A Strong Direct Product Theorem for Corruption and the Multiparty NOF Communication Complexity of Disjointness

Paul Beame[*]
University of Washington
Seattle, WA 98195-2350
beame@cs.washington.edu

Toniann Pitassi[†]
University of Toronto
Toronto, ON M5S 1A4
toni@cs.toronto.edu

Nathan Segerlind[‡]
University of Washington
Seattle, WA 98195-2350
nsegerli@cs.washington.edu

Avi Wigderson[§]
Institute for Advanced Study
Princeton, NJ
avi@ias.edu

July 5, 2006

Abstract

We prove that two-party randomized communication complexity satisfies a strong direct product property, so long as the communication lower bound is proved by a "corruption" or "one-sided discrepancy" method over a rectangular distribution. We use this to prove new $n^{\Omega(1)}$ lower bounds for 3-player number-on-the-forehead protocols in which the first player speaks once and then the other two players proceed arbitrarily. Using other techniques, we also establish an $\Omega(n^{1/(k-1)}/(k-1))$ lower bound for k-player randomized number-on-the-forehead protocols for the disjointness function in which all messages are broadcast simultaneously. A simple corollary of this is that general randomized number-on-the-forehead protocols require $\Omega(\log n/(k-1))$ bits of communication to compute the disjointness function.

1 Introduction

1.1 Number-on-the-forehead communication protocols

A fundamental problem in communication complexity is understanding the amount of communication necessary to compute the two-player disjointness function: Alice and Bob are each given a subset of $\{1, \ldots n\}$ and they want to determine whether or not they share a common element [2, 19, 31, 30]. A natural extension of two-party disjointness is k-party disjointness. In this set-up, there are k players, with sets $x_1, \ldots, x_k \subseteq \{1, \ldots n\}$, and the players want to determine whether or not the sets share a common element. To this end, the players exchange bits, and possibly make use of a shared source of randomness. They

[*] Supported by NSF grants CCR-0098066 and ITR-0219468

[†] Supported by an Ontario Premiere's Research Excellence Award, an NSERC grant, and the Institute for Advanced Study. Research done while at the Institute for Advanced Study.

[‡] Supported by NSF Postdoctoral Fellowship DMS-0303258. Research partially done while at the Institute for Advanced Study.

[§] Partially supported by NSF grant CCR-0324906

wish to compute the correct answer, or get the answer correct with probability at least two-thirds, while minimizing the number of bits exchanged.

What makes the multi-player problem especially interesting are the ways in which the players can share partial information about the inputs. We consider the "number-on-the-forehead" (NOF) model [13] in which the i'th player can see every input x_j for $j \neq i$. Metaphorically, it is as if the input x_i is on the forehead of player i. Contrast this with the well-studied "number-in-the-hand" (NIH) model, in which player i sees input x_i and no other inputs. Notice that in the number-in-the-hand model, the players share no information, whereas in the number-on-the-forehead model the players share a large amount of information. Disjointness has been studied extensively in the number-in-hand model largely because randomized lower bounds in this model provide lower bounds on the space complexity of randomized streaming algorithms that approximately compute frequency moments of a data set [1]. While the communication complexity of disjointness is almost completely characterized for the number-in-the-hand model [1, 32, 6, 7, 11], it is almost entirely open for the number-on-the-forehead model.

The number-on-the-forehead communication model is useful in theoretical computer science because phenomena such as circuits, branching programs, and propositional proofs can be transformed into number-on-the-forehead communication protocols. For this reason, establishing large enough communication lower bounds is a time-honored method for establishing lower bounds in other computational models. Most famously, linear lower bounds for $k = n^\epsilon$ players for any explicit function would yield explicit superpolynomial lower bounds for ACC circuits. (We emphasize that such bounds are not yet known, and how to establish communication bounds for a super-logarithmic number of players is probably the central question in the study of number-on-the-forehead protocols.) The communication complexity of the set-disjointness function also has interesting consequences. The first three authors of this paper show in [8] work that $\omega(\log^4 n)$ lower bounds for the k-party randomized number-on-the-forehead communication complexity of disjointness imply proof size lower bounds for a family of proof systems known as tree-like, degree $k-1$ threshold systems. Proving proof size lower bounds for these systems is a major open problem in propositional proof complexity. Such proof systems are quite powerful, and include the tree forms of systems such as the Chvátal-Gomory Cutting Planes proof system, and the Lovász-Schrijver proof systems. In [8], it also is shown that lower bounds of the form $\omega(\log^2 n (\log\log n)^2)$ for randomized three-party number-on-the-forehead communication of disjointness imply superpolynomial size lower bounds for Lovász-Schrijver proofs with polynomially-bounded coefficients.

Another motivation for the study of disjointness in the number-on-the-forehead model is to understand the power of non-determinism in this concrete computational model. Large enough communication lower bounds for disjointness imply a better separation between nondeterministic and deterministic (or randomized) multiparty number-on-the-forehead communication complexity[1] than the best currently known separation, which is barely super-constant [24].

With the exception of one barely-super-constant bound [13], known lower bounds for number-on-the-forehead communication complexity for more than two parties use the discrepancy method [5, 14, 29] in which it is shown that the function is nearly balanced on all large cylinder intersections. The discrepancy method completely fails when trying to prove communication lower bounds for disjointness under any distribution that gives even modest weight to intersecting inputs. This is because the disjointness function is constant on some very large cylinder intersections. Progress here seems to require a new kind of argument.

Prior to our work, little was known about the multi-player number-on-the-forehead communication complexity of the disjointness function. For two-party randomized protocols, it was known that the disjointness

[1] In the preliminary version of this paper [9] we claimed that, by extending the arguments in [2], the disjointness problem can be shown to be complete for the class k-NP^{cc}, the multiparty analogue of NP^{cc}. This claim does not seem to be correct.

function requires $\Theta(n)$ bits of communication to compute with constant error [19, 31]. For three or more players, the best protocol known for the k-party number-on-the-forehead disjointness problem is the protocol of Grolmusz [17] that uses $O(kn/2^k)$ bits of communication. (Grolmusz's protocol is designed for the generalized-inner-product function, however, the protocol works for the disjointness function with an obvious modification.) Prior to and independent of our work, Tesson had shown in an unpublished section of his doctoral dissertation [34] that the deterministic k-party number-on-the-forehead communication complexity of disjointness is $\Omega(\frac{\log n}{k})$. We obtain the following communication lower bounds for randomized number-on-the-forehead protocols:

1. Three-player protocols such that the first player speaks once and the other two players then proceed arbitrarily require $\Omega(n^{1/3})$ bits of communication to compute the disjointness function for deterministic computation or randomized computation with constant error. The only three-player number-on-the-forehead model for which an $n^{\Omega(1)}$ lower bound for disjointness was previously known is the *one-way model* in which the first player speaks, then the second player speaks, and finally the third player calculates the answer. A result of Wigderson (included in the appendix of a paper by Babai, Hayes and Kimmel [4]), shows that the one-way three-party number-on-the-forehead complexity of disjointness is $\Omega(n^{1/2})$. While the one-way model is weaker, the $\Omega(n^{1/2})$ bound is quantitatively better, so the two results are incomparable. (The bound as stated is for a layered pointer jumping problem which corresponds to the special case of the disjointness problem in which the first player's input is one of $\sqrt{n}$ disjoint subsets of $[n]$ of size $\sqrt{n}$, the second player's input has one element in each of these $\sqrt{n}$ blocks and the third player's input is an arbitrary vector of n bits.)

2. k-player protocols in which all players broadcast a single message simultaneously require $\Omega\left(n^{1/(k-1)}/(k-1)\right)$ bits of communication. This uses an argument based on that used by Babai, Gal, Kimmel and Lokam [3] to study other functions in the simultaneous messages model.

3. General k-player randomized number-on-the-forehead protocols require $\frac{\log_2 n}{k-1} - O(1)$ bits of communication to compute disjointness with constant error. This is slightly better than the unpublished bound by Tesson [34] since it is for randomized protocols rather than deterministic protocols (though it seems likely that his methods can be extended to the randomized case), and the constants in our bound seem to be sharper.

1.2 A Direct Product Theorem

Our lower bound for three-player, number-on-the-forehead, "first player speaks then dies" protocols is proved by using the three-player protocol to solve many independent instances of the two-player disjointness problem. We then make use of our core technical theorem, which says that for a broad class of functions f, whenever f requires b bits of communication by a two-player randomized protocol to be calculated correctly with probability $\delta < 1$, computing the answer for t independent instances of f using t' bits of communication for some t' that is $\Theta(tb)$ is correct with probability at most $\delta^{\Omega(t)}$. Results of this form are known as *strong direct product theorems*.

Direct sum and direct product theorems are a broad family of results relating the computational difficulty of computing a function on many different instances with the computational difficulty of computing the function on a single instance. Given a function $f : I \to O$, the function $f^t : I^t \to O^t$ is given by $f^t(x_1, \ldots, x_t) = (f(x_1), \ldots, f(x_t))$.

A complexity measure C, such as communication complexity or circuit size, satisfies a *direct sum property* if and only if $C(f^t) = \Omega(tC(f))$. Karchmer, Raz, and Wigderson [21] introduced the direct sum problem in two-party communication complexity in the context of search problems based on random functions. They showed that if a direct sum result holds for these search problems, then $\mathsf{NC}^1 \neq \mathsf{NC}^2$. Direct sum theorems are known for nondeterministic and co-nondeterministic two-party communication complexity and direct sum properties are known for bounded-round deterministic [20] and bounded-round distributional/randomized [18] two-party communication complexity. Recent information theory based techniques, information complexity [12, 6] and conditional information complexity [7], are useful because these measures satisfy direct sum properties under rectangular (or conditionally rectangular) distributions.

Direct product results relate the amount of error made by a computation of f^t to the amount of error made by a computation of f. More precisely, they relate the probability of success under a distribution μ^t to the probability of success under distribution μ. A good example of such a result is the Concatenation Lemma, a variant of Yao's XOR lemma: if all circuits of size $\leq s$ compute f correctly on at most a p fraction of inputs, then for all $\epsilon > 0$, circuits of size $\leq s\left(\epsilon/n\right)^{O(1)}$ compute f^t correctly on at most a $p^t + \epsilon$ fraction of inputs [16]. (Note that when ϵ is in the interesting range around p^t, f^t has a hardness guarantee only for circuits of size far less than the size for which computing f is hard.) Direct product results naturally concern distributional complexity, but by Yao's arguments relating distributional and randomized computation they imply results for randomized algorithms as well.

Strong direct product results combine the resource amplification of a direct sum result with the error amplification of a direct product result: If a computation using r resources gets the answer for f correct on at most a p measure of the inputs under distribution μ, then for some $r' = \Omega(tr)$ a computation using r' resources gets the answer for f^t correct on at most a $p^{\Omega(t)}$ measure of the inputs under distribution μ^t.

Few strong direct product results are known and strong direct product theorems do not hold for many interesting models of computation. In particular, Shaltiel has shown that distributional two-party communication complexity in general does not satisfy a strong direct product theorem [33]. However, for communication complexity under the uniform distribution, Shaltiel [33] proved that lower bounds obtained by the discrepancy method under the uniform distribution satisfy a strong direct product property in that for any 2-party protocol sending $r' = tr$ bits, the correlation of its output with the exclusive-or of the t binary outputs of f^t decays exponentially with t.

As with Shaltiel's result for discrepancy, the way we ensure that a strong direct product theorem holds is to make use of the method used to prove the communication lower bound. Lower bounds for the distributional (and thus randomized) two-party communication complexity of the disjointness function have been proved using the *corruption method*[2]. In general, a corruption bound shows that for a function f and distribution μ, for some frequently occurring value b in the range of f, on every not-very-tiny set of the form $A \times B$, at least an ϵ fraction of the elements map to answers different from b. In [22], Klauck formalized many ideas similar to the corruption bound, and showed that it is tightly connected to the amount of communication needed in MA^{cc} and AM^{cc} protocols. It is easy to see that, up to constant factors, lower bounds based on corruption are at least as large as those based on discrepancy. Moreover Babai, Frankl, and Simon [2] showed, using the two-party disjointness function, that lower bounds based on corruption can be exponentially better than those based on discrepancy [2].

Our theorem shows that when μ is a distribution on pairs (x, y) in which the distribution on x is independent of the distribution on y, communication bounds proved using the corruption method obey a strong direct product theorem. Our strong direct product theorem is incomparable with the discrepancy result of Shaltiel, because Shaltiel's result involves a more restrictive technique for obtaining lower bounds and a nar-

[2] Although corruption bounds are frequently used, there does not seem to be a consistent terminology for such bounds. The monograph by Kushilevitz and Nisan [24] uses the term "one-sided discrepancy". Klauck calls the method "ϵ-error complexity" [22].

rower class of distributions but requires less from the protocol in that it only has to predict the exclusive-or of the outputs of f^t rather than all of f^t. We also extend our strong direct product theorem to the case of approximate computation of f^t; essentially the same strong direct product bounds apply to protocols that compute any function g each of whose outputs has small Hamming distance from the corresponding output of f^t. We use this approximate version in deriving sharper bounds for the case of randomized 3-party protocols.

2 Background and Notation

2.1 Sets, Strings and Miscellaneous Notation

The set of integers $\{1, \ldots n\}$ is denoted $[n]$. We identify $\mathcal{P}([n])$ with $\{0,1\}^n$ by identifying sets with their characteristic vectors. We will refer to elements of $\{0,1\}^n$ interchangeably as sets or vectors. In this spirit, we write $x \cap y$ for the string whose i-th coordinate is 1 if and only if the i-th coordinate of both x and y are 1.

At times we use regular expression notation when specifying sets of strings over a finite alphabet such as $\{0,1\}$ or $\{p,q\}$. The empty string is written as Λ. When A and B are expressions for sets of strings, $AB = \{xy \mid x \in A, y \in B\}$, $A^i = \{x_1 \ldots x_i \mid x_1, \ldots x_i \in A\}$, $A^{\leq i} = \bigcup_{j \leq i} A^j$, $A^* = \bigcup_{k=0}^{\infty} A^k$, and $A \cup B$ is the set-theoretical union of A and B. The notation x^j denoting j repetitions of the string x could clash with the use of superscripts when naming variables. However, in this paper, the repetition notation is used only with elements of the alphabet, such as $0, 1, p, q$, or sets, and it is never used with symbols that are used variable names, such as x, y, z.

Let μ be a probability distribution on a set X. The *support of* μ is $\{x \in X \mid \mu(x) > 0\}$. When μ is a probability distribution on a product set $X \times Y$, μ is said to be a *rectangular* distribution if there exist distributions μ_X on X and μ_Y on Y so that for all $(x,y) \in X \times Y$, $\mu(x,y) = \mu_X(x) \cdot \mu_Y(y)$. The phrase *product distribution* is often used in the literature instead of rectangular distribution.

2.2 Communication Complexity

Number-on-the-forehead protocols are strategies by which a group of k players compute a function on $X_1 \times \ldots \times X_k$, $f(x_1, \ldots x_k)$, when each player i has access only to the inputs $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k$. In randomized protocols, in addition to their inputs players have access to a shared source of random bits. (This is the so-called public randomness model and is equivalent to a probability distribution over deterministic protocols.)

A protocol is *simultaneous* if each player's message depends only on the random bits and the inputs visible to that player, a protocol is *one-way* if each player speaks exactly once and the players do so in a fixed order. We identify each player in a number-on-the-forehead communication protocol with the name of the set from which the inputs on its forehead are drawn. We describe restrictions on communication order such as those above by a *communication pattern* P. Examples of communication patterns P we consider are

- $X_1 \rightarrow \ldots \rightarrow X_k$ indicating that the protocol is one-way in that players $X_1, \ldots, X_k$ each speak once in that order.
- $X_1 || \ldots || X_k$ indicating that players $X_1, \ldots, X_k$ each speak simultaneously and independently.

- $X_1 \leftrightarrow \ldots \leftrightarrow X_k$ indicating that the order of speaking is arbitrary. Since this is unrestricted computation, following standard notation we simply write that P is k to denote that it is unrestricted k-party computation.

These patterns can be combined using parentheses to create more complicated communication patterns. In particular, we denote the 3-party communication pattern in which "the first player speaks then dies" by $Z \to (Y \leftrightarrow X)$. (We use these set/player names so that communication between the last two parties has similar set names to standard two-party communication complexity.)

Formal definitions of such protocols are quite standard and may be found, for example, in [24]; we do not repeat them here.

Definition 2.1. *For a deterministic protocol Π, and input $\vec{x}$, let $\Pi(\vec{x})$ denote the output of the protocol on input $\vec{x}$ and let $c_\Pi(\vec{x})$ denote the sequence of bits communicated on that input. For randomized protocols the corresponding values are denoted $\Pi(\vec{x}, r)$ and $c_\Pi(\vec{x}, r)$ where r is the shared random string. For a given communication pattern P for a function f on $\vec{X}$ define*

- *the* deterministic communication complexity *of f, $D^P(f)$, to be the minimum over all deterministic protocols Π with pattern P and with $\Pi(\vec{x}) = f(\vec{x})$ for every $\vec{x}$, of $C(\Pi) = \max_{\vec{x}} |c_\Pi(\vec{x})|$.*

- *the* ϵ-error randomized communication complexity *of f, $R^P_\epsilon(f)$, to be the minimum over all randomized protocols Π with pattern P and with $\Pr_r[\Pi(\vec{x}, r) \neq f(\vec{x})] \leq \epsilon$ for every $\vec{x}$, of $C(\Pi) = \max_{\vec{x},r} |c_\Pi(\vec{x}, r)|$.*

- *for any probability distribution μ on $\vec{X}$, the* (μ, ϵ)-distributional communication complexity *of f, $D^{P,\epsilon}_\mu(f)$ to be the minimum over all deterministic protocols Π with pattern P and $\Pr_\mu[\Pi(\vec{x}) \neq f(\vec{x})] \leq \epsilon$ of $C(\Pi) = \max_{\vec{x}} |c_\Pi(\vec{x})|$.*

As usual in studying communication complexity we need the following definitions.

Definition 2.2. *A* combinatorial rectangle *R in $X \times Y$ is a set of the form $A \times B$ with $A \subseteq X$ and $B \subseteq Y$. An* i-cylinder *C on $U = X_1 \times \cdots \times X_k$ is a set of the form $\{(x_1, \ldots, x_k) \in U \mid g(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k) = 1\}$ for some function $g : X_1 \times \ldots \times X_k \to \{0, 1\}$. A* cylinder intersection *on $X_1 \times \cdots \times X_k$ is a set $E = \bigcap_{i=1}^k C_i$ where C_i is an i-cylinder on $X_1 \times \cdots \times X_k$. A cylinder-intersection in the product of k sets $X_1 \times \ldots \times X_k$ is called a* k-dimensional cylinder intersection.

Observe that a combinatorial rectangle is a two-dimensional cylinder intersection. We make use of the following standard results in communication complexity, cf. [24].

Proposition 2.3. *Let $k \geq 2$ be an integer, let $X_1, \ldots, X_k$ be nonempty sets, and let Π be a randomized k-party number-on-the-forehead protocol on $X_1 \times \ldots \times X_k$. For each setting of the random source $r \in \{0, 1\}^*$, and each $s \in \{0, 1\}^*$, $\{(x_1, \ldots, x_k) \in X_1 \times \ldots \times X_k \mid c_\Pi(x_1, \ldots, x_k, r) = s\}$ is a cylinder intersection.*

Proposition 2.4 (Yao's lemma). *Let P be a communication pattern on $\vec{X}$ and μ be a distribution on $\vec{X}$. For any f defined on $\vec{X}$ and $\epsilon > 0$, $R^P_\epsilon(f) = \max_\mu D^{P,\mu}_\epsilon(f)$.*

We will also use the following standard bounds on tails of the binomial distribution and the standard amplification results relating different error bounds in communication complexity that follow.

Proposition 2.5. *Let $0 \leq p \leq 1$ and $B(n, p)$ denote the binomial distribution. Then*

1. $\Pr[B(n,p) \leq pn/4] \leq 2^{-pn/2}$.

2. For $p < 1/2$*,* $\Pr[B(n,p) \geq n/2] \leq (4p(1-p))^{n/2}$.

Proof. The first bound follows from a standard Chernoff bound, $\Pr[B(n,p) \leq pn/4] \leq (\sqrt{2}/e^{3/4})^{pn} \leq 2^{-pn/2}$ and the second follows via

$$\Pr[B(n,p) \geq n/2] \leq \sum_{k=n/2}^{n} \binom{n}{k} p^k (1-p)^{n-k} \leq 2^n p^{n/2}(1-p)^{n/2} = (4p(1-p))^{n/2}.$$

□

Proposition 2.6. *There is a constant c such that for any* $0 < \epsilon' < \epsilon < 1/2$*, and any* $f : \vec{X} \to \{0,1\}$*, for any communication pattern* P*,* $R^P_{\epsilon'}(f) \leq c\frac{\log_{1/\epsilon}(1/\epsilon')}{(1-2\epsilon)^2} R^P_\epsilon(f)$*.*

Proof. Suppose first that $1/8 < \epsilon < 1/2$. Write $\delta = \frac{1}{2} - \epsilon$. Applying Proposition 2.5 with $p = \epsilon$ and $n = \lceil \frac{1}{\delta^2} \rceil = \lceil \frac{4}{(1-2\epsilon)^2} \rceil$ we obtain that $\Pr[B(n,\epsilon) \geq n/2] \leq (1-4\delta^2)^{n/2} \leq e^{-2} < 1/8$. Therefore if we define a new protocol P' that takes the majority of n independent runs of the original protocol we obtain an error at most $1/8$. For $\epsilon \leq 1/8$, $4\epsilon(1-\epsilon) < \epsilon^{1/3}$ and thus repeating any such protocol $6\log_{1/\epsilon}(1/\epsilon')$ times and taking the majority yields error at most ϵ'. Combining these two arguments yields the claim. □

Finally, we define the k-party disjointness function.

Definition 2.7. *The k-party disjointness function for* $X_1 = \cdots = X_k = \{0,1\}^n$ *is the function* $\text{DISJ}_{k,n} : X_1 \times \cdots \times X_k \to \{0,1\}$ *defined by* $\text{DISJ}_{k,n}(x_1,\ldots,x_k) = 1$ *if for there is some* $j \in [n]$ *such that* $x_{i,j} = 1$ *for all* $i \in [k]$ *and* $\text{DISJ}_{k,n}(x_1,\ldots,x_k) = 0$ *otherwise. (That is,* $\text{DISJ}_{k,n}(x_1,\ldots,x_k) = 1$ *if and only if* $x_1 \cap \ldots \cap x_k \neq \emptyset$*.) We drop the subscript* n *if it is understood from the context.*

This is a natural extension of the usual two-party disjointness function so we have kept the same terminology but when it evaluates to 0 it does not mean that the inputs $x_1,\ldots,x_n$ viewed as sets are mutually disjoint; instead it means that there is no common point of intersection among these sets. (Note that in the analysis of disjointness in the number-in-hand model (e.g. [1]) the lower bounds apply to either version of the problem. In the number-on-the-forehead model only the version of problem that we define is non-trivial.)

3 Discrepancy, Corruption, and Communication Complexity

Let $f : I \to O$. For $b \in O$, a subset $S \subseteq I$ is called *b-monochromatic* for f if and only if $f(s) = b$ for all $s \in S$ and is called *monochromatic* if and only if it is b-monochromatic for f for some $b \in O$. Let μ be a probability measure on I. For $b \in O$, a subset $S \subseteq I$ is called *ϵ-error b-monochromatic* for f under μ if and only if $\mu(S \setminus f^{-1}(b)) \leq \epsilon \cdot \mu(S)$. For $f : I \to \{0,1\}$, $b \in \{0,1\}$, and $S \subseteq I$ the *b-discrepancy* of f on S under μ,

$$\text{disc}^b_\mu(f,S) = \mu(S \cap f^{-1}(b)) - \mu(S \setminus f^{-1}(b)).$$

Let Γ be a collection of subsets of I and let $f : I \to O$.

$$\begin{aligned}
\mathrm{mono}^b_{\mu,\Gamma}(f) &= \max\{\mu(S) \mid S \in \Gamma \text{ is } b\text{-monochromatic}\}\\
\epsilon\text{-mono}^b_{\mu,\Gamma}(f) &= \max\{\mu(S) \mid S \in \Gamma \text{ is } \epsilon\text{-error } b\text{-monochromatic}\}\\
\mathrm{disc}^b_{\mu,\Gamma}(f) &= \max\{\mathrm{disc}^b_{\mu}(f,S) \mid S \in \Gamma\}\\
\mathrm{mono}_{\mu,\Gamma}(f) &= \max\{\mathrm{mono}^b_{\mu,\Gamma}(f) \mid b \in O\}\\
\mathrm{disc}_{\mu,\Gamma}(f) &= \max\{\mathrm{disc}^b_{\mu,\Gamma}(f) \mid b \in O\}
\end{aligned}$$

When μ is omitted from these notations, it is treated as the uniform distribution. When Γ is not specified, it is the set of k-dimensional cylinder intersections on the input space. In particular, Γ is the set of combinatorial rectangles when $k = 2$.

Proposition 3.1. *For any function $f : I \to \{0, 1\}$, distribution μ on I, $\Gamma \subseteq \mathcal{P}(I)$, $\epsilon < 1/2$, and $b \in \{0, 1\}$, $\mathrm{disc}^b_{\mu,\Gamma}(f) \ge (1 - 2\epsilon)(\epsilon\text{-mono}^b_{\mu,\Gamma}(f))$.*

Proof. Choose $S \in \Gamma$ so that $\mu(S) = \epsilon\text{-mono}^b_{\mu,\Gamma}(f)$ and $\mu(S \setminus f^{-1}(b)) \le \epsilon\mu(S)$. Then $\mathrm{disc}^b_{\mu,\Gamma}(f) \ge \mathrm{disc}^b_{\mu}(f, S) \ge (1 - 2\epsilon)\mu(S)$ as required. □

Let $N^2_1(f)$ and $N^2_0(f)$ be the two-party nondeterministic and co-nondeterministic communication complexities of a function $f : X \times Y \to O$. (That is, the logarithm of the minimum number of 1-monochromatic rectangles needed to cover $f^{-1}(1)$ and the logarithm of the minimum number of 0-monochromatic rectangles needed to cover $f^{-1}(0)$, respectively, cf. [24].) The following is a standard way to obtain two-party communication complexity lower bounds (cf. [24]):

Proposition 3.2. *Let Γ be the set of combinatorial rectangles on $X \times Y$. For any $f : X \times Y \to \{0, 1\}$ and for any probability measure μ on $X \times Y$,*

(a) $D^2(f) \ge \log_2(1/\mathrm{mono}_{\mu,\Gamma}(f))$,

(b) For $b \in \{0, 1\}$, $N^2_b(f) \ge \log_2(\mu(f^{-1}(b))/\mathrm{mono}^b_{\mu,\Gamma}(f))$.

The following are the standard discrepancy lower bounds for randomized communication complexity (see for example [24]).

Proposition 3.3 (Discrepancy Bound)**.** *Let Γ be the set of combinatorial rectangles on $X \times Y$. Let $f : X \times Y \to \{0, 1\}$, $\epsilon < 1/2$, and μ be any probability distribution on $X \times Y$.*

(a) $R^2_\epsilon(f) \ge D^{2,\mu}_\epsilon(f) \ge \log_2((1 - 2\epsilon)/\mathrm{disc}_{\mu,\Gamma}(f))$

(b) For $b \in \{0, 1\}$, $R^2_\epsilon(f) \ge D^{2,\mu}_\epsilon(f) \ge \log_2((\mu(f^{-1}(b)) - \epsilon)/\mathrm{disc}^b_{\mu,\Gamma}(f))$.

More generally, for $k \ge 2$, if $f : X_1 \times \cdots \times X_k \to \{0, 1\}$ and Γ is replaced by the set of cylinder intersections on $X_1 \times \cdots \times X_k$ then

$$R^k_\epsilon(f) \ge D^{k,\mu}_\epsilon(f) \ge \log_2((\mu(f^{-1}(b)) - \epsilon)/\mathrm{disc}^b_{\mu,\Gamma}(f)).$$

It is easy to see that the bound from part (a) can never be more than 1 plus the maximum of the two bounds from part (b). Without loss of generality, suppose that $\mu(f^{-1}(1)) \geq 1/2$. We have that

$$\frac{\mu(f^{-1}(1)) - \epsilon}{\mathrm{disc}^1_{\mu,\Gamma}(f)} \geq \frac{1/2 - \epsilon}{\max\{\mathrm{disc}^1_{\mu,\Gamma}(f), \mathrm{disc}^0_{\mu,\Gamma}(f)\}} = \frac{1}{2} \cdot \frac{1 - 2\epsilon}{\mathrm{disc}_{\mu,\Gamma}(f)}$$

The discrepancy bound works well for analyzing functions such as the inner product, the generalized inner product [5], and matrix multiplication [29]. However, it does not suffice to derive lower bounds for functions such as disjointness. A more general method that is used to prove two-party communication lower bounds for disjointness is the *corruption technique*. A corruption bound says that any sufficiently large rectangle cannot be fully b-monochromatic and makes errors on some fixed fraction of its inputs. Hence, we say that the rectangle is "corrupted". The corruption technique has been used implicitly many times before, and we formalize the principle below. For later discussions of corruption we find it convenient to use the following definition in its statement.

Definition 3.4. *For a collection Γ of subsets of I, distribution μ on I, function $f : I \to O$, $\epsilon > 0$ and $b \in O$ define*

$$\mathrm{corrbd}^b_{\mu,\Gamma}(f, \epsilon) = \log_2(1/(\epsilon\text{-}\mathrm{mono}^b_{\mu,\Gamma}(f))).$$

Lemma 3.5 (Corruption Bound)**.** *Let Γ be the set of combinatorial rectangles on $X \times Y$. Let $f : X \times Y \to O$, $O' \subset O$, $\epsilon \leq 1$, and μ be any probability distribution on $X \times Y$. For $\epsilon' < \epsilon \cdot \mu(f^{-1}(O'))$,*

$$\begin{aligned} R^2_{\epsilon'}(f) \geq D^{2,\mu}_{\epsilon'}(f) &\geq \min_{b \in O'} \log_2((\mu(f^{-1}(O')) - \epsilon'/\epsilon)/\epsilon\text{-}\mathrm{mono}^b_{\mu,\Gamma}(f)) \\ &= \min_{b \in O'} \mathrm{corrbd}^b_{\mu,\Gamma}(f, \epsilon) - \log_2(\frac{1}{\mu(f^{-1}(O') - \epsilon'/\epsilon)}). \end{aligned}$$

More generally, for $k \geq 2$, if $f : X_1 \times \cdots \times X_k \to O$ and Γ is the set of cylinder intersections on $X_1 \times \cdots \times X_k$ then the same lower bound applies to $R^k_{\epsilon'}(f)$.

Proof. We give the proof for $k = 2$; the argument for $k > 2$ is completely analogous. By Yao's lemma (Proposition 2.4), $R^2_{\epsilon'}(f) \geq \max_{\mu'} D^{2,\mu'}_{\epsilon'}(f) \geq D^{2,\mu}_{\epsilon'}(f)$. Consider any deterministic protocol Π of cost $D^{2,\mu}_{\epsilon'}(f)$ that computes f correctly on all but at most an ϵ' fraction of inputs under distribution μ. Consider the partition $\mathcal{R}$ of $X \times Y$ into rectangles induced by the protocol. Let $\gamma = \max_{b \in O'} \epsilon\text{-}\mathrm{mono}^b_{\mu,\Gamma}(f)$. For $b \in O'$, let

$$\alpha_b = \mu(\{x \mid \Pi(x) = b \text{ and } x \in \bigcup_{R \in \mathcal{R},\ \mu(R) \leq \gamma} R\}),$$

the total measure of inputs contained in rectangles of measure at most γ on which the protocol outputs b. There must be at least $\sum_{b \in O'} \alpha_b/\gamma$ such rectangles and thus $D^{2,\mu}_{\epsilon'}(f) \geq \log_2(\sum_{b \in O'} \alpha_b/\gamma)$.

We now bound $\sum_{b \in O'} \alpha_b$. For any $b \neq b' \in O$, let $\epsilon'_{b \to b'}$ the total measure of inputs on which the protocol answers b' when the correct answer is b. Clearly $\epsilon' = \sum_{b,b':b \neq b'} \epsilon'_{b \to b'}$. By definition, the protocol answers b on at least a $\mu(f^{-1}(b)) + \sum_{b' \neq b} \epsilon'_{b' \to b} - \sum_{b' \neq b} \epsilon'_{b \to b'}$ measure of the inputs. By the definition of γ and $\epsilon\text{-}\mathrm{mono}^b_{\mu,\Gamma}(f)$, any rectangle of measure larger than γ on which the protocol answers b must have at least an ϵ proportion of its total measure on which the correct answer is not b; i.e., an ϵ proportion of its measure contributes to $\sum_{b' \neq b} \epsilon'_{b' \to b}$. Thus in total for $b \in O$ we have

$$\sum_{b' \neq b} \epsilon'_{b' \to b} \geq \epsilon \cdot [\mu(f^{-1}(b)) + \sum_{b' \neq b} \epsilon'_{b' \to b} - \alpha_b - \sum_{b' \neq b} \epsilon'_{b \to b'}].$$

Rearranging, we have

$$\alpha_b \geq \mu(f^{-1}(b)) - \sum_{b' \neq b} \epsilon'_{b \to b'} - (1/\epsilon - 1) \sum_{b' \neq b} \epsilon'_{b' \to b}.$$

Summing this over all choices of $b \in O'$ we obtain

$$\begin{aligned}
\sum_{b \in O'} \alpha_b &\geq \sum_{b \in O'} \mu(f^{-1}(b)) - \sum_{b \in O'} \sum_{b' \neq b} \epsilon'_{b \to b'} - (1/\epsilon - 1) \sum_{b \in O'} \sum_{b' \neq b} \epsilon'_{b' \to b} \\
&= \mu(f^{-1}(O')) - (1/\epsilon) \sum_{b, b' \in O' : b \neq b'} \epsilon'_{b \to b'} - \sum_{b \in O'} \sum_{b' \notin O'} \epsilon'_{b \to b'} - (1/\epsilon - 1) \sum_{b \notin O'} \sum_{b' \in O'} \epsilon'_{b \to b'} \\
&\geq \mu(f^{-1}(O')) - (1/\epsilon) \sum_{b, b' : b \neq b'} \epsilon'_{b \to b'} \\
&= \mu(f^{-1}(O')) - \epsilon'/\epsilon
\end{aligned}$$

which yields the claimed lower bound. □

In the special case that the output set $O = \{0, 1\}$ we obtain the following corollary.

Corollary 3.6. *Let Γ be the set of combinatorial rectangles on $X \times Y$. For any $\epsilon < 1/2$ there is a constant $c_\epsilon > 0$ with $c_\epsilon = O(\frac{1}{(1-2\epsilon)^2})$ such that for $f : X \times Y \to \{0, 1\}$, μ any probability distribution on $X \times Y$, and $b \in \{0, 1\}$,*

$$\begin{aligned}
R^2_\epsilon(f) \geq D^{2,\mu}_\epsilon(f) &\geq c_\epsilon \log_2((\mu(f^{-1}(b)) - \epsilon)/\epsilon\text{-mono}^b_{\mu,\Gamma}(f)) \\
&= c_\epsilon[\text{corrbd}^b_\mu(f, \epsilon) - \log_2(\frac{1}{\mu(f^{-1}(b)) - \epsilon})]
\end{aligned}$$

and the same lower bound holds for the case of $R^k_\epsilon(f)$ where Γ is the corresponding set of cylinder intersections on $X_1 \times \cdots \times X_k$.

Proof. We reduce the protocol error to $\epsilon' = \epsilon^2$ using Proposition 2.6 and then apply Lemma 3.5 to obtain the claimed result. The bound on c_ϵ follows since $\log_{1/\epsilon}(1/\epsilon')$ is constant. □

Up to the multiplicative factor $c_\epsilon = O(\frac{1}{(1-2\epsilon)^2})$, the above bound is of the same form as that of Proposition 3.3 except that it uses corruption rather than the discrepancy. By Proposition 3.1, a corruption bound is applicable whenever a discrepancy bound is applicable but the reverse is not the case. (Disjointness is a counterexample.) So, up to a multiplicative constant factor and a small additive term at worst, corruption bounds are always superior to discrepancy bounds.

4 A Direct Product Theorem for Corruption under Rectangular Distributions

We now relate the corruption bound for f to the corruption bound for solving t disjoint instances of f.

Definition 4.1. *For a function $f : X \times Y \to \{0, 1\}$, define $f^t : X^t \times Y^t \to \{0, 1\}^t$ by $f^t(\vec{x}, \vec{y}) = (f(x_1, y_1), \ldots, f(x_t, y_t))$ where $\vec{x} = (x_1, \ldots, x_t)$ and $\vec{y} = (y_1, \ldots, y_t)$.*

Given a distribution μ on a set I, the distribution μ^t is the distribution on I^t with $\mu^t(x_1, \ldots x_t) = \prod_{j=1}^t \mu(x_j)$.

Theorem 4.2 (Direct Product for Corruption)**.** *Let $f : X \times Y \to \{0,1\}$ and μ be a rectangular probability distribution on $X \times Y$. Let $b \in \{0,1\}$, t be a positive integer, $m = \mathrm{corrbd}^b_\mu(f,\epsilon)$, and ϵ satisfy $1 > \epsilon > 12mt/2^{m/8}$.*

(a) Let $T_0 \subseteq \{1,\ldots,t\}$ with $|T_0| = t_0$ and define $V_{T_0} = \{\vec{v} \in \{0,1\}^t \mid v_i = b \text{ for all } i \in T_0\}$. If R is a combinatorial rectangle on $X^t \times Y^t$ with $\mu^t(R) \geq 2^{-t_0 m/6}$ then
$$\mu^t(R \cap (f^t)^{-1}(V_{T_0})) < (3/\epsilon)(1-\epsilon/2)^{t_0/2}\mu^t(R).$$

(b) In particular, if $\vec{v} \in \{0,1\}^t$ is a binary vector with at least t_0 many b's then
$$\mathrm{corrbd}^{\vec{v}}_{\mu^t}(f^t, 1-(3/\epsilon)(1-\epsilon/2)^{t_0/2}) \geq t_0 \cdot \mathrm{corrbd}^b_\mu(f,\epsilon)/6.$$

This theorem implies very strong error properties: any large rectangle on which a protocol P outputs a vector v with many b's has the correct answer on only an exponentially small fraction of the inputs under distribution μ^t. Up to small factors in the communication and the error this is as strong a theorem as one could hope for. Note that, because the corruption bound only measures the complexity when the the output is b, both the communication and error exponent in any such bound must scale with t_0 rather than t.

The general technique we use for our direct product bound follows a standard paradigm of iterated conditional probability analysis on the coordinates that allow one to prove Yao's XOR lemma [16], Raz's parallel repetition theorem [28], and bounds on the complexity savings given by 'help bits' [10, 25].

Definition 4.3. *Let $T \subseteq [t]$ and $U = [t] - T$. For $A \subseteq X^t$, let A_T be the set of projections of A on X^T. (If T is a singleton set $\{j\}$ then we write A_j for $A_{\{j\}}$.) For $x_U \in X^U$ and $A \subseteq X^t$ let $A(x_U)$ be the set of all $\vec{x}' \in A$ such that $x'_U = x_U$. For $B \subseteq Y^t$ and $y_U \in Y^U$ can define B_T and $B(x_U)$ similarly. Moreover, extend the definition for $S \subseteq X^t \times Y^t$ to S_T, the set of projections of S on $X^T \times Y^T$, and, for $(x_U, y_U) \in X^U \times Y^U$, to $S(x_U, y_U)$, the set of all $(\vec{x}', \vec{y}') \in S$ such that $x'_U = x_U$ and $y'_U = y_U$.*

Let μ be a distribution on $X \times Y$. For $T \subseteq [k]$ define μ^T on $X^T \times Y^T$ as the product μ^T on those coordinates. Define μ^T_X and μ^T_Y similarly so that μ^T is the cross product of μ^T_X and μ^T_Y.

Finally, we say that S is rectangular with respect to coordinates *T if and only if for every $(x_U, y_U) \in S_U$, $S(x_U, y_U)_T$ is a combinatorial rectangle in $X^T \times Y^T$.*

The following lemma is the main tool we need to prove the direct product property of corruption. Its proof is the sole reason that we need to restrict the distribution μ to be rectangular. Intuitively, it says that in any rectangle $A \times B$ on $X^k \times Y^k$, except for a small error set E, the set of inputs for which $f(x_1, y_1) = b$ is contained in the union of two disjoint well-structured sets (rectangular on the remaining coordinates) with the property that one has little variation in the first coordinate and the other is constant factor smaller than the set of inputs in $A \times B$ not in the first set. We will apply this repeatedly to prove Theorem 4.2 by carefully accounting for each of the t_0 coordinates on which the lemma can be applied, and observing that either the lack of variation or the reduction in size will be compounded many times.

Lemma 4.4 (Key Lemma)**.** *Let $f : X \times Y \to \{0,1\}$ and μ be a rectangular probability distribution on $X \times Y$. Let $b \in \{0,1\}$ and $m = \mathrm{corrbd}^b_\mu(f,\epsilon)$ for $\epsilon < 1$. Let $k \geq 1$ and $A \times B$ be a combinatorial rectangle in $X^k \times Y^k$. Let an integer $K' \geq 1$ be given and set $K = \lceil \log_{(1-\epsilon/6)} 2^{-K'} \rceil = \lceil -K'/\log_2(1-\epsilon/6) \rceil$. There are sets $P, Q, E \subseteq A \times B$ such that the set of inputs $(\vec{x}, \vec{y}) \in A \times B$ for which $f(x_1, y_1) = b$ is contained in $P \cup Q \cup E$ where*

1. *$\mu^k(E) \leq 2^{1-K'}$,*

2. *$\mu^k(Q) \leq (1-\epsilon/2)\mu^k(A \times B - P - E)$,*

3. $\mu(P_1) \le K^2 2^{-m}$.

Furthermore P*,* Q*, and* E *are rectangular on coordinates* $\{2, \dots, k\}$ *and* P_1*,* Q_1*, and* E_1 *are all disjoint.*

Proof. We would like to upper bound the fraction of inputs in $A \times B$ on which $f(x_1, y_1) = b$. The general idea of the proof involves considering the set of projections (x_1, y_1) of the elements of $A \times B$ on the first coordinate. This set forms a rectangle on $X \times Y$. By definition of $m = \mathrm{corrbd}_\mu^b(f, \epsilon)$, if this set has μ measure larger than 2^{-m} then $f(x_1, y_1) = b$ for at most a $1 - \epsilon$ fraction of the projected pairs (x_1, y_1).

However, because the different (x_1, y_1) occur with different frequencies in $A \times B$, the overall fraction of errors may be much smaller. To overcome this problem we group the elements of A and B based on the number of extensions their projections x_1 or y_1 have in A or B respectively. We choose the groups so that each is a rectangle and in any group there is very little variation in the number of extensions. For any one of these groups containing at least a 2^{-m} fraction of (x_1, y_1) pairs we can apply the corruption bound for f to upper-bound the fraction of inputs on which the function has output b. Any group that does not satisfy this must be small. To keep the number of groups small we first separate out one set consisting of those inputs where the number of extensions is tiny. In our argument, Q will be the union of the large groups, P will be the union of the small groups, and E will be the set of inputs with a tiny number of extensions.

Let A_1 be the set of projections of A on the first coordinate and B_1 be the set of projections of B on the first coordinate. Choose $\delta = \epsilon/6$ and let $T = \{2, \dots, k\}$. Sort the elements of A_1 based on the number of their extensions: For $1 \le i \le \lceil \log_{(1-\delta)} 2^{-K'} \rceil = \lceil -K'/\log_2(1-\delta) \rceil = K$ let $A_{1,i} = \{x_1 \in A_1 \mid i = \lceil \log_{(1-\delta)} \mu_X^T(A(x_1)_T) \rceil\}$ and $B_{1,i'} = \{y_1 \in B_1 \mid i' = \lceil \log_{(1-\delta)} \mu_Y^T(B(y_1)_T) \rceil\}$. Every point in $A_{1,i}$ has between a $(1-\delta)^{i-1}$ and $(1-\delta)^i$ measure of extensions in the T coordinates and the same holds for each $B_{1,i'}$. Let $A^{1,i} = \{\vec{x} \in A \mid x_1 \in A_{1,i}\}$ and $B^{1,i'} = \{\vec{y} \in B \mid y_1 \in B_{1,i'}\}$. Let $E = [(A - \bigcup_{i=1}^K A^{1,i}) \times B] \cup [A \times (B - \bigcup_{i'=1}^K B^{1,i})]$. We bound the size of E as follows: For each $x_1 \in A \setminus \bigcup_{i=1}^K A^{1,i}$, we have $\log_{1-\delta} \mu_X^T((A(x_1))_T)\rceil > K$, and therefore $\mu^k\left(\left(A \setminus \bigcup_{i=1}^K A^{1,i}\right) \times Y\right) \le (1-\delta)^K \le (1-\delta)^{\log_{1-\delta} 2^{-K'}} \le 2^{-K'}$. Similarly, $\mu^k\left(X \times \left(B \setminus \bigcup_{i=1}^k B^{1,i}\right)\right) \le 2^{-K'}$, and therefore $\mu^k(E) \le 2 \cdot 2^{-K'}$.

For $i, i' \le K$ let $R^{(i,i')} = A^{1,i} \times B^{1,i'}$ and then $A \times B = E \cup \bigcup_{i=1}^K \bigcup_{i'=1}^K R^{(i,i')}$. By definition $R_1^{(i,i')} = A_{1,i} \times B_{1,i'}$ is the projection of $R^{(i,i')}$ on the first coordinate. Every $(x_1, y_1) \in R_1^{(i,i')}$ has at most a $(1-\delta)^{i+i'-2}$ and at least a $(1-\delta)^{i+i'}$ measure of extensions in $R^{(i,i')}$ because:

$$\mu^T((R^{(i,i')}(x_1, y_1))_T) = \mu^T(((A \times B)(x_1, y_1))_T) = \mu^T(A(x_1)_T \times B(y_1)_T) = \mu_X^T(A(x_1)_T) \cdot \mu_Y^T(B(y_1)_T)$$

and for $(x_1, y_1) \in R_1^{(i,i')}$ the first quantity in the product is between $(1-\delta)^{i-1}$ and $(1-\delta)^i$ and the second is between $(1-\delta)^{i'-1}$ and $(1-\delta)^{i'}$. Furthermore, this guarantees that the measures of extensions for any two pairs $(x_1, y_1), (x_1', y_1') \in R_1^{(i,i')}$ have a ratio between 1 and $(1-\delta)^2 \ge 1 - 2\delta = 1 - \epsilon/3$.

Let $G = \{(i, i') \mid \mu(R_1^{(i,i')}) = \mu(A_{1,i} \times B_{1,i'}) \ge 2^{-m}\}$. Because $m = \mathrm{corrbd}_\mu^b(f, \epsilon)$, for every $(i, i') \in G$ we have

$$\mu(A_{1,i} \times B_{1,i'} \cap f^{-1}(b)) \le (1-\epsilon)\mu(A_{1,i} \times B_{1,i'}).$$

Let $Q^{(i,i')} = \{(\vec{x}, \vec{y}) \in R^{(i,i')} \mid f(x_1, y_1) = b\}$. Since elements in $R_1^{(i,i')} = A_{1,i} \times B_{1,i'}$ have a μ^T measure of extensions in $R^{(i,i')}$ between $(1 - \epsilon/3)$ and 1,

$$\mu^k(Q^{(i,i')}) \le (1-\epsilon)\mu^k(R^{(i,i')})/(1-\epsilon/3) \le (1-\epsilon/2)\mu^k(R^{(i,i')})$$

Let $Q = \bigcup_{(i,i')\in G} Q^{(i,i')}$ and $P = \bigcup_{(i,i')\notin G} R^{(i,i')}$. Then

$$\mu^k(Q) \le (1-\epsilon/2)\mu^k(\bigcup_{(i,i')\in G} R^{(i,i')}) = (1-\epsilon/2)\mu^k(A\times B - P - E)$$

Furthermore for the projection P_1 of P on the first coordinate, $\mu(P_1) = \mu(\bigcup_{(i,i')\in[K]^2\setminus G} A_{1,i}\times B_{1,i'}) < K^2 2^{-m}$. Observe that the conditions that determine whether an element $(\vec{x},\vec{y}) \in A\times B$ is in Q or P is based solely on the the (x_1,y_1) coordinates of $(\vec{x},\vec{y})$ so each of Q and P is rectangular with respect to $T = \{2,\ldots,k\}$. □

Proof of Theorem 4.2. We prove part (a); part (b) is an immediate corollary. Without loss of generality, we may assume that $b = 0$, and by symmetry we may assume that $T_0 = \{1,\ldots,t_0\}$. Let R be any rectangle on $X^t \times Y^t$. We will classify inputs in R based on the properties of their projections on each of the t_0 prefixes of their coordinates based on the trichotomy given by Lemma 4.4. Lemma 4.4 splits the set of inputs in any rectangle R based solely on their the first coordinate into a tiny error set E of inputs, a set P of inputs among which there are very few choices for the first coordinate and a set Q of the remaining inputs on which an output of 0 for that coordinate can be correct only on a $(1-\epsilon/2)$ fraction of inputs.

The sets of inputs corresponding to sets P and Q are iteratively subdivided using Lemma 4.4 based on the properties of their second coordinate, etc. For $j \le t_0$ we will group together all the tiny error sets E found at any point into a single error set which also will be tiny. For the remaining inputs the decomposition over the various coordinates leads to disjoint sets of inputs corresponding to the branches of a binary tree, depending on whether the input fell into the P or Q set at each application of Lemma 4.4. At each stage we either get a very small multiplicative factor in the upper bound on the total number of inputs possible because of the lack of variation in the coordinate (the case of set P) or we get a small multiplicative factor in the upper bound on the fraction of remaining inputs on which the answer of 0 can be correct (the case of set Q). For $\alpha \in \{p,q\}^{t_0}$ we will write S^α for the set of inputs such that for each $j \in [t_0]$, the input is in a P set at coordinate j when $\alpha_j = p$ and in a Q set at coordinate j when $\alpha_j = q$. Out of t_0 coordinates, one of p or q must occur at least $t_0/2$ times which will be good enough to derive the claimed bound.

For $\alpha \in \{p,q\}^j$ define $\#_p(\alpha)$ (resp. $\#_q(\alpha)$) to be the number of p's (resp. q's) in α. For $0 \le j \le t_0$ and $\alpha \in \{p,q\}^j$ we inductively define sets $S^\alpha, E^j \subseteq X^t \times Y^t$ satisfying the following properties:

1. $R \cap (f^t)^{-1}(V_{T_0}) \subseteq E^j \cup \bigcup_{\alpha\in\{p,q\}^j} S^\alpha$.
2. For every $\alpha \in \{p,q\}^j$, S^α is rectangular with respect to coordinates $j+1,\ldots,t$.
3. For $U = \{1,\ldots,j\}$, for all $\alpha,\beta \in \{p,q\}^j$, if $\alpha \ne \beta$ then $S^\alpha_U \cap S^\beta_U = \emptyset$.
4. For $\alpha \in \{p,q\}^{j-1}$, $\mu^t(S^{\alpha q}) \le (1-\epsilon/2)(\mu^t(S^\alpha) - \mu^t(S^{\alpha p}))$.
5. For $U = \{1,\ldots,j\}$, for all $\alpha \in \{p,q\}^j$, $\mu^U(S^\alpha_U) \le \lceil -mt/\log(1-\epsilon/6)\rceil^{2j} 2^{-\#_p(\alpha)m}$.
6. $\mu^t(E^j) \le 2j2^{-mt}$.

For the base case when $j = 0$: Define $S^\lambda = R$ and $E^0 = \emptyset$ where λ is the empty string. Clearly all the properties are satisfied. To inductively proceed from j to $j+1$, for each $\alpha \in \{p,q\}^j$ we apply Lemma 4.4 to build the sets $S^{\alpha p}$, $S^{\alpha q}$, and E^{j+1} from sets S^α and E^j as follows:

Let $\alpha \in \{p,q\}^j$. Let $U = \{1,\ldots,j\}$ and $T = [t] - U$. Since by property 2 for j, S^α is rectangular on T, for each $(x_U,y_U) \in S^\alpha_U$, the set $S^\alpha(x_U,y_U)_T$ can be expressed as $A_{x_U,y_U} \times B_{x_U,y_U}$. Apply Lemma 4.4

with $k = t - j$ and $K' = mt$ to $A_{x_U,y_U} \times B_{x_U,y_U}$ to obtain disjoint sets P_{x_U,y_U}, Q_{x_U,y_U}, and E_{x_U,y_U} that contain all projections of inputs in $(S^\alpha(x_U, y_U))_T$ on which the $j+1$-st output 0 is correct. Thus sets $P_{(x_U,y_U)} = \{(x_U, y_U)\} \times P_{x_U,y_U}$, $Q_{(x_U,y_U)} = \{(x_U, y_U)\} \times Q_{x_U,y_U}$, and $E_{(x_U,y_U)} = \{(x_U, y_U)\} \times E_{x_U,y_U}$ are disjoint and contain all inputs of $S^\alpha(x_U, y_U)$ on which the $j+1$-st output 0 is correct. Moreover, by Lemma 4.4 these sets are disjoint on coordinate $j+1$, rectangular on coordinates $j+2, \ldots, t$ and for $K = \lceil -mt/\log_2(1-\epsilon/6) \rceil$ satisfy:

$$\mu^T((E_{(x_U,y_U)})_T) \le 2^{1-mt} \tag{1}$$

$$\mu((P_{(x_U,y_U)})_{j+1}) \le K^2 2^{-m} \tag{2}$$

$$\mu^T((Q_{(x_U,y_U)})_T) \le (1-\epsilon/2)\mu^T(S^\alpha(x_U, y_U)_T - (P_{(x_U,y_U)})_T) \tag{3}$$

(Lemma 4.4 yields a slightly stronger bound than (3) but we only need the weaker bound.)

For $\alpha \in \{p, q\}^j$ define

$$S^{\alpha p} = \bigcup_{(x_U,y_U) \in S_U^\alpha} P_{(x_U,y_U)},$$

$$S^{\alpha q} = \bigcup_{(x_U,y_U) \in S_U^\alpha} Q_{(x_U,y_U)},$$

and define

$$E^{j+1} = E^j \cup \bigcup_{\alpha \in \{p,q\}^j} \bigcup_{(x_U,y_U) \in S_U^\alpha} E_{(x_U,y_U)}.$$

Properties 1, 2, and 3 for $j+1$ follow immediately from Lemma 4.4 and the properties 1–6 for j.

Now consider property 4:

$$\begin{aligned}
&\mu^t(S^{\alpha q}) \\
&= \mu^t(\bigcup_{(x_U,y_U) \in S_U^\alpha} Q_{(x_U,y_U)}) = \sum_{(x_U,y_U) \in S_U^\alpha} \mu^t(Q_{(x_U,y_U)}) \\
&= \sum_{(x_U,y_U) \in S_U^\alpha} \mu^U(\{(x_U, y_U)\})\mu^T(Q_{(x_U,y_U)}) \\
&\le \sum_{(x_U,y_U) \in S_U^\alpha} \mu^U(\{(x_U, y_U)\})(1-\epsilon/2)\mu^T(S^\alpha(x_U, y_U)_T - (P_{(x_U,y_U)})_T) \qquad \text{by (3)} \\
&= (1-\epsilon/2)\left(\sum_{(x_U,y_U) \in S_U^\alpha} \mu^U(\{(x_U, y_U)\})\mu^T(S^\alpha(x_U, y_U)_T - \sum_{(x_U,y_U) \in S_U^\alpha} \mu^U(\{(x_U, y_U)\})\mu^T(P_{(x_U,y_U)})_T) \right) \\
&= (1-\epsilon/2)(\mu^t(S^\alpha) - \mu^t(S^{\alpha p}))
\end{aligned}$$

which proves that property 4 is satisfied for $j+1$.

For the case of property 5 observe that for $\alpha \in \{p, q\}^j$,

$$\begin{aligned}
&\mu^{U \cup \{j+1\}}(S^{\alpha p}_{U \cup \{j+1\}}) \\
&= \mu^{U \cup \{j+1\}} \left(\bigcup_{(x_U,y_U) \in S_U^\alpha} (P_{(x_U,y_U)})_{U \cup \{j+1\}} \right)
\end{aligned}$$

$$= \sum_{(x_U,y_U)\in S^\alpha_U} \mu^{U\cup\{j+1\}}((P_{(x_U,y_U)})_{U\cup\{j+1\}})$$

(since the sets $P_{(x_U,y_U)}$ have distinct values in coordinates $U \cup \{j+1\}$)

$$\begin{aligned}
&= \sum_{(x_U,y_U)\in S^\alpha_U} \mu^U(\{(x_U,y_U)\})\cdot \mu((P_{(x_U,y_U)})_{j+1}) \\
&\leq \sum_{(x_U,y_U)\in S^\alpha_U} \mu^U(\{(x_U,y_U)\})\cdot K^2 2^{-m} \qquad \text{by (2)} \\
&\leq \mu^U(S^\alpha_U)\cdot K^2 2^{-m} \leq \ K^{2j}2^{-\#_p(\alpha)m}\cdot K^2 2^{-m} = \ K^{2(j+1)}2^{-\#_p(\alpha p)m}
\end{aligned}$$

and

$$\begin{aligned}
\mu^{U\cup\{j+1\}}(S^{\alpha q}_{U\cup\{j+1\}}) \leq \mu^{U\cup\{j+1\}}(S^{\alpha}_{U\cup\{j+1\}}) &\leq \mu^U(S^\alpha_U)\cdot\mu(S^\alpha_{j+1}) \\
&\leq \mu^U(S^\alpha_U) \leq K^{2j}2^{-\#_p(\alpha)m} = K^{2j}2^{-\#_p(\alpha q)m}.
\end{aligned}$$

Thus property 5 is satisfied for $j+1$.

Finally, for property 6,

$$\begin{aligned}
\mu^t(E^{j+1}) &= \mu^t(E^j \cup \bigcup_{\alpha\in\{p,q\}^j}\ \bigcup_{(x_U,y_U)\in S^\alpha_U} \mu^t(E_{(x_U,y_U)})) \\
&\leq \mu^t(E^j) + \sum_{\alpha\in\{p,q\}^j}\ \sum_{(x_U,y_U)\in S^\alpha_U} \mu^t(E_{(x_U,y_U)}) \\
&= \mu^t(E^j) + \sum_{\alpha\in\{p,q\}^j}\ \sum_{(x_U,y_U)\in S^\alpha_U} \mu^U(\{(x_U,y_U)\})\mu^T((E_{(x_U,y_U)})_T) \\
&\leq 2j2^{-mt} + \sum_{\alpha\in\{p,q\}^j}\ \sum_{(x_U,y_U)\in S^\alpha_U} \mu^U(\{(x_U,y_U)\})\mu^T((E_{(x_U,y_U)})_T)
\end{aligned}$$

and by (1), the definition of S^α_U, and the fact that the S^α_U for distinct α are disjoint this is

$$\begin{aligned}
&\leq 2j2^{-mt} + \mu^U(\bigcup_{\alpha\in\{p,q\}^j} S^\alpha_U)\, 2^{1-mt} \\
&\leq 2j2^{-mt} + 2^{1-mt} \leq 2(j+1)2^{-mt},
\end{aligned}$$

which proves that property 6 is satisfied for $j+1$.

All the properties required for the induction hypothesis are satisfied, therefore the recursive construction produces the desired sets. We now use all these properties to derive the upper bound on $\mu^t(R \cap (f^t)^{-1}(V_{T_0}))$:

By property 1, $R \cap (f^t)^{-1}(V_{T_0}) \subseteq E^{t_0} \cup \bigcup_{\alpha\in\{p,q\}^{t_0}} S^\alpha$. Therefore for $\alpha \in \{p,q\}^{t_0}$ with $\#_p(\alpha) \geq t_0/2$,

$$\mu^t(S^\alpha) \leq \mu^{\{1,\ldots,t_0\}}(S^\alpha_{\{1,\ldots,t_0\}}) \leq K^{2t_0}2^{-\#_p(\alpha)m} \leq K^{2t_0}2^{-t_0 m/2}$$

and therefore

$$\mu^t\left(\bigcup_{\alpha\in\{p,q\}^{t_0}:\#_p(\alpha)\geq t_0/2} S^\alpha\right) \leq 2^{t_0}K^{2t_0}2^{-t_0 m/2}$$

We now upper bound the total measure of S^α for $\#_p(\alpha) \leq t_0/2$.

CLAIM: For every $j \leq t_0$, $\mu^t(\bigcup_{\alpha\in\{p,q\}^{t_0}\colon\, \#_q(\alpha)=j} S^\alpha) \leq (1-\epsilon/2)^j \mu^t(R)$.

The claim is clearly true for $j=0$. For any $\alpha \in \{p,q\}^*$, by multiple applications of property 4,

$$\begin{aligned}\mu^t(\bigcup_{i\leq t_0-|\alpha|-1} S^{\alpha p^i q}) &= \sum_{i\leq t_0-|\alpha|-1} \mu^t(S^{\alpha p^i q})\\ &\leq \sum_{i\leq t_0-|\alpha|-1} (1-\epsilon/2)\left(\mu^t(S^{\alpha p^i}) - \mu^t(S^{\alpha p^{i+1}})\right) \leq (1-\epsilon/2)\mu^t(S^\alpha)\end{aligned}$$

since the sum telescopes. Let $Z_j = (p^*q)^j \cap \{p,q\}^{\leq t_0}$ be the set of all strings of length up to t_0 that end in a q and have a total of j q's. The above for $\alpha = \lambda$ implies that $\mu^t(\bigcup_{\beta\in Z_1} S^\beta) \leq (1-\epsilon/2)\mu^t(R)$. We can also apply the above to all $\alpha \in Z_j$ to yield that $\mu^t(\bigcup_{\beta\in Z_{j+1}} S^\beta) \leq (1-\epsilon/2)\mu^t(\bigcup_{\alpha\in Z_j} S^\alpha)$ and thus by induction that $\mu^t(\bigcup_{\alpha\in Z_j} S^\alpha) \leq (1-\epsilon/2)^j\mu^t(R)$. Finally, since $S^{\alpha p} \subseteq S^\alpha$ for any α we derive that

$$\mu^t\left(\bigcup_{\alpha\in\{p,q\}^{t_0}\colon\, \#_q(\alpha)=j} S^\alpha\right) = \mu^t(\bigcup_{\alpha\in Z_j} S^{\alpha p^{t_0-|\alpha|}}) \leq \mu^t\left(\bigcup_{\alpha\in Z_j} S^\alpha\right) \leq (1-\epsilon/2)^j\mu^t(R)$$

and the claim is proved.

Thus the total

$$\mu^t(\bigcup_{\alpha\in\{p,q\}^{t_0}\colon\, \#_p(\alpha)<t_0/2} S^\alpha) = \mu^t(\bigcup_{\alpha\in\{p,q\}^{t_0}\colon\, \#_q(\alpha)>t_0/2} S^\alpha) \leq (2/\epsilon)(1-\epsilon/2)^{t_0/2}\mu^t(R)$$

Putting it all together we have

$$\begin{aligned}\mu^t(R\cap(f^t)^{-1}(V_{T_0})) &\leq \mu^t(E^{t_0}) + \mu^t(\bigcup_{\alpha\in\{p,q\}^{t_0}} S^\alpha)\\ &= \mu^t(E^{t_0}) + \mu^t(\bigcup_{\substack{\alpha\in\{p,q\}^{t_0}\\ \#_p(\alpha)\geq t_0/2}} S^\alpha) + \mu^t(\bigcup_{\substack{\alpha\in\{p,q\}^{t_0}\\ \#_p(\alpha)<t_0/2}} S^\alpha)\\ &\leq 2t_0 2^{-mt} + 2^{t_0}K^{2t_0}2^{-t_0 m/2} + (2/\epsilon)(1-\epsilon/2)^{t_0/2}\mu^t(R).\end{aligned}$$

Since $-\log_2(1-\epsilon/6) > -\sqrt{2}\ln(1-\epsilon/6) \geq \sqrt{2}\epsilon/6$ and $\epsilon > 12mt/2^{m/8}$, $K = \lceil -mt/\log_2(1-\epsilon/6)\rceil < 2^{m/8}/2^{3/2}$ and therefore

$$2^{t_0}K^{2t_0}2^{-t_0 m/2} < 2^{-t_0 m/4}/2^{2t_0}.$$

Therefore, because the condition on ϵ implies that $m \geq 24$, if $\mu^t(R) \geq 2^{-t_0 m/6}$ then

$$\begin{aligned}\mu^t(R\cap(f^t)^{-1}(V_{T_0})) &< 2t_0 2^{-mt} + 2^{-t_0 m/4}/2^{2t_0} + (2/\epsilon)(1-\epsilon/2)^{t_0/2}\mu^t(R)\\ &< 2^{-t_0 m/4} + (2/\epsilon)(1-\epsilon/2)^{t_0/2}\mu^t(R)\\ &\leq 2^{-t_0 m/12}\mu^t(R) + (2/\epsilon)(1-\epsilon/2)^{t_0/2}\mu^t(R)\\ &\leq 2^{-t_0}\mu^t(R) + (2/\epsilon)(1-\epsilon/2)^{t_0/2}\mu^t(R)\\ &\leq (1-\epsilon/2)^{t_0/2}\mu^t(R) + (2/\epsilon)(1-\epsilon/2)^{t_0/2}\mu^t(R)\\ &\leq (3/\epsilon)(1-\epsilon/2)^{t_0/2}\mu^t(R)\end{aligned}$$

as required. □

The following is a direct product theorem for randomized communication complexity derived from corruption bounds on cross product distributions on rectangles.

Theorem 4.5. *Let $f : X \times Y \to \{0,1\}$ and let μ be a rectangular distribution on $X \times Y$. Let $b \in \{0,1\}$, $p = \mu(f^{-1}(b))$, and $\epsilon < p$ be given. There are constants $c, c' > 0$ and $\delta \le e^{-p\epsilon/144} < 1$ such that for any integer $t \le 2^{\mathrm{corrbd}^b_\mu(f,\epsilon)/16}/8$ such that $pt \ge 8$ and $\epsilon \ge \frac{18\ln(pt)}{pt}$,*

$$R^2_{1-\delta^t}(f^t) \ge D^{2,\mu^t}_{1-\delta^t}(f^t) \ge cpt \cdot \mathrm{corrbd}^b_\mu(f,\epsilon) - c'pt.$$

Proof. Assume without loss of generality that $b = 0$. Set $m = \mathrm{corrbd}^0_\mu(f,\epsilon)$. Set $O' = \{\vec{v} \in \{0,1\}^t \mid \vec{v} \text{ has} \ge pt/4 \text{ 0's}\}$, and let I_s be the set of all inputs $(\vec{x},\vec{y}) \in X^t \times Y^t$ such that $f^t(\vec{x},\vec{y})$ contains precisely s 0's. By definition $\mu^t(I_s) = \Pr[B(t,p) = s]$ where $B(t,p)$ is the binomial distribution that is the sum of t Bernoulli trials with success probability p. Therefore by a standard tail bound (Proposition 2.5) $\mu^t(\bigcup_{s<pt/4} I_s) \le 2^{-pt/2}$ and thus $\mu^t((f^t)^{-1}(O')) \ge 1 - 2^{-pt/2}$.

For simplicity, we will choose $c \le c'/72$ so that the bound will be trivial for $m \le 72$; we now assume that $m \ge 72$. Since $\frac{\ln(x)}{x}$ is a decreasing function of x for $x \ge 8$, $p \le 1$, $m \ge 72$, and $t \le 2^{m/16}/8$,

$$\frac{18\ln(pt)}{pt} \ge \frac{18\ln(t)}{t} \ge 144\ln(2)(m/16-3)2^{-m/16} > \frac{3}{2}m2^{-m/16} \ge 12mt2^{-m/8}.$$

It follows by hypothesis that $\epsilon > 12mt2^{-m/8}$ and so we may apply Theorem 4.2 with $t_0 = pt/4$. This shows that for every $\vec{v} \in O'$ we have $\mathrm{corrbd}^{\vec{v}}_{\mu^t}(f^t, 1-\gamma) \ge (pt/4)m/6 = ptm/24$ for $\gamma = (3/\epsilon)(1-\epsilon/2)^{pt/8}$.

Now define $\gamma' = (4/\epsilon)(1-\epsilon/2)^{pt/8}$ and let $g = f^t$. Because $\epsilon < p \le 1$, we have that

$$\begin{aligned}
\frac{1-\gamma'}{1-\gamma} &= 1 - \frac{(\gamma'-\gamma)}{1-\gamma} \\
&\le 1 - (\gamma'-\gamma) \\
&= 1 - (1/\epsilon)(1-\epsilon/2)^{pt/8} \\
&\le 1 - 2^{-pt/8} \\
&\le \mu(g^{-1}(O')).
\end{aligned}$$

Therefore, $1-\gamma' \le \mu(g^{-1}(O'))(1-\gamma)$ and we may apply Lemma 3.5 to obtain

$$R^2_{1-\gamma'}(g) \ge D^{2,\mu^t}_{1-\gamma'}(g) \ge \frac{ptm}{24} - \log_2(\frac{1}{\mu(g^{-1}(O') - (1-\gamma')/(1-\gamma))}).$$

Moreover,

$$\mu(g^{-1}(O')) - (1-\gamma')/(1-\gamma) \ge 1 - 2^{-pt/2} - (1 - 2^{-pt/8}) = 2^{-pt/8} - 2^{-pt/2} \ge 2^{-pt/2}$$

since $pt \ge 8$. Therefore $R^2_{1-\gamma'}(g) \ge D^{2,\mu^t}_{1-\gamma'}(g) \ge ptm/24 - pt/2$. Now since $\epsilon \ge \frac{18\ln(pt)}{pt}$,

$$\begin{aligned}
\gamma' &= (4/\epsilon)(1-\epsilon/2)^{pt/9}(1-\epsilon/2)^{pt/72} \\
&\le (4/\epsilon)e^{-\epsilon pt/18}(1-\epsilon/2)^{pt/72} \\
&\le \frac{4pt}{18\ln(pt)}e^{-\ln(pt)}(1-\epsilon/2)^{pt/72} \\
&\le (1-\epsilon/2)^{pt/72}.
\end{aligned}$$

Thus for $\delta = (1-\epsilon/2)^{p/72} \le e^{-p\epsilon/144} < 1$, we have $\gamma' \le \delta^t$ and choosing $c = 1/144$ and $c' = 1/2$ we obtain the claimed bound. (Note that by explicitly including an extra condition that $\epsilon > 12mt2^{-m/8}$ in the statement of the theorem we could have increased c to $1/24$.) □

We can show something even stronger than Theorem 4.5, namely that simply *approximating* f^t with significant probability requires a similar number of bits of communication.

Definition 4.6. *Let Δ be the usual Hamming distance on $\{0,1\}^t$. For $0 \le \alpha \le 1$ and $g, h : X^t \times Y^t \to \{0,1\}^t$ we say that g is an α-*approximation *of h if and only if for every $(\vec{x},\vec{y}) \in X^t \times Y^t$, $\Delta(g(\vec{x},\vec{y}), h(\vec{x},\vec{y})) \le \alpha t$; i.e. the function values differ on at most an α fraction of coordinates.*

Theorem 4.7. *Let $f : X \times Y \to \{0,1\}$ and let μ be a rectangular distribution on $X \times Y$. Let $b \in \{0,1\}$, $p = \mu(f^{-1}(b))$, and $0 < \epsilon < p$ be given. There are absolute constants $c, c', c'', c''', c'''' > 0$ such that for $0 < \alpha \le c'''\epsilon/\log_2(1/\epsilon)$, $\delta \le e^{-c''''\epsilon p} < 1$, and for any integer $t \le 2^{\mathrm{corrbd}^b_\mu(f,\epsilon)/16}/24$ such that $pt \ge 8$ and $\epsilon \ge \frac{c''\ln(pt)}{pt}$ and for any function $g : X^t \times Y^t \to \{0,1\}$ that is an αp approximation of f^t,*

$$R^2_{1-\delta^t}(g) \ge D^{2,\mu^t}_{1-\delta^t}(g) \ge cpt \cdot \mathrm{corrbd}^b_\mu(f,\epsilon) - c'pt.$$

Proof. The proof follows the outline of the proof of Theorem 4.5. Assume without loss of generality that $b = 0$ and set $m = \mathrm{corrbd}^0_\mu(f,\epsilon)$. Set $O' = \{\vec{v} \in \{0,1\}^t \mid \#_0(\vec{v}) \ge pt/4\}$. As above, $\mu^t((f^t)^{-1}(O')) \ge 1 - 2^{-pt/2}$. Let $O'' = \{\vec{v} \in \{0,1\}^t \mid \#_0(\vec{v}) \ge (1/4-\alpha)pt\}$. Since g is an αp approximation of f^t, $g^{-1}(O'') \supseteq (f^t)^{-1}(O')$ so $\mu^t(g^{-1}(O'')) \ge 1 - 2^{-pt/2}$.

Let $t_0 = (1/4 - 2\alpha)pt$. Since g is an αp approximation of f^t, for every input $(\vec{x},\vec{y}) \in O''$ the functions f^t and g agree on at least t_0 coordinates with value 0. Fix any $\vec{v} \in O''$. Let $S \subseteq \{1,\ldots,t\}$ be the set of 0 coordinates of $\vec{v}$ and $s = |S|$. Assume that $\alpha \le 1/24$; then $s \ge (1/4-\alpha)pt > pt/5$. Let $t_0 = s - \alpha pt$ which is $\ge s - 5\alpha s$.

Fix any rectangle R in $X^t \times Y^t$ with $\mu^t(R) \ge 2^{-t_0 m/6}$ We bound $\mu^t(g^{-1}(\vec{v}) \cap R)$. Let $(\vec{x},\vec{y}) \in g^{-1}(\vec{v})$. Since g is an αp approximation of f^t, $f^t(\vec{x},\vec{y})$ has value 0 on at least t_0 of the coordinates in S. There are at most $\binom{s}{5\alpha s} \le 2^{H_2(5\alpha)s}$ different ways to choose a set $T_0 \subseteq S$ of size t_0 where H_2 is the binary entropy function. For each set $T_0 \subseteq S$, by the properties of our parameters as in the previous proof, we can apply Theorem 4.2 to f (this time using part (a)) to show that

$$\begin{aligned}\mu^t((f^t)^{-1}(V_{T_0}) \cap R) &\le (3/\epsilon)(1-\epsilon/2)^{t_0/2}\mu^t(R)\\ &\le (3/\epsilon)(1-\epsilon/2)^{s(1-5\alpha)/2}\mu^t(R)\end{aligned}$$

where $V_{T_0} = \{\vec{v'} \in \{0,1\}^t \mid v'_i = 0 \text{ for all } i \in T_0\}$. By construction

$$g^{-1}(\vec{v}) \subseteq \bigcup_{T_0 \subseteq S,\ |T_0| = t_0} (f^t)^{-1}(V_{T_0}).$$

Therefore,

$$\begin{aligned}
\mu^t(g^{-1}(\vec{v}) \cap R) &\leq \sum_{T_0 \subseteq S,\ |T_0|=t_0} \mu^t((f^t)^{-1}(V_{T_0}) \cap R) \\
&\leq \sum_{T_0 \subseteq S,\ |T_0|=t_0} (3/\epsilon)(1-\epsilon/2)^{s(1-5\alpha)/2}\mu^t(R) \\
&\leq 2^{H_2(5\alpha)s}(3/\epsilon)(1-\epsilon/2)^{s(1-5\alpha)/2}\mu^t(R) \\
&= (3/\epsilon)[(1-\epsilon/2)^{(1-5\alpha)/2}2^{H_2(5\alpha)}]^s\mu^t(R) \\
&\leq (3/\epsilon)[(1-\epsilon/2)^{(1-5\alpha)/2}2^{H_2(5\alpha)}]^{pt/5}\mu^t(R).
\end{aligned}$$

Therefore we have $\mu^t(g^{-1}(O'')) \geq 1 - 2^{-pt/2}$ and for any $\vec{v} \in O''$ we have $\mathrm{corrbd}^{\vec{v}}_{\mu^t}(g, 1-\gamma) \geq t_0 m/6 = ptm/30$ for $\gamma \leq (3/\epsilon)[(1-\epsilon/2)^{(1-5\alpha)/2}2^{H_2(5\alpha)}]^{pt/5}$.

Now for $\alpha \leq c'''\epsilon/\log_2(1/\epsilon)$ for a sufficiently small constant $c''' > 0$, the quantity $(1-\epsilon/2)^{(1-5\alpha)/2}2^{H_2(5\alpha)}$ is at most $e^{-c^*\epsilon}$ for some constant $c^* > 0$. Then, by an analogous argument to one in the previous proof we may apply Lemma 3.5 to g and use our assumptions on the parameters to obtain that

$$R^2_{1-\gamma'}(g) \geq D^{2,\mu^t}_{1-\gamma'}(g) \geq cptm - c'pt$$

for suitable constants $c, c' > 0$ and for $\gamma' \leq \delta^t$ for some $\delta \leq e^{-c''''\epsilon p} < 1$. This proves the theorem. □

Disjointness

Recall the disjointness predicate $\textsc{Disj}_{2,n} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ such that $\textsc{Disj}_{2,n}(x, y) = 1$ if and only if $x \cap y \neq \emptyset$.

Let μ be the rectangular distribution on $X \times Y = \{0,1\}^n \times \{0,1\}^n$ given by $\Pr_\mu[x_i = 1] = \Pr_\mu[y_i = 1] = n^{-1/2}$ independently for $(x, y) \in X \times Y$. Babai, Frankl, and Simon [2] proved the following corruption lower bound on $\textsc{Disj}_{2,n}$ under distribution μ.

Proposition 4.8. *[Babai, Frankl, Simon[2]] Let μ be the rectangular distribution defined as above. Then $\mu(\textsc{Disj}^{-1}_{2,n}(0))$ is $\Omega(1)$ and for any sufficiently small constant $\epsilon > 0$, $\mathrm{corrbd}^0_\mu(\textsc{Disj}_{2,n}, \epsilon)$ is $\Omega(\sqrt{n})$.*

Combining the Proposition 4.8 with Theorems 4.2 and 4.5 gives the following corollary.

Corollary 4.9. *There is a $\delta < 1$ and a constant $c > 0$ such that for $t \leq 2^{c\sqrt{n}}$ the following hold:*

(a) Let μ be defined as above. There is a constant $c' > 0$ such that for any $\vec{v} \in \{0,1\}^t$ with $\#_0(\vec{v}) \geq t_0$, $\mathrm{corrbd}^{\vec{v}}_{\mu^t}(\textsc{Disj}^t_{2,n}, 1-\delta^t) \geq c' t_0 \sqrt{n}$.

(b) $R^2_{1-\delta^t}(\textsc{Disj}^t_{2,n})$ is $\Omega(t\sqrt{n})$.

Remark 1. *Using the direct sum property for conditional information complexity and the lower bound of [7], for fixed error $\epsilon < 1$ one can obtain the bound $R^2_\epsilon(\textsc{Disj}^t_{2,n})$ is $\Omega(tn)$. However this bound is incomparable to the above corollary because the direct product result guarantees that correctness is at most $(1-\epsilon)^{\Omega(t)}$ whereas the direct sum result only guarantees that correctness is at most $1 - \epsilon$.*

5 3-party Number-on-the-forehead Communication Complexity of Disjointness

We consider the computation of $\textsc{Disj}_{3,n}$ in two models, the randomized $Z \to (Y \leftrightarrow X)$ model and the general 3-party model.

5.1 $Z \to (Y \leftrightarrow X)$ Protocols

Nisan and Wigderson [26] suggested the study of 3-party one way communication complexity as a potential approach to obtaining size-depth trade-offs in circuit complexity. In particular, they proved lower bounds on the communication complexity of functions of the form $f(x, h, i) = h(x)_i$, where x is drawn from a set X, h from a family H of universal hash functions from X to $\{0, 1\}^n$, and i from $[n]$. Their lower bound argument also applies to $Z \to (Y \leftrightarrow X)$ protocols for $Z = [n]$ and $Y = H$. Using our new direct product results on corruption we apply a similar argument to yield lower bounds for $\textsc{Disj}_{3,n}$ in this model.

Theorem 5.1. *$D^{Z\to(Y\leftrightarrow X)}(\textsc{Disj}_{3,n})$ is $\Omega(n^{1/3})$ and for $\epsilon < 1/2$, $R_\epsilon^{Z\to(Y\leftrightarrow X)}(\textsc{Disj}_{3,n})$ is $\Omega((1-2\epsilon)^2 n^{1/3})$.*

Proof. We follow the general approach of [26] but use a direct product bound for corruption in place of a discrepancy bound for universal hash function families. Note that although the basic approach and bound of [26] is correct, there is an issue with the proof in [26] that is discussed and corrected below.

Fix any $Z \to (Y \leftrightarrow X)$ protocol P computing $\textsc{Disj}_{3,n}$ and let $C(P)$ be the total number of bits communicated in P. Let $t = n^{1/3}$. View each string x, y, z as a sequence of t blocks, $x_1, \ldots, x_t, y_1, \ldots, y_t, z_1, \ldots, z_t \in \{0, 1\}^{n/t}$.

Given P we first construct a $Z \to (Y \leftrightarrow X)$ protocol P' that computes $(\textsc{Disj}_{2,n/t}(x_1, y_1), \ldots, \textsc{Disj}_{2,n/t}(x_t, y_t))$ in which the Z-player sends $C(P)$ bits and the X and Y players together send $tC(P)$ bits: Consider runs of the protocol P with different choices of $z \in Z$, in particular with $z^j = 0^{(j-1)n/t}1^{n/t}0^{(t-j)n/t}$ for $j = 1, \ldots, t$. For $z = z^j$, $\textsc{Disj}_{3,n}(x, y, z) = \textsc{Disj}_{2,n/t}(x_j, y_j)$. Also observe that for each of these choices, the message $m_Z(x, y)$ sent by the Z-player is independent of the choice of z. On input (x, y), the new protocol P' simulates P on inputs (x, y, z^j) for $j = 1, \ldots, t$ except that, since the message sent by the Z-player is the same in each case, the Z-player sends this message only once. P' then outputs the tuple of results.

The function computed by P' does not depend on the choice of z, so it can be viewed as a two-player protocol with advice for computing $\textsc{Disj}^t_{2,n/t}(x, y)$. Define a protocol P'' in which the Z-player receives (x, y) as input as before but the X player only receives x and the Y player only receives y. (To conform with the standard two-player notation, we say that player X can see input x and player Y can see input y.) The Z player sends the message that he would under protocol P'. After the Z-player's communication of $C(P)$ bits, the X- and Y-players exchange $tC(P)$ bits in order to compute $\textsc{Disj}^t_{2,n/t}(x, y)$.

Consider the distribution ν on $X \times Y \times Z$ in which we choose z uniformly at random from $\{z^j \mid j \in [t]\}$, and independently set each bit of x and each bit of y to 0 with probability $1 - n^{-1/3}$ and to 1 with probability $n^{-1/3}$. Observe that the induced distribution on $X^t \times Y^t$ given by ν is $\mu^t_{n/t}$ where $\mu_{n/t} = \mu_{n^{2/3}}$ is the distribution μ used in Proposition 4.8 for input strings of length $n/t = n^{2/3}$. Let $p = \Pr_\nu[\textsc{Disj}_{2,n/t}(x_j, y_j) = 0] = \Pr_{\mu_{n/t}}[\textsc{Disj}_{2,n/t}(x_j, y_j) = 0]$, the probability that x and y intersect in block j (which is independent of j) and observe that $p = (1 - n^{-2/3})^{n^{2/3}} = \Omega(1)$.

Since the set of possible messages is prefix-free and $|m_z| \leq C(P)$, there is some m_z such that $\Pr_\nu[m_Z(x, y) = m_z] \geq 2^{-C(P)}$. Fix that m_z.

At this point in [9] we gave a direct argument using Theorem 4.2 to derive the claimed lower bound. Here, we apply Theorem 4.5 instead. Let $S_{m_z} \subseteq X \times Y$ be the set of inputs on which $m_Z(x, y) = m_z$. Define a deterministic 2-party protocol P''_{m_z} of complexity $t \cdot C(P)$ on $X \times Y$ that is given by protocol P'' with the advice given by communication $m_Z = m_z$ fixed. Since P'' is always correct, P''_{m_z} correctly computes $\textsc{Disj}^t_{2,n/t}$ on S_{m_z}. Now by our choice of m_z, the measure of S_{m_z} within $X \times Y$ satisfies

$$\mu^t_{n/t}(S_{m_z}) = \Pr_\nu[m_Z(x, y) = m_z] \geq 2^{-C(P)}$$

and thus P''_{m_z} correctly computes $\textsc{Disj}^t_{2,n/t}$ on a set with $\mu^t_{n/t}$ measure at least $2^{-C(P)}$. Let $\epsilon < p$ be a sufficiently small positive constant that Proposition 4.8 applies and that also satisfies $\epsilon \geq \frac{9\ln(pt)}{pt}$. By Proposition 4.8 and Theorem 4.5, there are constants c, c' and $\delta < 1$ such that

$$\begin{aligned} D^{2,\mu^t_{n/t}}_{1-\delta^t}(\textsc{Disj}^t_{2,n/t}) &\geq cpt \cdot \mathrm{corrbd}^0_{\mu_{n/t}}(\textsc{Disj}_{2,n/t}, \epsilon) - c'pt \\ &\geq c''t\sqrt{n/t} \end{aligned}$$

for some constant $c'' > 0$. This says that no algorithm that sends fewer than $c''t\sqrt{n/t}$ bits can correctly compute $\textsc{Disj}^t_{2,n/t}$ on at least a δ^t measure of inputs under $\mu^t_{n/t}$. Thus, either $2^{-C(P)} < \delta^t$ or $C(P''_{m_z}) = t \cdot C(P) \geq c''t\sqrt{n/t}$. It follows that $C(P)$ is $\Omega(\min\{t, \sqrt{n/t}\})$ which is $\Omega(n^{1/3})$ since $t = n^{1/3}$.

One can use a similar argument in the case of randomized complexity to derive a lower bound of the form $\Omega((1 - 2\epsilon)^2 n^{1/3}/\log n)$ by first applying Lemma 2.6 to reduce the probability of error below $1/(4t)$ and then applying Yao's lemma with distribution ν to obtain a protocol that correctly computes $\textsc{Disj}^t_{2,n/t}$ on at least $3/4$ of the $\mu^t_{n/t}$ measure of $X \times Y$ and then fix a popular communication m_z on which a 2-party protocol has large success to derive a bound as in the deterministic case. There is a $\Theta(\frac{\log t}{(1-2\epsilon)^2}) = \Theta(\frac{\log n}{(1-2\epsilon)^2})$ factor lost compared to the deterministic case due to the amount of amplification required.

Instead, in the case of ϵ error randomized complexity we apply an argument based on Theorem 4.7 instead of Theorem 4.5. Let $\alpha = c'''\epsilon/\log_2(1/\epsilon) > 0$ where $c''' > 0$ is the constant in Theorem 4.7. We apply Lemma 2.6 to reduce the error in randomized protocol P from ϵ to $\epsilon' = \alpha p/4$. This increases the communication complexity by a factor that is $O\left(\frac{1}{(1-2\epsilon)^2}\right)$. We then use Yao's lemma with the distribution ν to derive a deterministic protocol P^* with complexity $C(P^*)$ that is $O(\frac{1}{(1-2\epsilon)^2}C(P))$ and has error at most ϵ' over the distribution ν.

We apply the argument from the deterministic case with P^* replacing P to obtain a protocol P'' computing $\textsc{Disj}^t_{2,n/t}(x, y)$ in which the Z-player sends $C(P^*) = O(\frac{1}{(1-2\epsilon)^2}C(P))$ bits based on (x, y) and the X and Y players interact sending a total of $tC(P^*)$ bits based on x and y respectively. Now, in constrast with the simpler argument for randomized protocols sketched above, the error in P^* is too large to guarantee that the protocol P'' computes $\textsc{Disj}^t_{2,n/t}$ on any portion of the input space. However, we see that for most inputs P'' produces a good approximation of $\textsc{Disj}^t_{2,n/t}$. Let $G = \{(x, y) \in X \times Y \mid \Delta(P''(x, y), \textsc{Disj}^t_{2,n/t}(x, y)) \leq \alpha pt\}$. Since P^* has error at most $\epsilon' = \alpha p/4$ under ν and ν gives all t of the z^j equal measure independent of the probability it assigns to x and y, by Markov's inequality at most a 1/4 measure of (x, y) under ν have more than $4\epsilon' t = \alpha pt$ inputs z^j for which P'' on input (z^j, x, y) does not output $\textsc{Disj}_{2,n/t}(x_j, y_j)$. Therefore $\mu^t_{n/t}(G) \geq 3/4$.

For each binary string m of length at most $C(P)$, let $S_m = \{(x, y) \mid m_Z(x, y) = m\}$; these sets partition $X \times Y \supseteq G$. Let $M = \{m \mid \mu^t_{n/t}(S_m \cap G) \geq \mu^t_{n/t}(S_m)/2\}$. Since $\mu^t_{n/t}(G) \geq 3/4$, by Markov's inequality we have that $\mu^t_{n/t}(\bigcup_{m \in M} S_m) \geq 1/2$. Because there are only $2^{C(P^*)}$ choices of m, we may choose $m_z \in M$ so that $\mu^t_{n/t}(S_{m_z}) \geq 2^{-C(P^*)-1}$ and thus $\mu^t_{n/t}(S_{m_z} \cap G) \geq 2^{-C(P^*)-2}$. Fix this m_z.

As above we consider the deterministic 2-party protocol P''_{m_z} which has complexity $t \cdot C(P^*)$. By construction, for every input $(x, y) \in S_{m_z} \cap G$, we have $\Delta(P''_{m_z}(x, y), \text{DISJ}^t_{2,n/t}(x, y)) \leq \alpha p$. Thus there is a function g that is an αp approximation to $\text{DISJ}^t_{2,n/t}$ such that P''_{m_z} computes g on every input in $S_{m_z} \cap G$ which is a set of measure at least $2^{-C(P^*)-2}$ under $\mu^t_{n/t}$. Applying Theorem 4.7 instead of Theorem 4.5, by the same argument as in the deterministic case we have that either $2^{-C(P^*)-2} < \delta^t$ or $tC(P^*) \geq c''t\sqrt{n/t}$ and thus $C(P^*)$ is $\Omega(n^{1/3})$. Therefore $C(P)$ is $\Omega((1-2\epsilon)^2 n^{1/3})$ as required. □

5.2 General 3-party Number-on-the-forehead Computation

In this section we prove an $\Omega(\log n)$ lower bound on the unrestricted three-party number-on-the-forehead communication complexity of $\text{DISJ}_{3,n}$. Although this is not yet strong enough to imply lower bounds for lift-and-project proof systems it is of independent interest since it uses a multiparty number-on-the-forehead corruption bound that does not follow from a discrepancy bound.

Theorem 5.2. *For any $\epsilon < 1/2$, $R^3_\epsilon(\text{DISJ}_{3,n})$ is $\Omega((1-2\epsilon)^2 \log n)$.*

To prove this theorem we use the following simple characterization of three-dimensional cylinder intersections.

Proposition 5.3. *A set E is a three-dimensional cylinder intersection on $X \times Y \times Z$ if and only if there is a family of combinatorial rectangles $R_z \in \mathcal{P}(X) \times \mathcal{P}(Y)$, for $z \in Z$, and a set $S \subseteq X \times Y$ such that $E = \bigcup_{z \in Z}((R_z \cap S) \times \{z\})$.*

Proof. "If": Let E be a set of the form $E = \bigcup_{z \in Z}((R_z \cap S) \times \{z\})$. For each $z \in Z$, choose $X_z \subseteq X$ and $Y_z \subseteq Y$ so that $R_z = X_z \times Y_z$. Set $C_X = \{(x, y, z) \in X \times Y \times Z \mid y \in Y_z\}$, $C_Y = \{(x, y, z) \in X \times Y \times Z \mid x \in X_z\}$, $C_Z = \{(x, y, z) \in X \times Y \times Z \mid (x, y) \in S\}$. Clearly C_X is an X-cylinder, C_Y is a Y-cylinder, and C_Z is a Z-cylinder. Moreover, $(x, y, z) \in C_X \cap C_Y$ if and only if $(x, y) \in X_z \times Y_z = R_z$. Therefore, $(x, y, z) \in C_X \cap C_Y \cap C_Z$ if and only if $(x, y, z) \in (R_z \cap S) \times \{z\}$.

"Only if": Let E be a three-dimensional cylinder intersection. By definition, E is the intersection of an X-cylinder C_X, a Y-cylinder C_Y, and a Z-cylinder C_Z. For each $z \in Z$, let $X_z = \{x \mid \exists y\ (x, y, z) \in C_Y\}$ and $Y_z = \{y \mid \exists x\ (x, y, z) \in C_X\}$ and $R_z = X_z \times Y_z$. Because C_X is an X-cylinder and C_Y is a Y-cylinder, for each $z \in Z$, $(x, y, z) \in C_X \cap C_Y$ if and only $(x, y) \in X_z \times Y_z = R_z$. Write $C_Z = S \times Z$ for some $S \subseteq X \times Y$. We now have that $(x, y, z) \in C_X \cap C_Y \cap C_Z$ if and only if $(x, y, z) \in (R_z \times \{z\}) \cap (S \times Z) = (R_z \cap S) \times \{z\}$. □

Proof of Theorem 5.2. Let $t = n^{1/3}$. Define a distribution ν on $X \times Y \times Z$ as follows: Choose z uniformly at random from $\{z^j = 0^{(j-1)(n/t)} 1^{n/t} 0^{(t-j)n/t} \mid j \in [t]\}$, and independently set each bit of x and each bit of y to 0 with probability $1 - n^{-1/3}$ and to 1 with probability $n^{-1/3}$. Clearly $\nu(\text{DISJ}^{-1}_{3,n}(0)) = (1 - n^{-2/3})^{n^{2/3}} = \Omega(1)$. Set $p = \nu(\text{DISJ}^{-1}_{3,n}(0))$.

Let Γ be the set of all cylinder intersections on $X \times Y \times Z$. We prove that for all $\epsilon' < 1$, $\epsilon'\text{-mono}^0_{\nu,\Gamma}(\text{DISJ}_{3,n}) = O(n^{-1/3} \log n)$. The claimed lower bound then follows by applying Proposition 2.6 to reduce the error below $p/2$ and then applying Corollary 3.6 with $\epsilon' = p/2$.

Let $\epsilon' < 1$ be given. Let E be a cylinder intersection in $X \times Y \times Z$. Apply Proposition 5.3 and write $E = \bigcup_{z \in Z}((S \cap R_z) \times \{z\})$ for $S \subseteq X \times Y$ and R_z rectangles on $X \times Y$. Suppose that $\nu(\text{DISJ}^{-1}_{3,n}(1) \cap E) \leq \epsilon' \cdot \nu(E)$. It is sufficient prove that $\nu(E)$ is $O(n^{-1/3} \log n)$.

Because the support of ν is $\{0,1\}^n \times \{0,1\}^n \times \{z^j \mid j \in [t]\}$, we may assume without loss of generality that $E = \bigcup_{j=1}^{t} (S \cap R_{z^j}) \times \{z^j\}$. For each $j \in [t]$, all $x \in \{0,1\}^n$, all $y \in \{0,1\}^n$, set $x^j = x \cap z^j$ and $y^j = y \cap z^j$. For each $(x,y) \in S$ let $J_{(x,y)} \subseteq [t]$ be the set of $j \in [t]$ for which $(x,y) \in R_{z_j}$ and $\textsc{Disj}_{3,n}(x,y,z^j) = 0$. This implies that for all $j \in J(x,y)$, $\textsc{Disj}_{2,n/t}(x^j,y^j) = 0$. Let $t_0 = \lceil (1-\epsilon')\nu(E)t/2 \rceil$ and let $S' = \{(x,y) \in S \mid |J_{(x,y)}| \geq t_0\}$. Let $E' = \{(x,y,z^j) \in E \mid (x,y) \in S'\}$ be the set of elements of E whose (x,y) components are in S'. Notice that E' is a cylinder-intersection. Let μ be the measure induced on $X \times Y$ by ν.

$$\nu((E - E') \cap \textsc{Disj}_{3,n}^{-1}(0)) \leq \sum_{j=1}^{t} \nu\left(\left(\left(R_{z_j} \cap S\right) \times \{z_j\}\right) - \left(\left(R_{z_j} \cap S'\right) \times \{z_j\}\right)\right)$$

and since $\nu(S) = \sum_{(x,y)} \frac{|J_{(x,y)}|}{t} \mu(\{(x,y)\})$ this is

$$\begin{aligned} &\leq \frac{t_0 - 1}{t} \mu(S - S') \\ &< \frac{(1-\epsilon')\nu(E)t/2}{t} \mu(S) \\ &\leq (1-\epsilon')\nu(E)/2 \end{aligned}$$

By the error assumption for E, $\nu(E \cap \textsc{Disj}_{3,n}^{-1}(0)) \geq (1-\epsilon')\nu(E)$. Therefore $\nu(E') \geq \nu(E' \cap \textsc{Disj}_{3,n}^{-1}(0)) \geq (1-\epsilon')\nu(E)/2$ and thus $\mu(S') \geq \nu(E') \geq (1-\epsilon')\nu(E)/2$.

We now break up the rectangles R_{z_j} into smaller rectangles to partition E' into a family of sub-cylinder-intersections. For $j = 1, \ldots, t$ write $R_{z_j} = A_j \times B_j$ for $A_j \subseteq X$ and $B_j \subseteq Y$. For $\alpha, \beta \in \{0,1\}^t$ define the rectangle

$$R_{\alpha,\beta} = (\bigcap_{j:\alpha_j=1} A_j \cap \bigcap_{j:\alpha_j=0} \overline{A_j}) \times (\bigcap_{j:\beta_j=1} B_j \cap \bigcap_{j:\beta_j=0} \overline{B_j}).$$

Some simple facts follow immediately from the construction of the $R_{\alpha,\beta}$'s:

1. For $(\alpha,\beta) \neq (\alpha',\beta')$, $R_{\alpha,\beta} \cap R_{\alpha',\beta'} = \emptyset$
2. For each $j \in [t]$, $R_{z_j} = \bigcup_{\alpha:\alpha_j=1} \bigcup_{\beta:\beta_j=1} R_{\alpha,\beta}$
3. $E' = \bigcup_{j=1}^{t} \bigcup_{\alpha:\alpha_j=1} \bigcup_{\beta:\beta_j=1} (R_{\alpha,\beta} \cap S') \times \{z_j\}$
4. For all (α,β), and all $(x,y),(x',y') \in R_{\alpha,\beta}$, $\{j \in [t] \mid (x,y) \in R_{z_j}\} = \{j \in [t] \mid (x',y') \in R_{z_j}\}$

As a corollary of Property 4, each $R_{\alpha,\beta}$ has an associated set $J_{\alpha,\beta} \subseteq [t]$, $|J_{\alpha,\beta}| \geq t_0$, such that for all $(x,y) \in R_{\alpha,\beta} \cap S'$, and all $j \in J_{\alpha,\beta}$, $\textsc{Disj}_{3,n}(x,y,z^j) = 0$. This implies that for all $j \in J_{\alpha,\beta}$, and all $(x,y) \in R_{\alpha,\beta} \cap S'$, $\textsc{Disj}_{2,n/t}^{t}(x^j,y^j) = 0$.

By Corollary 4.9(a) there are some constants $c, \delta > 0$ and such that for any α, β if $\mu(R_{\alpha,\beta}) \geq 2^{-ct_0\sqrt{n/t}}$ then $\mu(R_{\alpha,\beta} \cap S') \leq \delta^{t_0} \mu(R_{\alpha,\beta})$. Since there are 2^{2t} choices of (α,β), by the union bound, at most $2^{2t - ct_0\sqrt{n/t}}$ measure of points in S' can be covered by rectangles $R_{\alpha,\beta}$ for which $\mu(R_{\alpha,\beta}) < 2^{-ct_0\sqrt{n/t}}$. Since the rectangles $R_{\alpha,\beta}$ covering S' are disjoint, by the corruption bound the total measure of the part of S' covered by rectangles $R_{\alpha,\beta}$ with $\mu(R_{\alpha,\beta}) \geq 2^{-ct_0\sqrt{n/t}}$ is at most δ^{t_0}. Therefore $\mu(S') \leq \delta^{t_0} + 2^{2t - ct_0\sqrt{n/t}}$ which, for $t = n^{1/3}$, is at most $\delta^{t_0} + 2^{-(ct_0-2)t}$. Therefore

$$(1-\epsilon')\nu(E)/2 \leq \delta^{t_0} + 2^{-(ct_0-2)t}.$$

By definition $t_0 \geq (1-\epsilon')\nu(E)t/2$. If $ct_0 < 3$ then $\mu(E)$ is $O(1/t) = O(n^{-1/3})$ and we are done. Otherwise, since $t_0 \leq t$ we have constants $c_1, c_2 > 0$ such that $\nu(E) \leq c_1 2^{-c_2\nu(E)t}$. Taking logarithms yields $\log_2 \nu(E) \leq -c_2\nu(E)t + c_3$ for some constant c_3. Thus $\frac{1}{\nu(E)} \log_2 \frac{1}{\nu(E)}$ is $\Omega(t)$ It follows that $\nu(E)$ is $O(\frac{\log t}{t}) = O(\frac{\log n}{n^{1/3}})$ as required. □

Observe that the corruption bound under the distribution used in the proof of Theorem 5.2 is asymptotically tight: The X or Y player sends $\lceil \log_2 t \rceil$ bits specifying the value of j and then the Z player computes $\text{DISJ}_{3,n}(x, y, z^j)$. There are natural distributions for which we doubt that the corruption bound of Theorem 5.2 is tight. For example, the distribution that independently sets each bit of each string, with each bit set to 1 with probability $n^{-1/3}$ and 0 with probability $1 - n^{-2/3}$. The $\Omega(\log n)$ corruption bound holds in this case as well, although the proof is a little more involved. Distributions such as this may have potential utility in deriving super-logarithmic lower bounds, although we have not yet been able use them to derive such bounds. The key limitation of the method of proof of Theorem 5.2 is the step in which we refine of the set of rectangles.

6 k-party Number-on-the-Forehead Communication Complexity

In this section, we establish an $\Omega(n^{1/(k-1)}/(k-1))$ lower bound for the case of randomized simultaneous communication and use this to derive an $\Omega((\log n)/(k-1))$ lower bound for the general randomized number-on-the-forehead model.

6.1 Simultaneous k-party Number-on-the-forehead Computation

The communication complexity of disjointness in the number-on-the-forehead simultaneous messages model can be analyzed using the techniques of Babai, Gal, Kimmel and Lokam [3]. Following [3] we directly analyze the complexity of this problem in the slightly stronger model in which one player, player k, receives simultaneous communication from the other players and outputs an answer based on their communication and input $x_k \in X_k$; clearly $R_\epsilon^{X_1\|\ldots\|X_k}(f) \geq R_\epsilon^{(X_1\|\ldots\|X_{k-1})\to X_k}(f)$.

The key idea of the approach in [3] is to find a small collection of possible inputs Q_i in each of the input sets $X_i = \{0,1\}^n$, for $i \in [k-1]$, with the property that taking all their combinations together yields a large number of different subproblems player k might need to solve. The only information that player k receives about x_k is from the other players so the information from all their possible messages must be enough to differentiate among these possibilities.

Definition 6.1. *For C and D subsets of $\{0,1\}^n$ write $C \sqcap D = \{x \cap y \mid x \in C,\ y \in D\}$.*

Proposition 6.2. *For $\ell \geq 1$ there exist $Q_1, \ldots, Q_\ell \subseteq \{0,1\}^n$ such that $|Q_i| = n^{1/\ell}$ and $Q_1 \sqcap \cdots \sqcap Q_\ell$ is the set of all singleton sets in $[n]$.*

Proof. Let $m = n^{1/\ell}$ and view $[n]$ as an ℓ-dimensional cube with sides of size m. Let $Q_i = \{Q_{i,1}, \ldots, Q_{i,m}\}$ be the partition of $[n]$ into subsets of size $m^{\ell-1}$ given by the m layers along the i-th dimension in this cube. Since the different sets within each Q_i are disjoint, all-nonempty sets in $Q_1 \sqcap \cdots \sqcap Q_\ell$ are disjoint. An element $j \in [n]$ can be indexed by its coordinates $(j_1, \ldots, j_\ell)$ in each of the ℓ dimensions of this cube. Clearly $\{j\} = Q_{1,j_1} \cap Q_{2,j_2} \cap \cdots \cap Q_{\ell,j_\ell}$. □

Let H be the binary entropy function and for $0 \leq \epsilon \leq 1$ define $H_2(\epsilon) = \epsilon \log_2 \frac{1}{\epsilon} + (1-\epsilon)\log_2 \frac{1}{1-\epsilon}$. Our argument uses basic properties of these functions that can be found for example in [15].

Theorem 6.3.
$$R_\epsilon^{(X_1||\cdots||X_{k-1})\to X_k}(f) \geq (1 - H_2(\epsilon))n^{1/(k-1)}/(k-1).$$

Proof. We apply Yao's lemma and analyze the complexity $C(P)$ of an ϵ-error deterministic protocol P under distribution μ given as follows: Apply Proposition 6.2 with $\ell = k-1$ to obtain sets $Q_1, \ldots, Q_{k-1} \subseteq \{0,1\}^n$ with $|Q_i| = m = n^{1/(k-1)}$ such that $Q_1 \sqcap \cdots \sqcap Q_{k-1}$ contains all singleton subsets of $[n]$. For each $j \in [n]$ we can identify a (unique) tuple $\vec{x}^j = (x_1^j, \ldots, x_{k-1}^j) \in Q_1 \times \cdots \times Q_{k-1}$ such that $\{j\} = x_1^j \cap \cdots \cap x_{k-1}^j$. Define distribution μ on $X_1 \times \ldots \times X_k$ by by choosing j uniformly at random from $[n]$ and independently choosing a uniformly random subset $x_k \subseteq [n]$ to produce the tuple $(x_1^j, \ldots, x_{k-1}^j, x_k)$.

Observe that for inputs in the support of μ, $\textsc{Disj}_{k,n}(\vec{x}^j, x_k) = 1$ if and only if $j \in x_k$. It follows that the vector $(\textsc{Disj}_{k,n}(\vec{x}^j, x_k))_{j\in[n]}$ completely determines x_k. If the protocol P were always correct, then we could encode x_k by listing all the possible messages that could be sent by players $1, \ldots, k-1$ for any of the possible extensions $\vec{x}^j$ on the first j coordinates since these would be sufficient to determine the values of $\{\textsc{Disj}_{k,n}(\vec{x}^j, x_k)\}_{j\in[n]}$ and thus the bits of x_k. Although there are $n = m^{k-1}$ different extensions of x_k, for each player $1, \ldots, k-1$, given x_k there are only $m^{k-2} = n^{1-1/(k-1)}$ different messages possible since player i's message does not depend on the i-th coordinate. Thus the total number of bits required would be at most $(k-1)n^{1-1/(k-1)}C(P)$ which must be at least n since they are sufficient to encode x_k and we would obtain $C(P) \geq n^{1-1/(k-1)}/(k-1)$.

Since P has error at most ϵ this vector $\vec{v}$ of possible messages is sufficient to determine each bit of x_k with error at most ϵ under distribution μ. Let $\mathbf{X}_k$ be random variable for the string x_k as selected by the distribution μ, and let $\vec{\mathbf{V}}$ be the random variable for the strings $\vec{v}$ as selected by μ. By Fano's inequality, for each $j \in [n]$, the entropy $\mathrm{H}(\mathbf{X}_{k,j} \mid \vec{\mathbf{V}}) \leq H_2(\epsilon)$. Thus by the sub-additivity of entropy, $\mathrm{H}(\mathbf{X}_k \mid \vec{\mathbf{V}}) \leq H_2(\epsilon)n$. Therefore

$$n = \mathrm{H}(\mathbf{X}_k) \leq \mathrm{H}(\vec{\mathbf{V}}) + \mathrm{H}(\mathbf{X}_k \mid \vec{\mathbf{V}}) \leq (k-1)n^{1-1/(k-1)}C(P) + H_2(\epsilon)n$$

Rearranging, we have $(k-1)n^{1-1/(k-1)}C(P) \geq (1 - H_2(\epsilon))n$ which yields the claimed bound. □

6.2 General k-party number-on-the-forehead Computation

We obtain lower bounds for general k-party number-on-the-forehead communication complexity as a simple consequence of Theorem 6.3 using a simulation of general protocols by simultaneous protocols.

Theorem 6.4. *For any $\epsilon < 1/2$, $R_\epsilon^k(\textsc{Disj}_{k,n})$ is $\frac{\log_2 n}{k-1} - O(1)$.*

Proof. Given an ϵ-error k-party number-on-the-forehead protocol P for $\textsc{Disj}_{k,n}$ of communication cost $R_\epsilon^k(\textsc{Disj}_{k,n})$, define a simultaneous protocol P' for $\textsc{Disj}_{k,n}$ as follows: Each player sends a vector of length $2^{R_\epsilon^k(\textsc{Disj}_{k,n})}$ of all bits that the player would have sent in protocol P for every prefix of communications in which it is his turn to speak. An application of Theorem 6.3 shows that:

$$2^{R_\epsilon^k(\textsc{Disj}_{k,n})} \geq (1 - H_2(\epsilon))n^{1/(k-1)}/(k-1)$$

and thus

$$R_\epsilon^k(\textsc{Disj}_{k,n}) \geq \log_2\Big((1 - H_2(\epsilon))n^{1/(k-1)}/(k-1)\Big) \geq \frac{\log_2 n}{k-1} - \log_2\left(\frac{k-1}{1 - H_2(\epsilon)}\right) = \Omega\left(\frac{\log n}{k-1}\right)$$

□

7 Discussion

Gievn the proximity of our $\Omega(\log n)$ lower bounds to the $\omega(\log^4 n)$ or $\omega(\log^2 n(\log\log n)^2)$ lower bounds required for the proof complexity consequences in [8], it might seem that we have come most of the way to our goal. However, an improvement from $\Omega(\log n)$ to $\omega(\log n)$ seems non-trivial at this time, and we are nowhere near to reconciling the lower bound of $\Omega(\log n)$ with the upper bound of $O(n/2^k)$.

Even for restricted models, such as one-way multi-player protocols, getting lower bounds for the communication complexity of $\text{DISJ}_{k,n}$ seems difficult. It is not at all clear how the bound in Theorem 5.1, or even the one-way lower bound in [4], could be extended to four or more players. Moreover, it is not at all clear how to prove a direct product theorem (or even direct sum theorem) for multi-player number-on-the-forehead communication complexity. An impediment to extending our bounds to this case is the failure of the three-party analogue of our method for Lemma 4.4. Even for a product distribution, the density of a three-dimensional cylinder intersection is not determined by the densities of the cylinders in a simple manner (as is the case for rectangles).

We have shown two different methods for deriving $\Omega(\log n)$ lower bounds on the general three-party number-on-the-forehead complexity of disjointness. One reason to consider both methods is that the properties from which they are derived seem to be incomparable. The proof of Theorem 5.2 yields bounds on corruption for large three-cylinder intersections that may be give useful insight into obtaining larger bounds. These bounds do not seem to follow from Theorem 6.4 but this has the advantage of a somewhat simpler proof and a result that applies more generally.

In our applications, for example in the proof of Theorem 5.1, we did not need the full power of a strong direct product theorem. The original protocol was converted into t independent runs, each with the same complexity C. We combined these into a single protocol with complexity tC and used the strong direct product theorem but, as Shaltiel (private communication) observed, it would have sufficed to maintain these as separate protocols each of which has access to the inputs of the others. This "forest of protocols" is precisely the kind of situation that occurs in arguments for Raz's Parallel Repetition Theorem for 2-prover protocols [28, 27]. In fact, Parnafes, Raz, and Wigderson [27] have extended the theorem from 2-prover protocols to communication complexity and refined the bounds to show that if a single protocol using C bits of communication succeeds with probability $\delta < 1$ on distribution μ then t protocols running on μ^t, each of which can see the others' inputs and uses C bits of communication, succeeds with probability at most $\delta^{\Omega(t/C)}$.

This result applies to arbitrary distributions μ. By applying this result to the $Z \to (Y \leftrightarrow X)$ model using a different value of t and a non-rectangular distribution μ yields an alternative proof of Theorem 5.1 that uses the stronger two-party disjointness lower bound of [31] rather than that of [2]. More precisely, using $t = n^{2/3}$ blocks of size $n^{1/3}$ and the distribution from [31] on $X_j \times Y_j$ in each block one can use $C = \Omega(n^{1/3})$ to derive success probability $\delta^{\Omega(t/C)} = \delta^{\Omega(n^{1/3})}$ and this can be substituted in the rest of our proof of Theorem 5.1.

Whether the C in the $\delta^{\Omega(t/C)}$ bound can be removed is an open question. An analogous term cannot be removed in the general 2-prover protocols of Raz [28] but it is open in the special case of communication complexity. Such a result would almost seem to be a strong direct product theorem for randomized computation, which Shaltiel has shown to be false [33], but, as Shaltiel has observed, it has the critical difference that the allocation of resources to each subproblem has a uniform bound C. Non-uniform allocation of resources to subproblems was the key method exploited to derive the counterexample in [33].

Finally, we note that independent of this work Klauck, Spalek, and de Wolf [23] derive similar bounds to Corollary 4.9(b) for two-party quantum communication complexity using the polynomial method.

8 Acknowledgments

We would like to thank some anonymous reviewers for useful suggestions. We would especially like to thank Ronen Shaltiel for pointing out the connections to [27]. We would like also to thank Troy Lee and Pascal Tesson for references to related work.

References

[1] N. Alon, Y. Matias, and M. Szegedy. The space complexity of approximating the frequency moments. *Journal of Computer and System Sciences*, 58(1):147–157, 1999.

[2] L. Babai, P. Frankl, and J. Simon. Complexity classes in communication complexity theory. In *27th Annual Symposium on Foundations of Computer Science*, pages 337–347, Toronto, Ontario, October 1986. IEEE.

[3] L. Babai, A. Gál, P. G. Kimmel, and S. V. Lokam. Communication complexity of simultaneous messages. *SIAM Journal on Computing*, 33(1):137–166, 2003.

[4] L. Babai, T. P. Hayes, and P. G. Kimmel. The cost of the missing bit: Communication complexity with help. *Combinatorica*, 21(4):455–488, 2001.

[5] L. Babai, N. Nisan, and M. Szegedy. Multiparty protocols, pseudorandom generators for logspace, and time-space trade-offs. *Journal of Computer and System Sciences*, 45(2):204–232, October 1992.

[6] Z. Bar-Yossef, T.S. Jayram, R. Kumar, and D. Sivakumar. Information theory methods in communication complexity. In *Proceedings Seventeenth Annual IEEE Conference on Computational Complexity*, pages 133–142, Montreal, PQ, Canada, May 2002.

[7] Z. Bar-Yossef, T.S. Jayram, R. Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *Journal of Computer and System Sciences*, 68(4):702–732, 2004.

[8] P. Beame, T. Pitassi, and N. Segerlind. Lower bounds for Lov'asz-Schrijver systems and beyond follow from multiparty communication complexity. In *Automata, Languages, and Programming: 32nd International Colloquium*, volume 3580 of *Lecture Notes in Computer Science*, pages 1176–1188, Lisbon, Portugal, July 2005. Springer-Verlag.

[9] P. Beame, T. Pitassi, N. Segerlind, and A. Wigderson. A direct sum theorem for corruption and the multiparty NOF communication complexity of set disjointness. In *Proceedings Twentieth Annual IEEE Conference on Computational Complexity*, pages 52–66, San Jose, CA, June 2005.

[10] J.-Y. Cai. Lower bounds for constant depth circuits in the presence of help bits. *Information Processing Letters*, 36(2):79–83, 1990.

[11] A. Chakrabarti, S. Khot, and X. Sun. Near-optimal lower bounds on the multi-party communication complexity of set disjointness. In *Proceedings Eighteenth Annual IEEE Conference on Computational Complexity*, pages 107–117, Aarhus, Denmark, July 2003.

[12] A. Chakrabarti, Y. Shi, A. Wirth, and A.C-C. Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *Proceedings 42nd Annual Symposium on Foundations of Computer Science*, pages 270–278, Las Vegas, Nevada, October 2001. IEEE.

[13] A. K. Chandra, M. L. Furst, and R. J. Lipton. Multi-party protocols. In *Proceedings of the Fifteenth Annual ACM Symposium on Theory of Computing*, pages 94–99, Boston, MA, April 1983.

[14] F. R. K. Chung and P. Tetali. Communication complexity and quasi-randomness. *SIAM Journal on Discrete Mathematics*, 6(1):110–123, 1993.

[15] T. Cover and J. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., New York, 1991.

[16] O. Goldreich, N. Nisan, and A. Wigderson. On Yao's XOR-lemma. Technical Report TR95-050, Electronic Colloquium in Computation Complexity, `http://www.eccc.uni-trier.de/eccc/`, 1995.

[17] V. Grolmusz. The BNS lower bound for multi-party protocols is nearly optimal. *Information and Computation*, 112(1):51–54, 1994.

[18] R. Jain, J. Radhakrishnan, and P. Sen. A direct sum theorem in communication complexity via message compression. In J. C. M. Baeten, J. K. Lenstra, J. Parrow, and G. J. Woeginger, editors, *Automata, Languages, and Programming: 30th International Colloquium*, volume 2719 of *Lecture Notes in Computer Science*, pages 300–315, Eindhoven, The Netherlands, July 2003. Springer-Verlag.

[19] B. Kalyanasundaram and Georg Schnitger. The probabilistic communication complexity of set intersection. In *Proceedings, Structure in Complexity Theory, Second Annual Conference*, pages 41–49, Cornell University, Ithaca, NY, June 1987. IEEE.

[20] M. Karchmer, E. Kushilevitz, and N. Nisan. Fractional covers and communication complexity. In *Proceedings, Structure in Complexity Theory, Seventh Annual Conference*, pages 262–274, Boston, MA, June 1992. IEEE.

[21] M. Karchmer, R. Raz, and A. Wigderson. Super-logarithmic depth lower bounds via direct sum in communication complexity. *Computational Complexity*, 5:191–204, 1995.

[22] H. Klauck. Rectangle size bounds and threshold covers in communication complexity. In *Proceedings Eighteenth Annual IEEE Conference on Computational Complexity*, pages 118–134, Aarhus, Denmark, July 2003.

[23] H. Klauck, R. Spalek, and R. de Wolf. Quantum and classical strong direct product theorems and optimal time-space tradeoffs. In *Proceedings 45th Annual Symposium on Foundations of Computer Science*, pages 12–21, Rome, Italy, October 2004. IEEE.

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

[25] N. Nisan, S. Rudich, and M. Saks. Products and help bits in decision trees. *SIAM Journal on Computing*, 28(3):1035–1050, 1999.

[26] N. Nisan and A. Wigderson. Rounds in communication complexity revisited. *SIAM Journal on Computing*, 22(1):211–219, 1993.

[27] I. Parnafes, R. Raz, and A. Widgerson. Direct product results and the GCD problem, in old and new communication models. In *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing*, pages 363–372, El Paso, TX, May 1997.

[28] R. Raz. A parallel repetition theorem. *SIAM Journal on Computing*, 27(1):763–803, 1998.

[29] R. Raz. The BNS-Chung criterion for multi-party communication complexity. *Computational Complexity*, 9:113–122, 2000.

[30] R. Raz and A. Wigderson. Monotone circuits for matching require linear depth. *Journal of the ACM*, 39(3):736–744, July 1992.

[31] A. A. Razborov. On the distributional complexity of disjointness. *Theoretical Computer Science*, 106(2):385–390, 1992.

[32] M. E. Saks and X. Sun. Space lower bounds for distance approximation in the data stream model. In *Proceedings of the Thirty-Fourth Annual ACM Symposium on Theory of Computing*, pages 360–369, Montreal, Quebec, Canada, May 2002.

[33] R. Shaltiel. Towards proving strong direct product theorems. In *Proceedings Sixteenth Annual IEEE Conference on Computational Complexity*, pages 107–117, Chicago, IL, June 2001.

[34] P. Tesson. *Communication Complexity Questions Related to Finite Monoids and Semigroups*. PhD thesis, McGill University, 2002.