proof of Theorem 8 shows, in some concrete sense, that $\textsc{Gt}$ is the "only" advantage $\mathsf{USBP}^{\mathsf{cc}}$ has over $\mathsf{SBP}^{\mathsf{cc}}$. Theorem 8 is analogous to, but more complicated than, Proposition 35 since both say that w.l.o.g. α is not too small in the SBP models.

Discussion of Theorem 9*.* The equivalence of $\mathsf{WAPP}^{\mathsf{cc}}$ and $\mathsf{UWAPP}^{\mathsf{cc}}$ implies the equivalence of the smooth rectangle bound (see Fact 25) and approximate nonnegative rank (see Fact 29), which was already known for total functions [37]. Our Theorem 9 implies that the equivalence holds even for partial functions, which was crucially used in the proof of Corollary 7. The situation is summarized in Figure 4.

6.1. The truncation lemma. The following lemma is a key component in the proofs of Theorems 8 and 9.

DEFINITION 41. *For a nonnegative matrix* M*, we define its* truncation $\overline{M}$ *to be the same matrix but where each entry* > 1 *is replaced with* 1*.*

Lemma 42 (truncation lemma). *For every $2^n \times 2^n$ nonnegative rank-1 matrix M and every d there exists an $O(d+\log n)$-communication public-randomness protocol Π such that for every (x, y) we have* $\mathrm{acc}_\Pi(x, y) \in \overline{M}_{x,y} \pm 2^{-d}$.

We describe some intuition for the proof. We can write $M_{x,y} = u_x v_y$, where $u_x, v_y \geq 0$. First, note that if all entries of M are at most 1, then $\mathrm{acc}_\Pi(x, y) = M_{x,y}$ can be achieved in a zero-communication manner: scaling all u_x's by some factor and scaling all v_y's by the inverse factor, we may assume that all $u_x, v_y \leq 1$; then Alice can accept with probability u_x and Bob can independently accept with probability v_y. Truncation makes all the entries at most 1 but may destroy the rank-1 property. Also note that, in general, for the nontruncated entries there may be no "global scaling" for which the zero-communication approach works: there may be some entries with $u_x v_y < 1$ but $u_x > 1$, and other entries with $u_x v_y < 1$ but $v_y > 1$. Roughly speaking, we instead think in terms of "local scaling" that depends on (x, y).

As a starting point, consider a protocol where Alice sends u_x to Bob, who then declares acceptance with probability $\min(u_x v_y, 1)$. We cannot afford to communicate u_x exactly, so we settle for an approximation. We express u_x and v_y in "scientific notation" with an appropriate base and round the mantissa of u_x to have limited precision. The exponent of u_x, however, may be too expensive to communicate, but since u_x, v_y are multiplied, all that matters is the sum of their exponents. Determining the sum of the exponents exactly may be too expensive, but the crux of the argument is that we only need to consider a limited number of cases. If the sum of the exponents is small, then the matrix entry is very close to 0 and we can reject without knowing the exact sum. If the sum of the exponents is large, then the matrix entry is guaranteed to be truncated and we can accept. Provided the base is large enough, there are only a few "inbetween" cases. Determining which case holds can be reduced to a greater-than problem, which can be solved with error exponentially small in d using communication $O(d + \log n)$.

We now give the formal proof.

Proof of Lemma 42. Let $M_{x,y} = u_x v_y$, where $u_x, v_y \geq 0$, and define $\delta \coloneqq 2^{-d}/2$ and $B \coloneqq 1/\delta$.

Henceforth we fix an input (x, y). For convenience we let all notation be relative to (x, y), so we start by defining $u \coloneqq u_x$ and $v \coloneqq v_y$, and note that $\overline{M}_{x,y} = \min(uv, 1)$. Assuming $u > 0$, define $i \coloneqq \lceil \log_B u \rceil$ (so $u \in (B^{i-1}, B^i]$) and $a \coloneqq u/B^i$ (so $a \in (\delta, 1]$). Similarly, assuming $v > 0$, define $j \coloneqq \lceil \log_B v \rceil$ (so $v \in (B^{j-1}, B^j]$) and $b \coloneqq v/B^j$ (so $b \in (\delta, 1]$). Note that $uv = abB^{i+j} \in (B^{i+j-2}, B^{i+j}]$. The protocol Π is as follows. (Line 4 is underspecified but we will address that later.)

1: If $u = 0$ or $v = 0$ then reject
2: Alice sends Bob $\widetilde{a} \in a \pm \delta^2$ (and ensuring $\widetilde{a} \leq 1$) using $O(d)$ bits
3: Bob computes $p \coloneqq \widetilde{a} \cdot b$
4: Determine with probability at least $1-\delta$ which of the following four cases holds:
5: If $i + j < 0$ then reject
6: If $i + j = 0$ then accept with probability p
7: If $i + j = 1$ then accept with probability $\min(pB, 1)$
8: If $i + j > 1$ then accept

We first argue correctness. Assume $u, v > 0$. We have $ab \in (\widetilde{a} \pm \delta^2)b \subseteq p \pm \delta^2$ (using $b \leq 1$) and thus $uv \in (p \pm \delta^2)B^{i+j}$. Pretending for the moment that line

4 succeeds with probability 1, we can verify that in all four cases the acceptance probability would be $\in \overline{M}_{x,y} \pm \delta$:

5: If $i+j<0$ then $0 \in \overline{M}_{x,y} \pm \delta$ since $uv \le B^{i+j} \le \delta$.
6: If $i+j=0$ then $p \in \overline{M}_{x,y} \pm \delta$ since $uv \in (p \pm \delta^2)B^{i+j} \subseteq p \pm \delta$.
7: If $i+j=1$ then $\min(pB,1) \in \overline{M}_{x,y} \pm \delta$ since $uv \in (p \pm \delta^2)B^{i+j} \subseteq pB \pm \delta$.
8: If $i+j>1$ then $1 = \overline{M}_{x,y}$ since $uv > B^{i+j-2} \ge 1$.

The error probability of line 4 only affects the overall acceptance probability by $\pm\delta$, so $\mathrm{acc}_\Pi(x,y) \in \overline{M}_{x,y} \pm 2\delta \subseteq \overline{M}_{x,y} \pm 2^{-d}$.

The communication cost is $O(d)$ except for line 4. Line 4 can be implemented with three tests: $i+j \ge 0$, $i+j \ge 1$, $i+j \ge 2$, each having error probability $\delta/3$. These tests are all implemented in the same way, so we just describe how to test whether $i+j \ge 0$. In other words, if we let T denote the indicator matrix for $i+j \ge 0$, then we want to compute T with error probability $\delta/3$ and communication $O(d+\log n)$. If we assume the rows are sorted in decreasing order of u and the columns are sorted in decreasing order of v, then each row and each column of T consists of 1's followed by 0's. To compute T, we may assume w.l.o.g. it has no duplicate rows and no duplicate columns, in which case it is a greater-than matrix (of size at most $2^n \times 2^n$) with the 1's in the upper-left triangle, possibly with the all-0 row deleted and/or the all-0 column deleted. The greater-than function can be computed with any error probability $\gamma > 0$ and communication $O(\log(n/\gamma))$ by running the standard protocol [38, p. 170] for $O(\log(n/\gamma))$ many steps. ∎

Remark 43. We note that the $O(d+\log n)$ communication bound in Lemma 42 is optimal, assuming $n \ge d$. Indeed, define a nonnegative rank-1 matrix M by $M_{x,y} := (2^{-d})^{x-y}$, where x and y are viewed as nonnegative n-bit integers. Consider any protocol Π with $\mathrm{acc}_\Pi(x,y) \in \overline{M}_{x,y} \pm 2^{-d}$, and note that it determines with error probability $2^{-(d-1)}$ whether $x \le y$. The latter is known to require $\Omega(\log n)$ communication (even for constant d) [65]. Also, by a union bound there exists an outcome of the randomness for which Π determines whether $x \le y$ for all pairs $x, y < 2^{d/2-1}$ (of which there are 2^{d-2}), which requires $\Omega(d)$ communication by the deterministic lower bound for greater than on $(d/2-1)$-bit integers.

6.2. Proofs of unrestricted–restricted equivalences. We now give the (very similar) proofs of Theorems 8 and 9 using the truncation lemma. We make use of the following basic fact.

FACT 44. *Given a private-randomness protocol* Π *of communication cost c, label the accepting transcripts as* $\tau \in \{1,2,\ldots,2^c\}$. *Then for each accepting transcript* τ *there exists a nonnegative rank-*1 *matrix* N^τ *such that the following holds. For each* (x,y)*, the probability of getting transcript* τ *on input* (x,y) *is* $N^\tau_{x,y}$*, and thus* $\mathrm{acc}_\Pi(x,y) = \sum_{\tau=1}^{2^c} N^\tau_{x,y}$.

For both proofs, the goal is to show that any protocol (of type $\mathsf{USBP}^{\mathsf{cc}}$ or $\mathsf{UWAPP}^{\mathsf{cc}}_\epsilon$) can be converted into another protocol (of type $\mathsf{SBP}^{\mathsf{cc}}$ or $\mathsf{WAPP}^{\mathsf{cc}}_\delta$, respectively) of comparable cost. We transform an α-correct protocol of cost c, where α might be prohibitively small, into a (roughly) 2^{-c}-correct protocol without increasing the communication by too much. We use Fact 44 to express the acceptance probabilities as a sum of nonnegative rank-1 matrices. The basic intuition is to divide everything by α to get a "1-correct" matrix sum; however, this new sum may not correspond to acceptance probabilities of a protocol. To achieve the latter, we truncate each summand (which does not hurt the correctness, and which makes each summand correspond to acceptance probabilities from the truncation lemma), then multiply each summand by 2^{-c} (which essentially changes the correctness parameter from 1 to 2^{-c}, and which corresponds to picking a uniformly random summand).

Proof of Theorem 8. Fix a cost-c $\mathsf{USBP}^{\mathsf{cc}}$ protocol Π for F with associated $\alpha > 0$ and associated matrices N^τ from Fact 44. Thus $\sum_\tau N^\tau_{x,y}$ is $\geq \alpha$ if $F(x,y) = 1$ and $\leq \alpha/2$ if $F(x,y) = 0$. We claim that the following public-randomness protocol Π' witnesses $\mathsf{SBP}^{\mathsf{cc}}(F) \leq O(c + \log n)$:

1: Pick $\tau \in \{1, 2, \ldots, 2^c\}$ uniformly at random
2: Run the protocol from Lemma 42 with $M^\tau \coloneqq \frac{1}{\alpha} N^\tau$ and $d \coloneqq c + 3$

We first argue correctness. We have

$$\mathrm{acc}_{\Pi'}(x,y) \in 2^{-c} \textstyle\sum_\tau \big(\overline{M}^\tau_{x,y} \pm 2^{-d}\big) = 2^{-c}\big(\sum_\tau \overline{M}^\tau_{x,y} \pm 2^{-3}\big).$$

If $F(x,y) = 0$ then $\sum_\tau \overline{M}^\tau_{x,y} \leq \sum_\tau \frac{1}{\alpha} N^\tau_{x,y} \leq 1/2$ and thus $\mathrm{acc}_{\Pi'}(x,y) \leq (5/8)2^{-c}$. Now suppose $F(x,y) = 1$. If $M^\tau_{x,y} \leq 1$ for all τ then $\sum_\tau \overline{M}^\tau_{x,y} = \sum_\tau \frac{1}{\alpha} N^\tau_{x,y} \geq 1$, and if not then we also have $\sum_\tau \overline{M}^\tau_{x,y} \geq \max_\tau \overline{M}^\tau_{x,y} = 1$. In either case, $\mathrm{acc}_{\Pi'}(x,y) \geq (7/8)2^{-c}$. Since there is a constant factor gap between the acceptance probabilities on 1-inputs and 0-inputs, we can use and-amplification in a standard way [25] to bring the gap to a factor of 2 while increasing the cost by only a constant factor. Since the communication cost of Π' is $O(d + \log n) = O(c + \log n)$, and the associated α' value is $2^{-O(c)}$, the overall cost is $O(c + \log n)$. ∎

Proof of Theorem 9. Fix a cost-c $\mathsf{UWAPP}^{\mathsf{cc}}_\epsilon$ protocol Π for F with associated $\alpha > 0$ and associated matrices N^τ from Fact 44. Thus $\sum_\tau N^\tau_{x,y}$ is $\in [(1-\epsilon)\alpha, \alpha]$ if $F(x,y) = 1$ and $\in [0, \epsilon\alpha]$ if $F(x,y) = 0$. We claim that the following public-randomness protocol Π' witnesses $\mathsf{WAPP}^{\mathsf{cc}}_\delta(F) \leq O(c + \log(n/\Delta))$, where $\Delta \coloneqq (\delta - \epsilon)/2$:

1: Pick $\tau \in \{1, 2, \ldots, 2^c\}$ uniformly at random
2: Run the protocol from Lemma 42 with $M^\tau \coloneqq \frac{1}{\alpha} N^\tau$ and $d \coloneqq c + \lceil \log(1/\Delta) \rceil$

We first argue correctness. We have

$$\mathrm{acc}_{\Pi'}(x,y) \in 2^{-c} \textstyle\sum_\tau \big(\overline{M}^\tau_{x,y} \pm 2^{-d}\big) \subseteq 2^{-c}\big(\sum_\tau \overline{M}^\tau_{x,y} \pm \Delta\big).$$

Define $\alpha' \coloneqq 2^{-c}(1 + \Delta)$. If $F(x,y) = 0$ then $\sum_\tau \overline{M}^\tau_{x,y} \leq \sum_\tau \frac{1}{\alpha} N^\tau_{x,y} \leq \epsilon$ and thus $\mathrm{acc}_{\Pi'}(x,y) \in [0, 2^{-c}(\epsilon + \Delta)] \subseteq [0, \delta\alpha']$. Now suppose $F(x,y) = 1$. Then $M^\tau_{x,y} \leq 1$ for all τ (otherwise $\mathrm{acc}_\Pi(x,y) = \sum_\tau \alpha M^\tau_{x,y} > \alpha$). Hence $\sum_\tau \overline{M}^\tau_{x,y} = \sum_\tau \frac{1}{\alpha} N^\tau_{x,y} \in [1 - \epsilon, 1]$, and thus $\mathrm{acc}_{\Pi'}(x,y) \in [2^{-c}(1 - \epsilon - \Delta), 2^{-c}(1 + \Delta)] \subseteq [(1-\delta)\alpha', \alpha']$. So Π' is a $\mathsf{WAPP}^{\mathsf{cc}}_\delta$ protocol for F of cost $O(d + \log n) + \log(1/\alpha') \leq O(c + \log(n/\Delta))$. ∎

Remark 45. In the proof of Theorem 9, note that if F is total then Lemma 42 is not needed: The entries of each M^τ are all bounded by 1, and thus $M^\tau_{x,y}$ can be written as $u_x v_y$, where $u_x, v_y \in [0,1]$. Hence to accept with probability $M^\tau_{x,y}$, Alice can accept with probability u_x and Bob can accept with probability v_y. This incurs no loss in the ϵ parameter and has communication cost 2, witnessing that $\mathsf{WAPP}^{\mathsf{cc}}_\epsilon(F) \leq \mathsf{UWAPP}^{\mathsf{cc}}_\epsilon(F) + 2$ if F is total.

Appendix A. Additional proofs.

A.1. Proof of Fact 25. $\mathsf{srec}^1_\epsilon(F)$ is defined as the log of the optimum value of the following linear program, which has a variable w_R for each rectangle R.

$$\begin{array}{rll} \text{minimize} & \sum_R w_R & \\ \text{subject to} & \sum_{R\,:\,(x,y)\in R} w_R \in [1-\epsilon, 1] & \forall (x,y) \in F^{-1}(1), \\ & \sum_{R\,:\,(x,y)\in R} w_R \in [0,\epsilon] & \forall (x,y) \in F^{-1}(0), \\ & w_R \geq 0 & \forall R. \end{array}$$

We first show the first inequality. Given a cost-c $\mathsf{WAPP}^{\mathsf{cc}}_\epsilon$ protocol for F, put it in the normal form given by Fact 24 so that $\alpha = 2^{-c}$ and each outcome of the randomness is a rectangle. For each rectangle R, let $w_R \coloneqq p_R/\alpha$, where p_R is the probability of R in the normal form protocol. This is a feasible solution with objective value $1/\alpha$, so $\mathsf{srec}^1_\epsilon(F) \leq \log(1/\alpha) = c$. We now show the second inequality. Given an optimal solution, let $\alpha \coloneqq 1/\sum_R w_R$ and consider a protocol that selects rectangle R with probability αw_R. This is an α-correct $\mathsf{WAPP}^{\mathsf{cc}}_\epsilon$ protocol for F of cost $2 + \mathsf{srec}^1_\epsilon(F)$.

A.2. Proof of Fact 29. We first show the first inequality. Fix a cost-c $\mathsf{UWAPP}^{\mathsf{cc}}_\epsilon$ protocol Π for F with associated $\alpha > 0$ and associated matrices N^τ from Fact 44. Thus $\sum_\tau N^\tau_{x,y}$ is $\in [(1-\epsilon)\alpha, \alpha]$ if $F(x,y) = 1$ and $\in [0, \epsilon\alpha]$ if $F(x,y) = 0$. Hence letting $M \coloneqq \sum_\tau \frac{1}{\alpha} N^\tau$, we have $M_{x,y} \in F(x,y) \pm \epsilon$ for all $(x,y) \in \operatorname{dom} F$ and $\operatorname{rank}^+(M) \leq 2^c$.

We now show the second inequality. Suppose M is such that $M_{x,y} \in F(x,y) \pm \epsilon/2$ for all $(x,y) \in \operatorname{dom} F$ and $r \coloneqq \operatorname{rank}^+(M)$ is witnessed by $M = UV$, and let t be the maximum entry in U, V. We claim that the following private-randomness protocol Π witnesses $\mathsf{UWAPP}^{\mathsf{cc}}_\epsilon(F) \leq \lceil \log r \rceil + 2$:

1: Alice picks $i \in \{1, 2, \ldots, r\}$ uniformly at random and sends it to Bob
2: Alice accepts with probability $U_{x,i}/t$ and sends her decision to Bob
3: Bob accepts with probability $V_{i,y}/t$ and sends his decision to Alice
4: Accept iff both Alice and Bob accept

We have $\operatorname{acc}_\Pi(x,y) = \frac{1}{r}\sum_i U_{x,i}V_{i,y}/t^2 = M_{x,y}/rt^2$. Let $\alpha \coloneqq (1+\epsilon/2)/rt^2$. If $F(x,y) = 1$ then $\operatorname{acc}_\Pi(x,y) \in [(1-\epsilon/2)/rt^2, (1+\epsilon/2)/rt^2] \subseteq [(1-\epsilon)\alpha, \alpha]$. If $F(x,y) = 0$ then $\operatorname{acc}_\Pi(x,y) \in [0, (\epsilon/2)/rt^2] \subseteq [0, \epsilon\alpha]$. Thus the protocol is correct with respect to α.

A.3. Proof of Fact 39. We use the notation $(t)_m$ for the falling factorial $t(t-1)\cdots(t-m+1)$. The acceptance probability is

$$\frac{\binom{n-d}{w-i}}{\binom{n}{w}} = \frac{(n-d)_{w-i}}{(w-i)!} \cdot \frac{w!}{(n)_w} = \frac{(w)_i}{(n)_w \,/\, (n-d)_{w-i}}.$$

We claim that
(i) $w^i \cdot (1 - o(1)) \leq (w)_i \leq w^i$,
(ii) $n^w \cdot (1 - o(1)) \leq (n)_w \leq n^w$,
(iii) $n^{w-i} \cdot (1 - o(1)) \leq (n-d)_{w-i} \leq n^{w-i}$.
Then the acceptance probability is in

$$\frac{w^i}{n^w \,/\, n^{w-i}} \cdot (1 \pm o(1)) = (w/n)^i \cdot (1 \pm o(1)).$$

The three upper bounds are trivial. For the lower bound in (i), we have

$$\begin{aligned}(w)_i &= w^i \cdot (1-\tfrac{0}{w})(1-\tfrac{1}{w})\cdots(1-\tfrac{i-1}{w})\\ &\geq w^i \cdot 4^{-0/w}4^{-1/w}\cdots 4^{-(i-1)/w}\\ &= w^i \cdot 4^{-i(i-1)/2w}\\ &\geq w^i \cdot (1-o(1))\end{aligned}$$

since $i \leq d \leq o(\sqrt{w})$. The lower bound in (ii) follows similarly using $w \leq o(\sqrt{n})$. For (iii), we have

$$(n-d)_{w-i} \geq (n-d)^{w-i} \cdot (1-o(1)) = n^{w-i} \cdot (1-o(1)) \cdot (1-d/n)^{w-i}$$

as above using $w-i \leq o(\sqrt{n-d})$, and we have $(1-d/n)^{w-i} \geq (4^{-d/n})^w \geq 1-o(1)$ since $d < w \leq o(\sqrt{n})$.

A.4. Proof of Fact 38. We first prove (i) $\Rightarrow$ (ii). Assume (i), and consider μ_1 distributed over $f^{-1}(1)$ and μ_0 distributed over $f^{-1}(0)$. We have for $\boldsymbol{h} \sim \mathcal{H}$ and $\boldsymbol{z_i} \sim \mu_i$ that

$$\begin{aligned}\mathbf{E}_{\boldsymbol{h}}\, \mathbf{Pr}_{\boldsymbol{z_1}}[\,\boldsymbol{h}(\boldsymbol{z_1})=1\,] &= \mathbf{Pr}_{\boldsymbol{h},\boldsymbol{z_1}}[\,\boldsymbol{h}(\boldsymbol{z_1})=1\,]\\ &\geq \min\nolimits_{z_1\in f^{-1}(1)} \mathbf{Pr}_{\boldsymbol{h}}[\,\boldsymbol{h}(z_1)=1\,]\\ &> 2\cdot \max\nolimits_{z_0\in f^{-1}(0)} \mathbf{Pr}_{\boldsymbol{h}}[\,\boldsymbol{h}(z_0)=1\,]\\ &\geq 2\cdot \mathbf{Pr}_{\boldsymbol{h},\boldsymbol{z_0}}[\,\boldsymbol{h}(\boldsymbol{z_0})=1\,]\\ &= \mathbf{E}_{\boldsymbol{h}}\ 2\cdot \mathbf{Pr}_{\boldsymbol{z_0}}[\,\boldsymbol{h}(\boldsymbol{z_0})=1\,].\end{aligned}$$

If $\mathbf{Pr}_{\boldsymbol{z_1}}[\,h(\boldsymbol{z_1})=1\,] \leq 2\cdot \mathbf{Pr}_{\boldsymbol{z_0}}[\,h(\boldsymbol{z_0})=1\,]$ for all h, then the above would be false.

We now prove (ii) $\Rightarrow$ (i). Assume (ii), and define α_{μ_1,μ_0} to be the maximum of $\mathbf{Pr}_{\boldsymbol{z_1}\sim\mu_1}[\,h(\boldsymbol{z_1})=1\,]$ over all h such that $\mathbf{Pr}_{\boldsymbol{z_1}\sim\mu_1}[\,h(\boldsymbol{z_1})=1\,] > 2\cdot \mathbf{Pr}_{\boldsymbol{z_0}\sim\mu_0}[\,h(\boldsymbol{z_0})=1\,]$. It is not difficult to see that the function $(\mu_1,\mu_0) \mapsto \alpha_{\mu_1,\mu_0}$ is lower semicontinuous, since if we change (μ_1,μ_0) infinitesimally then $\mathbf{Pr}_{\boldsymbol{z_1}\sim\mu_1}[\,h(\boldsymbol{z_1})=1\,] > 2\cdot \mathbf{Pr}_{\boldsymbol{z_0}\sim\mu_0}[\,h(\boldsymbol{z_0})=1\,]$ still holds for the (previously) optimum h, and the left side of the inequality only changes infinitesimally (but another h may become “available” and raise the value of α_{μ_1,μ_0}, hence the function is not upper semicontinuous). It is a basic fact of analysis that a lower semicontinuous function on a compact set attains its infimum. Since the set of (μ_1,μ_0) pairs is compact, and since $\alpha_{\mu_1,\mu_0} > 0$ for all (μ_1,μ_0), we have $\inf_{\mu_1,\mu_0}\alpha_{\mu_1,\mu_0} > 0$. Let α^* be any real such that $0 < \alpha^* < \inf_{\mu_1,\mu_0}\alpha_{\mu_1,\mu_0}$. Hence we have $\alpha_{\mu_1,\mu_0} > \alpha^*$ for all (μ_1,μ_0).

Let M be the matrix with rows indexed by Z and columns indexed by $\mathscr{H}$, such that $M_{z,h} \coloneqq h(z)$. Then for every (μ_1,μ_0) there exists an h such that $\mathbf{E}_{\boldsymbol{z_1}\sim\mu_1} M_{\boldsymbol{z_1},h} > \alpha^*$ and $\mathbf{E}_{\boldsymbol{z_1}\sim\mu_1} M_{\boldsymbol{z_1},h} > 2\cdot \mathbf{E}_{\boldsymbol{z_0}\sim\mu_0} M_{\boldsymbol{z_0},h}$. Let M' be the matrix with rows indexed by Z and (infinitely many) columns indexed by $\mathscr{H}\times[0,1]$, such that $M'_{z,(h,s)} \coloneqq s\cdot h(z)$. Then for every (μ_1,μ_0) there exists an (h,s) such that $\mathbf{E}_{\boldsymbol{z_1}\sim\mu_1} M'_{\boldsymbol{z_1},(h,s)} > \alpha^*$ and $\mathbf{E}_{\boldsymbol{z_0}\sim\mu_0} M'_{\boldsymbol{z_0},(h,s)} < \alpha^*/2$ (by choosing s to be slightly greater than $\alpha^*/\,\mathbf{E}_{\boldsymbol{z_1}\sim\mu_1} M_{\boldsymbol{z_1},h}$). Let $A\colon \mathbb{R}\to\mathbb{R}$ be the affine transformation $A(x) \coloneqq (1-x)\cdot \frac{\alpha^*}{1-\alpha^*/2}$. Let M'' be the matrix indexed like M', such that $M''_{z,(h,s)} \coloneqq M'_{z,(h,s)}$ if $f(z)=1$, and $M''_{z,(h,s)} \coloneqq A\big(M'_{z,(h,s)}\big)$ if $f(z)=0$. Then for every (μ_1,μ_0) there exists an (h,s) such that $\mathbf{E}_{\boldsymbol{z_1}\sim\mu_1} M''_{\boldsymbol{z_1},(h,s)} > \alpha^*$ and, by linearity of expectation, $\mathbf{E}_{\boldsymbol{z_0}\sim\mu_0} M''_{\boldsymbol{z_0},(h,s)} = A\big(\mathbf{E}_{\boldsymbol{z_0}\sim\mu_0} M'_{\boldsymbol{z_0},(h,s)}\big) > \big(1-\alpha^*/2\big)\cdot\frac{\alpha^*}{1-\alpha^*/2} = \alpha^*$.

We claim that for every distribution μ over Z there exists an (h,s) such that $\mathbf{E}_{\boldsymbol{z}\sim\mu} M''_{\boldsymbol{z},(h,s)} > \alpha^*$. If $\mu(f^{-1}(1)) > 0$ and $\mu(f^{-1}(0)) > 0$ then this follows from the above using $\mu_1 = (\mu \mid f^{-1}(1))$ and $\mu_0 = (\mu \mid f^{-1}(0))$. Otherwise if, say, $\mu(f^{-1}(0)) = 0$ (similarly if $\mu(f^{-1}(1)) = 0$) then we can let $\mu_1 = \mu$ and μ_0 be an arbitrary distribution over $f^{-1}(0)$, and apply the above.

Now by the minimax theorem (a continuous version as used in [61]) the two-player zero-sum game given by M'' (with payoffs to the column player) has value $> \alpha^*$, and thus there exists a distribution $\mathcal{H}'$ over $\mathscr{H} \times [0,1]$ such that for all $z \in Z$, $\mathbf{E}_{(\boldsymbol{h},\boldsymbol{s})\sim\mathcal{H}'} M''_{z,(\boldsymbol{h},\boldsymbol{s})} > \alpha^*$. Thus for all $z_1 \in f^{-1}(1)$ we have $\mathbf{E}_{(\boldsymbol{h},\boldsymbol{s})\sim\mathcal{H}'} M'_{z_1,(\boldsymbol{h},\boldsymbol{s})} > \alpha^*$, and for all $z_0 \in f^{-1}(0)$ by linearity of expectation we have $\mathbf{E}_{(\boldsymbol{h},\boldsymbol{s})\sim\mathcal{H}'} M'_{z_0,(\boldsymbol{h},\boldsymbol{s})} = A^{-1}(\mathbf{E}_{(\boldsymbol{h},\boldsymbol{s})\sim\mathcal{H}'} M''_{z_0,(\boldsymbol{h},\boldsymbol{s})}) < 1 - \alpha^* \cdot \frac{1-\alpha^*/2}{\alpha^*} = \alpha^*/2$.

For $h \in \mathscr{H}$, if we define p_h to be the expectation under $\mathcal{H}'$ of the function that outputs s on inputs (h,s) and outputs 0 otherwise, then for all z we have $\mathbf{E}_{(\boldsymbol{h},\boldsymbol{s})\sim\mathcal{H}'} M'_{z,(\boldsymbol{h},\boldsymbol{s})} = \sum_h p_h \cdot M_{z,h}$. Finally, we define the distribution $\mathcal{H}$ over $\mathscr{H}$ so the probability of h is p_h/P, where $P \coloneqq \sum_h p_h$. Then for all z we have $\mathbf{Pr}_{\boldsymbol{h}\sim\mathcal{H}}[\,\boldsymbol{h}(z) = 1\,] = \frac{1}{P} \cdot \mathbf{E}_{(\boldsymbol{h},\boldsymbol{s})\sim\mathcal{H}'} M'_{z,(\boldsymbol{h},\boldsymbol{s})}$. Thus for all $z_1 \in f^{-1}(1)$ we have $\mathbf{Pr}_{\boldsymbol{h}\sim\mathcal{H}}[\,\boldsymbol{h}(z_1) = 1\,] > \alpha^*/P$, and for all $z_0 \in f^{-1}(0)$ we have $\mathbf{Pr}_{\boldsymbol{h}\sim\mathcal{H}}[\,\boldsymbol{h}(z_0) = 1\,] < \alpha^*/2P$, and hence (i) holds.

Acknowledgments. We thank Troy Lee and Toniann Pitassi for discussions. We are also grateful to anonymous reviewers (of both STOC and SICOMP) for thoughtful comments.

REFERENCES

[1] S. Aaronson, *Quantum computing, postselection, and probabilistic polynomial-time*, R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci., 461 (2005), pp. 3473–3482, doi:10.1098/rspa.2005.1546.

[2] S. Aaronson and A. Wigderson, *Algebrization: A new barrier in complexity theory*, ACM Trans. Comput. Theory, 1 (2009), 2, doi:10.1145/1490270.1490272.

[3] A. Ada and A. Chattopadhyay, *Multiparty Communication Complexity of Disjointness*, Technical report TR08-002, Electronic Colloquium on Computational Complexity (ECCC), 2008, http://eccc.hpi-web.de/report/2008/002/.

[4] N. Alon, *Problems and results in extremal combinatorics*–I, Discrete Math., 273 (2003), pp. 31–53, doi:10.1016/S0012-365X(03)00227-9.

[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*, Technical report TR16-072, Electronic Colloquium on Computational Complexity (ECCC), 2016, http://eccc.hpi-web.de/report/2016/072/.

[6] L. Babai, P. Frankl, and J. Simon, *Complexity classes in communication complexity theory*, in Proceedings of the 27th Symposium on Foundations of Computer Science (FOCS), IEEE Computer Society, Los Alamitos, CA, 1986, pp. 337–347, doi:10.1109/SFCS.1986.15.

[7] 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, doi:10.1007/s00037-007-0220-2.

[8] E. Böhler, C. Glasser, and D. Meister, *Error-bounded probabilistic computations between MA and AM*, J. Comput. System Sci., 72 (2006), pp. 1043–1076, doi:10.1016/j.jcss.2006.05.001.

[9] H. Buhrman, N. Vereshchagin, and R. de Wolf, *On computation and communication with small bias*, in Proceedings of the 22nd Conference on Computational Complexity (CCC), IEEE Computer Society, Los Alamitos, CA, 2007, pp. 24–32, doi:10.1109/CCC.2007.18.

[10] A. Chakrabarti, R. Kondapally, and Z. Wang, *Information complexity versus corruption and applications to orthogonality and gap-hamming*, in Proceedings of the 16th International Workshop on Randomization and Computation (RANDOM), Lecture Notes in Comput. Sci. 7408, Springer, Basel, 2012, pp. 483–494, doi:10.1007/978-3-642-32512-0_41.

[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, doi:10.1137/120861072.

[12] S. O. Chan, J. Lee, P. Raghavendra, and D. Steurer, *Approximate constraint satisfaction requires large LP relaxations*, in Proceedings of the 54th Symposium on

Foundations of Computer Science (FOCS), IEEE, Piscataway, NJ, 2013, pp. 350–359, doi:10.1109/FOCS.2013.45.

[13] A. Chattopadhyay, *Circuits, Communication and Polynomials*, Ph.D. thesis, McGill University, Montreal, Canada, 2008.

[14] B. Chor and O. Goldreich, *Unbiased bits from sources of weak randomness and probabilistic communication complexity*, SIAM J. Comput., 17 (1988), pp. 230–261, doi:10.1137/0217015.

[15] S. Fenner, *PP-lowness and a simple definition of AWPP*, Theory Comput. Syst., 36 (2003), pp. 199–212, doi:10.1007/s00224-002-1089-8.

[16] D. Gavinsky and S. Lovett, *En route to the log-rank conjecture: New reductions and equivalent formulations*, in Proceedings of the 41st International Colloquium on Automata, Languages, and Programming (ICALP), Lecture Notes in Comput. Sci. 8572, Springer, Berlin, 2014, pp. 514–524, doi:10.1007/978-3-662-43948-7_43.

[17] D. Gavinsky and A. Sherstov, *A separation of NP and coNP in multiparty communication complexity*, Theory Comput., 6 (2010), pp. 227–245, doi:10.4086/toc.2010.v006a010.

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

[19] M. Göös and T. Jayram, *A composition theorem for conical juntas*, in Proceedings of the 31st Computational Complexity Conference (CCC), Schloss Dagstuhl, Dagstuhl, Germany, 2016, to appear, http://eccc.hpi-web.de/report/2015/167/.

[20] M. Göös, T. Jayram, T. Pitassi, and T. Watson, *Randomized Communication vs. Partition Number*, Technical report TR15-169, Electronic Colloquium on Computational Complexity (ECCC), 2015, http://eccc.hpi-web.de/report/2015/169/.

[21] M. Göös, S. Lovett, R. Meka, T. Watson, and D. Zuckerman, *Rectangles are nonnegative juntas*, in Proceedings of the 47th Symposium on Theory of Computing (STOC), ACM, New York, 2015, pp. 257–266, doi:10.1145/2746539.2746596.

[22] M. Göös, T. Pitassi, and T. Watson, *Deterministic communication vs. partition number*, in Proceedings of the 56th Symposium on Foundations of Computer Science (FOCS), IEEE, Piscataway, NJ, 2015, pp. 1077–1088, doi:10.1109/FOCS.2015.70.

[23] M. Göös, T. Pitassi, and T. Watson, *The landscape of communication complexity classes*, in Proceedings of the 43rd International Colloquium on Automata, Languages, and Programming (ICALP), Schloss Dagstuhl, Dagstuhl, Germany, 2016, to appear http://eccc.hpi-web.de/report/2015/049/.

[24] M. Göös, T. Pitassi, and T. Watson, *Zero-information protocols and unambiguity in Arthur–Merlin communication*, Algorithmica, (2016), doi:10.1007/s00453-015-0104-9.

[25] M. Göös and T. Watson, *Communication complexity of set-disjointness for all probabilities*, in Proceedings of the 18th International Workshop on Randomization and Computation (RANDOM), Schloss Dagstuhl, Dagstuhl, Germany, 2014, pp. 721–736, doi:10.4230/LIPIcs.APPROX-RANDOM.2014.721.

[26] T. Gur and R. Raz, *Arthur–Merlin streaming complexity*, Inform. Comput., 243 (2015), pp. 145–165, doi:10.1016/j.ic.2014.12.011.

[27] Y. Han, L. A. Hemaspaandra, and T. Thierauf, *Threshold computation and cryptographic security*, SIAM J. Comput., 26 (1997), pp. 59–78, doi:10.1137/S0097539792240467.

[28] P. Harsha and R. Jain, *A strong direct product theorem for the tribes function via the smooth-rectangle bound*, in Proceedings of the 33rd Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS), Schloss Dagstuhl, Dagstuhl, Germany, 2013, pp. 141–152, doi:10.4230/LIPIcs.FSTTCS.2013.141.

[29] R. Jain and H. Klauck, *The partition bound for classical communication complexity and query complexity*, in Proceedings of the 25th Conference on Computational Complexity (CCC), IEEE Computer Society, Los Alamitos, CA, 2010, pp. 247–258, doi:10.1109/CCC.2010.31.

[30] R. Jain and P. Yao, *A Strong Direct Product Theorem in terms of the Smooth Rectangle Bound*, Technical report, arXiv, 2012, preprint, arxiv.org/abs/1209.0263, 2012.

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

[32] 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, doi:10.1137/130928273.

[33] H. Klauck, *Rectangle size bounds and threshold covers in communication complexity*, in Proceedings of the 18th Conference on Computational Complexity (CCC), IEEE Computer Society, Los Alamitos, CA, 2003, pp. 118–134, doi:10.1109/CCC.2003.1214415.

[34] H. Klauck, *Lower bounds for quantum communication complexity*, SIAM J. Comput., 37 (2007), pp. 20–46, doi:10.1137/S0097539702405620.

[35] H. Klauck, *A strong direct product theorem for disjointness*, in Proceedings of the 42nd Symposium on Theory of Computing (STOC), ACM, New York, 2010, pp. 77–86, doi:10.1145/1806689.1806702.

[36] H. Klauck, *On Arthur Merlin games in communication complexity*, in Proceedings of the 26th Conference on Computational Complexity (CCC), IEEE, Piscataway, NJ, 2011, pp. 189–199, doi:10.1109/CCC.2011.33.

[37] G. Kol, S. Moran, A. Shpilka, and A. Yehudayoff, *Approximate nonnegative rank is equivalent to the smooth rectangle bound*, in Proceedings of the 41st International Colloquium on Automata, Languages, and Programming (ICALP), Lecture Notes in Comput. Sci. 8572, Springer, Berlin, 2014, pp. 701–712, doi:10.1007/978-3-662-43948-7_58.

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

[39] T. Lee and A. Shraibman, *An approximation algorithm for approximation rank*, in Proceedings of the 24th Conference on Computational Complexity (CCC), IEEE, Piscataway, NJ, 2009, pp. 351–357, doi:10.1109/CCC.2009.25.

[40] T. Lee and A. Shraibman, *Disjointness is hard in the multiparty number-on-the-forehead model*, Comput. Complexity, 18 (2009), pp. 309–336, doi:10.1007/s00037-009-0276-2.

[41] T. Lee, A. Shraibman, and R. Špalek, *A direct product theorem for discrepancy*, in Proceedings of the 23rd Conference on Computational Complexity (CCC), IEEE Computer Society, Los Alamitos, CA, 2008, pp. 71–80, doi:10.1109/CCC.2008.25.

[42] T. Lee and S. Zhang, *Composition theorems in communication complexity*, in Proceedings of the 37th International Colloquium on Automata, Languages, and Programming (ICALP), Lecture Notes in Comput. Sci. 6198, Springer, Berlin, 2010, pp. 475–489, doi:10.1007/978-3-642-14165-2_41.

[43] I. Newman, *Private vs. common random bits in communication complexity*, Inform. Process. Lett., 39 (1991), pp. 67–71, doi:10.1016/0020-0190(91)90157-D.

[44] R. O'Donnell, *Analysis of Boolean Functions*, Cambridge University Press, New York, 2014, http://www.analysisofbooleanfunctions.org.

[45] R. Paturi and J. Simon, *Probabilistic communication complexity*, J. Comput. System Sci., 33 (1986), pp. 106–123, doi:10.1016/0022-0000(86)90046-2.

[46] A. Rao and A. Yehudayoff, *Simplified lower bounds on the multiparty communication complexity of disjointness*, in Proceedings of the 30th Computational Complexity Conference (CCC), Schloss Dagstuhl, Dagstuhl, Germany, 2015, pp. 88–101, doi:10.4230/LIPIcs.CCC.2015.88.

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

[48] R. Raz and A. Shpilka, *On the power of quantum proofs*, in Proceedings of the 19th Conference on Computational Complexity (CCC), IEEE Computer Society, Los Alamitos, CA, 2004, pp. 260–274, doi:10.1109/CCC.2004.1313849.

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

[50] A. Razborov, *Quantum communication complexity of symmetric predicates*, Izv. Math., 67 (2003), pp. 145–159, doi:10.1070/IM2003v067n01ABEH000422.

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

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

[53] A. Sherstov, *Communication lower bounds using dual polynomials*, Bull. Eur. Assoc. Theor. Comput. Sci. EATCS, 95 (2008), pp. 59–93.

[54] A. A. Sherstov, *Separating AC^0 from depth-2 majority circuits*, SIAM J. Comput., 38 (2009), pp. 2113–2129, doi:10.1137/08071421X.

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

[56] A. Sherstov, *The unbounded-error communication complexity of symmetric functions*, Combinatorica, 31 (2011), pp. 583–614, doi:10.1007/s00493-011-2580-0.

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

[58] A. Sherstov, *The multiparty communication complexity of set disjointness*, in Proceedings of the 44th Symposium on Theory of Computing (STOC), ACM, New York, 2012, pp. 525–548, doi:10.1145/2213977.2214026.

[59] A. Sherstov, *Communication lower bounds using directional derivatives*, J. ACM, 61 (2014), 34, doi:10.1145/2629334.

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

[61] L. Trevisan, M. Tulsiani, and S. Vadhan, *Regularity, boosting, and efficiently simulating every high-entropy distribution*, in Proceedings of the 24th Conference on Computational Complexity (CCC), IEEE, Piscataway, NJ, 2009, pp. 126–136, doi:10.1109/CCC.2009.41.

[62] S. Vadhan, *Pseudorandomness*, Found. Trends Theoret. Comput. Sci. 7, 2012, NOW, Boston, doi:10.1561/0400000010.

[63] U. Vazirani, *Randomness, Adversaries and Computation*, Ph.D. thesis, University of California, Berkeley, CA,1986.

[64] 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, doi:10.4086/cjtcs.2012.001.

[65] E. Viola, *The communication complexity of addition*, Combinatorica, 35 (2015), pp. 703–747, doi:10.1007/s00493-014-3078-3.

[66] T. Watson, *Nonnegative rank vs. binary rank*, Chicago J. Theoret. Comput. Sci., 2016 (2016), pp. 1–13, doi:10.4086/cjtcs.2016.002.

[67] A. Yao, *Lower bounds by probabilistic arguments*, in Proceedings of the 24th Symposium on Foundations of Computer Science (FOCS), IEEE Computer Society, Los Alamitos, CA, 1983, pp. 420–428, doi:10.1109/SFCS.1983.30.