comput. complex.
© Springer International Publishing AG,
part of Springer Nature 2018
https://doi.org/10.1007/s00037-018-0166-6

computational complexity

THE LANDSCAPE OF COMMUNICATION COMPLEXITY CLASSES

Mika Göös, Toniann Pitassi,
and Thomas Watson

Abstract. We prove several results which, together with prior work, provide a nearly-complete picture of the relationships among classical communication complexity classes between P and PSPACE, short of proving lower bounds against classes for which no explicit lower bounds were already known. Our article also serves as an up-to-date survey on the state of structural communication complexity.

Among our new results we show that $\mathsf{MA} \not\subseteq \mathsf{ZPP}^{\mathsf{NP}[1]}$, that is, Merlin–Arthur proof systems cannot be simulated by zero-sided error randomized protocols with one NP query. Here the class $\mathsf{ZPP}^{\mathsf{NP}[1]}$ has the property that generalizing it in the slightest ways would make it contain $\mathsf{AM} \cap \mathsf{coAM}$, for which it is notoriously open to prove any explicit lower bounds. We also prove that $\mathsf{US} \not\subseteq \mathsf{ZPP}^{\mathsf{NP}[1]}$, where US is the class whose canonically complete problem is the variant of set-disjointness where yes-instances are uniquely intersecting. We also prove that $\mathsf{US} \not\subseteq \mathsf{coDP}$, where DP is the class of differences of two NP sets. Finally, we explore an intriguing open issue: Are rank-1 matrices inherently more powerful than rectangles in communication complexity? We prove a new separation concerning PP that sheds light on this issue and strengthens some previously known separations.

Keywords. landscape, communication, complexity, classes

Subject classification. 68Q15

Published online: 22 March 2018

1. Introduction

Complexity classes form the infrastructure of classical complexity theory. They are used to express the power of models of computation, characterize the complexities of important computational problems, and catalyze proofs of other results. A central project is to ascertain the full, intricate landscape of relationships among complexity classes.

Beginning with Babai *et al.* (1986), there has been a lot of research on the analogues of classical (Turing machine) complexity classes in two-party communication complexity. The analogue of P (the class of decision problems solvable in polynomial time) is the class of functions $F\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ for which Alice and Bob, given x and y, respectively, can evaluate $F(x,y)$ with a protocol that uses polylogarithmically many bits of communication. For other classical complexity classes representing other models of computation, one can generally define, in a canonical way, associated communication complexity classes representing associated models of communication. There are many motivations for studying the relationships (inclusions and non-inclusions) between these communication complexity classes.

- A holy grail of classical complexity is to prove separations of classes between P and PSPACE. Separations relative to oracles can often be viewed as class separations in the restricted setting of *query complexity*; see Vereshchagin (1999) for an excellent survey. Communication complexity can be viewed as a restricted (but generally less restricted than query complexity) setting for which lower bounds are more difficult to obtain. Such separations in restricted settings are sometimes construed as evidence for the classical separations, or at least as barriers to refuting the classical separations. A stronger form of relativization barriers is known as *algebrization* Aaronson & Wigderson (2009), which directly employs communication complexity class separations.
- Proving lower bounds against strong communication complexity classes has applications to other areas of theoretical computer science. One of the most notorious open problems in

communication complexity is to prove lower bounds against the analogue of the polynomial hierarchy (PH) for any explicit two-party function. Proving PH lower bounds is a necessary step for obtaining strong rank rigidity lower bounds Lokam (2001, 2009); Razborov (1989); Wunderlich (2012) (as well as margin complexity rigidity lower bounds Linial & Shraibman (2009)), which in turn are related to circuit complexity Valiant (1977). Lower bounds against PH are also related to graph complexity Jukna (2006); Pudlák *et al.* (1988). It even remains open to prove communication lower bounds against the subclass of PH known as AM (Arthur–Merlin games) for any explicit function (which would be relevant to streaming delegation Chakrabarti *et al.* (2014a,b, 2015); Gur & Raz (2015); Klauck & Prakash (2013, 2014)).

- Communication complexity has a menagerie of techniques for proving lower bounds (among the oldest being discrepancy and corruption). These techniques often provide lower bounds against powerful communication complexity classes, and in some cases turn out to be *equivalent* to the communication measures corresponding to those classes (e.g., discrepancy is equivalent to PP communication Klauck (2007), and corruption is equivalent to SBP communication Göös & Watson (2016)). See Göös *et al.* (2016a) for more background on this. Thus, by studying complexity classes, as a byproduct we study the relative strength of lower bound techniques.

- The various models of communication corresponding to complexity classes are mathematically interesting because protocols in these models can be viewed as succinct representations of boolean matrices. The study of classes exposes natural combinatorial questions about how succinctly matrices can be represented.

We contribute to the exploration of the communication complexity landscape by filling in many of the remaining gaps in the known relationships among (non-quantum) classes, and discovering new techniques and insights along the way. At a glance, the state of affairs (including our new results) is summarized in Figure 1.1, which shows a map of known inclusions and non-inclusions between pairs of communication classes. In Section 2 we state our results more precisely and provide some intuition for the proofs. In Section 3, we provide a comprehensive survey of all the nontrivial (non-)inclusions among the traditional classes depicted in Figure 1.1. This updates previous surveys by Babai, Frankl, and Simon Babai *et al.* (1986) and Halstenberg and Reischuk Halstenberg & Reischuk (1990).

We refer to Jukna (2012); Kushilevitz & Nisan (1997) for background on communication complexity. In Appendix B we provide a catalog of communication complexity class definitions; throughout the text, we provide definitions on a "need-to-know" basis. If $\mathcal{C}$ is the name of a model (e.g., P for deterministic or NP for nondeterministic), we follow the convention of using $\mathcal{C}$ to denote both a complexity class and the corresponding complexity measure: $\mathcal{C}(F)$ denotes the minimum cost of a correct protocol for the (possibly partial) two-party function F in model $\mathcal{C}$, and $\mathcal{C}$ denotes the class of all (families of) partial functions F with $\mathcal{C}(F) \leq \text{poly}(\log n)$. These classes can also be defined for total functions, in which case the relations between the classes are occasionally different than—or not known to be the same as—the partial function case. We do not consider classes of search problems.

2. Our contributions

Several of our results concern two-party composed functions, so we introduce some general notation for this. A *composed function* is of the form $f \circ g^m$ where $f\colon \{0,1\}^m \to \{0,1\}$ is a (possibly partial) *outer function* and $g\colon \{0,1\}^b \times \{0,1\}^b \to \{0,1\}$ is an *inner function* also called a *gadget*. We write $F \coloneqq f \circ g^m\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ where $n \coloneqq m \cdot b$. We view the inputs to Alice and Bob as $x, y \in (\{0,1\}^b)^m$, which are partitioned into *blocks* $x_i, y_i \in \{0,1\}^b$ for $i \in [m]$. The goal is to compute $F(x,y) \coloneqq f(g(x_1,y_1),\ldots,g(x_m,y_m))$.

2.1. $\mathsf{MA} \not\subseteq \mathsf{ZPP}^{\mathsf{NP}[1]}$. A Merlin–Arthur ($\mathsf{MA}$) communication protocol is a proof system in which a nondeterministic party called Merlin sends a proof string (depending on the input) to Alice and Bob (collectively constituting Arthur), who then execute a ran-

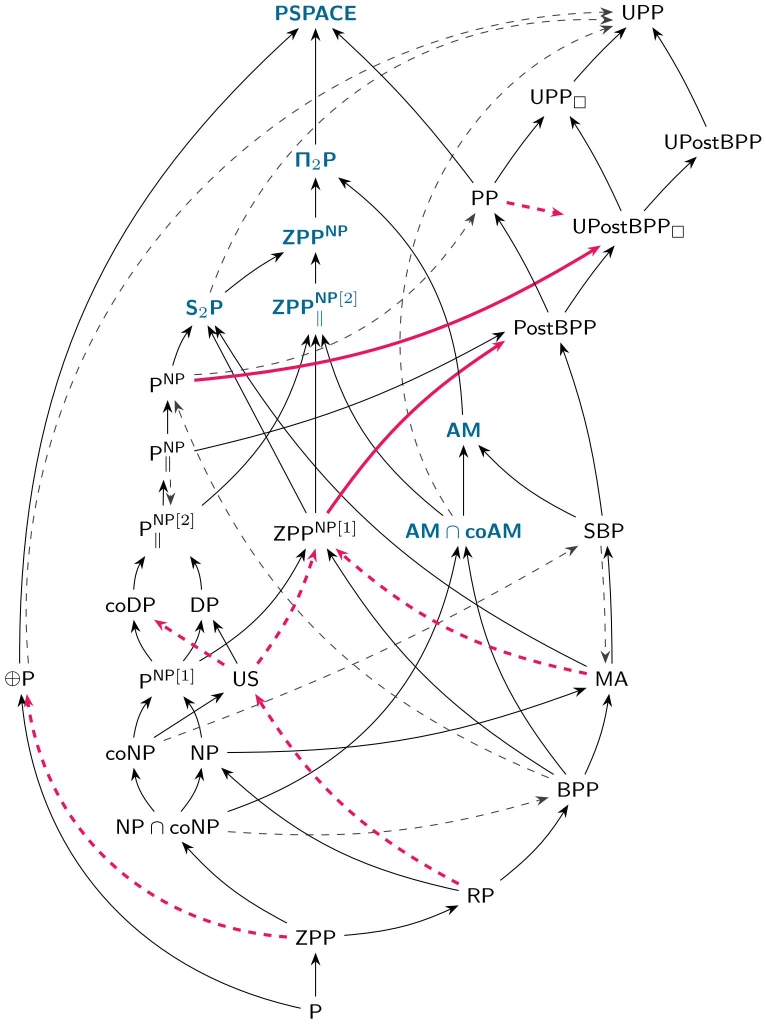

Figure 1.1: $\mathcal{C}_1 \rightarrow \mathcal{C}_2$ denotes $\mathcal{C}_1 \subseteq \mathcal{C}_2$, and $\mathcal{C}_1 \dashrightarrow \mathcal{C}_2$ denotes $\mathcal{C}_1 \not\subseteq \mathcal{C}_2$. **Bold red** arrows indicate new results. **Bold blue** indicates classes for which no explicit lower bounds are known (color figure online).

domized protocol to verify the proof. Merlin–Arthur communication protocols have been studied many times Aaronson & Wigderson (2009); Gavinsky & Sherstov (2010); Gur & Raz (2015); Gur & Rothblum (2015); Klauck (2003, 2011); Raz & Shpilka (2004), starting with the work of Klauck Klauck (2003), who gave a $\Omega(\sqrt{n})$ lower bound on the MA communication complexity of set-disjointness. In contrast, for the related (and stronger) model of Arthur–Merlin (AM) communication protocols, in which Merlin's proof string may depend on Alice's and Bob's randomness, no nontrivial lower bound is known for any explicit function, and such lower bounds have become very sought-after in the recent literature Chakrabarti *et al.* (2015); Klauck & Prakash (2014); Linial & Shraibman (2009); Papakonstantinou *et al.* (2014).

Our first result concerns the relationship between MA and another class, $\mathsf{ZPP}^{\mathsf{NP}[1]}$, which is a slightly obscure but intriguing character with many curious properties. A ZPP-type protocol is randomized and may output the correct answer or $\perp$ (representing "don't know"), and must output the correct answer with high probability on every input; granting the protocol access to one query to an NP oracle yields $\mathsf{ZPP}^{\mathsf{NP}[1]}$. It is not a priori clear that the model is robust with respect to the choice of threshold for the success probability, since standard amplification by repetition would increase the number of NP oracle queries. However, it was shown in Chang & Purini (2008) that $\mathsf{ZPP}^{\mathsf{NP}[1]}$ does indeed admit efficient amplification as long as the success probability is $> 1/2$ (the proof for time-bounded complexity also works for communication complexity); hence we define the model with success probability some constant $> 1/2$, say $3/4$.

If we allowed $\mathsf{ZPP}^{\mathsf{NP}[1]}$ to have success probability $< 1/2$, the class would change drastically: it would contain $\mathsf{AM} \cap \mathsf{coAM}$ (see Section 3), and hence proving explicit lower bounds for the communication version would yield breakthrough AM communication lower bounds. Granting the model access to two nonadaptive NP queries (and requiring success probability $> 1/2$) would also encompass $\mathsf{AM} \cap \mathsf{coAM}$. Thus, in a sense, $\mathsf{ZPP}^{\mathsf{NP}[1]}$ represents a boundary beyond which AM lower bounds would be the next step. The class $\mathsf{ZPP}^{\mathsf{NP}[1]}$ is also sandwiched between BPP and $\mathsf{S}_2\mathsf{P}$ Cai &

Chakaravarthy (2006); $\mathsf{S_2P}$ is a subclass of the polynomial hierarchy that has not been studied before in communication complexity (the definition appears in Appendix B), and no nontrivial lower bounds against it are known for any explicit function. This is another sense in which $\mathsf{ZPP}^{\mathsf{NP}[1]}$ constitutes a new frontier toward the elusive goal of proving explicit PH communication lower bounds. We also mention that $\mathsf{ZPP}^{\mathsf{NP}[1]}$ shows up frequently in the literature on the "two queries problem" (e.g., if $\mathsf{P}_{\|}^{\mathsf{NP}[2]} \subseteq \mathsf{ZPP}^{\mathsf{NP}[1]}$ then $\mathsf{PH} = \mathsf{S_2P}$ Tripathi (2010)).

We prove that $\mathsf{MA} \not\subseteq \mathsf{ZPP}^{\mathsf{NP}[1]}$ in the setting of communication complexity. This can be interpreted as saying that one-round non-interactive[1] proof systems cannot be made to have zero-sided error, even if the proof is generalized to an NP oracle query that depends on the randomness.

Before officially stating the theorem, we give the relevant formal definitions. An MA communication protocol computing $F\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ consists of a randomized two-party protocol which takes as input, in addition to the usual inputs x and y, a proof string (witness) $w \in \{0,1\}^k$ that is visible to both Alice and Bob. The completeness criterion is that for every $(x,y) \in F^{-1}(1)$ there exists a w such that the protocol accepts with probability at least $3/4$, and the soundness criterion is that for every $(x,y) \in F^{-1}(0)$ and every w, the protocol rejects with probability at least $3/4$. The cost is the witness length k plus the length of the subsequent transcript between Alice and Bob.

A $\mathsf{ZPP}^{\mathsf{NP}[1]}$ protocol Π computing F is a distribution over $\mathsf{P}^{\mathsf{NP}[1]}$-type protocols, each of which is of the following form: There is a deterministic protocol where for each leaf v having associated rectangle R_v, there is also an associated collection of "witness rectangles" $\{S_{v,w} \subseteq R_v \;:\; w \in \{0,1\}^k\}$ and an associated "output function" $o_v\colon \{0,1\} \to \{0,1,\perp\}$. The output of the $\mathsf{P}^{\mathsf{NP}[1]}$-type protocol on input (x,y) is obtained by running the deterministic part to reach a leaf v, then applying o_v to the indicator of whether $(x,y) \in \bigcup_w S_{v,w}$. The correctness criterion is that for every $(x,y) \in F^{-1}$ (where we use F^{-1} to denote the domain of F),

[1] Here, the term non-interactive means that Alice and Bob cannot interact with Merlin other than receiving the proof string.

$\mathbb{P}\big[\Pi(x,y) \in \{F(x,y), \perp\}\big] = 1$ and $\mathbb{P}\big[\Pi(x,y) = F(x,y)\big] \geq 3/4$ (where $\mathbb{P}$ denotes probability). The cost is the witness length k plus the maximum communication cost of the deterministic part of any of the constituent $\mathsf{P}^{\mathsf{NP}[1]}$-type protocols. The result of Chang & Purini (2008) shows that changing the success probability from 3/4 to any other constant strictly between 1/2 and 1 would only change the measure $\mathsf{ZPP}^{\mathsf{NP}[1]}(F)$ by a constant factor.

We prove a lower bound for the block-equality[2] function $\textsc{Block-Eq} \coloneqq \textsc{Or} \circ \textsc{Eq}^m$ where the input to $\textsc{Or}$ is $m \coloneqq \sqrt{n}$ bits, and each input to $\textsc{Eq}$ is $b \coloneqq \sqrt{n}$ bits. In other words, writing $x \coloneqq x_1 \cdots x_{\sqrt{n}} \in (\{0,1\}^{\sqrt{n}})^{\sqrt{n}}$ and $y \coloneqq y_1 \cdots y_{\sqrt{n}} \in (\{0,1\}^{\sqrt{n}})^{\sqrt{n}}$, we have $\textsc{Block-Eq}(x,y) = 1$ iff $x_i = y_i$ for some i. Note that $\textsc{Block-Eq} \in \mathsf{MA}$ since i can be nondeterministically guessed by Merlin, and then $x_i = y_i$ can be verified using a randomized protocol for $\textsc{Eq}$. (It was first noticed in Lam & Ruzzo (1992) that $\textsc{Block-Eq} \in \mathsf{\Sigma_2 P} \cap \mathsf{\Pi_2 P}$, which is a superset of MA.)

Theorem 2.1. $\mathsf{ZPP}^{\mathsf{NP}[1]}(\textsc{Block-Eq}) = \Theta(\sqrt{n})$, *and hence* $\mathsf{MA} \not\subseteq \mathsf{ZPP}^{\mathsf{NP}[1]}$.

To prove Theorem 2.1 (Section 4), we apply a new lower bound technique that combines the corruption bound with the 1-monochromatic rectangle size bound and asserts that they hold *simultaneously* (under the same distribution over inputs). We prove that, perhaps surprisingly, this combined technique gives a lower bound for $\mathsf{ZPP}^{\mathsf{NP}[1]}$ (though neither of the individual bounds suffices).

To apply our technique to $\textsc{Block-Eq}$, we first note that it is straightforward to achieve the two bounds separately: the 1-monochromatic rectangle size bound follows by simple counting, and the corruption bound follows by using Razborov's corruption lemma for the set-intersection function $\textsc{Inter}$ Razborov (1992) together with a simple reduction from $\textsc{Inter}$ to $\textsc{Block-Eq}$. However, the latter does *not* result in a distribution satisfying the 1-monochromatic rectangle size bound for $\textsc{Block-Eq}$. To fix this problem, we argue that if we average Razborov's distribution over all ways of implementing the reduction (of which there are

[2] The complement of block-equality is often known as list-non-equality.

many), then the corruption bound is still satisfied, and now the 1-monochromatic rectangle size bound is also satisfied.

2.2. $\mathsf{US} \not\subseteq \mathsf{ZPP}^{\mathsf{NP}[1]}$. For the set-intersection function $\textsc{Inter}$, Alice and Bob are each given a subset of $[n]$ (and we identify the subset with its characteristic vector, a length-n bit string), and the goal is to output 1 when the sets are intersecting and 0 when they are disjoint.[3] Phrased as a composed function, $\textsc{Inter} \coloneqq \textsc{Or} \circ \textsc{And}^n$ (for single-bit $\textsc{And}$). This is the canonical NP-complete problem in communication complexity, holding a comparable status to satisfiability, the canonical NP-complete problem in time-bounded complexity.

In the literature, “unique-set-intersection” commonly refers to the partial function version of $\textsc{Inter}$ where the intersection is promised to have size 0 or 1. We propose a change in terminology, in order to be consistent with the following corresponding terminology from time-bounded complexity (see, e.g., Blass & Gurevich (1982); Chang *et al.* (1995); Valiant & Vazirani (1986)): Unique-satisfiability is the problem of determining whether the number of satisfying assignments of a formula is exactly 1, and is complete for the complexity class called US. Unambiguous-satisfiability is the problem of determining whether the number of satisfying assignments of a formula is 0 or 1 under the promise that one of these cases holds, and is complete for the complexity class called UP.

Therefore, we make the following definitions: Unique-set-intersection is the total function $\textsc{Unique-Inter}\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ that maps (x,y) to 1 iff $|x \cap y| = 1$, i.e., $\textsc{Unique-Inter} \coloneqq \textsc{Unique-Or} \circ \textsc{And}^n$ where $\textsc{Unique-Or}(z) = 1$ iff the Hamming weight of z is 1. Unambiguous-set-intersection is the partial function $\textsc{Unambig-Inter}\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ that maps (x,y) to $|x \cap y|$ if the latter is in $\{0,1\}$, i.e., $\textsc{Unambig-Inter} \coloneqq \textsc{Unambig-Or} \circ \textsc{And}^n$ where $\textsc{Unambig-Or}(z)$ equals the Hamming weight of z if the latter is in $\{0,1\}$.

Note that $\textsc{Unique-Inter}$ is US-complete, where a cost-k US communication protocol is defined as a collection of rectangles

[3] We let “set-disjointness” refer to the complementary function where 1-inputs are disjoint.

$\{R_w \subseteq \{0,1\}^n \times \{0,1\}^n \ : \ w \in \{0,1\}^k\}$, where on input (x, y) the output of the protocol is 1 iff (x, y) is in R_w for exactly one w.

THEOREM 2.2. *$\mathsf{ZPP}^{\mathsf{NP}[1]}$(UNIQUE-INTER) $= \Theta(n)$, and hence $\mathsf{US} \not\subseteq \mathsf{ZPP}^{\mathsf{NP}[1]}$.*

We give two proofs of Theorem 2.2. Both proofs show that Theorem 2.2 holds even under the promise that the input sets intersect in at most two coordinates. Also, in both proofs, handling $\mathsf{ZPP}^{\mathsf{NP}[1]}$ instead of $\mathsf{P}^{\mathsf{NP}[1]}$ incurs almost no extra complication.

The first proof (Section 4) employs the same lower bound technique as in Theorem 2.1, but where we use Razborov's corruption lemma Razborov (1992) directly (and we must do a little analysis to verify the 1-monochromatic rectangle size bound). The optional second proof (relegated to Appendix A) uses information complexity tools (including an adaptation of the "partial information cost" approach from Jayram *et al.* (2003)) and, although longer to write, has some minor advantages over the first proof: It is more self-contained, as it does not rely on the corruption lemma (only on some basic facts that are standard in information complexity). Also, it directly handles success probability $1/2 + \epsilon$ (for any constant $\epsilon > 0$) without relying on the amplification result of Chang & Purini (2008) (whereas the first proof assumes success probability 0.999).

2.3. $\mathsf{US} \not\subseteq$ coDP. The class DP was introduced in Papadimitriou & Yannakakis (1984) to capture the complexity of certain exact versions of optimization problems. A set (of all 1-inputs of a function) is in DP iff it is the difference between two NP sets. The classes P, NP, and DP are the 0^{th}, 1^{st}, and 2^{nd} (respectively) levels of the so-called boolean hierarchy.

We have $\mathsf{US} \subseteq \mathsf{DP}$ since to check that there is exactly one witness, we can use an NP computation to check that there is at least one witness, and another to check that there are at least two witnesses, and require that the first computation returns 1 and the second returns 0. However, it is unlikely that $\mathsf{US} \subseteq \mathsf{coDP}$: Chang *et al.* (1995) showed that this inclusion cannot hold in the classical time-bounded setting unless the polynomial hierarchy collapses. This result does not yield an unconditional communication separation, since it is unknown whether the polynomial hierarchy collapses. Nevertheless, we show that indeed $\mathsf{US} \not\subseteq \mathsf{coDP}$ in communication complexity.

Formally, a cost-k coDP communication protocol is defined as a pair of collections of rectangles, $\{S_w \subseteq \{0,1\}^n \times \{0,1\}^n : w \in \{0,1\}^k\}$ and $\{T_w \subseteq \{0,1\}^n \times \{0,1\}^n : w \in \{0,1\}^k\}$, where on input (x,y) the output is 0 iff $(x,y) \in \bigcup_w S_w \smallsetminus \bigcup_w T_w$.

Theorem 2.3. $\mathsf{coDP}(\textsc{Unique-Inter}) = \Theta(n)$, *and hence* $\mathsf{US} \not\subseteq \mathsf{coDP}$.

To prove Theorem 2.3 (Section 4), we show that the same lower bound technique we introduced for $\mathsf{ZPP}^{\mathsf{NP}[1]}$ (the combination of the corruption bound and the 1-monochromatic rectangle size bound) *also* lower bounds coDP complexity. Thus we can simply reuse the application of the technique to Unique-Inter from Theorem 2.2. (Reusing the application to Block-Eq from Theorem 2.1 would show that Block-Eq $\notin \mathsf{coDP}$, but this already follows from the facts that Block-Eq $\notin \mathsf{P}^{\mathsf{NP}}$ Impagliazzo & Williams (2010) and $\mathsf{coDP} \subseteq \mathsf{P}^{\mathsf{NP}}$.)

2.4. $\mathsf{ZPP}^{\mathsf{NP}[1]} \subseteq \mathsf{PostBPP}$.

Consider bounded-error randomized computations (like in BPP) but with *postselection*: the output may come from $\{0,1,\perp\}$ and must be correct with high probability *conditioned* on not outputting $\perp$ (and the probability of this conditioning event must be positive). The complexity class corresponding to this model was originally called $\mathsf{BPP}_{\mathsf{path}}$ Han *et al.* (1997), but the name $\mathsf{PostBPP}$ (inspired by Aaronson (2005)) has gained popularity in the recent literature (Göös *et al.* (2016a) is one example) and seems more appropriate, so we use it instead.

According to modern conventions, the standard way to define the cost of a $\mathsf{PostBPP}$ communication protocol for F would be as the communication cost plus $\log(1/\alpha)$, where α is the minimum over all $(x,y) \in F^{-1}$ of the probability of not outputting $\perp$. (Allowing public randomness and not charging for α would enable $\mathsf{PostBPP}$ protocols to compute every function with constant cost.) Similarly, the cost of a PP (i.e., unbounded-error randomized) protocol would be the communication cost plus $\log(1/\epsilon)$ where $1/2+\epsilon$ is the minimum over all $(x,y) \in F^{-1}$ of the probability of outputting the correct answer.

However, for reasons that will become clear in Section 2.5, we choose to revert to the original convention of Babai *et al.* (1986) and define $\mathsf{PostBPP}$ and PP in a slightly different but equivalent way: we do not charge for α or ϵ but we require the public randomness to be uniformly distributed over $\{0,1\}^k$ and we charge for k. For both $\mathsf{PostBPP}$ and PP, this cost measure is equivalent to the above "modern" definition within a constant factor and additive $O(\log n)$ term, by standard sparsification of the public randomness Newman (1991).

Formally, we define a $\mathsf{PostBPP}$ communication protocol Π for F in the following succinct way: For each outcome of the public randomness (which is uniformly distributed over $\{0,1\}^k$) there is a deterministic protocol outputting values in $\{0,1,\bot\}$. For each $(x,y) \in F^{-1}$ we must have $\mathbb{P}\big[\Pi(x,y) = F(x,y)\big] > 2 \cdot \mathbb{P}\big[\Pi(x,y) = 1 - F(x,y)\big]$. The cost is the randomness length k plus the maximum communication cost of any of the constituent deterministic protocols.

A priori it is not clear that any explicit lower bounds for $\mathsf{ZPP}^{\mathsf{NP}[1]}$ follow from prior work. The following result shows that in fact they do, since many explicit lower bounds for $\mathsf{PostBPP}$ were known (see Section 3).

THEOREM 2.4. *$\mathsf{PostBPP}(F) \leq O\big(\mathsf{ZPP}^{\mathsf{NP}[1]}(F) + \log n\big)$ for all F, and hence $\mathsf{ZPP}^{\mathsf{NP}[1]} \subseteq \mathsf{PostBPP}$.*

It turns out that Theorem 2.4 can be derived from the lower bound technique we develop for $\mathsf{ZPP}^{\mathsf{NP}[1]}$ in Section 4; however, that approach is more complicated than necessary and, more importantly, is specific to communication complexity. We give a proof of Theorem 2.4 (Section 5) using a black-box simulation that also works for time-bounded complexity, without exploiting any special properties of communication.

Intuitively, the worst case for simulating a $\mathsf{ZPP}^{\mathsf{NP}[1]}$ protocol is the following situation: Whenever the NP oracle responds "0" the protocol outputs the right answer, and whenever the NP oracle responds "1" the protocol outputs $\perp$ but would have output the wrong answer if the response were "0." In this situation, pretending the oracle always responds "0" would yield a BPP protocol (this is where we crucially need the success probability to be $> 1/2$). To handle more general situations, we must also randomly guess and verify a witness for the NP query, outputting $\perp$ if the witness is invalid.

2.5. Open issue: Rank-1 vs. rectangles The classes $\mathsf{PostBPP}$ and PP can be further generalized by allowing the use of private randomness, which does not count toward the cost. This gives rise to the so-called unrestricted probabilities classes $\mathsf{UPostBPP}$ (which was defined, but not extensively studied, in Göös *et al.* (2016a)) and UPP (which is well-studied Forster (2002); Paturi & Simon (1986); Razborov & Sherstov (2010); Sherstov (2011b)). In $\mathsf{UPostBPP}$ and UPP we can dispense with public randomness altogether as the public coins could be tossed privately by Alice and then sent to Bob.

Combinatorially, $\mathsf{PostBPP}$ and PP protocols of cost c induce a distribution over 2^c labeled rectangles (rank-1 matrices with 0-1 entries) each occurring with a "restricted" probability of at least 2^{-c} (Observation B.21 and Observation B.26). In the case of $\mathsf{UPostBPP}$ and UPP there is a similar characterization with rectangles replaced by nonnegative rank-1 matrices (Observation B.22 and Observation B.27). A natural question arises:

> **Informal question:** *Are rank-*1 *matrices inherently more powerful than rectangles in communication complexity?*

While it has been shown that, e.g., $\mathsf{PP} \neq \mathsf{UPP}$ Buhrman *et al.* (2007); Sherstov (2008), the known examples of functions $F \in \mathsf{UPP} \setminus \mathsf{PP}$ can actually be computed without exploiting the full power of private randomness (their rank-1 property): we can use a UPP protocol whose associated rank-1 matrices are still rectangles, but occurring with *unrestricted*, possibly tiny, probability. We conclude that "PP vs. UPP" is not the right way to formalize our informal question (and the existing proofs for $\mathsf{PP} \neq \mathsf{UPP}$ do not incidentally answer our question).

A better formalization is as follows. We define new communication classes, $\mathsf{UPostBPP}_{\Box} \subseteq \mathsf{UPostBPP}$ and $\mathsf{UPP}_{\Box} \subseteq \mathsf{UPP}$, in the same way as $\mathsf{PostBPP}$ and PP, except allowing the public randomness to be arbitrarily distributed over $\{0,1\}^k$ (still charging for k and not for α or ϵ). Combinatorially, we have a distribution over 2^k labeled rectangles, but with no restrictions on their probabilities. (The "rectangle" subscript $\Box$ refers to the fact that these classes can be characterized similarly to $\mathsf{UPostBPP}$ and UPP but with rectangles playing the role of rank-1 matrices; see Appendix B.3.) Our informal question can now be formalized as follows:

> **Formal question:** *Do we have* $\mathsf{UPostBPP} = \mathsf{UPostBPP}_{\Box}$*?* *How about* $\mathsf{UPP} = \mathsf{UPP}_{\Box}$*?*

The seemingly minor syntactic generalization introduced in the definitions of the $\Box$-classes makes a huge difference: We observe (Section 7) that $\mathsf{P}^{\mathsf{NP}} \subseteq \mathsf{UPostBPP}_{\Box}$,[4] whereas it is known that $\mathsf{PostBPP}$ and P^{NP} are incomparable (see Section 3). Hence $\mathsf{UPostBPP}_{\Box}$ is a strict superset of both $\mathsf{PostBPP}$ and P^{NP}. This leaves us with no known examples of functions to witness a separation for our "rank-1 vs. rectangle" question; currently the best gap is $\mathsf{UPostBPP}(F) \leq O(1)$ vs. $\mathsf{UPostBPP}_{\Box}(F) \geq \Omega(\log n)$ where F is the usual Greater-Than function defined by $F(x,y) = 1$ iff $x > y$ when $x, y \in [2^n]$ are viewed as numbers. There is also no clear analogue of the "rank-1 vs. rectangle" distinction in query complexity, so a separation of the two notions in communication complexity might require interesting techniques. In fact, in the context of SBP (subclass of $\mathsf{PostBPP}$), it can be shown that rank-1 matrices do not add any power over mere rectangles Göös *et al.* (2016a).

2.6. $\mathsf{PP} \not\subseteq \mathsf{UPostBPP}_{\Box}$.

Our final result is to develop and apply a useful lower bound method for the class $\mathsf{UPostBPP}_{\Box}$ introduced above. $\mathsf{PostBPP}$ already has a tight rectangle-based lower bound technique, which was dubbed "extended discrepancy" in Gavinsky

[4] This inclusion also holds for time-bounded complexity. In defining the time-bounded version of $\mathsf{UPostBPP}_{\Box}$, we would allow the distribution of the random string to depend nonuniformly on the input length n, though for the inclusion of P^{NP}, the distribution is computable in exponential time given the string 1^n.

& Lovett (2014) but was used earlier in Klauck (2003) to show that $\mathsf{PP} \not\subseteq \mathsf{PostBPP}$. We strengthen the latter result to show that $\mathsf{PP} \not\subseteq \mathsf{UPostBPP}_\square$. (Showing $\mathsf{PP} \not\subseteq \mathsf{UPostBPP}$ remains open.) In our proof, we make use of the main theorem from Göös *et al.* (2016a), which applies to composed functions where the gadget is as follows.

DEFINITION 2.5. *The* confounding *gadget g is defined by $g(x_i, y_i) := \langle x_i, y_i\rangle \bmod 2$, where $x_i, y_i \in \{0,1\}^b$ and the block length b is $b(m) := 100 \log m$.*

We introduce the confounded-majority function, defined as $\textsc{Conf-Maj} := f \circ g^m$ where f is the majority function and g is the confounding gadget. Note that $\textsc{Conf-Maj}$ has input length $n := m \cdot b = m \cdot 100 \log m$ and is in PP since Alice and Bob can pick $i \in [m]$ uniformly at random and then exchange $b + 1 \leq O(\log n)$ bits to evaluate $g(x_i, y_i)$.

THEOREM 2.6. *$\mathsf{UPostBPP}_\square(\textsc{Conf-Maj}) = \Theta(n)$, and hence $\mathsf{PP} \not\subseteq \mathsf{UPostBPP}_\square$.*

To prove Theorem 2.6 (Section 6) we introduce a lower bound technique for $\mathsf{UPostBPP}_\square$ that strengthens the extended discrepancy bound (for $\mathsf{PostBPP}$) by requiring it to hold under a product distribution over inputs (analogously to how Papakonstantinou *et al.* (2014) showed that the "monochromatic rectangle size bound under product distributions" gives a lower bound for P^{NP}). However, only a $\Omega(\sqrt{n \log n})$ lower bound for $\textsc{Conf-Maj}$ follows using this technique, so to get the $\Omega(n)$ lower bound in Theorem 2.6, we generalize the technique further by allowing a rectangle's *size* to be measured with respect to some product distribution while its *error* is measured with respect to some other (arbitrary) distribution. (This is very analogous to the idea of relative discrepancy Fontes *et al.* (2016); Ganor *et al.* (2016).) To apply our general lower bound technique to $\textsc{Conf-Maj}$, we employ the communication-to-query machinery from Göös *et al.* (2016a) in a new, somewhat indirect way.

Finally, we mention another intriguing property of $\mathsf{UPostBPP}_\square$: By our lower bound technique and the results of Gavinsky & Lovett

(2014) it follows immediately that to prove the Log Rank Conjecture, i.e., that $\mathsf{P}(F) \leq \operatorname{poly}(\log \operatorname{rank}(F))$ for all total boolean matrices F, it suffices to prove the same with $\mathsf{UPostBPP}_{\square}$ instead of P. See Section 6 for more details.

3. Cartography

In this section we explore in detail the known (non-)inclusions shown on the map in Figure 1.1. We have not drawn any redundant arrows in the map: other relationships can be inferred from those shown; e.g., if $\mathcal{C}_1 \not\subseteq \mathcal{C}_2$ and $\mathcal{C}_1 \subseteq \mathcal{C}_3$ and $\mathcal{C}_4 \subseteq \mathcal{C}_2$, then $\mathcal{C}_3 \not\subseteq \mathcal{C}_4$.

Inclusions. The following inclusions also hold for time-bounded (Turing machine) complexity, as they do not exploit any special properties of communication. Also recall that all our classes consist of partial functions (promise problems); in particular, none of these inclusions exploit special properties of total functions.

$\mathsf{BPP} \subseteq \mathsf{ZPP}^{\mathsf{NP}[1]}$: This was first shown implicitly in Goldreich & Zuckerman (2011); Nisan & Wigderson (1994). A particularly clean and elegant argument was given in Cai & Chakaravarthy (2006).

$\mathsf{ZPP}^{\mathsf{NP}[1]} \subseteq \mathsf{PostBPP}$: This is our Theorem 2.4.

$\mathsf{P}_{\parallel}^{\mathsf{NP}} \subseteq \mathsf{PostBPP}$: This was shown in Han *et al.* (1997). In fact, this was strengthened to $\mathsf{P}_{\parallel}^{\mathsf{SBP}} = \mathsf{PostBPP}$ in O'Donnell & Say (2016).

$\mathsf{P}^{\mathsf{NP}} \subseteq \mathsf{UPostBPP}_{\square}$: We sketch the proof of this in Section 7 (Observation 7.1).

$\mathsf{SBP} \subseteq \mathsf{AM}$: This was shown in Goldwasser & Sipser (1986) (despite the fact that the class SBP was not defined and named until later Böhler *et al.* (2006)).

$\mathsf{AM} \cap \mathsf{coAM} \subseteq \mathsf{ZPP}_{\parallel}^{\mathsf{NP}[2]}$: This follows from the well-known facts that $\mathsf{AM} = \mathsf{coR} \cdot \mathsf{NP}$ and that $\mathsf{ZPP} =$

$\mathsf{RP} \cap \mathsf{coRP}$ relativizes. (We do not know whether $\mathsf{ZPP}_{\parallel}^{\mathsf{NP}[2]}$ admits efficient amplification, but for concreteness we define it with success probability 3/4.) The same argument shows that $\mathsf{AM} \cap \mathsf{coAM}$ would be in $\mathsf{ZPP}^{\mathsf{NP}[1]}$ if we allowed the latter to have success probability a constant less than 1/2.

$\mathsf{P}^{\mathsf{NP}} \subseteq \mathsf{S_2P}$: This was shown in Canetti (1996); Russell & Sundaram (1998).

$\mathsf{MA} \subseteq \mathsf{S_2P}$: This was shown in Russell & Sundaram (1998). (It was shown in Canetti (1996) that $\mathsf{BPP} \subseteq \mathsf{S_2P}$.)

$\mathsf{ZPP}^{\mathsf{NP}[1]} \subseteq \mathsf{S_2P}$: This was shown in Cai & Chakaravarthy (2006).

$\mathsf{S_2P} \subseteq \mathsf{ZPP}^{\mathsf{NP}}$: This was shown in Cai (2007). See also Fortnow *et al.* (2008).

Non-inclusions. For a non-inclusion $\mathcal{C}_1 \not\subseteq \mathcal{C}_2$, the result is strengthened if we show that some total function is in $\mathcal{C}_1$ but not in $\mathcal{C}_2$. All the following non-inclusions are known to hold for a total function, except in cases where we say otherwise.

$\mathsf{MA} \not\subseteq \mathsf{ZPP}^{\mathsf{NP}[1]}$: This is our Theorem 2.1.

$\mathsf{US} \not\subseteq \mathsf{ZPP}^{\mathsf{NP}[1]}$: This is our Theorem 2.2.

$\mathsf{US} \not\subseteq \mathsf{coDP}$: This is our Theorem 2.3.

$\mathsf{RP} \not\subseteq \mathsf{US}$: This is fairly simple to show, but was not recorded in the literature before, so we take the opportunity to do so in Section 7 (Observation 7.2).

$\mathsf{NP} \cap \mathsf{coNP} \not\subseteq \mathsf{BPP}$: This was shown in Klauck (2003). Of course, it is known that $\mathsf{NP} \cap \mathsf{coNP} = \mathsf{P}$ for total functions, so the function witnessing this is necessarily partial.

$\mathsf{coNP} \not\subseteq \mathsf{SBP}$: This was shown in Göös & Watson (2016) using the corruption lemma of Razborov (1992).

$\mathsf{P}_{\|}^{\mathsf{NP}[q+1]} \not\subseteq \mathsf{P}_{\|}^{\mathsf{NP}[q]}$: This holds for all constants $q \geq 0$ Halstenberg & Reischuk (1990). It is also known that $\mathsf{P}^{\mathsf{NP}[q]} = \mathsf{P}_{\|}^{\mathsf{NP}[2^q-1]}$ for all constants $q \geq 0$ Beigel (1991), and hence $\mathsf{P}^{\mathsf{NP}[q+1]} \not\subseteq \mathsf{P}^{\mathsf{NP}[q]}$.

$\mathsf{BPP} \not\subseteq \mathsf{P}^{\mathsf{NP}}$: This was implicitly shown in Papakonstantinou *et al.* (2014), though only for a partial function (the variant of gap-Hamming-distance with a constant relative gap). Progress toward witnessing $\mathsf{BPP} \not\subseteq \mathsf{P}^{\mathsf{NP}}$ by a total function can be made in two directions: finding a total function not in P^{NP} that is in a small superclass of BPP, and finding a total function in BPP that is not in a large subclass of P^{NP}. For the former, $\mathsf{MA} \not\subseteq \mathsf{P}^{\mathsf{NP}}$ is witnessed by Block-Eq Impagliazzo & Williams (2010). For the latter, $\mathsf{BPP} \not\subseteq \mathsf{P}_{\|}^{\mathsf{NP}}$ is witnessed by Greater-Than Halstenberg & Reischuk (1990).[5]

$\mathsf{SBP} \not\subseteq \mathsf{MA}$: This was shown in Göös *et al.* (2016a), though only for a partial function.

$\mathsf{P}^{\mathsf{NP}} \not\subseteq \mathsf{PP}$: This was shown in Buhrman *et al.* (2007); there it was only stated that $\mathsf{UPP} \not\subseteq \mathsf{PP}$, but the function witnessing this is, in fact, in P^{NP}.

$\mathsf{PP} \not\subseteq \mathsf{UPostBPP}_{\square}$: This is our Theorem 2.6. Previously, $\mathsf{PP} \not\subseteq \mathsf{PostBPP}$ was shown in Klauck (2003), and $\mathsf{PP} \not\subseteq \mathsf{P}^{\mathsf{NP}}$ was known since the negation would imply $\oplus\mathsf{P} \subseteq \mathsf{P}^{\mathsf{NP}}$ by binary search, and a fairly simple proof that $\oplus\mathsf{P} \not\subseteq \mathsf{P}^{\mathsf{NP}}$ was given in Papakonstantinou *et al.* (2014).

[5] It was only claimed in Halstenberg & Reischuk (1990) that Greater-Than $\notin \mathsf{P}_{\|}^{\mathsf{NP}[q]}$ for any constant q, but in fact their proof shows that Greater-Than $\notin \mathsf{P}_{\|}^{\mathsf{NP}}$.

$\mathsf{ZPP} \not\subseteq \oplus\mathsf{P}$: This is fairly simple to show, but was not recorded in the literature before, so we take the opportunity to do so in Section 7 (Observation 7.3). Of course, it is known that $\mathsf{ZPP} = \mathsf{P}$ for total functions, so the function witnessing this is necessarily partial. The non-equality total function NEQ witnesses $\mathsf{RP} \not\subseteq \oplus\mathsf{P}$.

$\oplus\mathsf{P} \not\subseteq \mathsf{UPP}$: This was shown in Forster (2002).

$\mathsf{AM} \cap \mathsf{coAM} \not\subseteq \mathsf{UPP}$: This was shown in Bouland *et al.* (2017), in fact for a certain subclass $\mathsf{NISZK} \subseteq \mathsf{AM} \cap \mathsf{coAM}$ (non-interactive statistical zero-knowledge), though only for a partial function. Previously, $\mathsf{AM} \cap \mathsf{coAM} \not\subseteq \mathsf{PP}$ was shown for a partial function in Klauck (2011) (by combining the results of Sherstov (2011a); Vereshchagin (1995)). Also, $\Pi_2\mathsf{P} \not\subseteq \mathsf{UPP}$ was shown for a total function in Razborov & Sherstov (2010).

$\mathsf{S}_2\mathsf{P} \not\subseteq \mathsf{UPP}$: This was shown in Chattopadhyay & Mande (2017), in fact for the subclass $\mathsf{P}^{\mathsf{MA}} \subseteq \mathsf{S}_2\mathsf{P}$.

Open issues. In summary, everything is now known about the relations between pairs of classes in Figure 1.1, except for the following conjectured non-inclusions:

- $\mathsf{PP} \not\subseteq \mathsf{UPostBPP}$ (or even $\mathsf{UPP} \not\subseteq \mathsf{UPostBPP}$),
- $\mathsf{UPostBPP} \not\subseteq \mathsf{UPP}_\square$ (or even $\mathsf{UPP} \not\subseteq \mathsf{UPP}_\square$ or $\mathsf{UPostBPP} \not\subseteq \mathsf{UPostBPP}_\square$),

and except for conjectured non-inclusions that would entail explicit AM lower bounds[6] or explicit $\mathsf{S}_2\mathsf{P}$ lower bounds:

- $\mathsf{coNP} \not\subseteq \mathsf{AM}$ (or even $\mathsf{PSPACE} \not\subseteq \mathsf{AM} \cap \mathsf{coAM}$ or $\mathsf{UPP} \not\subseteq \mathsf{AM} \cap \mathsf{coAM}$),

[6] Note that if we had an $\mathsf{AM} \cap \mathsf{coAM}$ lower bound for an explicit function F, then we would also have an AM lower bound for the explicit function that maps $((b, x), y) \mapsto F(x, y) \oplus b$ where $b \in \{0, 1\}$.

- $\mathsf{SBP} \not\subseteq \Sigma_2\mathsf{P}$ (or even $\mathsf{PSPACE} \not\subseteq \mathsf{S}_2\mathsf{P}$ or $\mathsf{UPP} \not\subseteq \mathsf{S}_2\mathsf{P}$),
- $\oplus\mathsf{P} \not\subseteq \Pi_2\mathsf{P}$,
- $\mathsf{MA} \not\subseteq \mathsf{ZPP}_{\|}^{\mathsf{NP}[2]}$,
- $\mathsf{P}_{\|}^{\mathsf{NP}} \not\subseteq \mathsf{ZPP}_{\|}^{\mathsf{NP}[2]}$,
- $\mathsf{AM} \cap \mathsf{coAM} \not\subseteq \mathsf{S}_2\mathsf{P}$,
- $\mathsf{UPostBPP}_{\square} \not\subseteq \mathsf{PSPACE}$,

and except for showing the following non-inclusions for total functions:

- $\mathsf{BPP} \not\subseteq \mathsf{P}^{\mathsf{NP}}$ (or even $\mathsf{ZPP}_{\|}^{\mathsf{NP}[2]} \not\subseteq \mathsf{P}^{\mathsf{NP}}$),
- $\mathsf{SBP} \not\subseteq \mathsf{MA}$ (or even $\mathsf{AM} \not\subseteq \mathsf{MA}$),
- $\mathsf{AM} \cap \mathsf{coAM} \not\subseteq \mathsf{UPP}$ (or even $\mathsf{ZPP}_{\|}^{\mathsf{NP}[2]} \not\subseteq \mathsf{PostBPP}$).

4. Lower bounds for block-equality and unique-set-intersection

We now describe a technique for lower bounding both $\mathsf{ZPP}^{\mathsf{NP}[1]}$ and coDP communication.

Lemma 4.1. *Suppose μ_0 is a distribution over $F^{-1}(0)$, μ_1 is a distribution over $F^{-1}(1)$, and C is a constant such that for every rectangle $R \subseteq \{0,1\}^n \times \{0,1\}^n$, $\mu_0(R) \leq C \cdot \mu_1(R) + \delta$, and if R is 1-monochromatic (i.e., contains no 0-inputs) then $\mu_1(R) \leq \delta$. Then*

(i) $\mathsf{ZPP}^{\mathsf{NP}[1]}(F) \geq \Omega(\log(1/\delta))$,

(ii) $\mathsf{coDP}(F) \geq \Omega(\log(1/\delta))$.

The first half of the technique ($\mu_0(R) \leq C \cdot \mu_1(R) + \delta$) is the corruption bound (which is a tight lower bound technique for so-called coSBP Göös & Watson (2016)), and the other half is the 1-monochromatic rectangle size bound (which is a tight lower bound technique for NP (Kushilevitz & Nisan 1997, §2.4)). The combined technique gives a lower bound for both $\mathsf{ZPP}^{\mathsf{NP}[1]}$ and coDP, even though neither of these classes appears to be a "combination" of coSBP and NP.

We prove parts (i) and (ii) of Lemma 4.1 in Section 4.1 and Section 4.2. Then we apply the technique to Block-Eq in Section 4.3 (thus proving Theorem 2.1), and finally we apply the technique to Unique-Inter in Section 4.4 (thus proving Theorem 2.2 and Theorem 2.3).

4.1. Proof of Lemma 4.1.(i) Suppose for contradiction there is a cost-$o(\log(1/\delta))$ $\mathsf{ZPP}^{\mathsf{NP}[1]}$ protocol Π computing F. Then in particular we have $\delta \leq o(1)$. By the amplification result of Chang & Purini (2008), we may assume $\mathbb{P}\big[\Pi(x, y) = \bot\big] \leq 1/10C$ for all $(x, y) \in F^{-1}$. By Markov's inequality and a union bound, we may fix a $\mathsf{P}^{\mathsf{NP}[1]}$-type protocol Π^* in the support of Π such that $\mathbb{P}_{(x,y)\sim\mu_0}\big[\Pi^*(x, y) = \bot\big] \leq 1/5C$ and $\mathbb{P}_{(x,y)\sim\mu_1}\big[\Pi^*(x, y) = \bot\big] \leq 1/5C$. Let the notation $k, R_v, S_{v,w}, o_v$ be with respect to Π^* (see the definition of $\mathsf{ZPP}^{\mathsf{NP}[1]}$ in Section 2.1), and note that without loss of generality, each o_v is non-constant (otherwise we could redefine $S_{v,w} = \emptyset$ for all w and redefine $o_v(1)$ arbitrarily).

For $b \in \{0, 1, \bot\}$, define $W_b := \bigcup_{v,w\,:\,o_v(1)=b} S_{v,w}$ as the set of "witnessed" inputs (the NP oracle responds "1") on which Π^* outputs b, and define $N_b := \bigcup_{v\,:\,o_v(0)=b} \big(R_v \smallsetminus \bigcup_w S_{v,w}\big)$ as the set of "non-witnessed" inputs (the NP oracle responds "0") on which Π^* outputs b. Note that $\{W_0, N_0, W_1, N_1, W_\bot, N_\bot\}$ partitions $\{0, 1\}^n \times \{0, 1\}^n$. By assumption, $\mu_0(W_\bot \cup N_\bot) \leq 1/5C$ and $\mu_1(W_\bot \cup N_\bot) \leq 1/5C$. By the correctness of Π, for $b \in \{0, 1\}$ we have $(W_b \cup N_b) \cap F^{-1}(1 - b) = \emptyset$.

Claim 4.2. $\mu_0(W_0) \leq 1/4$.

Claim 4.3. $\mu_0(N_0) \leq 1/4$.

This provides the contradiction since then $\mu_0\big(\{0, 1\}^n \times \{0, 1\}^n\big) = \mu_0(W_0) + \mu_0(N_0) + \mu_0(W_1 \cup N_1) + \mu_0(W_\bot \cup N_\bot) \leq 1/4 + 1/4 + 0 + 1/5C < 1$.

Proof (Proof of Claim 4.2). For each v, w such that $o_v(1) = 0$, we have $\mu_1(S_{v,w}) = 0$ and hence $\mu_0(S_{v,w}) \leq \delta$. Thus by a union bound, $\mu_0(W_0) \leq \sum_{v,w\,:\,o_v(1)=0} \mu_0(S_{v,w}) \leq 2^{o(\log(1/\delta))} \cdot \delta \leq \delta^{1-o(1)} \leq 1/4$. □

Proof (Proof of Claim 4.3). If v is such that $o_v(0) = 0$, then we have

$$\mu_0\big(R_v \smallsetminus \bigcup_w S_{v,w}\big) \;\le\; \mu_0(R_v) \;\le\; C\cdot\mu_1(R_v)+\delta \;=\; C\cdot\mu_1\big(\bigcup_w S_{v,w}\big)+\delta$$

by the fact that $\big(R_v \smallsetminus \bigcup_w S_{v,w}\big) \cap F^{-1}(1) = \emptyset$. Also, since each o_v is non-constant, we have

$$\begin{aligned}\sum_{v\,:\,o_v(0)=0} \mu_1\big(\bigcup_w S_{v,w}\big) &= \sum_{v\,:\,o_v(0)=0,\,o_v(1)=\perp} \mu_1\big(\bigcup_w S_{v,w}\big)\\ &\qquad + \sum_{v\,:\,o_v(0)=0,\,o_v(1)=1} \mu_1\big(\bigcup_w S_{v,w}\big)\\ &\le\; \mu_1(W_\perp) + \sum_{v,w\,:\,o_v(1)=1} \mu_1(S_{v,w})\\ &\le\; \mu_1(W_\perp \cup N_\perp) + 2^{o(\log(1/\delta))}\cdot\delta\\ &\le\; 1/5C + \delta^{1-o(1)}\end{aligned}$$

where the third line follows since $S_{v,w}$ is 1-monochromatic if $o_v(1) = 1$. Combining these, we have

$$\begin{aligned}\mu_0(N_0) &= \sum_{v\,:\,o_v(0)=0} \mu_0\big(R_v \smallsetminus \bigcup_w S_{v,w}\big)\\ &\le\; \sum_{v\,:\,o_v(0)=0} \Big(C\cdot\mu_1\big(\bigcup_w S_{v,w}\big)+\delta\Big)\\ &\le\; C\cdot\Big(\sum_{v\,:\,o_v(0)=0} \mu_1\big(\bigcup_w S_{v,w}\big)\Big) + 2^{o(\log(1/\delta))}\cdot\delta\\ &\le\; C\cdot\big(1/5C+\delta^{1-o(1)}\big)+\delta^{1-o(1)}\\ &\le\; 1/4.\end{aligned}$$

□

4.2. Proof of Lemma 4.1.(ii) Suppose for contradiction there is a cost-k coDP protocol Π computing F where $k \le o(\log(1/\delta))$. Then in particular we have $\delta \le o(1)$. We have a pair of collections of rectangles, $\{S_w : w \in \{0,1\}^k\}$ and $\{T_w : w \in \{0,1\}^k\}$, such that if $F(x,y) = 0$ then $(x,y) \in \bigcup_w S_w$ and $(x,y) \notin \bigcup_w T_w$, and if $F(x,y) = 1$ then $(x,y) \notin \bigcup_w S_w$ or $(x,y) \in \bigcup_w T_w$. Since $\mu_0\big(\bigcup_w S_w\big) = 1$, there exists a w^* such that $\mu_0(S_{w^*}) \ge 2^{-k} \ge \delta^{1/3}$ and hence $\mu_1(S_{w^*}) \ge \frac{1}{C}\cdot(\delta^{1/3}-\delta) \ge \delta^{1/2}$. Since $S_{w^*} \cap F^{-1}(1) \subseteq \bigcup_w T_w$, there exists a w' such that $\mu_1(T_{w'}) \ge \mu_1\big(S_{w^*} \cap F^{-1}(1)\big)\cdot 2^{-k} > \delta^{1/2}\cdot\delta^{1/2} = \delta$. But $T_{w'}$ is 1-monochromatic since $F^{-1}(0) \cap \bigcup_w T_w = \emptyset$, so this is a contradiction.

4.3. Proof of Theorem 2.1. Let μ_0 be the uniform distribution over $\textsc{Block-Eq}^{-1}(0)$, and let μ_1 be the uniform distribution over the subset of $\textsc{Block-Eq}^{-1}(1)$ consisting of all (x, y) for which $x_i = y_i$ for a unique i.

Lemma 4.4. *$\mu_0(R) \leq 45 \cdot \mu_1(R) + 2^{-\Omega(\sqrt{n})}$ holds for every rectangle $R \subseteq \{0,1\}^n \times \{0,1\}^n$.*

Lemma 4.5. *$\mu_1(R) \leq 2^{-\Omega(\sqrt{n})}$ holds for every 1-monochromatic rectangle R of* Block-Eq.

Together, Lemma 4.4 and Lemma 4.5 show that the hypothesis of Lemma 4.1 holds with $F :=$ Block-Eq, $C := 45$, and $\delta := 2^{-\Omega(\sqrt{n})}$. The lower bound in Theorem 2.1 follows. For the upper bound, in fact $\mathsf{ZPP}(\textsc{Block-Eq}) \leq O(\sqrt{n})$ holds (Kushilevitz & Nisan 1997, §4.1.1) (though it is slightly quicker to see that $\mathsf{NP}(\textsc{Block-Eq}) \leq O(\sqrt{n})$ holds by guessing i and deterministically verifying that $x_i = y_i$).

For the proofs of the lemmas, we define $m := \sqrt{n}$ and $b := \sqrt{n}$ (as in the notation for the decomposition $\textsc{Block-Eq} := \textsc{Or} \circ \textsc{Eq}^m$ where Eq takes b-bit inputs).

Proof (Proof of Lemma 4.4). For $x^0, x^1, y^0, y^1 \in \{0,1\}^b$, we say the tuple (x^0, x^1, y^0, y^1) is *valid* iff $x^0 \neq y^0$, $x^0 \neq y^1$, $x^1 \neq y^0$, and $x^1 = y^1$. We say

$$\Xi \;:=\; \big((x_1^0, x_1^1, y_1^0, y_1^1), \ldots, (x_m^0, x_m^1, y_m^0, y_m^1)\big)$$

is valid iff it is a tuple of valid tuples. If Ξ is valid then the injection $\Phi_\Xi \colon \{0,1\}^m \times \{0,1\}^m \to \{0,1\}^n \times \{0,1\}^n$ defined by

$$\Phi_\Xi(u, v) \;:=\; \big(x_1^{u_1} \cdots x_m^{u_m}, y_1^{v_1} \cdots y_m^{v_m}\big)$$

is a reduction from $\textsc{Inter} := \textsc{Or} \circ \textsc{And}^m$ (for single-bit And) to Block-Eq:

$$\textsc{Inter}(u, v) \;=\; \textsc{Block-Eq}\big(\Phi_\Xi(u, v)\big).$$

(In other words, the image of Φ_Ξ, as a submatrix of the Block-Eq matrix, is a copy of the Inter matrix.)

Define $\textsc{Unambig-Inter} \coloneqq \textsc{Unambig-Or} \circ \textsc{And}^m$ where the partial function $\textsc{Unambig-Or}$ is $\textsc{Or}$ restricted to the domain of strings of Hamming weight 0 or 1. That is, $\textsc{Unambig-Inter}^{-1}(0)$ consists of all pairs of disjoint sets, and $\textsc{Unambig-Inter}^{-1}(1)$ consists of all pairs of uniquely intersecting sets.

Lemma 4.6 (Razborov 1992). *There exist a distribution ν_0 over $\textsc{Unambig-Inter}^{-1}(0)$ and a distribution ν_1 over $\textsc{Unambig-Inter}^{-1}(1)$ such that $\nu_0(R) \leq 45 \cdot \nu_1(R) + 2^{-\Omega(m)}$ holds for every rectangle $R \subseteq \{0,1\}^m \times \{0,1\}^m$. Moreover, the uniquely intersecting coordinate in ν_1 is uniformly distributed.*

Letting $\mathbb{E}$ denote expectation, we claim that for $a \in \{0,1\}$ we have $\mu_a = \mathbb{E}_\Xi \, \Phi_\Xi(\nu_a)$ where a valid Ξ is chosen uniformly at random independently of ν_a. In other words, μ_a equals the distribution obtained by choosing Ξ, then independently taking a sample from ν_a, then applying Φ_Ξ to the sample (i.e., the uniform mixture of the distributions $\Phi_\Xi(\nu_a)$). We only argue that $\mu_1 = \mathbb{E}_\Xi \, \Phi_\Xi(\nu_1)$ (the argument for $\mu_0 = \mathbb{E}_\Xi \, \Phi_\Xi(\nu_0)$ is essentially the same). In fact, we make the stronger claim that for every $(u,v) \in \textsc{Unambig-Inter}^{-1}(1)$, say with $u_i = v_i = 1$, the distribution $\mathbb{E}_\Xi \, \Phi_\Xi(u,v)$ is uniform over the subset of $\textsc{Block-Eq}^{-1}(1)$ consisting of all (x,y) for which $x_i = y_i$ and $x_j \neq y_j$ for all $j \neq i$. The original claim follows from this since the uniquely intersecting coordinate i is uniformly distributed. The stronger claim follows immediately from the facts that the coordinates of Ξ are independent, that (x_i^1, y_i^1) is uniformly distributed over $\textsc{Eq}^{-1}(1)$, and that for $j \neq i$, (x_j^0, y_j^0), (x_j^0, y_j^1), and (x_j^1, y_j^0) are all marginally uniformly distributed over $\textsc{Eq}^{-1}(0)$. The claim is established.

Now for every rectangle $R \subseteq \{0,1\}^n \times \{0,1\}^n$, if we let $\Phi_\Xi^{-1}(R)$ denote the rectangle of all points in $\{0,1\}^m \times \{0,1\}^m$ that map into R under Φ_Ξ, then we have

$$\begin{aligned}
\mu_0(R) \quad &= \mathbb{E}_\Xi \big(\Phi_\Xi(\nu_0)(R)\big) \\
&= \mathbb{E}_\Xi \, \nu_0\big(\Phi_\Xi^{-1}(R)\big) \\
&\leq \mathbb{E}_\Xi \Big(45 \cdot \nu_1\big(\Phi_\Xi^{-1}(R)\big) + 2^{-\Omega(m)}\Big) \\
&= 45 \cdot \mathbb{E}_\Xi \, \nu_1\big(\Phi_\Xi^{-1}(R)\big) + 2^{-\Omega(m)}
\end{aligned}$$

$$\begin{aligned} &= 45 \cdot \mathbb{E}_{\Xi}\left(\Phi_{\Xi}(\nu_1)(R)\right) + 2^{-\Omega(m)} \\ &= 45 \cdot \mu_1(R) + 2^{-\Omega(\sqrt{n})}. \end{aligned}$$

□

Proof (Proof of Lemma 4.5). Note that μ_1 is uniform over a set of size

$$m\cdot 2^b\cdot(2^{2b}-2^b)^{m-1} = m\cdot 2^b\cdot 2^{2b(m-1)}\cdot(1-2^{-b})^{m-1} \geq \Omega(m\cdot 2^b\cdot 2^{2b(m-1)}).$$

If $R \coloneqq A \times B$ is 1-monochromatic, then $|A| \leq m \cdot 2^{b(m-1)}$ (since for any $y \in B$ there are at most $m \cdot (2^b)^{m-1}$ many x's for which Block-Eq$(x, y) = 1$), and similarly $|B| \leq m \cdot 2^{b(m-1)}$, and hence $|R| \leq m^2 \cdot 2^{2b(m-1)}$. It follows that

$$\mu_1(R) \;\leq\; \frac{m^2 \cdot 2^{2b(m-1)}}{\Omega(m \cdot 2^b \cdot 2^{2b(m-1)})} \;\leq\; O(m \cdot 2^{-b}) \;\leq\; 2^{-\Omega(\sqrt{n})}.$$

□

4.4. Proof of Theorem 2.2 and Theorem 2.3. We again use the corruption lemma from Razborov (1992), but now we need to take a closer look at the distribution over 1-inputs. Let $n = 4\ell - 1$. Let μ_0 be the distribution over Unique-Inter$^{-1}(0)$ that samples uniformly random disjoint sets of size ℓ, and let μ_1 be the distribution over Unique-Inter$^{-1}(1)$ that samples uniformly random uniquely intersecting sets of size ℓ.

Lemma 4.7 (Razborov 1992). *$\mu_0(R) \leq 45 \cdot \mu_1(R) + 2^{-\Omega(n)}$ holds for every rectangle $R \subseteq \{0,1\}^n \times \{0,1\}^n$.*

Lemma 4.8. *$\mu_1(R) \leq 2^{-\Omega(n)}$ holds for every 1-monochromatic rectangle R of* Unique-Inter.

Together, Lemma 4.7 and Lemma 4.8 show that the hypothesis of Lemma 4.1 holds with $F \coloneqq$ Unique-Inter, $C \coloneqq 45$, and $\delta \coloneqq 2^{-\Omega(n)}$. Theorem 2.2 and Theorem 2.3 follow.

Proof (Proof of Lemma 4.8). For each $i \in [n]$ let us define the rectangle $R_i \coloneqq \{(x, y) \in R : x_i = y_i = 1\}$, and note that the R_i's partition R. For each i we have $|R_i| \leq 2^{n-1}$ since every $(x, y) \in R_i$ is disjoint on the coordinates $[n] \smallsetminus \{i\}$.[7] Hence $|R| \leq n2^{n-1} \leq 2^{(1+o(1))n}$.

Note that μ_1 can be sampled by the following process.

1. Pick a uniformly random $i \in [n]$.
2. Pick a uniformly random $H \subseteq [n] \smallsetminus \{i\}$ of size $2\ell - 2$. There are $\binom{n-1}{2\ell-2} = \Theta(2^n/\sqrt{n})$ choices.
3. Pick a uniformly random partition $H = H_1 \cup H_2$ into sets of size $\ell - 1$. There are $\binom{2\ell-2}{\ell-1} = \Theta(2^{0.5n}/\sqrt{n})$ choices.
4. Let $x \coloneqq \{i\} \cup H_1$ and $y \coloneqq \{i\} \cup H_2$.

Hence μ_1 is uniform over its support of size

$$n \cdot \Theta(2^n/\sqrt{n}) \cdot \Theta(2^{0.5n}/\sqrt{n}) \;=\; \Theta(2^{1.5n}) \geq 2^{(1.5-o(1))n}.$$

It follows that $\mu_1(R) \leq 2^{(1+o(1))n}/2^{(1.5-o(1))n} \leq 2^{-\Omega(n)}$. □

5. Proof of Theorem 2.4

Consider an optimal $\mathsf{ZPP}^{\mathsf{NP}[1]}$ protocol Π for F with deterministic communication cost c and witness length k. By standard sparsification, we may assume the public randomness is uniformly distributed over $\{0,1\}^{O(\log n)}$. We let Π^r denote the $\mathsf{P}^{\mathsf{NP}[1]}$-type protocol induced by an outcome $r \in \{0,1\}^{O(\log n)}$, and we let the notation $R^r_v, S^r_{v,w}, o^r_v$ be with respect to Π^r (see the definition of $\mathsf{ZPP}^{\mathsf{NP}[1]}$ in Section 2.1). We claim that the following protocol $\widetilde{\Pi}$ is a $\mathsf{PostBPP}$ protocol for F of cost $O\big(\mathsf{ZPP}^{\mathsf{NP}[1]}(F) + \log n\big)$.

[7] By Kaibel & Weltge (2015), the bound $|R_i| \leq 2^{n-1}$ also holds assuming every input in R has intersection size either 1 or ≥ 3. Using this, it follows that Theorem 2.2 and Theorem 2.3 hold under the promise that at most two coordinates intersect.

```
   Input: $(x, y)$
   Output: $\in \{0, 1, \perp\}$
1  pick $r$ uniformly at random
2  run the deterministic part of $\Pi^r(x, y)$ to a leaf $v^r$
3  pick $a \in \{0, 1\}$ uniformly at random
4  if $a = 1$ then
5      pick $w \in \{0, 1\}^k$ uniformly at random
6      if $(x, y) \in S^r_{v^r,w}$ then output $o^r_{v^r}(1)$ else output $\perp$
7  else if $a = 0$ then
8      output $o^r_{v^r}(0)$ with probability $2^{-(k+2)}$
9      output $\perp$ with the remaining probability
10 end
```

$\widetilde{\Pi}$ has communication cost $c+O(1)$ and randomness cost $O(\log n) + k + O(1)$ and hence cost $O(c + k + \log n)$. We now argue the correctness. Let $\widetilde{\Pi}^r$ denote $\widetilde{\Pi}$ with a particular r chosen on line 1. Fix an input (x, y), and let $\chi^r \in \{0, 1\}$ indicate whether $(x, y) \in \bigcup_w S^r_{v^r,w}$ (i.e., the NP oracle's response). Let $A \coloneqq \{r : \chi^r = 1\}$, $\overline{A} \coloneqq \{r : \chi^r = 0\}$, $B \coloneqq \big\{r : o^r_{v^r}(\chi^r) = F(x, y)\big\}$, and $\overline{B} \coloneqq \big\{r : o^r_{v^r}(\chi^r) = \perp\big\}$. We have

$$\mathbb{P}\big[\widetilde{\Pi}^r(x, y) = F(x, y)\big] \;\geq\; \begin{cases} 2^{-(k+1)} & \text{if } r \in A \cap B \\ 0 & \text{if } r \in A \cap \overline{B} \\ 2^{-(k+3)} & \text{if } r \in \overline{A} \cap B \\ 0 & \text{if } r \in \overline{A} \cap \overline{B} \end{cases}$$

and

$$\mathbb{P}\big[\widetilde{\Pi}^r(x, y) = 1 - F(x, y)\big] \;\leq\; \begin{cases} 2^{-(k+3)} & \text{if } r \in A \\ 0 & \text{if } r \in \overline{A} \end{cases}.$$

Thus since $\mathbb{P}[r \in B] \geq 3/4$ we have

$$\begin{aligned} \mathbb{P}\big[\widetilde{\Pi}(x, y) = F(x, y)\big] \;&\geq\; 2^{-(k+3)} \cdot \mathbb{P}[r \in B] \\ &\quad + \big(2^{-(k+1)} - 2^{-(k+3)}\big) \cdot \mathbb{P}[r \in A \cap B] \\ &=\; 2^{-(k+3)} \cdot \big(\mathbb{P}[r \in B] + 3 \cdot \mathbb{P}[r \in A \cap B]\big) \end{aligned}$$

$$
\begin{aligned}
&\geq 2^{-(k+3)} \cdot \bigl(3 \cdot \mathbb{P}[r \in \overline{B}] + 3 \cdot \mathbb{P}[r \in A \cap B]\bigr) \\
&\geq 3 \cdot 2^{-(k+3)} \cdot \mathbb{P}[r \in A] \\
&\geq 3 \cdot \mathbb{P}\bigl[\widetilde{\Pi}(x,y) = 1 - F(x,y)\bigr]
\end{aligned}
$$

so $\widetilde{\Pi}$ is a correct $\mathsf{PostBPP}$ protocol for F.

6. Lower bound for majority

In this section we prove Theorem 2.6. We first give a general lower bound technique for $\mathsf{UPostBPP}_\square$ in Section 6.1, then we describe the machinery we borrow from Göös *et al.* (2016a) in Section 6.2, and finally we give the proof of Theorem 2.6 in Section 6.3.

6.1. Lower bound technique.

DEFINITION 6.1. *For* $F\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, $R \subseteq \{0,1\}^n \times \{0,1\}^n$, *and* μ *a distribution over* $\{0,1\}^n \times \{0,1\}^n$, *we say* R *is* μ-unbiased *(with respect to* F*) if* $\frac{1}{2} \cdot \mu\bigl(R \cap F^{-1}(0)\bigr) \leq \mu\bigl(R \cap F^{-1}(1)\bigr) \leq 2 \cdot \mu\bigl(R \cap F^{-1}(0)\bigr)$, *and is* μ-biased *otherwise.*

LEMMA 6.2. *Suppose* μ *is a distribution over* F^{-1} *and* ρ *is a product distribution over* $\{0,1\}^n \times \{0,1\}^n$ *such that for every rectangle* $R \subseteq \{0,1\}^n \times \{0,1\}^n$, *if* $\rho(R) \geq \delta$ *then* R *is* μ*-unbiased (with respect to* F*), and if* $\rho(R) \geq 1/2$ *then* $\mu(R) > 0$*. Then* $\mathsf{UPostBPP}_\square(F) \geq \Omega(\log(1/\delta))$*.*

The case where $\mu = \rho$ is equivalent to extended discrepancy Gavinsky & Lovett (2014) under product distributions, and leads to the lower bound $\mathsf{UPostBPP}_\square(\textsc{Conf-Maj}) \geq \Omega(\sqrt{n \log n})$ (details omitted). The more general form is needed to get the $\Omega(n)$ lower bound. The results of Gavinsky & Lovett (2014) show that for total F, $\mathsf{P}(F) \leq \mathrm{poly}(\log n)$ follows from the assumptions that F's matrix has $\mathrm{poly}(\log n)$ rank (over the reals) and that every rectangle S has a subrectangle that has measure $\geq 2^{-\mathrm{poly}(\log n)}$ and is biased (both with respect to the uniform distribution over S). By letting $\rho = \mu$ be uniform over an arbitrary S in Lemma 6.2, the latter property follows from the existence of a $\mathrm{poly}(\log n)$-cost

$\mathsf{UPostBPP}_\square$ protocol. Hence to prove the Log Rank Conjecture, it suffices to prove the same with $\mathsf{UPostBPP}_\square$ instead of P.

Proof (Proof of Lemma 6.2). Suppose Π is a cost-k $\mathsf{UPostBPP}_\square$ protocol for F. By Observation B.21 (in Appendix B), we may assume Π is just a distribution over 2^k many $\{0,1\}$-labeled rectangles. For R a rectangle and $o \in \{0,1\}$, we let $\pi_{R,o}$ denote the probability of getting R with label o under Π.

We start by recording the following observation: For every distribution ν over F^{-1} there exists an (R,o) such that $\pi_{R,o} > 0$ and R is ν-biased. This follows by considering the 2-player 0-sum game where the row strategies are inputs $(x,y) \in F^{-1}$, the column strategies are (R,o) pairs with $\pi_{R,o} > 0$, and the payoff to the column player is 1 if $(x,y) \in R$ and $F(x,y) = o$, is -2 if $(x,y) \in R$ and $F(x,y) = 1-o$, and is 0 if $(x,y) \notin R$. The mixed column strategy π demonstrates that the game has positive value, and hence for every mixed row strategy there exists a pure column strategy for which the expected payoff to the column player is positive. This implies the observation. Only the straightforward direction of the Minimax Theorem is used.

Consider the following procedure.

```
1 let $Q_0 \coloneqq \{0,1\}^n \times \{0,1\}^n$
2 for $i = 1, 2, \ldots$ do
3     let $R_i$ be a $(\mu \mid Q_{i-1})$-biased rectangle such that
      $\pi_{R_i,o_i} > 0$ for some $o_i$
4     let $A_i \times B_i \coloneqq R_i \cap Q_{i-1}$
5     let $Q_i \coloneqq Q_{i-1} \setminus$
      $\big($either $A_i \times \{0,1\}^n$ or $\{0,1\}^n \times B_i$, whichever is smaller
       under $\rho\big)$
6 until $\rho(Q_i) < 1/2$
```

We show by induction on i that lines 3 and 4 always succeed, Q_i is a rectangle, $\rho(Q_i) \geq 1 - i \cdot \sqrt{\delta}$, and the R_j's for $j \in \{1, \ldots, i\}$ are all distinct from each other and disjoint from Q_i. The base case $i = 0$ is trivial, so assume this holds for $i-1$. Since Q_{i-1} is a rectangle and $\rho(Q_{i-1}) \geq 1/2$ by line 6, we have $\mu(Q_{i-1}) > 0$ by assumption (with $R \coloneqq Q_{i-1}$) and hence the conditioning on line

3 is valid. By the above observation (with $\nu \coloneqq (\mu \mid Q_{i-1})$), line 3 succeeds. Since Q_{i-1} is a rectangle, so are $R_i \cap Q_{i-1}$ (hence line 4 succeeds) and Q_i. Since R_i is $(\mu \mid Q_{i-1})$-biased, we have that $R_i \cap Q_{i-1}$ is μ-biased and hence $\rho(R_i \cap Q_{i-1}) < \delta$ by assumption (with $R \coloneqq R_i \cap Q_{i-1}$). Since ρ is a product distribution, either $\rho\big(A_i \times \{0,1\}^n\big) < \sqrt{\delta}$ or $\rho\big(\{0,1\}^n \times B_i\big) < \sqrt{\delta}$. Hence $\rho(Q_i) \geq \rho(Q_{i-1}) - \sqrt{\delta} \geq 1 - i \cdot \sqrt{\delta}$. Since $R_i \cap Q_{i-1}$ is μ-biased, R_i is not disjoint from Q_{i-1} and hence R_i is distinct from the R_j's for $j \in \{1, \ldots, i-1\}$ (since the latter are all disjoint from Q_{i-1}). Since $Q_i \subseteq Q_{i-1}$, the R_j's for $j \in \{1, \ldots, i-1\}$ are also disjoint from Q_i. Finally, line 5 ensures that R_i is disjoint from Q_i, since $Q_i \subseteq Q_{i-1} \smallsetminus (A_i \times B_i) = Q_{i-1} \smallsetminus R_i$. This completes the induction step.

Since the R_i's are all distinct and $\pi_{R_i,o_i} > 0$, there are at most 2^k iterations. Let i^* be the final value of i. By line 6, we have $1/2 > \rho(Q_{i^*}) \geq 1 - i^* \cdot \sqrt{\delta}$ and hence $2^k \geq i^* > 1/2\sqrt{\delta}$. Thus $k > \frac{1}{2} \cdot \log(1/\delta) - 1$. □

6.2. Conjunction rectangles. We now state the "Packing with Conjunctions Theorem" from Göös *et al.* (2016a), which is the technical heart of the main "Junta Theorem" from that paper. The theorem makes no reference to the outer function f; it is simply a statement about the function $G \coloneqq g^m$ where g is the confounding gadget with block length b (Definition 2.5).

Definition 6.3. *Two distributions over $\{0,1\}^m$ are ϵ-*close *if for every $z \in \{0,1\}^m$, the probabilities of z under the two distributions are within a factor $(1 \pm \epsilon)$ of each other.*

Definition 6.4. *A rectangle S is a (d, ϵ)-*conjunction rectangle *if there exists a width-d conjunction $h\colon \{0,1\}^m \to \{0,1\}$ (i.e., h 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 the distributions over $\{0,1\}^m$ obtained in the following two ways are ϵ-close:*

- *picking a uniformly random $z \in \{0,1\}^m$ and a uniformly random $(x, y) \in G^{-1}(z)$ and conditioning on $(x, y) \in S$,*
- *picking a uniformly random $z \in h^{-1}(1)$.*

Definition 6.5. *A distribution ν over $\{0,1\}^n \times \{0,1\}^n$ is the* lift *of a distribution ξ over $\{0,1\}^m$ if $\nu(x,y) = \xi(z)/|G^{-1}(z)|$ where $z := G(x,y)$. Note that a lifted distribution is a convex combination of distributions that are uniform over a set $G^{-1}(z)$.*

Theorem 6.6. *For $\epsilon := 1/100$, and for every $d \geq 0$, every lifted distribution ν, and every rectangle R with $\nu(R) \geq 2^{-db/20}$, there exist disjoint (d,ϵ)-conjunction subrectangles $S_1, S_2, \ldots \subseteq R$ such that $\nu\big(\bigcup_i S_i \mid R\big) \geq 1 - \epsilon$.*

The proof in Göös *et al.* (2016a) actually gives $\epsilon := 2^{-\Theta(b)}$, but we only need $\epsilon := 1/100$.

6.3. Proof of Theorem 2.6. For convenience, assume m is odd. We have Conf-Maj $:= f \circ G$ where $G := g^m$. Let $M := \big\{z \in \{0,1\}^m : |z| \in \{\lfloor m/2 \rfloor, \lceil m/2 \rceil\}\big\}$ (the "middle layers" of the Hamming cube), and let $L := G^{-1}(M)$ (the "lifted version" of the set M). Let the distribution μ be the lift of the uniform distribution over M (so μ is supported on L), and let ρ be the uniform distribution over $\{0,1\}^n \times \{0,1\}^n$ (which is a product distribution). We argue the following two claims, both of which exploit Theorem 6.6. Recall that $b := 100 \log m$ and $n := m \cdot b$.

Claim 6.7. *For every rectangle R, if $\rho(R) \geq 0.99999^n$ then $\mu(R) \geq 0.997^n$.*

Claim 6.8. *For every rectangle R, if $\mu(R) \geq 0.997^n$ then R is μ-unbiased.*

Theorem 2.6 follows because the assumptions of Lemma 6.2 hold with $\delta := 0.99999^n$: The first part (if $\rho(R) \geq \delta$ then R is μ-unbiased) holds by Claim 6.7 and Claim 6.8. The second part (if $\rho(R) \geq 1/2$ then $\mu(R) > 0$) holds by Claim 6.7 alone.

Observation 6.9. *Let ρ' be the lift of the uniform distribution over $\{0,1\}^m$, and note the following.*

(i) *ρ and ρ' are $(1/2)$-close. (This is straightforward to verify using $b \coloneqq 100\log m$ and the fact that $|g^{-1}(0)|, |g^{-1}(1)| \in 2^{2b-1} \pm 2^{b-1}$.)*

(ii) *$\mu(\cdot) = \rho'(\cdot \mid L)$.*

(iii) *The first distribution in Definition 6.4 picks z with probability $\rho'\big(G^{-1}(z) \mid S\big)$; hence this value is in $(1\pm\epsilon)\cdot\mathbb{P}_{z'\in h^{-1}(1)}[z' = z]$ where the notation $\mathbb{P}_{z'\in h^{-1}(1)}$ means a uniformly random choice.*

In the proof of Claim 6.7 we use the following fact, which holds by Stirling approximations.

FACT 6.10. *For all $s \geq t$ we have*

$$\begin{aligned}\binom{s}{t} &= \Theta\left(\frac{1}{\sqrt{s}}\cdot\left(\frac{s}{t}\right)^{t+1/2}\cdot\left(\frac{s}{s-t}\right)^{s-t+1/2}\right)\\ &\geq \Omega\left(\frac{1}{\sqrt{s}}\cdot\left(\frac{s}{\max(t,\, s-t)}\right)^{s}\right).\end{aligned}$$

PROOF (Proof of Claim 6.7). Assuming $\rho(R) \geq 0.99999^n$, we have $\rho'(R) \geq 0.99998^n$. Apply Theorem 6.6 with $\nu \coloneqq \rho'$ and $d \coloneqq m/1000$ (noting that $0.99998^n \geq 2^{-db/20}$) to get disjoint (d,ϵ)-conjunction subrectangles $S_1, S_2, \ldots \subseteq R$ with associated conjunctions $h_1, h_2, \ldots$, such that $\rho'\big(\bigcup_i S_i \mid R\big) \geq 1-\epsilon$ (where $\epsilon \coloneqq 1/100$). For each i, assuming for convenience that h_i depends on exactly d variables, exactly j of which are positive literals, we have

$$\begin{aligned}\big|h_i^{-1}(1)\cap M\big| &= \binom{m-d}{\lfloor m/2\rfloor - j} + \binom{m-d}{\lceil m/2\rceil - j}\\ &\geq \binom{m-d}{\lfloor m/2\rfloor - j}\\ &\geq \Omega\left(\frac{1}{\sqrt{m}}\cdot\left(\frac{m-d}{\lceil m/2\rceil}\right)^{m-d}\right)\\ &\geq 2^{m-d}\cdot\Omega\left(\frac{1}{\sqrt{m}}\cdot 0.9987^{m-d}\right)\\ &\geq 2^{m-d}\cdot 0.9985^{m}\end{aligned}$$

where the third line follows by Fact 6.10 and the fourth line follows by $d := m/1000$. Thus we have $\mathbb{P}_{z\in h_i^{-1}(1)}[z \in M] = \frac{|h_i^{-1}(1)\cap M|}{2^{m-d}} \geq 0.9985^m$, and hence

$$\begin{aligned}
\rho'(L \mid S_i) &= \textstyle\sum_{z\in M} \rho'\big(G^{-1}(z) \mid S_i\big) \\
&\geq \textstyle\sum_{z\in M} (1-\epsilon)\cdot \mathbb{P}_{z'\in h_i^{-1}(1)}[z' = z] \\
&= (1-\epsilon)\cdot \mathbb{P}_{z\in h_i^{-1}(1)}[z \in M] \\
&\geq 0.9983^m
\end{aligned}$$

where the second line follows since S_i is a (d,ϵ)-conjunction rectangle. Then we have

$$\begin{aligned}
\rho'(L \mid R) &\geq \sum\nolimits_i \rho'(L \cap S_i \mid R) \\
&= \sum\nolimits_i \rho'(L \mid S_i)\cdot \rho'(S_i \mid R) \\
&\geq \sum\nolimits_i 0.9983^m \cdot \rho'(S_i \mid R) \\
&= 0.9983^m \cdot \rho'\big(\bigcup\nolimits_i S_i \mid R\big) \\
&\geq 0.998^m
\end{aligned}$$

where the last line follows by $\rho'\big(\bigcup_i S_i \mid R\big) \geq 1-\epsilon$, and finally

$$\begin{aligned}
\mu(R) &= \rho'(R \mid L) \\
&\geq \rho'(L \mid R)\cdot \rho'(R) \\
&\geq 0.998^m \cdot 0.99998^n \\
&\geq 0.997^n.
\end{aligned}$$

□

Proof (Proof of Claim 6.8). Apply Theorem 6.6 with $\nu := \mu$ and $d := m/10$ (noting that $0.997^n \geq 2^{-db/20}$) to get disjoint (d,ϵ)-conjunction subrectangles $S_1, S_2, \ldots \subseteq R$ with associated conjunctions $h_1, h_2, \ldots$, such that $\mu\big(\bigcup_i S_i \mid R\big) \geq 1-\epsilon$ (where $\epsilon := 1/100$). Recall that $f\colon \{0,1\}^m \to \{0,1\}$ is the majority function. For each i, assuming for convenience that h_i depends on exactly d variables, exactly j of which are positive literals, we have

$$\frac{\mathbb{P}_{z\in h_i^{-1}(1)}\big[z \in f^{-1}(0)\cap M\big]}{\mathbb{P}_{z\in h_i^{-1}(1)}\big[z \in f^{-1}(1)\cap M\big]} = \frac{\binom{m-d}{\lfloor m/2\rfloor - j}}{\binom{m-d}{\lceil m/2\rceil - j}}$$

$$\begin{aligned} &= \frac{\lceil m/2\rceil - j}{\lceil m/2\rceil - d + j} \\ &= 1 + \frac{d-2j}{\lceil m/2\rceil - d + j} \\ &\in \left[\tfrac{3}{4}, \tfrac{4}{3}\right] \end{aligned}$$

since $d \coloneqq m/10$. Now fix any output $o \in \{0,1\}$, and let $E_o \coloneqq \textsc{Conf-Maj}^{-1}(o) = G^{-1}(f^{-1}(o))$. We have

$$\begin{aligned} \rho'(E_o \cap L \mid S_i) &= \textstyle\sum_{z \in f^{-1}(o)\cap M} \rho'\big(G^{-1}(z) \mid S_i\big) \\ &\geq \textstyle\sum_{z \in f^{-1}(o)\cap M} (1-\epsilon)\cdot \mathbb{P}_{z' \in h_i^{-1}(1)}[z' = z] \\ &= (1-\epsilon)\cdot \mathbb{P}_{z \in h_i^{-1}(1)}\big[z \in f^{-1}(o)\cap M\big] \\ &\geq (1-\epsilon)\cdot \tfrac{3}{4}\cdot \mathbb{P}_{z \in h_i^{-1}(1)}\big[z \in f^{-1}(1-o)\cap M\big] \\ &\geq (1-\epsilon)\cdot \tfrac{3}{4}\cdot (1-\epsilon)\cdot \rho'(E_{1-o}\cap L \mid S_i) \\ &\geq \tfrac{2}{3}\cdot \rho'(E_{1-o}\cap L \mid S_i). \end{aligned}$$

If $\mu(S_i) > 0$ (equivalently, $\rho'(L \mid S_i) > 0$) then

$$\begin{aligned} \mu(E_o \mid S_i) &= \rho'(E_o \mid L \cap S_i) \\ &= \rho'(E_o \cap L \mid S_i)\,/\,\rho'(L \mid S_i) \\ &\geq \tfrac{2}{3}\cdot \rho'(E_{1-o}\cap L \mid S_i)\,/\,\rho'(L \mid S_i) \\ &= \tfrac{2}{3}\cdot \mu(E_{1-o} \mid S_i) \end{aligned}$$

and hence $\mu(E_o \mid S_i) \geq \frac{2}{5}$. Now we have

$$\begin{aligned} \mu(E_o \mid R) &\geq \textstyle\sum_i \mu(E_o \cap S_i \mid R) \\ &= \textstyle\sum_{i\,:\,\mu(S_i)>0} \mu(E_o \mid S_i)\cdot \mu(S_i \mid R) \\ &\geq \textstyle\sum_{i\,:\,\mu(S_i)>0} \tfrac{2}{5}\cdot \mu(S_i \mid R) \\ &= \tfrac{2}{5}\cdot \mu\big(\textstyle\bigcup_i S_i \mid R\big) \\ &\geq \tfrac{1}{3}. \end{aligned}$$

where the last line follows by $\mu\big(\bigcup_i S_i \mid R\big) \geq 1-\epsilon$. Thus $\mu(E_o \mid R) \geq \frac{1}{2}\cdot \mu(E_{1-o} \mid R)$ and hence $\mu(R \cap E_o) \geq \frac{1}{2}\cdot \mu(R \cap E_{1-o})$. Since this holds for either $o \in \{0,1\}$, R is μ-unbiased with respect to $\textsc{Conf-Maj}$. □

7. Additional observations

Observation 7.1. $\mathsf{UPostBPP}_{\square}(F) \leq O\big(\mathsf{P}^{\mathsf{NP}}(F)\big)$ *for all* F, *and hence* $\mathsf{P}^{\mathsf{NP}} \subseteq \mathsf{UPostBPP}_{\square}$.

Proof (Proof sketch). It is a classical fact that if we consider "super-witnesses" that consist of a string of purported responses to the NP oracle queries along with purported witnesses for all the queries for which the purported response is "1," and if we order the super-witnesses reverse-lexicographically by the string of oracle responses (the witnesses do not matter for the ordering), then the output of a P^{NP} computation is determined by the first super-witness for which all the purported oracle query witnesses check out. (This fact was phrased as an "overlay" characterization in Papakonstantinou *et al.* (2014) and was also used in the proof that $\mathsf{P}^{\mathsf{NP}} \subseteq \mathsf{S}_2\mathsf{P}$ Canetti (1996); Russell & Sundaram (1998).) To get a $\mathsf{UPostBPP}_{\square}$ protocol, we can pick a random super-witness with probabilities geometrically decreasing according to the order, output $\perp$ if one of the purported witnesses does not check out, and otherwise produce the same output as the computation path given the purported oracle responses. □

Let Neq be the non-equality function, which is in RP.

Observation 7.2. $\mathsf{US}(\textsc{Neq}) = \Theta(n)$, *and hence* $\mathsf{RP} \not\subseteq \mathsf{US}$.

Proof. The matrix of Neq is the complement of the identity matrix. Consider a collection of rectangles that touches each off-diagonal entry exactly once, and touches each diagonal entry either zero times or at least twice. If we sum these rectangles as 0-1 matrices over the reals, the resulting matrix M has all off-diagonal entries $= 1$ and all diagonal entries $\neq 1$. Subtracting the all-1's matrix from M results in a diagonal matrix with all nonzero diagonal entries, which has full rank. Thus M has rank at least $2^n - 1$ since the all-1's matrix has rank 1. However, the number of rectangles upper bounds the rank since each rectangle has rank 1. □

Consider the partial function Which-Eq: $\{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ where the two inputs are each partitioned into two strings of length $n/2$, $x := x_0x_1$ and $y := y_0y_1$, such that $\textsc{Which-Eq}(x,y) = 0$ iff $x_0 = y_0$ and $x_1 \neq y_1$, and $\textsc{Which-Eq}(x,y) = 1$ iff $x_0 \neq y_0$ and $x_1 = y_1$. Note that $\textsc{Which-Eq} \in \mathsf{ZPP}$.

Observation 7.3. *$\oplus\mathsf{P}(\textsc{Which-Eq}) = \Theta(n)$, and hence $\mathsf{ZPP} \not\subseteq \oplus\mathsf{P}$.*

Proof. Consider any total boolean matrix M that agrees with $\textsc{Which-Eq}$ on the latter's domain (i.e., on the set of all inputs (x,y) on which $\textsc{Which-Eq}$ is defined, meaning that exactly one of $x_0 = y_0$ or $x_1 = y_1$ holds). We claim that M contains a $2^{n/2-1} \times 2^{n/2-1}$ identity or complement-of-identity submatrix; hence M has rank at least $2^{n/2-1} - 1$ over $GF(2)$. If $M(z,z) = 1$ for at least half of all $z \in \{0,1\}^n$, fix z_0 so that $M(z_0z_1, z_0z_1) = 1$ for all z_1 in some $Z_1 \subseteq \{0,1\}^{n/2}$ of size $2^{n/2-1}$, and note that for $x_1, y_1 \in Z_1$, $M(z_0x_1, z_0y_1)$ indicates whether $x_1 = y_1$. On the other hand, if $M(z,z) = 0$ for at least half of all $z \in \{0,1\}^n$, fix z_1 so that $M(z_0z_1, z_0z_1) = 0$ for all z_0 in some $Z_0 \subseteq \{0,1\}^{n/2}$ of size $2^{n/2-1}$, and note that for $x_0, y_0 \in Z_0$, $M(x_0z_1, y_0z_1)$ indicates whether $x_0 \neq y_0$. □

8. Conclusion and open questions

It is open to prove that any explicit function is not in $\mathsf{S_2P}$; we wish to highlight this as a new frontier (presumably incomparable to the $\mathsf{AM} \cap \mathsf{coAM}$ frontier) toward proving explicit lower bounds for the communication polynomial hierarchy.

Is Lemma 6.2 a tight lower bound technique for $\mathsf{UPostBPP}_\square$? (This is related to the question of whether the lower bound technique given in Papakonstantinou *et al.* (2014) for P^{NP} is tight, which has been resolved in the negative Göös *et al.* (2017).) It is also open to prove a $\mathsf{UPostBPP}_\square$ lower bound for the majority function lifted with a constant-size gadget (which, without loss of generality, would be $\textsc{And}$ or $\textsc{Xor}$). Finally, we mention that for some known results, there is room for quantitative improvement; e.g., is there an $F \in \mathsf{MA}$ such that $\mathsf{ZPP}^{\mathsf{NP}[1]}(F) \geq \omega(\sqrt{n})$?

Our survey in Section 3 lists all the open problems that fall directly within the scope of this paper. Although we aimed for our survey to be fairly comprehensive, there are some further topics concerning communication complexity classes that we have not addressed. Our discussion has excluded classes involving limited ambiguity (such as UP, FewP Grolmusz & Tardos (2003); Karchmer *et al.* (1994); Klauck (2010), and UAM Göös *et al.* (2016b)), more exotic counting classes (such as Few, APP,[8] WPP, AWPP, WAPP, LWPP, SPP, $\mathsf{C_{=}P}$, and $\mathsf{Mod}_m\mathsf{P}$ Damm *et al.* (2004) for integers $m > 2$), classes defined using the dot operator (such as $\mathsf{BP} \cdot \mathsf{UP}$ Klauck (2010), $\mathsf{U} \cdot \mathsf{BPP}$, and $\mathsf{N} \cdot \mathsf{BPP}$ which may differ from MA), and classes with oracles other than NP. One can also ask about more complicated relationships among the classes (e.g., concerning intersections of different classes, although we have mentioned $\mathsf{NP} \cap \mathsf{coNP}$, $\mathsf{AM} \cap \mathsf{coAM}$, and $\Sigma_2\mathsf{P} \cap \Pi_2\mathsf{P}$), and about closure properties (e.g., it is open whether UPP is closed under intersection). Finally, we have not considered average-case models, quantum models, multi-party models, variable partition models, round-restricted models, asymmetric models, search problems, or functions with non-boolean codomains.

A. Appendix: Information complexity proof of Theorem 2.2

In this appendix we provide an alternate proof of Theorem 2.2 using information complexity tools.

A.1. Preliminaries. In this proof it is more convenient to consider the private-randomness version of $\mathsf{ZPP}^{\mathsf{NP}[1]}$, in which a protocol consists of a single $\mathsf{P}^{\mathsf{NP}[1]}$-type protocol over the domain $\big(\{0,1\}^n \times \{0,1\}^q\big) \times \big(\{0,1\}^n \times \{0,1\}^q\big)$ (for some q), and on input (x, y) the protocol is applied to $\big((x, r_x), (y, r_y)\big)$ where $r_x, r_y \in \{0,1\}^q$ are chosen independently uniformly at random. This model is equivalent to the public-randomness version, within a constant factor and additive $O(\log n)$ term in the cost, by standard sparsification of randomness and the fact that the success probability can be amplified as long as it is a constant greater than $1/2$ Chang & Purini

[8] Not to be confused with the measure APP from Klauck (2003), which is equivalent to $\mathsf{PostBPP}$.

(2008).

Throughout this appendix, we use bold letters for random variables, $\mathbb{P}$ for probability, $\mathbb{E}$ for expectation, $\mathbb{H}$ for Shannon entropy, $\mathbb{I}$ for mutual information, H for Hellinger distance, and Δ for statistical (total variation) distance. Recall that if $\boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2$ are distributions over a finite set Ω, then $H^2(\boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2) \coloneqq 1 - \sum_{\omega \in \Omega} \sqrt{\boldsymbol{\Psi}_1(\omega)\boldsymbol{\Psi}_2(\omega)}$ and $\Delta(\boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2) \coloneqq \frac{1}{2} \sum_{\omega \in \Omega} \big|\boldsymbol{\Psi}_1(\omega) - \boldsymbol{\Psi}_2(\omega)\big|$. We use the following (by-now standard) lemmas Bar-Yossef *et al.* (2004); Lin (1991).

Lemma A.1. *Suppose $\boldsymbol{\Psi}, \boldsymbol{\Lambda}$ are jointly distributed random variables and $\boldsymbol{\Lambda}$ is uniform over two outcomes, say $\{1, 2\}$. Then $H^2(\boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2) \le \mathbb{I}(\boldsymbol{\Psi} \; ; \boldsymbol{\Lambda})$ where $\boldsymbol{\Psi}_\Lambda \coloneqq (\boldsymbol{\Psi} \mid \boldsymbol{\Lambda} = \Lambda)$ for $\Lambda \in \{1, 2\}$.*

Lemma A.2. *If $\boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2$ are distributions, then*

$$H^2(\boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2) \;\le\; \Delta(\boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2) \;\le\; \sqrt{2} H(\boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2).$$

A.2. Proof of Theorem 2.2. Suppose for contradiction there is a cost-$o(n)$ private-randomness $\mathsf{ZPP}^{\mathsf{NP}[1]}$ protocol Π computing Unique-Inter with success probability $1/2 + \epsilon$ (for any constant $\epsilon > 0$). Let the notation $k, R_v, S_{v,w}, o_v$ be with respect to Π (similarly to the definition of $\mathsf{ZPP}^{\mathsf{NP}[1]}$ in Section 2.1), and let c be the communication cost of the deterministic part of Π. Consider the following jointly distributed random variables.

- Let $\boldsymbol{i}$ be uniform over $[n]$.
- Let $\boldsymbol{z} \coloneqq \boldsymbol{z}_1 \cdots \boldsymbol{z}_n$ be distributed as follows. For each $j \in [n]$ (independently), if $j = \boldsymbol{i}$ then $\boldsymbol{z}_j$ is uniform over the two outcomes $\big\{\{00\}, \{11\}\big\}$, and if $j \ne \boldsymbol{i}$ then $\boldsymbol{z}_j$ is uniform over the two outcomes $\big\{\{00, 10\}, \{00, 01\}\big\}$.
- Let $\boldsymbol{x} \coloneqq \boldsymbol{x}_1 \cdots \boldsymbol{x}_n$ and $\boldsymbol{y} \coloneqq \boldsymbol{y}_1 \cdots \boldsymbol{y}_n$ be distributed as follows. For each $j \in [n]$ (independently), $\boldsymbol{x}_j \boldsymbol{y}_j$ is uniform over the elements of the outcome of $\boldsymbol{z}_j$.
- Let $\boldsymbol{r}_x, \boldsymbol{r}_y$ be the private random strings, which are independent of $\boldsymbol{x}, \boldsymbol{y}$.
- Let $\boldsymbol{v} \in \{0, 1\}^c$ be the leaf reached (i.e., the deterministic transcript) of $\Pi\big((\boldsymbol{x}, \boldsymbol{r}_x), (\boldsymbol{y}, \boldsymbol{r}_y)\big)$.

- Let $\boldsymbol{\chi} \in \{0,1\}$ indicate whether $\big((\boldsymbol{x},\boldsymbol{r}_x),(\boldsymbol{y},\boldsymbol{r}_y)\big) \in \bigcup_w S_{\boldsymbol{v},w}$ (i.e., the NP oracle's response).
- Let $\boldsymbol{w} \in \{\varepsilon\} \cup \{0,1\}^k$ (where ε is the empty string) be distributed as follows. If $\boldsymbol{\chi} = 0$ then $\boldsymbol{w} := \varepsilon$. If $\boldsymbol{\chi} = 1$ then let $\boldsymbol{w} \in \{0,1\}^k$ be chosen arbitrarily such that $\big((\boldsymbol{x},\boldsymbol{r}_x),(\boldsymbol{y},\boldsymbol{r}_y)\big) \in S_{\boldsymbol{v},\boldsymbol{w}}$.
- Let $\boldsymbol{\Pi} := \boldsymbol{vw}$, which is distributed over $\{0,1\}^c \cup \{0,1\}^{c+k}$. Note that $\boldsymbol{\Pi}$ is a deterministic function of $\big((\boldsymbol{x},\boldsymbol{r}_x),(\boldsymbol{y},\boldsymbol{r}_y)\big)$.

Let $\boldsymbol{x}_{-\boldsymbol{i}}, \boldsymbol{y}_{-\boldsymbol{i}}, \boldsymbol{z}_{-\boldsymbol{i}}$ denote the restrictions of $\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}$ to coordinates in $[n] \smallsetminus \{\boldsymbol{i}\}$. We have

$$\mathbb{I}\big(\boldsymbol{\Pi} \;;\; \boldsymbol{x}_{-\boldsymbol{i}}, \boldsymbol{y}_{-\boldsymbol{i}} \;\big|\; \boldsymbol{i},\boldsymbol{z}\big) \;\leq\; \mathbb{H}\big(\boldsymbol{\Pi} \;\big|\; \boldsymbol{i},\boldsymbol{z}\big) \;\leq\; c+k \;\leq\; o(n).$$

By the standard direct sum property for mutual information Bar-Yossef *et al.* (2004); Jayram *et al.* (2003), if $\boldsymbol{j}$ is uniform over $[n] \smallsetminus \{\boldsymbol{i}\}$ (and independent of the other random variables, conditioned on $\boldsymbol{i}$) then

$$\mathbb{I}\big(\boldsymbol{\Pi} \;;\; \boldsymbol{x}_{\boldsymbol{j}}, \boldsymbol{y}_{\boldsymbol{j}} \;\big|\; \boldsymbol{i},\boldsymbol{j},\boldsymbol{z}\big) \;\leq\; \tfrac{1}{n-1} \cdot \mathbb{I}\big(\boldsymbol{\Pi} \;;\; \boldsymbol{x}_{-\boldsymbol{i}}, \boldsymbol{y}_{-\boldsymbol{i}} \;\big|\; \boldsymbol{i},\boldsymbol{z}\big) \;\leq\; o(1).$$

Define two more random variables (which are deterministic functions of $(\boldsymbol{i},\boldsymbol{j},\boldsymbol{z})$) as follows: $\boldsymbol{h} := \{\boldsymbol{i},\boldsymbol{j}\}$ and

$$\boldsymbol{g} := \begin{cases} \text{HEADS} & \text{if } \boldsymbol{i} < \boldsymbol{j} \text{ and } \boldsymbol{z}_{\boldsymbol{i}}{=}\{11\}, \text{ or if } \boldsymbol{i} > \boldsymbol{j} \text{ and } \boldsymbol{z}_{\boldsymbol{i}} = \{00\} \\ \text{TAILS} & \text{if } \boldsymbol{i} < \boldsymbol{j} \text{ and } \boldsymbol{z}_{\boldsymbol{i}} = \{00\}, \text{ or if } \boldsymbol{i} > \boldsymbol{j} \text{ and } \boldsymbol{z}_{\boldsymbol{i}} = \{11\} \end{cases}.$$

Let $\boldsymbol{x}_{-\boldsymbol{h}}, \boldsymbol{y}_{-\boldsymbol{h}}, \boldsymbol{z}_{-\boldsymbol{h}}$ denote the restrictions of $\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}$ to coordinates in $[n] \smallsetminus \boldsymbol{h}$. Since we have $\mathbb{I}\big(\boldsymbol{\Pi} \;;\; \boldsymbol{x}_{\boldsymbol{j}}, \boldsymbol{y}_{\boldsymbol{j}} \;\big|\; \boldsymbol{g},\boldsymbol{h},\boldsymbol{i},\boldsymbol{j},\boldsymbol{z}\big) \leq o(1)$, there must exist outcomes $g^* \in \{\text{HEADS},\text{TAILS}\}$, $h^* \in \binom{[n]}{2}$, and $z^*_{-h^*}$ such that if we let E denote the event $\big(\boldsymbol{g} = g^*,\ \boldsymbol{h} = h^*,\ \boldsymbol{z}_{-\boldsymbol{h}} = z^*_{-h^*}\big)$ then

$$\mathbb{I}\big(\boldsymbol{\Pi} \;;\; \boldsymbol{x}_{\boldsymbol{j}}, \boldsymbol{y}_{\boldsymbol{j}} \;\big|\; E,\boldsymbol{i},\boldsymbol{j},\boldsymbol{z}_{\boldsymbol{h}}\big) \;\leq\; o(1). \tag{A.3}$$

Assume $g^* = \text{HEADS}$ and $h^* = \{1,2\}$ (the other cases are analogous). As illustrated in Figure A.1, define A $:= (00,00)$, B $:= (10,00)$, C $:= (00,10)$, D $:= (10,10)$, E $:= (11,10)$, F $:= (10,11)$. Conditioned on E, there are the following four equally likely outcomes of $(\boldsymbol{i},\boldsymbol{j},\boldsymbol{z}_{\boldsymbol{h}})$.

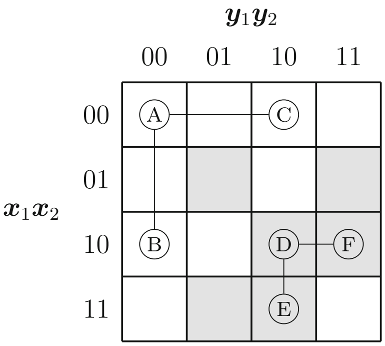
$\boldsymbol{y}_1\boldsymbol{y}_2$
00
01
10
11
$\boldsymbol{x}_1\boldsymbol{x}_2$
00
01
10
11
A
C
B
D
F
E

Figure A.1: The shaded rectangles are all 1-inputs, and the unshaded rectangles are all 0-inputs.

> $\big(\boldsymbol{i}=2,\ \boldsymbol{j}=1,\ \boldsymbol{z_i}=\{00\},\ \boldsymbol{z_j}=\{00,10\}\big)$ so $(\boldsymbol{x}_1\boldsymbol{x}_2,\boldsymbol{y}_1\boldsymbol{y}_2)$ is uniform over $\{\text{A},\text{B}\}$.
>
> $\big(\boldsymbol{i}=2,\ \boldsymbol{j}=1,\ \boldsymbol{z_i}=\{00\},\ \boldsymbol{z_j}=\{00,01\}\big)$ so $(\boldsymbol{x}_1\boldsymbol{x}_2,\boldsymbol{y}_1\boldsymbol{y}_2)$ is uniform over $\{\text{A},\text{C}\}$.
>
> $\big(\boldsymbol{i}=1,\ \boldsymbol{j}=2,\ \boldsymbol{z_i}=\{11\},\ \boldsymbol{z_j}=\{00,10\}\big)$ so $(\boldsymbol{x}_1\boldsymbol{x}_2,\boldsymbol{y}_1\boldsymbol{y}_2)$ is uniform over $\{\text{D},\text{E}\}$.
>
> $\big(\boldsymbol{i}=1,\ \boldsymbol{j}=2,\ \boldsymbol{z_i}=\{11\},\ \boldsymbol{z_j}=\{00,01\}\big)$ so $(\boldsymbol{x}_1\boldsymbol{x}_2,\boldsymbol{y}_1\boldsymbol{y}_2)$ is uniform over $\{\text{D},\text{F}\}$.

Letting $\boldsymbol{e}\in\{\text{AB},\text{AC},\text{DE},\text{DF}\}$ be the random variable indicating which of these cases holds, ((A.3)) says that $\mathbb{I}\big(\boldsymbol{\Pi}\ ;\ \boldsymbol{x_j},\boldsymbol{y_j}\ \mid\ E,\boldsymbol{e}\big)\leq o(1)$. Thus for each outcome e we have

$$\mathbb{I}\big(\boldsymbol{\Pi}\ ;\ \boldsymbol{x_j},\boldsymbol{y_j}\ \mid\ E,\,\boldsymbol{e}=e\big)\ \leq\ o(1).$$

Conditioned on $(E,\,\boldsymbol{e}=e)$, $(\boldsymbol{x_j},\boldsymbol{y_j})$ is uniform over two outcomes, so we can apply Lemma A.1 with $\boldsymbol{\Psi}\coloneqq(\boldsymbol{\Pi}\ \mid\ E,\,\boldsymbol{e}=e)$ and $\boldsymbol{\Lambda}\coloneqq(\boldsymbol{x_j},\boldsymbol{y_j}\ \mid\ E,\,\boldsymbol{e}=e)$.

Hence, if for $s\in\{\text{A},\text{B},\text{C},\text{D},\text{E},\text{F}\}$ we let E_s denote the event $\big((\boldsymbol{x}_1\boldsymbol{x}_2,\boldsymbol{y}_1\boldsymbol{y}_2)=s,\ \boldsymbol{z}_{-h^*}=z^*_{-h^*}\big)$ and we define the distribution $\boldsymbol{\Pi}_s\coloneqq(\boldsymbol{\Pi}\ \mid\ E_s)$, then (noting that $\boldsymbol{\Pi}_{\text{A}}$ is distributed identically to $\big(\boldsymbol{\Pi}\ \mid\ E,\,\boldsymbol{e}=\text{AB},\,\boldsymbol{x}_1\boldsymbol{y}_1=00\big)$ and similarly for the other possibilities) we have $H(\boldsymbol{\Pi}_{\text{A}},\boldsymbol{\Pi}_{\text{B}})$, $H(\boldsymbol{\Pi}_{\text{A}},\boldsymbol{\Pi}_{\text{C}})$, $H(\boldsymbol{\Pi}_{\text{D}},\boldsymbol{\Pi}_{\text{E}})$, $H(\boldsymbol{\Pi}_{\text{D}},\boldsymbol{\Pi}_{\text{F}})\leq o(1)$. Assuming that "close" means "within Hellinger distance $o(1)$" or equivalently "within statistical distance $o(1)$" (by

Lemma A.2), by the triangle inequality, $\mathbf{\Pi}_{\mathrm{A}}, \mathbf{\Pi}_{\mathrm{B}}, \mathbf{\Pi}_{\mathrm{C}}$ are all close and $\mathbf{\Pi}_{\mathrm{D}}, \mathbf{\Pi}_{\mathrm{E}}, \mathbf{\Pi}_{\mathrm{F}}$ are all close. In particular, the same holds for the distributions $\boldsymbol{v}_s := (\boldsymbol{v} \mid E_s)$ (equivalently, $\boldsymbol{v}_s$ is the marginal of the first c bits of $\mathbf{\Pi}_s$): $\boldsymbol{v}_{\mathrm{A}}, \boldsymbol{v}_{\mathrm{B}}, \boldsymbol{v}_{\mathrm{C}}$ are all close and $\boldsymbol{v}_{\mathrm{D}}, \boldsymbol{v}_{\mathrm{E}}, \boldsymbol{v}_{\mathrm{F}}$ are all close.

Note that $(\boldsymbol{x}, \boldsymbol{y} \mid E_s)$ is uniform over a rectangle consisting only of 0-inputs if $s \in \{\mathrm{A}, \mathrm{B}, \mathrm{C}\}$ and only of 1-inputs if $s \in \{\mathrm{D}, \mathrm{E}, \mathrm{F}\}$. Since for every leaf v, the event $\boldsymbol{v} = v$ consists of a rectangle in the domain of $\big((\boldsymbol{x}, \boldsymbol{r}_x), (\boldsymbol{y}, \boldsymbol{r}_y)\big)$, we have $\mathbb{P}[\boldsymbol{v}_{\mathrm{A}} = v] \cdot \mathbb{P}[\boldsymbol{v}_{\mathrm{D}} = v] = \mathbb{P}[\boldsymbol{v}_{\mathrm{B}} = v] \cdot \mathbb{P}[\boldsymbol{v}_{\mathrm{C}} = v]$. This implies that $H(\boldsymbol{v}_{\mathrm{A}}, \boldsymbol{v}_{\mathrm{D}}) = H(\boldsymbol{v}_{\mathrm{B}}, \boldsymbol{v}_{\mathrm{C}}) \leq o(1)$, and hence $\boldsymbol{v}_{\mathrm{A}}, \boldsymbol{v}_{\mathrm{B}}, \boldsymbol{v}_{\mathrm{C}}, \boldsymbol{v}_{\mathrm{D}}, \boldsymbol{v}_{\mathrm{E}}, \boldsymbol{v}_{\mathrm{F}}$ are all close.

Let $E_{s,v}$ denote the intersection of the event $\boldsymbol{v} = v$ with E_s. Let $\boldsymbol{w}_{s,v}$ denote the distribution $(\boldsymbol{w} \mid E_{s,v})$ assuming $\mathbb{P}[E_{s,v}] > 0$ (equivalently, assuming $\mathbb{P}[\boldsymbol{v}_s = v] > 0$).

Claim A.4. *There exists a leaf v^* such that the following all hold.*

- $\mathbb{P}[E_{s,v^*}] > 0$ *for all* $s \in \{\mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{D}, \mathrm{E}, \mathrm{F}\}$.
- $\mathbb{P}\big[o_{v^*}(\boldsymbol{\chi}) = \bot \mid E_{\mathrm{D},v^*}\big] < 1$.
- $\mathbb{P}\big[o_{v^*}(\boldsymbol{\chi}) = \bot \mid E_{\mathrm{A},v^*}\big] < 1$.
- $\Delta\big(\boldsymbol{w}_{\mathrm{B},v^*}, \boldsymbol{w}_{\mathrm{C},v^*}\big) < 1$.
- $\Delta\big(\boldsymbol{w}_{\mathrm{E},v^*}, \boldsymbol{w}_{\mathrm{F},v^*}\big) < 1$.

Proof. Since

$$\mathbb{E}_{v \sim \boldsymbol{v}_{\mathrm{D}}} \mathbb{P}\big[o_v(\boldsymbol{\chi}) = \bot \mid E_{\mathrm{D},v}\big] \;=\; \mathbb{P}\big[o_{\boldsymbol{v}}(\boldsymbol{\chi}) = \bot \mid E_{\mathrm{D}}\big] \;\leq\; 1/2 - \epsilon,$$

we have

(A.5)
$$\mathbb{P}_{v \sim \boldsymbol{v}_{\mathrm{D}}}\Big[\mathbb{P}\big[o_v(\boldsymbol{\chi}) = \bot \mid E_{\mathrm{D},v}\big] < 1 \;\text{ and }\; \mathbb{P}[E_{\mathrm{D},v}] > 0\Big] \;\geq\; 1/2 + \epsilon$$

by Markov's inequality. Similarly, ((A.5)) holds with A in place of D, and thus

(A.6)
$$\mathbb{P}_{v \sim \boldsymbol{v}_{\mathrm{D}}}\Big[\mathbb{P}\big[o_v(\boldsymbol{\chi}) = \bot \mid E_{\mathrm{A},v}\big] < 1 \;\text{ and }\; \mathbb{P}[E_{\mathrm{A},v}] > 0\Big] \;\geq\; 1/2 + \epsilon - o(1)$$

since $\Delta(\boldsymbol{v}_{\mathrm{A}}, \boldsymbol{v}_{\mathrm{D}}) \leq o(1)$. Next, we show that

(A.7)
$$\mathbb{P}_{v\sim\boldsymbol{v}_{\mathrm{D}}}\Big[\Delta\big(\boldsymbol{w}_{\mathrm{B},v}, \boldsymbol{w}_{\mathrm{C},v}\big) < 1 \ \text{ and } \ \mathbb{P}[E_{\mathrm{B},v}] > 0 \ \text{ and } \ \mathbb{P}[E_{\mathrm{C},v}] > 0\Big] \geq \ 1 - o(1)$$

holds. Similarly,

(A.8)
$$\mathbb{P}_{v\sim\boldsymbol{v}_{\mathrm{D}}}\Big[\Delta\big(\boldsymbol{w}_{\mathrm{E},v}, \boldsymbol{w}_{\mathrm{F},v}\big) < 1 \ \text{ and } \ \mathbb{P}[E_{\mathrm{E},v}] > 0 \ \text{ and } \ \mathbb{P}[E_{\mathrm{F},v}] > 0\Big] \geq \ 1 - o(1)$$

will hold. The claim then follows from ((A.5)), ((A.6)), ((A.7)), and ((A.8)) by a union bound over $v \sim \boldsymbol{v}_{\mathrm{D}}$. It remains to show ((A.7)). Let

$$V \coloneqq \Big\{v : \Delta\big(\boldsymbol{w}_{\mathrm{B},v}, \boldsymbol{w}_{\mathrm{C},v}\big) = 1 \ \text{ and } \ \mathbb{P}[E_{\mathrm{B},v}] > 0 \ \text{ and } \ \mathbb{P}[E_{\mathrm{C},v}] > 0\Big\}$$

and let $T \coloneqq \big\{vw \ : \ v \in V \text{ and } w \in \operatorname{supp}(\boldsymbol{w}_{\mathrm{B},v})\big\}$. Note that $\mathbb{P}[\boldsymbol{\Pi}_{\mathrm{C}} \in T] = 0$ since $\operatorname{supp}(\boldsymbol{w}_{\mathrm{B},v}) \cap \operatorname{supp}(\boldsymbol{w}_{\mathrm{C},v}) = \emptyset$ for each $v \in V$. Thus $\mathbb{P}[\boldsymbol{v}_{\mathrm{B}} \in V] = \mathbb{P}[\boldsymbol{\Pi}_{\mathrm{B}} \in T] \leq 0 + \Delta(\boldsymbol{\Pi}_{\mathrm{B}}, \boldsymbol{\Pi}_{\mathrm{C}}) \leq o(1)$. It follows that $\mathbb{P}[\boldsymbol{v}_{\mathrm{D}} \in V] \leq o(1) + \Delta(\boldsymbol{v}_{\mathrm{D}}, \boldsymbol{v}_{\mathrm{B}}) \leq o(1)$. We also have $\mathbb{P}_{v\sim\boldsymbol{v}_{\mathrm{D}}}\big[\mathbb{P}[E_{\mathrm{B},v}] = 0\big] \leq \Delta(\boldsymbol{v}_{\mathrm{B}}, \boldsymbol{v}_{\mathrm{D}}) \leq o(1)$ and $\mathbb{P}_{v\sim\boldsymbol{v}_{\mathrm{D}}}\big[\mathbb{P}[E_{\mathrm{C},v}] = 0\big] \leq \Delta(\boldsymbol{v}_{\mathrm{C}}, \boldsymbol{v}_{\mathrm{D}}) \leq o(1)$. Hence the left side of ((A.7)) is at least

$$1 - \mathbb{P}_{v\sim\boldsymbol{v}_{\mathrm{D}}}[v \in V] - \mathbb{P}_{v\sim\boldsymbol{v}_{\mathrm{D}}}\big[\mathbb{P}[E_{\mathrm{B},v}]=0\big] - \mathbb{P}_{v\sim\boldsymbol{v}_{\mathrm{D}}}\big[\mathbb{P}[E_{\mathrm{C},v}] = 0\big] \geq 1 - o(1).$$

□

By the correctness of Π, we have $\mathbb{P}\big[o_{v^*}(\boldsymbol{\chi}) = 0 \ \mid \ E_{\mathrm{A},v^*}\big] > 0$ and $\mathbb{P}\big[o_{v^*}(\boldsymbol{\chi}) = 1 \ \mid \ E_{\mathrm{D},v^*}\big] > 0$. Thus 0 and 1 are both possible outputs of o_{v^*}, and hence ⊥ is not a possible output of o_{v^*}; i.e., since there are only two possible inputs to o_{v^*}, namely 0 and 1, one of them must map to 0 and the other to 1 (and neither to ⊥). In what follows, note that E_s can be viewed as a subset of the domain of $\big((\boldsymbol{x}, \boldsymbol{r}_x), (\boldsymbol{y}, \boldsymbol{r}_y)\big)$.

First suppose $o_{v^*}(1) = 0$ and $o_{v^*}(0) = 1$. For all $s \in \{\mathrm{A}, \mathrm{B}, \mathrm{C}\}$, we actually have $\mathbb{P}\big[o_{v^*}(\boldsymbol{\chi}) = 0 \ \mid \ E_{s,v^*}\big] = 1$ and hence $\mathbb{P}\big[\boldsymbol{\chi} =$

$1 \mid E_{s,v^*}\big] = 1$ and hence $\mathbb{P}\big[\boldsymbol{w}_{s,v^*} \neq \varepsilon\big] = 1$. Since $\Delta\big(\boldsymbol{w}_{\mathrm{B},v^*}, \boldsymbol{w}_{\mathrm{C},v^*}\big) < 1$, this implies that there exists a $w^* \in \{0,1\}^k$ such that $\mathbb{P}\big[\boldsymbol{w}_{\mathrm{B},v^*} = w^*\big] > 0$ and $\mathbb{P}\big[\boldsymbol{w}_{\mathrm{C},v^*} = w^*\big] > 0$. Hence there exist a $\big((x, r_x), (y, r_y)\big) \in S_{v^*,w^*} \cap E_{\mathrm{B}}$ and a $\big((x', r'_x), (y', r'_y)\big) \in S_{v^*,w^*} \cap E_{\mathrm{C}}$. Since S_{v^*,w^*} is a rectangle, $\big((x, r_x), (y', r'_y)\big) \in S_{v^*,w^*}$ and hence Π outputs $o_{v^*}(1) = 0$. This contradicts the correctness since (x, y') is a 1-input (having $x \cap y' = \{1\}$ and lying in the D cell).

On the other hand, suppose $o_{v^*}(1) = 1$ and $o_{v^*}(0) = 0$. The argument is very similar: For all $s \in \{\mathrm{D}, \mathrm{E}, \mathrm{F}\}$, we actually have $\mathbb{P}\big[o_{v^*}(\boldsymbol{\chi}) = 1 \mid E_{s,v^*}\big] = 1$ and hence $\mathbb{P}\big[\boldsymbol{\chi} = 1 \mid E_{s,v^*}\big] = 1$ and hence $\mathbb{P}\big[\boldsymbol{w}_{s,v^*} \neq \varepsilon\big] = 1$. Since $\Delta\big(\boldsymbol{w}_{\mathrm{E},v^*}, \boldsymbol{w}_{\mathrm{F},v^*}\big) < 1$, this implies that there exists a $w^* \in \{0,1\}^k$ such that $\mathbb{P}\big[\boldsymbol{w}_{\mathrm{E},v^*} = w^*\big] > 0$ and $\mathbb{P}\big[\boldsymbol{w}_{\mathrm{F},v^*} = w^*\big] > 0$. Hence there exist a $\big((x, r_x), (y, r_y)\big) \in S_{v^*,w^*} \cap E_{\mathrm{E}}$ and a $\big((x', r'_x), (y', r'_y)\big) \in S_{v^*,w^*} \cap E_{\mathrm{F}}$. Since S_{v^*,w^*} is a rectangle, $\big((x, r_x), (y', r'_y)\big) \in S_{v^*,w^*}$ and hence Π outputs $o_{v^*}(1) = 1$. This contradicts the correctness since (x, y') is a 0-input (having $x \cap y' = \{1, 2\}$ and lying in the bottom-right cell in Figure A.1).

B. Appendix: Catalog of communication complexity classes

We now provide formal definitions of all the communication complexity classes considered in Section 3. If $\mathcal{C}$ is the name of a model and $F\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is a partial function, then we let $\mathcal{C}(F)$ denote the minimum cost of a correct protocol for F in model $\mathcal{C}$, and we also let $\mathcal{C}$ denote the class of all (families of) partial functions F with $\mathcal{C}(F) \leq \mathrm{poly}(\log n)$. We let $\mathsf{co}\mathcal{C}(F) \coloneqq \mathcal{C}(\neg F)$.

For example, $\mathsf{P}(F)$ is the minimum cost of a deterministic protocol for F, and P is the set of partial functions with $\mathrm{poly}(\log n)$-cost deterministic protocols. We group the remaining models into four categories (corresponding to the four subsections): the NP query hierarchy, bounded-error randomized models, models with postselection or unbounded error, and models with alternation.

In the definitions that follow, we always use Π to denote a protocol, F to denote an arbitrary partial function, and (x, y) to denote an arbitrary input in the domain of F (the models are worst case, so the correctness criteria always hold for all such (x, y)). All randomized models are assumed to have public randomness except when noted otherwise.

B.1. The NP query hierarchy.

Definition B.1. (NP)

Syntax: Π *is a collection of rectangles* $\big\{R_w \;:\; w \in \{0,1\}^k\big\}$, *and* Π *outputs* 1 *or* 0 *indicating whether* $(x,y) \in \bigcup_w R_w$.

Correctness: $\Pi(x,y) = F(x,y)$.

Cost: k.

Definition B.2. (US)

Syntax: Π *is a collection of rectangles* $\big\{R_w \;:\; w \in \{0,1\}^k\big\}$, *and* Π *outputs* 1 *or* 0 *indicating whether the number of* w*'s such that* $(x,y) \in R_w$ *is exactly one.*

Correctness: $\Pi(x,y) = F(x,y)$.

Cost: k.

Definition B.3. (DP)

Syntax: Π *is a pair of collections of rectangles,* $\big\{S_w \;:\; w \in \{0,1\}^k\big\}$ *and* $\big\{T_w \;:\; w \in \{0,1\}^k\big\}$, *and* Π *outputs* 1 *or* 0 *indicating whether* $(x,y) \in \bigcup_w S_w \smallsetminus \bigcup_w T_w$.

Correctness: $\Pi(x,y) = F(x,y)$.

Cost: k.

Definition B.4. ($\mathsf{P}_{\|}^{\mathsf{NP}[q]}$ *for constant* q)

Syntax: Π *is a deterministic protocol where for each leaf* v *with associated rectangle* R_v, *there are* q *associated collections of subrectangles* $\big\{S_{v,i,w} \subseteq R_v \;:\; w \in \{0,1\}^k\big\}$ $(i \in [q])$ *and an associated output function* $o_v \colon \{0,1\}^q \to \{0,1\}$ *that is applied to the indicators of whether* $(x,y) \in \bigcup_w S_{v,i,w}$ *for each* i.

Correctness: $\Pi(x,y) = F(x,y)$.

Cost: k+*the communication cost of the deterministic part.*

Definition B.5. (P^{NP})

Syntax: Π *is a protocol tree where each internal node* v *is labeled with either (i) a* 1*-bit function of Alice's or of Bob's input in the usual way, or (ii) an "*NP *query" consisting of a collection of rectangles* $\{S_{v,w} : w \in \{0,1\}^{k_v}\}$*, where the indicator of whether* $(x,y) \in \bigcup_w S_{v,w}$ *determines which child to descend to in the protocol tree. The output of* Π *is determined by the leaf reached.*

Correctness: $\Pi(x,y) = F(x,y)$.

Cost: The maximum over all root-to-leaf paths of the following: the length of the path plus the sum of k_v *over all type-(ii) nodes* v *on the path.*

Definition B.6. ($\mathsf{P}^{\mathsf{NP}[q]}$ *for constant* q)

Syntax: Π *is a* P^{NP}*-type protocol where there are at most* q NP *queries on each root-to-leaf path.*

Correctness: $\Pi(x,y) = F(x,y)$.

Cost: Same as Definition B.5. Affecting the cost only by a constant factor, it can be assumed that all NP *queries happen at the end and all have the same witness length* k_v.

Definition B.7. ($\mathsf{P}_{\|}^{\mathsf{NP}}$)

Syntax: Π *is a* P^{NP}*-type protocol where the result of each* NP *query is not revealed until the last query on any path down the tree. Thus, each type-(ii) node has* 1 *child if it has a type-(ii) descendant, and has* 2^q *children if it has no type-(ii) descendants (where* q *is the number of type-(ii) nodes on that path). Hence without loss of generality, all the* NP *queries are consecutive.*

Correctness: $\Pi(x,y) = F(x,y)$.

Cost: Same as Definition B.5.

B.2. Bounded-error randomized models.

Definition B.8. (ZPP)

Syntax: Π *is a distribution over deterministic protocols outputting values in* $\{0,1,\perp\}$.

Correctness: $\mathbb{P}\big[\Pi(x,y) \in \{F(x,y),\perp\}\big] = 1$ *and* $\mathbb{P}\big[\Pi(x,y) = F(x,y)\big] \geq 3/4$.

Cost: The maximum communication cost of any constituent deterministic protocol.

Definition B.9. (RP)

Syntax: Π *is a distribution over deterministic protocols outputting values in* $\{0,1\}$.

Correctness: If $F(x,y)=1$ *then* $\mathbb{P}[\Pi(x,y)=1] \geq 1/2$.
If $F(x,y)=0$ *then* $\mathbb{P}[\Pi(x,y)=0] = 1$.

Cost: The maximum communication cost of any constituent deterministic protocol.

Definition B.10. (BPP)

Syntax: Π *is a distribution over deterministic protocols outputting values in* $\{0,1\}$.

Correctness: $\mathbb{P}\big[\Pi(x,y) = F(x,y)\big] \geq 3/4$.

Cost: The maximum communication cost of any constituent deterministic protocol.

Definition B.11. (MA)

Syntax: Π *is a distribution over deterministic protocols that take an additional input* $w \in \{0,1\}^k$, *which is visible to both Alice and Bob.*

Correctness: Completeness: if $F(x,y)=1$ *then*
$\exists w : \mathbb{P}\big[\Pi(x,y,w)=1\big] \geq 3/4$.
Soundness: if $F(x,y)=0$ *then*
$\forall w : \mathbb{P}\big[\Pi(x,y,w)=0\big] \geq 3/4$.

Cost: k*+the maximum communication cost of any constituent deterministic protocol.*

Definition B.12. (AM)

Syntax: Π *is a distribution over nondeterministic* (NP*-type) protocols.*

Correctness: $\mathbb{P}\big[\Pi(x,y) = F(x,y)\big] \geq 3/4$.

Cost: *The maximum cost of any constituent nondeterministic protocol.*

Definition B.13. ($\mathsf{ZPP}_{\|}^{\mathsf{NP}[q]}$ *for constant* q)

Syntax: Π *is a distribution over* $\mathsf{P}_{\|}^{\mathsf{NP}[q]}$*-type protocols outputting values in* $\{0,1,\perp\}$.

Correctness: *Same as Definition B.8.*

Cost: *The maximum cost of any constituent* $\mathsf{P}_{\|}^{\mathsf{NP}[q]}$*-type protocol.*

Definition B.14. ($\mathsf{ZPP}_{\|}^{\mathsf{NP}}$)

Syntax: Π *is a distribution over* $\mathsf{P}_{\|}^{\mathsf{NP}}$*-type protocols outputting values in* $\{0,1,\perp\}$.

Correctness: *Same as Definition B.8.*

Cost: *The maximum cost of any constituent* $\mathsf{P}_{\|}^{\mathsf{NP}}$*-type protocol.*

Definition B.15. ($\mathsf{ZPP}^{\mathsf{NP}[q]}$ *for constant* q)

Syntax: Π *is a distribution over* $\mathsf{P}^{\mathsf{NP}[q]}$*-type protocols outputting values in* $\{0,1,\perp\}$.

Correctness: *Same as Definition B.8.*

Cost: *The maximum cost of any constituent* $\mathsf{P}^{\mathsf{NP}[q]}$*-type protocol.*

Definition B.16. ($\mathsf{ZPP}^{\mathsf{NP}}$)

Syntax: Π *is a distribution over* P^{NP}*-type protocols outputting values in* $\{0,1,\perp\}$.

Correctness: *Same as Definition B.8.*

Cost: *The maximum cost of any constituent* P^{NP}*-type protocol.*

B.3. Models with postselection or unbounded error.

Although SBP is not defined in terms of postselection or unbounded error, we include the definition here since it provides a nice segue.

Definition B.17. (SBP)

Syntax: Π *has public randomness uniformly distributed over* $\{0,1\}^k$*, with each outcome having an associated deterministic protocol outputting values in* $\{0,1\}$.

Correctness: $\min_{(x,y)\in F^{-1}(1)} \mathbb{P}[\Pi(x,y)=1] > 2 \cdot \max_{(x,y)\in F^{-1}(0)} \mathbb{P}[\Pi(x,y) = 1]$.

Cost: k*+the maximum communication cost of any constituent deterministic protocol.*

SBP communication complexity is known to be equivalent to the corruption bound Göös & Watson (2016).

Definition B.18. ($\mathsf{PostBPP}$)

Syntax: Π *has public randomness uniformly distributed over* $\{0,1\}^k$*, with each outcome having an associated deterministic protocol outputting values in* $\{0,1,\bot\}$.

Correctness: $\mathbb{P}\big[\Pi(x,y) = F(x,y)\big] > 2 \cdot \mathbb{P}\big[\Pi(x,y) = 1 - F(x,y)\big]$.

Cost: k*+the maximum communication cost of any constituent deterministic protocol.*

$\mathsf{PostBPP}$ communication complexity is known to be equivalent to the extended discrepancy bound Gavinsky & Lovett (2014).

Definition B.19. ($\mathsf{UPostBPP}_\square$)

Syntax: Same as Definition B.18, except the public randomness is arbitrarily distributed over $\{0,1\}^k$.

Correctness: Same as Definition B.18.

Cost: Same as Definition B.18.

Definition B.20. ($\mathsf{UPostBPP}$)

Syntax: Π *is a private-randomness protocol outputting values in* $\{0, 1, \bot\}$.

Correctness: *Same as Definition B.18.*

Cost: *The communication cost of the underlying deterministic protocol.*

We have $\mathsf{UPostBPP}(F) \leq \mathsf{UPostBPP}_{\Box}(F) \leq \mathsf{PostBPP}(F)$ for all F, and hence $\mathsf{PostBPP} \subseteq \mathsf{UPostBPP}_{\Box} \subseteq \mathsf{UPostBPP}$.

Observation B.21. *Without loss of generality, in a* $\mathsf{PostBPP}$ *or* $\mathsf{UPostBPP}_{\Box}$ *protocol, each of the constituent deterministic protocols consists of a single rectangle (with fixed output* 0 *or* 1 *on inputs in the rectangle, and output* $\bot$ *on inputs outside the rectangle).*

Proof. We may modify a $\mathsf{PostBPP}$ or $\mathsf{UPostBPP}_{\Box}$ protocol so that after choosing the original public randomness, it then picks a uniformly random leaf rectangle (of which we assume there are exactly 2^c) from the associated deterministic protocol, outputs the same value on inputs in the rectangle, and outputs $\bot$ on all inputs outside the rectangle. The correctness is unaffected. The number of random bits becomes $k+c$, and the communication cost becomes 2, so the overall cost becomes $k+c+2$. If after the transformation, any rectangle has label $\bot$, we can instead assume it is an empty rectangle with non-$\bot$ label. $\Box$

Observation B.22.
$\mathsf{UPostBPP}(F) \in \min\big(\log \operatorname{rank}_+(M^0) + \log \operatorname{rank}_+(M^1)\big) \pm O(1)$ *where* rank_+ *denotes nonnegative rank, and the minimum is over nonnegative real matrices* M^0, M^1 *(indexed by inputs) such that for each* $(x, y) \in F^{-1}$, $M^{F(x,y)}_{x,y} > 2 \cdot M^{1-F(x,y)}_{x,y}$.

The argument for Observation B.22 is essentially the same as the argument from Paturi & Simon (1986) that UPP complexity is equivalent to log of sign-rank (Observation B.27 below).

Definition B.23. (PP)

Syntax: Π *has public randomness uniformly distributed over* $\{0,1\}^k$*, with each outcome having an associated deterministic protocol outputting values in* $\{0,1\}$.

Correctness: $\mathbb{P}\big[\Pi(x,y) = F(x,y)\big] > 1/2$.

Cost: k*+the maximum communication cost of any constituent deterministic protocol.*

PP communication complexity is known to be equivalent to the discrepancy bound Klauck (2007).

Definition B.24. ($\mathsf{UPP}_\square$)

Syntax: Same as Definition B.23, except the public randomness is arbitrarily distributed over $\{0,1\}^k$.

Correctness: Same as Definition B.23.

Cost: Same as Definition B.23.

Definition B.25. (UPP)

Syntax: Π *is a private-randomness protocol outputting values in* $\{0,1\}$.

Correctness: Same as Definition B.23.

Cost: The communication cost of the underlying deterministic protocol.

We have $\mathsf{UPP}(F) \leq \mathsf{UPP}_\square(F) \leq \mathsf{PP}(F)$ for all F, and hence $\mathsf{PP} \subseteq \mathsf{UPP}_\square \subseteq \mathsf{UPP}$.

Observation B.26. *Without loss of generality, in a* PP *or* $\mathsf{UPP}_\square$ *protocol, each of the constituent deterministic protocols consists of a single rectangle (with output only depending on whether the input is in the rectangle).*

Proof. We may modify a PP or $\mathsf{UPP}_\square$ protocol so that after choosing the original public randomness, it then picks a uniformly random leaf rectangle (of which we assume there are exactly 2^c) from the associated deterministic protocol, outputs the same value on inputs in the rectangle, and flips a coin to determine the output on all inputs outside the rectangle. The correctness is unaffected. The number of random bits becomes $k+c+1$, and the communication cost becomes 2, so the overall cost becomes $k+c+3$. □

Observation B.27.
$\mathsf{UPP}(F) \in \min\bigl(\log \operatorname{rank}_+(M^0) + \log \operatorname{rank}_+(M^1)\bigr) \pm O(1)$ *where* rank_+ *denotes nonnegative rank, and the minimum is over nonnegative real matrices* M^0, M^1 *(indexed by inputs) such that for each* $(x,y) \in F^{-1}$, $M^{F(x,y)}_{x,y} > M^{1-F(x,y)}_{x,y}$.

B.4. Models with alternation.

Definition B.28. ($\mathsf{S_2P}$)

Syntax: Π *is a matrix with rows indexed by* $w^0 \in \{0,1\}^k$ *and columns indexed by* $w^1 \in \{0,1\}^k$, *with each entry* (w^0, w^1) *having an associated deterministic protocol* Π_{w^0,w^1} *outputting values in* $\{0,1\}$.

Correctness: *If* $F(x,y) = 1$ *then* $\exists w^1\ \forall w^0 : \Pi_{w^0,w^1}(x,y) = 1$.
If $F(x,y) = 0$ *then* $\exists w^0\ \forall w^1 : \Pi_{w^0,w^1}(x,y) = 0$.

Cost: $k+$*the maximum communication cost of any constituent deterministic protocol.*

Definition B.29. ($\mathsf{\Sigma_\ell P}$ *for constant* ℓ)

Syntax: Π *is a complete* 2^k*-ary tree of depth* ℓ *(root-to-leaf paths have* ℓ *edges) representing a formula with alternating layers of* Or *and* And *gates, with an* Or *gate at the root, and where each leaf is the indicator for a rectangle (if* ℓ *is odd) or the complement of a rectangle (if* ℓ *is even).*

Correctness: $\Pi(x,y) = F(x,y)$.

Cost: k.

Definition B.30. ($\mathsf{\Pi_\ell P}$ *for constant* ℓ)

Syntax: Π *is a complete* 2^k*-ary tree of depth* ℓ *(root-to-leaf paths have* ℓ *edges) representing a formula with alternating layers of* And *and* Or *gates, with an* And *gate at the root, and where each leaf is the indicator for a rectangle (if* ℓ *is even) or the complement of a rectangle (if* ℓ *is odd).*

Correctness: $\Pi(x,y) = F(x,y)$.

Cost: k.

The class PH is defined as $\bigcup_\ell \Sigma_\ell\mathsf{P} = \bigcup_\ell \Pi_\ell\mathsf{P}$ (where the union is over constants ℓ).

Definition B.31. (PSPACE)

Syntax: Π *is a formula where each leaf is the indicator for a rectangle.*

Correctness: $\Pi(x, y) = F(x, y)$.

Cost: *The* log *of the size of the formula.*

Although $\oplus\mathsf{P}$ is not defined in terms of alternation, we include the definition in this subsection since in a sense, it is at least as powerful as alternation: $\mathsf{PH} \subseteq \mathsf{BP} \cdot \oplus\mathsf{P}$ Toda (1991).

Definition B.32. ($\oplus\mathsf{P}$)

Syntax: Π *is a collection of rectangles* $\{R_w \ : \ w \in \{0,1\}^k\}$, *and* Π *outputs* 1 *or* 0 *indicating whether the number of* w*'s such that* $(x, y) \in R_w$ *is odd.*

Correctness: $\Pi(x, y) = F(x, y)$.

Cost: k.

Observation B.33. $\oplus\mathsf{P}(F) \in \log \operatorname{rank}(F) \pm O(1)$ *where the rank is over* $GF(2)$.

Acknowledgements

We thank anonymous referees for helpful remarks. This work was supported by NSERC funding and by NSF grant CCF-1657377.

References

Scott Aaronson (2005). Quantum Computing, Postselection, and Probabilistic Polynomial-Time. *Proceedings of the Royal Society A* **461**(2063), 3473–3482.

Scott Aaronson & Avi Wigderson (2009). Algebrization: A New Barrier in Complexity Theory. *ACM Transactions on Computation Theory* **1**(1).

László Babai, Peter Frankl & Janos Simon (1986). Complexity Classes in Communication Complexity Theory. In *Proceedings of the*

27th Symposium on Foundations of Computer Science (FOCS), 337–347. IEEE.

Ziv Bar-Yossef, T.S. Jayram, Ravi Kumar & D. Sivakumar (2004). An Information Statistics Approach to Data Stream and Communication Complexity. *Journal of Computer and System Sciences* **68**(4), 702–732.

Richard Beigel (1991). Bounded Queries to SAT and the Boolean Hierarchy. *Theoretical Computer Science* **84**(2), 199–223.

Andreas Blass & Yuri Gurevich (1982). On the Unique Satisfiability Problem. *Information and Control* **55**(1–3), 80–88.

Elmar Böhler, Christian Glasser & Daniel Meister (2006). Error-Bounded Probabilistic Computations Between MA and AM. *Journal of Computer and System Sciences* **72**(6), 1043–1076.

Adam Bouland, Lijie Chen, Dhiraj Holden, Justin Thaler & Prashant Vasudevan (2017). On the Power of Statistical Zero Knowledge. In *Proceedings of the 58th Symposium on Foundations of Computer Science*, 708–719. IEEE.

Harry Buhrman, Nikolai Vereshchagin & Ronald de Wolf (2007). On Computation and Communication with Small Bias. In *Proceedings of the 22nd Conference on Computational Complexity (CCC)*, 24–32. IEEE.

Jin-Yi Cai (2007). $\mathrm{S_2P} \subseteq \mathrm{ZPP^{NP}}$. *Journal of Computer and System Sciences* **73**(1), 25–35.

Jin-Yi Cai & Venkatesan Chakaravarthy (2006). On Zero Error Algorithms Having Oracle Access to One Query. *Journal of Combinatorial Optimization* **11**(2), 189–202.

Ran Canetti (1996). More on BPP and the Polynomial-Time Hierarchy. *Information Processing Letters* **57**(5), 237–241.

Amit Chakrabarti, Graham Cormode, Navin Goyal & Justin Thaler (2014a). Annotations for Sparse Data Streams. In *Proceedings of the 25th Symposium on Discrete Algorithms (SODA)*, 687–706. ACM-SIAM.

Amit Chakrabarti, Graham Cormode, Andrew McGregor & Justin Thaler (2014b). Annotations in Data Streams. *ACM Transactions on Algorithms* **11**(1), 7.

Amit Chakrabarti, Graham Cormode, Andrew McGregor, Justin Thaler & Suresh Venkatasubramanian (2015). Verifiable Stream Computation and Arthur–Merlin Communication. In *Proceedings of the 30th Computational Complexity Conference (CCC)*, 217–243. Schloss Dagstuhl.

Richard Chang, Jim Kadin & Pankaj Rohatgi (1995). On Unique Satisfiability and the Threshold Behavior of Randomized Reductions. *Journal of Computer and System Sciences* **50**(3), 359–373.

Richard Chang & Suresh Purini (2008). Amplifying $\mathrm{ZPP}^{\mathrm{SAT}[1]}$ and the Two Queries Problem. In *Proceedings of the 23rd Conference on Computational Complexity (CCC)*, 41–52. IEEE.

Arkadev Chattopadhyay & Nikhil Mande (2017). Weights at the Bottom Matter When the Top is Heavy. Technical Report TR17-083, Electronic Colloquium on Computational Complexity (ECCC). URL http://eccc.weizmann.ac.il/report/2017/083/.

Carsten Damm, Matthias Krause, Christoph Meinel & Stephan Waack (2004). On Relations Between Counting Communication Complexity Classes. *Journal of Computer and System Sciences* **69**(2), 259–280.

Lila Fontes, Rahul Jain, Iordanis Kerenidis, Sophie Laplante, Mathieu Laurière & Jérémie Roland (2016). Relative Discrepancy Does Not Separate Information and Communication Complexity. *ACM Transactions on Computation Theory* **9**(1), 4:1–4:15.

Jürgen Forster (2002). A Linear Lower Bound on the Unbounded Error Probabilistic Communication Complexity. *Journal of Computer and System Sciences* **65**(4), 612–625.

Lance Fortnow, Russell Impagliazzo, Valentine Kabanets & Christopher Umans (2008). On the Complexity of Succinct Zero-Sum Games. *Computational Complexity* **17**(3), 353–376.

Anat Ganor, Gillat Kol & Ran Raz (2016). Exponential Separation of Information and Communication for Boolean Functions. *Journal of the ACM* **63**(5), 46:1–46:31.

Dmitry Gavinsky & Shachar Lovett (2014). 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)*, 514–524. Springer.

Dmitry Gavinsky & Alexander Sherstov (2010). A Separation of NP and coNP in Multiparty Communication Complexity. *Theory of Computing* **6**(1), 227–245.

Oded Goldreich & David Zuckerman (2011). Another Proof That BPP $\subseteq$ PH (and More). In *Studies in Complexity and Cryptography*, 40–53. Springer.

Shafi Goldwasser & Michael Sipser (1986). Private Coins versus Public Coins in Interactive Proof Systems. In *Proceedings of the 18th Symposium on Theory of Computing (STOC)*, 59–68. ACM.

Mika Göös, Pritish Kamath, Toniann Pitassi & Thomas Watson (2017). Query-to-Communication Lifting for P^{NP}. In *Proceedings of the 32nd Computational Complexity Conference (CCC)*, 12:1–12:16. Schloss Dagstuhl.

Mika Göös, Shachar Lovett, Raghu Meka, Thomas Watson & David Zuckerman (2016a). Rectangles Are Nonnegative Juntas. *SIAM Journal on Computing* **45**(5), 1835–1869.

Mika Göös, Toniann Pitassi & Thomas Watson (2016b). Zero-Information Protocols and Unambiguity in Arthur–Merlin Communication. *Algorithmica* **76**(3), 684–719. Special issue on information complexity and applications.

Mika Göös & Thomas Watson (2016). Communication Complexity of Set-Disjointness for All Probabilities. *Theory of Computing* **12**(1), 1–23. Special issue for selected papers from APPROX–RANDOM 2014.

Vince Grolmusz & Gábor Tardos (2003). A Note on Non-Deterministic Communication Complexity with Few Witnesses. *Theory of Computing Systems* **36**(4), 387–391.

Tom Gur & Ran Raz (2015). Arthur–Merlin Streaming Complexity. *Information and Computation* **243**, 145–165.

Tom Gur & Ron Rothblum (2015). Non-Interactive Proofs of Proximity. In *Proceedings of the 6th Innovations in Theoretical Computer Science Conference (ITCS)*, 133–142. ACM.

Bernd Halstenberg & Rüdiger Reischuk (1990). Relations Between Communication Complexity Classes. *Journal of Computer and System Sciences* **41**(3), 402–429.

Yenjo Han, Lane Hemaspaandra & Thomas Thierauf (1997). Threshold Computation and Cryptographic Security. *SIAM Journal on Computing* **26**(1), 59–78.

Russell Impagliazzo & Ryan Williams (2010). Communication Complexity with Synchronized Clocks. In *Proceedings of the 25th Conference on Computational Complexity (CCC)*, 259–269. IEEE.

T.S. Jayram, Ravi Kumar & D. Sivakumar (2003). Two Applications of Information Complexity. In *Proceedings of the 35th Symposium on Theory of Computing (STOC)*, 673–682. ACM.

Stasys Jukna (2006). On Graph Complexity. *Combinatorics, Probability, & Computing* **15**(6), 855–876.

Stasys Jukna (2012). *Boolean Function Complexity: Advances and Frontiers*, volume 27 of *Algorithms and Combinatorics*. Springer.

Volker Kaibel & Stefan Weltge (2015). A Short Proof that the Extension Complexity of the Correlation Polytope Grows Exponentially. *Discrete & Computational Geometry* **53**(2), 397–401.

Mauricio Karchmer, Ilan Newman, Michael Saks & Avi Wigderson (1994). Non-Deterministic Communication Complexity with Few Witnesses. *Journal of Computer and System Sciences* **49**(2), 247–257.

Hartmut Klauck (2003). Rectangle Size Bounds and Threshold Covers in Communication Complexity. In *Proceedings of the 18th Conference on Computational Complexity (CCC)*, 118–134. IEEE.

Hartmut Klauck (2007). Lower Bounds for Quantum Communication Complexity. *SIAM Journal on Computing* **37**(1), 20–46.

Hartmut Klauck (2010). A Strong Direct Product Theorem for Disjointness. In *Proceedings of the 42nd Symposium on Theory of Computing (STOC)*, 77–86. ACM.

Hartmut Klauck (2011). On Arthur Merlin Games in Communication Complexity. In *Proceedings of the 26th Conference on Computational Complexity (CCC)*, 189–199. IEEE.

Hartmut Klauck & Ved Prakash (2013). Streaming Computations with a Loquacious Prover. In *Proceedings of the 4th Innovations in Theoretical Computer Science Conference (ITCS)*, 305–320. ACM.

Hartmut Klauck & Ved Prakash (2014). An Improved Interactive Streaming Algorithm for the Distinct Elements Problem. In *Proceedings of the 41st International Colloquium on Automata, Languages, and Programming (ICALP)*, 919–930. Springer.

Eyal Kushilevitz & Noam Nisan (1997). *Communication Complexity.* Cambridge University Press.

Tak Wah Lam & Walter Ruzzo (1992). Results on Communication Complexity Classes. *Journal of Computer and System Sciences* **44**(2), 324–342.

Jianhua Lin (1991). Divergence Measures Based on the Shannon Entropy. *IEEE Transactions on Information Theory* **37**(1), 145–151. ISSN 0018-9448.

Nathan Linial & Adi Shraibman (2009). Learning Complexity vs Communication Complexity. *Combinatorics, Probability, & Computing* **18**(1–2), 227–245.

Satyanarayana Lokam (2001). Spectral Methods for Matrix Rigidity with Applications to Size-Depth Trade-offs and Communication Complexity. *Journal of Computer and System Sciences* **63**(3), 449–473.

Satyanarayana Lokam (2009). Complexity Lower Bounds using Linear Algebra. *Foundations and Trends in Theoretical Computer Science* **4**(1–2), 1–155.

Ilan Newman (1991). Private vs. Common Random Bits in Communication Complexity. *Information Processing Letters* **39**(2), 67–71.

NOAM NISAN & AVI WIGDERSON (1994). Hardness vs. Randomness. *Journal of Computer and System Sciences* **49**(2), 149–167.

RYAN O'DONNELL & A. C. CEM SAY (2016). The Weakness of CTC Qubits and the Power of Approximate Counting. Technical Report TR16-147, Electronic Colloquium on Computational Complexity (ECCC). URL http://eccc.hpi-web.de/report/2016/147/.

CHRISTOS PAPADIMITRIOU & MIHALIS YANNAKAKIS (1984). The Complexity of Facets (and Some Facets of Complexity). *Journal of Computer and System Sciences* **28**(2), 244–259.

PERIKLIS PAPAKONSTANTINOU, DOMINIK SCHEDER & HAO SONG (2014). Overlays and Limited Memory Communication. In *Proceedings of the 29th Conference on Computational Complexity (CCC)*, 298–308. IEEE.

RAMAMOHAN PATURI & JANOS SIMON (1986). Probabilistic Communication Complexity. *Journal of Computer and System Sciences* **33**(1), 106–123.

PAVEL PUDLÁK, VOJTECH RÖDL & PETR SAVICKÝ (1988). Graph Complexity. *Acta Informatica* **25**(5), 515–535.

RAN RAZ & AMIR SHPILKA (2004). On the Power of Quantum Proofs. In *Proceedings of the 19th Conference on Computational Complexity (CCC)*, 260–274. IEEE.

ALEXANDER RAZBOROV (1989). On Rigid Matrices. Technical report, Steklov Mathematical Institute. In Russian.

ALEXANDER RAZBOROV (1992). On the Distributional Complexity of Disjointness. *Theoretical Computer Science* **106**(2), 385–390.

ALEXANDER RAZBOROV & ALEXANDER SHERSTOV (2010). The Sign-Rank of AC^0. *SIAM Journal on Computing* **39**(5), 1833–1855.

ALEXANDER RUSSELL & RAVI SUNDARAM (1998). Symmetric Alternation Captures BPP. *Computational Complexity* **7**(2), 152–162.

ALEXANDER SHERSTOV (2008). Halfspace Matrices. *Computational Complexity* **17**(2), 149–178.

Alexander Sherstov (2011a). The Pattern Matrix Method. *SIAM Journal on Computing* **40**(6), 1969–2000.

Alexander Sherstov (2011b). The Unbounded-Error Communication Complexity of Symmetric Functions. *Combinatorica* **31**(5), 583–614.

Seinosuke Toda (1991). PP is as Hard as the Polynomial-Time Hierarchy. *SIAM Journal on Computing* **20**(5), 865–877.

Rahul Tripathi (2010). The 1-Versus-2 Queries Problem Revisited. *Theory of Computing Systems* **46**(2), 193–221.

Leslie Valiant (1977). Graph-Theoretic Arguments in Low-Level Complexity. In *Proceedings of the 6th Symposium on Mathematical Foundations of Computer Science (MFCS)*, 162–176. Springer.

Leslie Valiant & Vijay Vazirani (1986). NP is as Easy as Detecting Unique Solutions. *Theoretical Computer Science* **47**(3), 85–93.

Nikolai Vereshchagin (1995). Lower Bounds for Perceptrons Solving some Separation Problems and Oracle Separation of AM from PP. In *Proceedings of the 3rd Israel Symposium on Theory of Computing and Systems (ISTCS)*, 46–51. IEEE.

Nikolai Vereshchagin (1999). Relativizability in Complexity Theory. In *Provability, Complexity, Grammars*, volume 192 of *AMS Translations, Series 2*, 87–172. American Mathematical Society.

Henning Wunderlich (2012). On a Theorem of Razborov. *Computational Complexity* **21**(3), 431–477.

Manuscript received 2 May 2016

Mika Göös
Harvard University
Cambridge
USA
mika@seas.harvard.edu

Toniann Pitassi
University of Toronto
Toronto
Canada
toni@cs.toronto.edu

Thomas Watson
University of Memphis
Memphis
USA
Thomas.Watson@memphis.edu