Electronic Colloquium on Computational Complexity, Report No. 66 (2019)

A Lower Bound for Sampling Disjoint Sets

Thomas Watson*

May 2, 2019

Abstract

Suppose Alice and Bob each start with private randomness and no other input, and they wish to engage in a protocol in which Alice ends up with a set $x \subseteq [n]$ and Bob ends up with a set $y \subseteq [n]$, such that (x, y) is uniformly distributed over all pairs of disjoint sets. We prove that for some constant $\varepsilon > 0$, this requires $\Omega(n)$ communication even to get within statistical distance ε of the target distribution. Previously, Ambainis, Schulman, Ta-Shma, Vazirani, and Wigderson (FOCS 1998) proved that $\Omega(\sqrt{n})$ communication is required to get within ε of the uniform distribution over all pairs of disjoint sets of size $\sqrt{n}$.

1 Introduction

In most traditional computational problems, the goal is to take an input and produce the "correct" output, or produce one of a set of acceptable outputs. In a *sampling* problem, on the other hand, the goal is to generate a random sample from a specified probability distribution D, or at least from a distribution that is close to D. There has been a surge of interest in studying sampling problems from a complexity theory perspective [ASTS+03, GGN10, Vio12a, Aar14, LV12, DW12, Vio14, BIL12, Vio12b, JSWZ13, Wat14, BCS14, Wat16, Vio16, Wat18, Vio18]. Unlike more traditional computational problems, sampling problems do not necessarily need to have any real input, besides the uniformly random bits fed into a sampling algorithm.

One commonly studied type of target distribution is "input–output pairs" of a function f, i.e., $(D, f(D))$ where D is perhaps the uniform distribution over inputs to f. Using an algorithm for computing f, one can sample $(D, f(D))$ by first sampling from D, then evaluating f on that input. However, for some functions f, generating an input jointly with the corresponding output may be computationally easier than evaluating f on an adversarially-chosen input. Thus in general, sampling lower bounds tend to be more challenging to prove than lower bounds for functions.

Many of the above-cited works focus on concrete computational models such as low-depth circuits. We consider the model of 2-party communication complexity, for which comparatively less is known about sampling. Which problem should we study? Well, the single most important function in communication complexity is Set-Disjointness, in which Alice gets a set $x \subseteq [n]$, Bob gets a set $y \subseteq [n]$, and the goal is to determine whether $x \cap y = \emptyset$. Identifying the sets with their characteristic bit strings, this can be viewed as $\text{Disj}\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ where $\text{Disj}(x, y) = 1$ iff $x \wedge y = 0^n$. The applications of communication bounds for Set-Disjointness are far too numerous to list, but they span areas such as streaming, circuit complexity, proof complexity, data

*Department of Computer Science, University of Memphis. Supported by NSF grant CCF-1657377.

ISSN 1433-8092

structures, combinatorial optimization, fine-grained complexity, and game theory. Because of its central role, Set-Disjointness has become the de facto testbed for proving new types of communication bounds. This function has been studied in the contexts of randomized [BFS86, KS92, Raz92, BYJKS04, BGK15] and quantum [BCW98, HdW02, Raz03, AA05, She11, SZ09] protocols; multiparty number-in-hand [BYJKS04, CKS03, Gro09, Jay09, BEO+13, BO15] and number-on-forehead [Gro94, Tes03, BPSW06, She11, CA08, LS09, BH09, She16, She14, RY15, PS17] models; Merlin–Arthur and related models [Kla03, AW09, GS10, GW16, GPW16, ARW17, Rub18, Che18]; with a bounded number of rounds of interaction [KNTZ07, JRS03, WW15, BGK+18, BO17]; with bounds on the sizes of the sets [HW07, KW09, Pat11, DKS12, BGMdW13, ST13]; very precise relationships between communication and error probability [BGPW13, BM13, GW16, FHLY17, DFHL18]; when the goal is to find the intersection [BCK+14, Gav16, Wat18, ACK19]; in space-bounded, online, and streaming models [KP14, BKM18, AWY18]; and direct product theorems [KSdW07, BPSW06, BRdW08, JKN08, Kla10, She12, She16, She14]. We contribute one more result to this thorough assault on Set-Disjointness.

Here is the definition of our 2-party sampling model: Let D be a probability distribution over $\{0,1\}^n \times \{0,1\}^n$; we also think of D as a matrix with rows and columns both indexed by $\{0,1\}^n$ where $D_{x,y}$ is the probability of outcome (x, y). We define $\mathsf{Samp}(D)$ as the minimum communication cost of any protocol where Alice and Bob each start with private randomness and no other input, and at the end Alice outputs some $x \in \{0,1\}^n$ and Bob outputs some $y \in \{0,1\}^n$ such that (x, y) is distributed according to D. Note that $\mathsf{Samp}(D) = 0$ iff D is a product distribution (x and y are independent), and $\mathsf{Samp}(D) \leq n$ for all D (since Alice can privately sample (x, y) and send y to Bob). Allowing public randomness would not make sense since Alice and Bob could read a properly-distributed (x, y) off of the randomness without communicating. We define $\mathsf{Samp}_\varepsilon(D)$ as the minimum of $\mathsf{Samp}(D')$ over all distributions D' with $\Delta(D, D') \leq \varepsilon$, where Δ denotes statistical (total variation) distance, defined as

$$\Delta(D, D') \;\coloneqq\; \max_{\text{event } E} \big|\mathbb{P}_D[E] - \mathbb{P}_{D'}[E]\big| \;=\; \max_{\text{event } E} \big(\mathbb{P}_D[E] - \mathbb{P}_{D'}[E]\big) \;=\; \tfrac{1}{2} \sum_{\text{outcome } o} \big|\mathbb{P}_D[o] - \mathbb{P}_{D'}[o]\big|.$$

1.1 A story

Our story begins with [ASTS+03], which proved that $\mathsf{Samp}_\varepsilon\big((D, \textsc{Disj}(D))\big) \geq \Omega(\sqrt{n})$ for some constant $\varepsilon > 0$, where D is uniform over the set of all pairs of sets of size $\sqrt{n}$ (note that this D is a product distribution and is approximately balanced between 0-inputs and 1-inputs of $\textsc{Disj}$). The main tool in the proof was a lemma that was originally employed in [BFS86] to prove an $\Omega(\sqrt{n})$ bound on the randomized communication complexity of *computing* $\textsc{Disj}$. The latter bound was improved to $\Omega(n)$ via several different proofs [KS92, Raz92, BYJKS04], which leads to a natural question: Can we improve the sampling bound of [ASTS+03] to $\Omega(n)$ by using the techniques of [KS92, Raz92, BYJKS04] instead of [BFS86]?

For starters, the answer is "no" for the particular D considered in [ASTS+03]—there is a trivial exact protocol with $O(\sqrt{n}\log n)$ communication since it only takes that many bits to specify a set of size $\sqrt{n}$. What about other interesting distributions D? The following illuminates the situation.

Observation 1. *For any D and constants $\varepsilon > \delta > 0$, if $\mathsf{Samp}_\varepsilon\big((D, \textsc{Disj}(D))\big) \geq \omega(\sqrt{n}\log n)$ then $\mathsf{Samp}_\delta(D) \geq \Omega\big(\mathsf{Samp}_\varepsilon\big((D, \textsc{Disj}(D))\big)\big)$.*

Proof. It suffices to show $\mathsf{Samp}_\varepsilon\big((D, \textsc{Disj}(D))\big) \leq \mathsf{Samp}_\delta(D) + O(\sqrt{n}\log n)$. First, note that for any sampling protocol, if we condition on a particular transcript then the output distribution becomes product (Alice and Bob are independent after they stop communicating). Second, [BFS86] also proved that for every product distribution and every constant $\gamma > 0$, there exists a deterministic protocol that uses $O(\sqrt{n}\log n)$ bits of communication and computes $\textsc{Disj}$ with error probability $\leq \gamma$ on a random input from the distibution. Now to ε-sample $(D, \textsc{Disj}(D))$, Alice and Bob can δ-sample D to obtain (x, y), and then conditioned on that sampler's transcript, they can run the average-case protocol from [BFS86] for the corresponding product distribution with error $\varepsilon - \delta$. A simple calculation shows this indeed gives statistical distance ε. ☐

The upshot is that to get an improved bound, *the hardness of sampling* $(D, \textsc{Disj}(D))$ *would come entirely from the hardness of just sampling* D. Thus such a result would not really be "about" the Set-Disjointness function, it would be about the distribution on inputs. Instead of abandoning this line of inquiry, we realize that if D itself is somehow defined in terms of $\textsc{Disj}$, then a bound for sampling D would still be saying something about the complexity of Set-Disjointness. In fact, the proof in [ASTS+03] actually shows something stronger than the previously-stated result: If D is instead defined as the uniform distribution over pairs of *disjoint* sets of size $\sqrt{n}$ (1-inputs of $\textsc{Disj}$), then $\mathsf{Samp}_\varepsilon(D) \geq \Omega(\sqrt{n})$. After this pivot, we are now facing a direction in which we can hope for an improvement. We prove that by removing the restriction on the sizes of the sets, the sampling problem becomes maximally hard.

Theorem 1. *Let U be the uniform distribution over the set of all $(x, y) \in \{0,1\}^n \times \{0,1\}^n$ with $x \wedge y = 0^n$. Then $\mathsf{Samp}_{0.01}(U) = \Omega(n)$.*

The proof from [ASTS+03] was a relatively short application of the technique from [BFS86], but for Theorem 1, harnessing known techniques for proving linear communication lower bounds turns out to be significantly more involved.

For calibration, the uniform distribution over *all* (x, y) achieves statistical distance $1 - 0.75^n$ from U since there are 4^n inputs and 3^n disjoint inputs. We can do a little better: Suppose for each coordinate independently, Alice picks 0 with probability $\sqrt{1/3}$ and picks 1 with probability $1-\sqrt{1/3}$, and Bob does the same. This again involves no communication, and it achieves statistical distance $1 - \left(2\sqrt{1/3} - 1/3\right)^n \leq 1 - 0.82^n$ from U. It is plausible that achieving statistical distance $1 - 2^{-o(n)}$ requires $\Omega(n)$ communication. The distance 0.01 in Theorem 1 can be relaxed slightly, but substantially new ideas will be needed to handle distances that approach 1, or even a large constant such as 0.5. Very roughly speaking, our argument has a "direct sum" flavor showing that most coordinates contribute $\Omega(1)$ to the communication cost, whereas ideally we would like a "direct product"-type argument that also shows most coordinates contribute a constant factor to $1 -$ the statistical distance. In the setting of circuits, considerable effort has gone into improving the statistical distances that can be handled by sampling lower bounds [DW12, BIL12, Vio18].

1.2 Interpreting the result

We first observe that our sampling model is equivalent to two other models. One of these we call (for lack of a better word) "synthesizing" the distribution D: Alice and Bob get inputs $x, y \in \{0,1\}^n$ respectively, in addition to their private randomness, and their goal is to accept with probability exactly $D_{x,y}$. We let $\mathsf{Synth}(D)$ denote the minimum communication cost of any synthesizing protocol for D, and $\mathsf{Synth}_\varepsilon(D)$ denote the minimum of $\mathsf{Synth}(D')$ over all D' with $\Delta(D, D') \leq \varepsilon$. The other model is the nonnegative rank of a matrix: $\text{rank}_+(D)$ is defined as the minimum k for which D can be written as a sum of k many nonnegative rank-1 matrices.

Observation 2. *For every distribution D, the following are all within $\pm O(1)$ of each other:*

$$\mathsf{Samp}(D), \quad \mathsf{Synth}(D), \quad \log \operatorname{rank}_+(D).$$

Proof. $\mathsf{Synth}(D) \le \mathsf{Samp}(D) + 2$ since a synthesizing protocol can just run a sampling protocol and accept iff the result equals the given input (x, y).

$\log \operatorname{rank}_+(D) \le \mathsf{Synth}(D)$ since for each transcript of a synthesizing protocol, the matrix that records the probability of getting that transcript on each particular input has rank 1; summing these matrices over all accepting transcripts yields a nonnegative rank decomposition of D.

To see that $\mathsf{Samp}(D) \le \lceil \log \operatorname{rank}_+(D) \rceil$, suppose $D = M^{(1)} + M^{(2)} + \cdots + M^{(k)}$ is a sum of nonnegative rank-1 matrices. For each i, by scaling we can write $M^{(i)}_{x,y} = p_i \, u^{(i)}_x v^{(i)}_y$ for some distributions $u^{(i)}$ and $v^{(i)}$ over $\{0,1\}^n$, where p_i is the sum of all entries of $M^{(i)}$. Since D is a distribution, $p \coloneqq (p_1, \ldots, p_k)$ is a distribution over $[k]$. To sample from D, Alice can privately sample $i \sim p$ and send it to Bob using $\lceil \log k \rceil$ bits, then Alice can sample $x \sim u^{(i)}$ and Bob can independently sample $y \sim v^{(i)}$ with no further communication. □

By this characterization, Theorem 1 can be viewed as a lower bound on the approximate nonnegative rank of the Disj matrix, where the approximation is in ℓ_1 (which has an average-case flavor). In the recent literature, "approximate nonnegative rank" generally refers to approximation in ℓ_∞ (which is a worst-case requirement), and this model is equivalent to the so-called smooth rectangle bound and WAPP communication complexity [JK10, KMSY19, GLM+16].

1.3 Overview of the proof

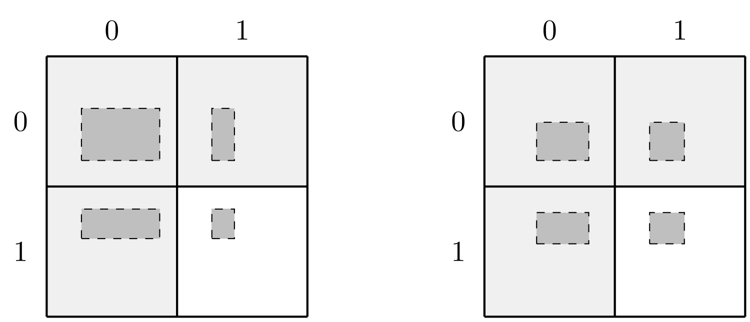
We use the Synth characterization from Observation 2 in our proof of Theorem 1. We also use the *information complexity* method that was pioneered in [CSWY01, BYJKS04] (as an alternative to the earlier *corruption* method [Raz92]) for proving the randomized $\Omega(n)$ bound for computing Disj. At a high level, the information complexity approach is to consider a probability space with a random input (from a distribution of our choosing) and with the random transcript generated by a protocol on that input, and to use the fact that the communication cost is lower bounded by the Shannon entropy of the transcript, which in turn is lower bounded by the "information cost": the mutual information between the transcript and the input. The key is that as long as the n input coordinates are independent of each other, the information cost obeys a direct sum property: It is at least the sum of the contributions of the n coordinates. Thus an $\Omega(n)$ bound follows by showing that the mutual information between the transcript and a typical coordinate is $\Omega(1)$.

How shall we implement this approach, given a synthesizing protocol for a distribution that is close to the U from Theorem 1? First we should decide which input distribution to measure information cost with respect to. It is natural to speculate that since the aim is to prove a lower bound for approximately synthesizing U, we should use U itself as the input distribution. That idea is somewhat of a red herring. We do use U itself for measuring information cost, but the reason has more to do with the fact that statistical (or ℓ_1) distance is a certain sum—rather than a weighted sum—over inputs. (This forms an obstacle to using the corruption lemma from [Raz92] as a black box, since the latter is tailored to an input distribution that is different from U.)

When looking at an individual coordinate's contribution to the information cost, we need the input to come from a product distribution, in order to exploit the fact that each transcript corresponds to a combinatorial rectangle. The standard technique is to decompose the input distribution into a mixture of product distributions (like what a sampling protocol does—but now this is purely for analysis purposes) and consider a typical component of this mixture. Then, we can use a standard lemma (from [BYJKS04]) for relating the mutual information to the statistical distance between transcript distributions on different inputs; however, we need to somewhat generalize this tool to handle the input distributions that arise from decomposing our U.

The next issue to tackle is that a synthesizing protocol rejects most inputs with extremely high probability, so the rejecting transcripts may not carry much information about the input (the information cost could be very low if we measure w.r.t. a random transcript in the naive way). Most of the "action" happens within the approximately 3^{-n} probability of acceptance on a typical 1-input. For this reason, our probability space will use a random transcript *conditioned* on the protocol accepting. This introduces two related sources of difficulties: It distorts the "product structure" we usually rely on for analyzing the behavior of a transcript across different inputs, and it interferes with the standard trick of "absorbing" the other $n - 1$ coordinates of the randomly chosen input into the protocol's private randomness (specifically, sampling an input and then a random accepting transcript on that input, is *not* the same as sampling an input, running the protocol, then conditioning the whole experiment on acceptance). A substantial portion of the technical effort goes into alleviating these issues.

Anyway, to give the gist of the overall structure of the argument, let us visualize a single "representative" transcript and ignore the complications mentioned in the previous paragraph. Focusing on a single input coordinate (Disj with $n = 1$ is just Nand), we think of the lower-right cell as a 0-input and the other three cells as 1-inputs. The area within each cell represents the protocol's private randomness along with "the rest" of the random input (besides the coordinate under the spotlight). If the transcript's rectangle occupies too much area in the upper-left cell (as shown on the left), it would be contributing to the protocol accepting 1-inputs with too high of probability. Otherwise, the rectangle is forced to occupy a relatively not-too-small area in the lower-right cell (as shown on the right). Accepting on some 0-inputs can be OK, but here is the catch: There are $n/3$ times as many uniquely-intersecting inputs as there are disjoint inputs (for the full Disj function), and it turns out this not-too-small acceptance probability would get "replicated" across many of these intersecting inputs. The sum of the acceptance probabilities would then be too great for the protocol to be synthesizing any distribution at all, much less one close to U.



2 Proof

Before proving Theorem 1, we give some preliminaries. We assume familiarity with the basics of communication complexity [KN97] and information theory [CT06]. A protocol Π is assumed to have private randomness, and we let $CC(\Pi)$ denote the worst-case communication cost. We use $\mathbb{P}$ for probability, $\mathbb{E}$ for expectation, $\mathbb{H}$ for Shannon entropy, $\mathbb{I}$ for mutual information, $\mathbb{D}$ for relative entropy, and Δ for statistical distance. We use bold letters to denote random variables, and non-bold letters for particular outcomes.

Fact 1. *Mutual information and relative entropy satisfy the following standard properties:*

- *Direct sum:* $\mathbb{I}(\boldsymbol{a} \,;\, \boldsymbol{b}_1 \cdots \boldsymbol{b}_n) \geq \mathbb{I}(\boldsymbol{a} \,;\, \boldsymbol{b}_1) + \cdots + \mathbb{I}(\boldsymbol{a} \,;\, \boldsymbol{b}_n)$ *if* $\boldsymbol{b}_1 \cdots \boldsymbol{b}_n$ *are fully independent.*
- *Alternative definition:* $\mathbb{I}(\boldsymbol{a} \,;\, \boldsymbol{b}) = \mathbb{E}_{a \sim \boldsymbol{a}} \mathbb{D}\big((\boldsymbol{b} \,|\, \boldsymbol{a} = a) \,\big\|\, \boldsymbol{b}\big)$.
- *Pinsker's inequality:* $\mathbb{D}(\boldsymbol{a} \,\|\, \boldsymbol{b}) \geq \frac{2}{\ln 2} \Delta(\boldsymbol{a}, \boldsymbol{b})^2$.

To make the paper slightly more self-contained, we include the quick proof of the direct sum property in § A. Pinsker's inequality has several proofs available in several sources, such as [DP09].

2.1 Proof of Theorem 1

Letting U be as in Theorem 1, suppose for contradiction $\mathsf{Samp}(D) \leq 0.0001n - 2$ for some distribution D with $\Delta(U, D) \leq 0.01$. By Observation 2, $\mathsf{Synth}(D) \leq 0.0001n$, so consider a synthesizing protocol Π for D with $CC(\Pi) \leq 0.0001n$. As a technical convenience, we may assume Π has been infinitesimally perturbed to ensure the acceptance probability is positive on each input;[1] this allows us to avoid special cases for handling conditioning on 0-probability events.

We build a probability space by decomposing U into a mixture of product distributions. For each $j \in [n]$ independently: Let $\boldsymbol{w}_j$ be uniform over $\{\text{LEFT}, \text{RIGHT}\}$.

Conditioned on $\boldsymbol{w}_j = \text{LEFT}$, let

$$\boldsymbol{x}_j \boldsymbol{y}_j := \begin{cases} 00 & \text{with probability } 1/3 \\ 10 & \text{with probability } 2/3 \end{cases}.$$

Conditioned on $\boldsymbol{w}_j = \text{RIGHT}$, let

$$\boldsymbol{x}_j \boldsymbol{y}_j := \begin{cases} 00 & \text{with probability } 1/3 \\ 01 & \text{with probability } 2/3 \end{cases}.$$

Note that the marginal distribution of $(\boldsymbol{x}, \boldsymbol{y})$ is U, but $\boldsymbol{x}$ and $\boldsymbol{y}$ are independent conditioned on $\boldsymbol{w}$. Conditioned on $(\boldsymbol{x}, \boldsymbol{y}) = (x, y)$, let $\boldsymbol{\tau}$ be a random transcript of $\Pi(x, y)$ conditioned on acceptance. Finally, let $\boldsymbol{i}$ be uniform over $[n]$ and independent of the other random variables. In summary, $(\boldsymbol{w}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\tau}, \boldsymbol{i})$ are jointly distributed over $\{\text{LEFT}, \text{RIGHT}\}^n \times \{0,1\}^n \times \{0,1\}^n \times \{0,1\}^{CC(\Pi)} \times [n]$.

Definition 1. *An outcome* $(i, w_{-i}) \in [n] \times \{\text{LEFT}, \text{RIGHT}\}^{[n] \setminus \{i\}}$ *is called* good *iff both:*

(1) $\mathbb{I}\big(\boldsymbol{\tau} \,;\, \boldsymbol{x}_i \boldsymbol{y}_i \,\big|\, \boldsymbol{i} = i,\, \boldsymbol{w} = w\big) \leq 0.0008$ *("cost")*
for each $w_i \in \{\text{LEFT}, \text{RIGHT}\}$, *and*

(2) $\mathbb{E}\big[\max(3^{-n} - D_{\boldsymbol{x},\boldsymbol{y}}, 0) \,\big|\, \boldsymbol{i} = i,\, \boldsymbol{w}_{-i} = w_{-i},\, \boldsymbol{x}_i \boldsymbol{y}_i = x_i y_i\big] \leq 0.12 \cdot 3^{-n}$ *("correctness")*
for each $x_i y_i \in \{00, 01, 10\}$.

Claim 1. $(i, w_{-i}) \sim (\boldsymbol{i}, \boldsymbol{w}_{-\boldsymbol{i}})$ *is good with probability at least* 0.5.

Proof. By a union bound, it suffices to show that *(1)* and *(2)* individually hold with probability at least 0.75 each.

[1] For an infinitesimal $\iota > 0$, accept with probability $4^{-n}\iota$, reject with probability $(1 - 4^{-n})\iota$, and otherwise run the original synthesizing protocol. This adds only 2 bits of communication, and it affects the statistical distance to the target distribution by at most ι.

For fixed i and w, abbreviate $\mathbb{I}\big(\boldsymbol{\tau}\,;\,\boldsymbol{x}_i\boldsymbol{y}_i \,\big|\, \boldsymbol{i}=i,\,\boldsymbol{w}=w\big)$ as $I_{i,w} \geq 0$. For each w,

$$\mathbb{E}_{i\sim\boldsymbol{i}}[I_{i,w}] \;\leq\; \tfrac{1}{n}\,\mathbb{I}\big(\boldsymbol{\tau}\,;\,\boldsymbol{x}\boldsymbol{y}\,\big|\,\boldsymbol{w}=w\big) \;\leq\; \tfrac{1}{n}\,\mathbb{H}(\boldsymbol{\tau}\,|\,\boldsymbol{w}=w) \;\leq\; \tfrac{1}{n}\,CC(\Pi) \;\leq\; 0.0001$$

by the first bullet from Fact 1 using $\boldsymbol{a} \coloneqq (\boldsymbol{\tau}\,|\,\boldsymbol{w}=w)$ and $\boldsymbol{b}_i \coloneqq (\boldsymbol{x}_i\boldsymbol{y}_i\,|\,\boldsymbol{w}=w)$. Now

$$\mathbb{E}_{(i,w_{-i})\sim(\boldsymbol{i},\boldsymbol{w}_{-\boldsymbol{i}})}\,\mathbb{E}_{w_i\sim\boldsymbol{w}_i}[I_{i,w}] \;=\; \mathbb{E}_{w\sim\boldsymbol{w}}\,\mathbb{E}_{i\sim\boldsymbol{i}}[I_{i,w}] \;\leq\; 0.0001.$$

By Markov's inequality, with probability at least 0.75 over $(i,w_{-i}) \sim (\boldsymbol{i},\boldsymbol{w}_{-\boldsymbol{i}})$ we have that $\mathbb{E}_{w_i\sim\boldsymbol{w}_i}[I_{i,w}] \leq 0.0004$, in which case $\max_{w_i}(I_{i,w}) \leq 2\,\mathbb{E}_{w_i\sim\boldsymbol{w}_i}[I_{i,w}] \leq 0.0008$ and thus *(1)* holds.

For fixed i, w_{-i}, and x_iy_i, abbreviate $\mathbb{E}\big[\max(3^{-n}-D_{\boldsymbol{x},\boldsymbol{y}},0)\,\big|\,\boldsymbol{i}=i,\,\boldsymbol{w}_{-i}=w_{-i},\,\boldsymbol{x}_i\boldsymbol{y}_i=x_iy_i\big]$ as $\delta_{i,w_{-i},x_iy_i} \geq 0$. Now

$$\begin{aligned}\mathbb{E}_{(i,w_{-i})\sim(\boldsymbol{i},\boldsymbol{w}_{-\boldsymbol{i}})}\,\mathbb{E}_{x_iy_i\sim\boldsymbol{x}_i\boldsymbol{y}_i}[\delta_{i,w_{-i},x_iy_i}] \;&=\; \mathbb{E}\big[\max(3^{-n}-D_{\boldsymbol{x},\boldsymbol{y}},0)\big] \;=\; 3^{-n}\textstyle\sum_{x,y}\max(U_{x,y}-D_{x,y},0)\\ &=\; 3^{-n}\,\Delta(U,D) \;\leq\; 0.01\cdot 3^{-n}.\end{aligned}$$

By Markov's inequality, with probability at least 0.75 over $(i,w_{-i}) \sim (\boldsymbol{i},\boldsymbol{w}_{-\boldsymbol{i}})$ we have that $\mathbb{E}_{x_iy_i\sim\boldsymbol{x}_i\boldsymbol{y}_i}[\delta_{i,w_{-i},x_iy_i}] \;\leq\; 0.04\cdot 3^{-n}$, in which case $\max_{x_iy_i}[\delta_{i,w_{-i},x_iy_i}] \;\leq\; 3\,\mathbb{E}_{x_iy_i\sim\boldsymbol{x}_i\boldsymbol{y}_i}[\delta_{i,w_{-i},x_iy_i}] \;\leq\; 0.12\cdot 3^{-n}$ and thus *(2)* holds. □

Lemma 1. *For each good* (i,w_{-i})*, either:*

(i) $\mathbb{E}\big[D_{\boldsymbol{x},\boldsymbol{y}}\,\big|\,\boldsymbol{i}=i,\,\boldsymbol{w}_{-i}=w_{-i}\big] \geq 5\cdot 3^{-n}$*, or*

(ii) $\mathbb{E}\big[D_{\widehat{\boldsymbol{x}},\widehat{\boldsymbol{y}}}\,\big|\,\boldsymbol{i}=i,\,\boldsymbol{w}_{-i}=w_{-i}\big] \geq 0.000001\cdot 3^{-n}$

where the random variables $\widehat{\boldsymbol{x}},\widehat{\boldsymbol{y}}$ *are the same as* $\boldsymbol{x},\boldsymbol{y}$ *except* $\widehat{\boldsymbol{x}}_{\boldsymbol{i}}\widehat{\boldsymbol{y}}_{\boldsymbol{i}}$ *is fixed to* 11.

Lemma 1 is the technical heart of the argument; we prove it in § 2.2. Note that the marginal distribution of $(\widehat{\boldsymbol{x}},\widehat{\boldsymbol{y}})$ is uniform over the set of all $(x,y) \in \{0,1\}^n \times \{0,1\}^n$ with $|x\wedge y| = 1$, where $|\cdot|$ denotes Hamming weight (i.e., x and y represent uniquely intersecting sets).

Combining Claim 1 and Lemma 1 shows that over $(i,w_{-i}) \sim (\boldsymbol{i},\boldsymbol{w}_{-\boldsymbol{i}})$, either *(i)* holds with probability at least 0.25 or *(ii)* holds with probability at least 0.25. In the former case,

$$\mathbb{E}[D_{\boldsymbol{x},\boldsymbol{y}}] \;\geq\; \mathbb{P}[\textit{(i)}\text{ holds}]\cdot\mathbb{E}\big[D_{\boldsymbol{x},\boldsymbol{y}}\,\big|\,\textit{(i)}\text{ holds}\big] \;\geq\; 0.25\cdot 5\cdot 3^{-n} \;>\; 3^{-n}$$

and thus $\sum_{x,y\,:\,|x\wedge y|=0} D_{x,y} = 3^n\,\mathbb{E}[D_{\boldsymbol{x},\boldsymbol{y}}] > 1$. In the latter case,

$$\mathbb{E}[D_{\widehat{\boldsymbol{x}},\widehat{\boldsymbol{y}}}] \;\geq\; \mathbb{P}[\textit{(ii)}\text{ holds}]\cdot\mathbb{E}\big[D_{\widehat{\boldsymbol{x}},\widehat{\boldsymbol{y}}}\,\big|\,\textit{(ii)}\text{ holds}\big] \;\geq\; 0.25\cdot 0.000001\cdot 3^{-n} \;>\; 1/(n3^{n-1})$$

and thus $\sum_{x,y\,:\,|x\wedge y|=1} D_{x,y} = n3^{n-1}\,\mathbb{E}[D_{\widehat{\boldsymbol{x}},\widehat{\boldsymbol{y}}}] > 1$. Either case contradicts the assumption that D is a distribution. This finishes the proof of Theorem 1, except for the proof of Lemma 1.

2.2 Proof of Lemma 1

Fix a good (i,w_{-i}). For convenience, we henceforth assume $i=1$ and we elide the conditioning on $\boldsymbol{i}=1$, $\boldsymbol{w}_{-1}=w_{-1}$ in the notation. Thus our whole probability space now consists of $(\boldsymbol{w}_1,\boldsymbol{x},\boldsymbol{y},\boldsymbol{\tau})$ which is actually distributed as $\big(\boldsymbol{w}_1,\boldsymbol{x},\boldsymbol{y},\boldsymbol{\tau}\,\big|\,\boldsymbol{i}=1,\,\boldsymbol{w}_{-1}=w_{-1}\big)$ in the original notation. Also, $(\widehat{\boldsymbol{x}},\widehat{\boldsymbol{y}}) \coloneqq (1\boldsymbol{x}_{-1},1\boldsymbol{y}_{-1})$.

With this convention, the definition of good becomes

(1) $\mathbb{I}\big(\boldsymbol{\tau}\ ;\ \boldsymbol{x}_1\boldsymbol{y}_1 \,\big|\, \boldsymbol{w}_1 = w_1\big) \leq 0.0008$ *("cost")*
for each $w_1 \in \{\text{LEFT}, \text{RIGHT}\}$*, and*

(2) $\mathbb{E}\big[\max(3^{-n} - D_{\boldsymbol{x},\boldsymbol{y}}, 0) \,\big|\, \boldsymbol{x}_1\boldsymbol{y}_1 = x_1y_1\big] \leq 0.12 \cdot 3^{-n}$ *("correctness")*
for each $x_1y_1 \in \{00, 01, 10\}$

and the statement of Lemma 1 becomes

(i) $\mathbb{E}[D_{\boldsymbol{x},\boldsymbol{y}}] \geq 5 \cdot 3^{-n}$*, or*

(ii) $\mathbb{E}[D_{\widehat{\boldsymbol{x}},\widehat{\boldsymbol{y}}}] \geq 0.000001 \cdot 3^{-n}$.

Claim 2. *If (1) holds then* $\Delta\big((\boldsymbol{\tau} \,|\, \boldsymbol{x}_1\boldsymbol{y}_1 = 00), (\boldsymbol{\tau} \,|\, \boldsymbol{x}_1\boldsymbol{y}_1 = x_1y_1)\big) \leq 0.05$ *for each* $x_1y_1 \in \{01, 10\}$.

To prove Claim 2, we use the following tool relating statistical distance and mutual information.

Lemma 2. *Let* $\boldsymbol{a}, \boldsymbol{b}$ *be jointly distributed, with* $\boldsymbol{b}$ *having support* $\{0,1\}$*. Then*

$$\Delta\big((\boldsymbol{a} \,|\, \boldsymbol{b} = 0), (\boldsymbol{a} \,|\, \boldsymbol{b} = 1)\big) \ \leq\ \sqrt{\mathbb{I}(\boldsymbol{a}\ ;\ \boldsymbol{b}) \,/\, \big(3 \cdot \mathbb{P}[\boldsymbol{b} = 0] \cdot \mathbb{P}[\boldsymbol{b} = 1]\big)}.$$

The special case of Lemma 2 where $\boldsymbol{b}$ is uniform over $\{0,1\}$ was known, dating back to [BYJKS04] (using [Lin91]). We give a proof of the general case in § 2.3, via a somewhat new and streamlined argument that avoids the intermediate use of Hellinger distance.

Proof of Claim 2. First, *(1)* tells us $\mathbb{I}\big(\boldsymbol{\tau}\ ;\ \boldsymbol{x}_1 \,\big|\, \boldsymbol{w}_1 = \text{LEFT}\big) \leq 0.0008$ (since $\boldsymbol{y}_1$ is always 0 conditioned on $\boldsymbol{w}_1 = \text{LEFT}$), and applying Lemma 2 with $(\boldsymbol{a}, \boldsymbol{b}) \coloneqq \big(\boldsymbol{\tau}, \boldsymbol{x}_1 \,\big|\, \boldsymbol{w}_1 = \text{LEFT}\big)$ gives

$$\Delta\big((\boldsymbol{\tau} \,|\, \boldsymbol{x}_1\boldsymbol{y}_1 = 00), (\boldsymbol{\tau} \,|\, \boldsymbol{x}_1\boldsymbol{y}_1 = 10)\big) \ \leq\ \sqrt{\mathbb{I}\big(\boldsymbol{\tau}\ ;\ \boldsymbol{x}_1 \,\big|\, \boldsymbol{w}_1 = \text{LEFT}\big) \,/\, \big(3 \cdot \tfrac{1}{3} \cdot \tfrac{2}{3}\big)} \ \leq\ \sqrt{0.0008} \cdot 1.5 \ \leq\ 0.05.$$

Similarly, *(1)* tells us $\mathbb{I}\big(\boldsymbol{\tau}\ ;\ \boldsymbol{y}_1 \,\big|\, \boldsymbol{w}_1 = \text{RIGHT}\big) \leq 0.0008$ (since $\boldsymbol{x}_1$ is always 0 conditioned on $\boldsymbol{w}_1 = \text{RIGHT}$), and applying Lemma 2 with $(\boldsymbol{a}, \boldsymbol{b}) \coloneqq \big(\boldsymbol{\tau}, \boldsymbol{y}_1 \,\big|\, \boldsymbol{w}_1 = \text{RIGHT}\big)$ gives

$$\Delta\big((\boldsymbol{\tau} \,|\, \boldsymbol{x}_1\boldsymbol{y}_1 = 00), (\boldsymbol{\tau} \,|\, \boldsymbol{x}_1\boldsymbol{y}_1 = 01)\big) \ \leq\ \sqrt{\mathbb{I}\big(\boldsymbol{\tau}\ ;\ \boldsymbol{y}_1 \,\big|\, \boldsymbol{w}_1 = \text{RIGHT}\big) \,/\, \big(3 \cdot \tfrac{1}{3} \cdot \tfrac{2}{3}\big)} \ \leq\ \sqrt{0.0008} \cdot 1.5 \ \leq\ 0.05.$$

This proves the claim. □

Let A be the set of all accepting transcripts of Π. For $\tau \in A$ and $(x,y) \in \{0,1\}^n \times \{0,1\}^n$ and $x_1y_1 \in \{0,1\}^2$, define

$$p_{x,y}(\tau) \ \coloneqq\ \mathbb{P}\big[\Pi(x,y) \text{ generates } \tau\big] \qquad p_{x_1y_1}(\tau) \ \coloneqq\ \mathbb{E}_{x_{-1}y_{-1} \sim \boldsymbol{x}_{-1}\boldsymbol{y}_{-1}}[p_{x,y}(\tau)]$$

$$q_{x,y}(\tau) \ \coloneqq\ \mathbb{P}\big[\Pi(x,y) \text{ generates } \tau \,\big|\, \Pi(x,y) \text{ accepts}\big] \qquad q_{x_1y_1}(\tau) \ \coloneqq\ \mathbb{E}_{x_{-1}y_{-1} \sim \boldsymbol{x}_{-1}\boldsymbol{y}_{-1}}[q_{x,y}(\tau)]$$

where the probabilities in the left column are over the private randomness of Alice and Bob; in particular, $\sum_{\tau \in A} p_{x,y}(\tau) = \mathbb{P}[\Pi(x,y) \text{ accepts}] = D_{x,y}$ and $p_{x,y}(\tau) = D_{x,y} \cdot q_{x,y}(\tau)$. In terms of our probability space $(\boldsymbol{w}_1, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\tau})$, we have:

for each $x_1y_1 \in \{00, 01, 10\}$:

$$\begin{aligned} p_{x_1y_1}(\tau) &= \mathbb{E}\big[p_{\boldsymbol{x},\boldsymbol{y}}(\tau) \,\big|\, \boldsymbol{x}_1\boldsymbol{y}_1 = x_1y_1\big] \\ \textstyle\sum_{\tau \in A} p_{x_1y_1}(\tau) &= \mathbb{E}\big[D_{\boldsymbol{x},\boldsymbol{y}} \,\big|\, \boldsymbol{x}_1\boldsymbol{y}_1 = x_1y_1\big] \\ q_{x_1y_1}(\tau) &= \mathbb{E}\big[q_{\boldsymbol{x},\boldsymbol{y}}(\tau) \,\big|\, \boldsymbol{x}_1\boldsymbol{y}_1 = x_1y_1\big] \\ &= \mathbb{P}\big[\boldsymbol{\tau} = \tau \,\big|\, \boldsymbol{x}_1\boldsymbol{y}_1 = x_1y_1\big] \end{aligned}$$

for $x_1y_1 = 11$:

$$\begin{aligned} p_{11}(\tau) &= \mathbb{E}[p_{\widehat{\boldsymbol{x}},\widehat{\boldsymbol{y}}}(\tau)] \\ \textstyle\sum_{\tau \in A} p_{11}(\tau) &= \mathbb{E}[D_{\widehat{\boldsymbol{x}},\widehat{\boldsymbol{y}}}] \end{aligned}$$

We postpone the proofs of the following two claims to the end of this subsection.

Claim 3. *If (2) holds then* $\mathbb{P}\big[p_{x_1y_1}(\boldsymbol{\tau})/q_{x_1y_1}(\boldsymbol{\tau}) \geq 0.03 \cdot 3^{-n} \,\big|\, \boldsymbol{x}_1\boldsymbol{y}_1 = x_1y_1\big] \geq 0.8$ *for each* $x_1y_1 \in \{01, 10\}$.

Claim 4. *If (i) does not hold then* $\mathbb{P}\big[p_{00}(\boldsymbol{\tau})/q_{00}(\boldsymbol{\tau}) \leq 75 \cdot 3^{-n} \,\big|\, \boldsymbol{x}_1\boldsymbol{y}_1 = 00\big] \geq 0.8$.

We now show how to combine Claim 2, Claim 3, and Claim 4 to prove that if *(1)* and *(2)* hold and *(i)* does not hold, then *(ii)* holds. Defining

$$\begin{aligned} T_{x_1y_1} &\coloneqq \big\{\tau \in A \,:\, p_{x_1y_1}(\tau)/q_{x_1y_1}(\tau) \geq 0.03 \cdot 3^{-n}\big\} && \text{for each } x_1y_1 \in \{01, 10\} \\ T_{00} &\coloneqq \big\{\tau \in A \,:\, p_{00}(\tau)/q_{00}(\tau) \leq 75 \cdot 3^{-n}\big\} && \text{for } x_1y_1 = 00 \\ T &\coloneqq T_{00} \cap T_{01} \cap T_{10} \end{aligned}$$

we have for each $x_1y_1 \in \{01, 10\}$,

$$\mathbb{P}\big[\boldsymbol{\tau} \in T_{x_1y_1} \,\big|\, \boldsymbol{x}_1\boldsymbol{y}_1 = 00\big] \;\geq\; \mathbb{P}\big[\boldsymbol{\tau} \in T_{x_1y_1} \,\big|\, \boldsymbol{x}_1\boldsymbol{y}_1 = x_1y_1\big] - 0.05 \;\geq\; 0.75$$

by Claim 2 and Claim 3, and $\mathbb{P}\big[\boldsymbol{\tau} \in T_{00} \,\big|\, \boldsymbol{x}_1\boldsymbol{y}_1 = 00\big] \geq 0.8$ by Claim 4, so by a union bound,

$$\textstyle\sum_{\tau \in T} q_{00}(\tau) \;=\; \mathbb{P}\big[\boldsymbol{\tau} \in T \,\big|\, \boldsymbol{x}_1\boldsymbol{y}_1 = 00\big] \;\geq\; 0.3. \tag{$\dagger$}$$

For each $x_1y_1 \in \{01, 10\}$ we define $d_{x_1y_1}(\tau) \coloneqq \big|q_{00}(\tau) - q_{x_1y_1}(\tau)\big|$ so that by Claim 2,

$$\textstyle\sum_{\tau \in A} d_{x_1y_1}(\tau) \;=\; 2\,\Delta\big((\boldsymbol{\tau} \,|\, \boldsymbol{x}_1\boldsymbol{y}_1 = 00), (\boldsymbol{\tau} \,|\, \boldsymbol{x}_1\boldsymbol{y}_1 = x_1y_1)\big) \;\leq\; 0.1. \tag{$\ddagger$}$$

Since $\boldsymbol{x}_{-1}, \boldsymbol{y}_{-1}$ are independent (implicitly conditioned on $\boldsymbol{w}_{-1} = w_{-1}$), we have $p_{00}(\tau) \cdot p_{11}(\tau) = p_{01}(\tau) \cdot p_{10}(\tau)$ by the rectangular nature of any transcript τ. We would like to rewrite this as $p_{11}(\tau) = p_{01}(\tau) \cdot p_{10}(\tau) \,/\, p_{00}(\tau)$ but we must be careful about division by 0. Adopting the convention $0/0 \coloneqq 0$, we can write

$$p_{11}(\tau) \;\geq\; p_{01}(\tau) \cdot p_{10}(\tau) \,/\, p_{00}(\tau). \tag{$*$}$$

We also note that for each x_1y_1, $p_{x_1y_1}(\tau) = 0$ iff $q_{x_1y_1}(\tau) = 0$. To convert between the "multiplicative" structure of transcripts as in $(*)$ and the "additive" structure of statistical distance, we appeal to the following basic fact, which has been used several times in recent works [GW16, GPW16, GJW18]. For completeness, we reproduce the argument in § A.

Fact 2. *For every* $\tau \in A$, $q_{01}(\tau) \cdot q_{10}(\tau) \,/\, q_{00}(\tau) \geq q_{00}(\tau) - d_{01}(\tau) - d_{10}(\tau)$.

At last we come to the punchline:

$$\begin{aligned} \mathbb{E}[D_{\widehat{\boldsymbol{x}},\widehat{\boldsymbol{y}}}] \;&=\; \sum_{\tau \in A} p_{11}(\tau) \;\geq\; \sum_{\tau \in T} p_{11}(\tau) \;\geq\; \sum_{\tau \in T} \frac{p_{01}(\tau) \cdot p_{10}(\tau)}{p_{00}(\tau)} \\ &=\; \sum_{\tau \in T} \frac{\frac{p_{01}(\tau)}{q_{01}(\tau)} \cdot \frac{p_{10}(\tau)}{q_{10}(\tau)}}{\frac{p_{00}(\tau)}{q_{00}(\tau)}} \cdot \frac{q_{01}(\tau) \cdot q_{10}(\tau)}{q_{00}(\tau)} \\ &\geq\; \sum_{\tau \in T} \frac{(0.03 \cdot 3^{-n}) \cdot (0.03 \cdot 3^{-n})}{75 \cdot 3^{-n}} \cdot \big(q_{00}(\tau) - d_{01}(\tau) - d_{10}(\tau)\big) \\ &\geq\; 0.00001 \cdot 3^{-n} \cdot \Bigg(\sum_{\tau \in T} q_{00}(\tau) - \sum_{\tau \in A} d_{01}(\tau) - \sum_{\tau \in A} d_{10}(\tau)\Bigg) \end{aligned}$$

$$\geq\ 0.00001 \cdot 3^{-n} \cdot \big(0.3 - 0.1 - 0.1\big) \ = \ 0.000001 \cdot 3^{-n}$$

where the third line uses Fact 2, and the last line follows by (†) and (‡). Thus *(ii)* holds. This finishes the proof of Lemma 1, except for the proofs of Claim 3, Claim 4, and Lemma 2.

Proof of Claim 3. To slightly declutter notation, we write the argument for $x_1y_1 = 01$ (nothing is different for $x_1y_1 = 10$). Assuming *(2)* holds, we have $\mathbb{E}\big[\max(3^{-n} - D_{\boldsymbol{x},\boldsymbol{y}}, 0) \,\big|\, \boldsymbol{x}_1\boldsymbol{y}_1 = 01\big] \leq 0.12 \cdot 3^{-n}$. We define S as the set of all (x, y) in the support of $(\boldsymbol{x}, \boldsymbol{y})$ conditioned on $\boldsymbol{x}_1\boldsymbol{y}_1 = 01$ (and implicitly on $\boldsymbol{w}_{-1} = w_{-1}$) such that $D_{x,y} \leq 0.16 \cdot 3^{-n}$ ("bad inputs"). By Markov's inequality,

$$\mathbb{P}\big[(\boldsymbol{x},\boldsymbol{y}) \in S \,\big|\, \boldsymbol{x}_1\boldsymbol{y}_1 = 01\big] \ = \ \mathbb{P}\big[\max(3^{-n} - D_{\boldsymbol{x},\boldsymbol{y}}, 0) \geq 0.84 \cdot 3^{-n} \,\big|\, \boldsymbol{x}_1\boldsymbol{y}_1 = 01\big] \ \leq \ 1/7 \ \leq \ 0.15.$$

We define B as the set of all $\tau \in A$ such that $\mathbb{P}\big[(\boldsymbol{x},\boldsymbol{y}) \in S \,\big|\, \boldsymbol{\tau} = \tau,\, \boldsymbol{x}_1\boldsymbol{y}_1 = 01\big] \geq 0.8$ ("bad transcripts"). We must have $\mathbb{P}\big[\boldsymbol{\tau} \in B \,\big|\, \boldsymbol{x}_1\boldsymbol{y}_1 = 01\big] \leq 0.2$ since otherwise

$$\begin{aligned}
\mathbb{P}\big[(\boldsymbol{x},\boldsymbol{y}) \in S \,\big|\, \boldsymbol{x}_1\boldsymbol{y}_1 = 01\big] \ &\geq \ \mathbb{P}\big[(\boldsymbol{x},\boldsymbol{y}) \in S \text{ and } \boldsymbol{\tau} \in B \,\big|\, \boldsymbol{x}_1\boldsymbol{y}_1 = 01\big] \\
&= \ \mathbb{P}\big[\boldsymbol{\tau} \in B \,\big|\, \boldsymbol{x}_1\boldsymbol{y}_1 = 01\big] \cdot \mathbb{P}\big[(\boldsymbol{x},\boldsymbol{y}) \in S \,\big|\, \boldsymbol{\tau} \in B,\, \boldsymbol{x}_1\boldsymbol{y}_1 = 01\big] \\
&\geq \ 0.2 \cdot 0.8 \ = \ 0.16 \ > \ 0.15.
\end{aligned}$$

Let $\chi_{x,y}$ be the indicator for $(x, y) \notin S$, so $D_{x,y} \geq 0.16 \cdot 3^{-n} \cdot \chi_{x,y}$. For each $\tau \in A \smallsetminus B$ we have

$$\begin{aligned}
\mathbb{E}\big[\chi_{\boldsymbol{x},\boldsymbol{y}} \cdot q_{\boldsymbol{x},\boldsymbol{y}}(\tau) \,\big|\, \boldsymbol{x}_1\boldsymbol{y}_1 = 01\big] \ &= \ \textstyle\sum_{(x,y)\notin S} \mathbb{P}\big[\boldsymbol{xy} = xy \,\big|\, \boldsymbol{x}_1\boldsymbol{y}_1 = 01\big] \cdot \mathbb{P}\big[\boldsymbol{\tau} = \tau \,\big|\, \boldsymbol{xy} = xy\big] \\
&= \ \mathbb{P}\big[(\boldsymbol{x},\boldsymbol{y}) \notin S \text{ and } \boldsymbol{\tau} = \tau \,\big|\, \boldsymbol{x}_1\boldsymbol{y}_1 = 01\big] \\
&= \ \mathbb{P}\big[(\boldsymbol{x},\boldsymbol{y}) \notin S \,\big|\, \boldsymbol{\tau} = \tau,\, \boldsymbol{x}_1\boldsymbol{y}_1 = 01\big] \cdot \mathbb{P}\big[\boldsymbol{\tau} = \tau \,\big|\, \boldsymbol{x}_1\boldsymbol{y}_1 = 01\big] \\
&\geq \ 0.2 \cdot q_{01}(\tau)
\end{aligned}$$

and thus

$$\begin{aligned}
p_{01}(\tau) \ &= \ \mathbb{E}\big[p_{\boldsymbol{x},\boldsymbol{y}}(\tau) \,\big|\, \boldsymbol{x}_1\boldsymbol{y}_1 = 01\big] \ = \ \mathbb{E}\big[D_{\boldsymbol{x},\boldsymbol{y}} \cdot q_{\boldsymbol{x},\boldsymbol{y}}(\tau) \,\big|\, \boldsymbol{x}_1\boldsymbol{y}_1 = 01\big] \\
&\geq \ 0.16 \cdot 3^{-n} \cdot \mathbb{E}\big[\chi_{\boldsymbol{x},\boldsymbol{y}} \cdot q_{\boldsymbol{x},\boldsymbol{y}}(\tau) \,\big|\, \boldsymbol{x}_1\boldsymbol{y}_1 = 01\big] \ \geq \ 0.16 \cdot 3^{-n} \cdot 0.2 \cdot q_{01}(\tau) \ \geq \ 0.03 \cdot 3^{-n} \cdot q_{01}(\tau).
\end{aligned}$$

In summary, $\mathbb{P}\big[p_{01}(\boldsymbol{\tau})/q_{01}(\boldsymbol{\tau}) \geq 0.03 \cdot 3^{-n} \,\big|\, \boldsymbol{x}_1\boldsymbol{y}_1 = 01\big] \geq \mathbb{P}\big[\boldsymbol{\tau} \notin B \,\big|\, \boldsymbol{x}_1\boldsymbol{y}_1 = 01\big] \geq 0.8$. □

Proof of Claim 4. Assume $\mathbb{E}\big[D_{\boldsymbol{x},\boldsymbol{y}} \,\big|\, \boldsymbol{x}_1\boldsymbol{y}_1 = 00\big] \leq 15 \cdot 3^{-n}$ since otherwise *(i)* would hold because

$$\mathbb{E}[D_{\boldsymbol{x},\boldsymbol{y}}] \ = \ \mathbb{E}_{x_1y_1 \sim \boldsymbol{x}_1\boldsymbol{y}_1} \mathbb{E}\big[D_{\boldsymbol{x},\boldsymbol{y}} \,\big|\, \boldsymbol{x}_1\boldsymbol{y}_1 = x_1y_1\big] \ \geq \ \tfrac{1}{3}\mathbb{E}\big[D_{\boldsymbol{x},\boldsymbol{y}} \,\big|\, \boldsymbol{x}_1\boldsymbol{y}_1 = 00\big] \ \geq \ 5 \cdot 3^{-n}.$$

Now

$$\mathbb{E}\left[\frac{p_{00}(\boldsymbol{\tau})}{q_{00}(\boldsymbol{\tau})} \,\middle|\, \boldsymbol{x}_1\boldsymbol{y}_1 = 00\right] \ = \ \sum_{\tau \in A} q_{00}(\tau) \cdot \frac{p_{00}(\tau)}{q_{00}(\tau)} \ = \ \sum_{\tau \in A} p_{00}(\tau) \ = \ \mathbb{E}\big[D_{\boldsymbol{x},\boldsymbol{y}} \,\big|\, \boldsymbol{x}_1\boldsymbol{y}_1 = 00\big] \ \leq \ 15 \cdot 3^{-n}.$$

Thus $\mathbb{P}\big[p_{00}(\boldsymbol{\tau})/q_{00}(\boldsymbol{\tau}) \leq 75 \cdot 3^{-n} \,\big|\, \boldsymbol{x}_1\boldsymbol{y}_1 = 00\big] \geq 0.8$ follows by Markov's inequality. □

2.3 Proof of Lemma 2

Restatement of Lemma 2. *Let* $\boldsymbol{a}, \boldsymbol{b}$ *be jointly distributed, with* $\boldsymbol{b}$ *having support* $\{0,1\}$*. Then*

$$\Delta\big((\boldsymbol{a} \,|\, \boldsymbol{b}=0), (\boldsymbol{a} \,|\, \boldsymbol{b}=1)\big) \;\leq\; \sqrt{\mathbb{I}(\boldsymbol{a} \,;\, \boldsymbol{b}) \,/\, \big(3 \cdot \mathbb{P}[\boldsymbol{b}=0] \cdot \mathbb{P}[\boldsymbol{b}=1]\big)}.$$

We first note that for every outcome a in the support of $\boldsymbol{a}$,

$$\begin{aligned}
& \big|\mathbb{P}[\boldsymbol{a}=a \,|\, \boldsymbol{b}=0] - \mathbb{P}[\boldsymbol{a}=a]\big| \\
=\; & \Big|\mathbb{P}[\boldsymbol{a}=a \,|\, \boldsymbol{b}=0] - \big(\mathbb{P}[\boldsymbol{a}=a \,|\, \boldsymbol{b}=0] \cdot \mathbb{P}[\boldsymbol{b}=0] + \mathbb{P}[\boldsymbol{a}=a \,|\, \boldsymbol{b}=1] \cdot \mathbb{P}[\boldsymbol{b}=1]\big)\Big| \\
=\; & \big|\mathbb{P}[\boldsymbol{a}=a \,|\, \boldsymbol{b}=0] \cdot (1 - \mathbb{P}[\boldsymbol{b}=0]) - \mathbb{P}[\boldsymbol{a}=a \,|\, \boldsymbol{b}=1] \cdot \mathbb{P}[\boldsymbol{b}=1]\big| \\
=\; & \mathbb{P}[\boldsymbol{b}=1] \cdot \big|\mathbb{P}[\boldsymbol{a}=a \,|\, \boldsymbol{b}=0] - \mathbb{P}[\boldsymbol{a}=a \,|\, \boldsymbol{b}=1]\big|
\end{aligned}$$

and thus

$$\begin{aligned}
\Delta\big((\boldsymbol{b} \,|\, \boldsymbol{a}=a), \boldsymbol{b}\big) \;&=\; \big|\mathbb{P}[\boldsymbol{b}=0 \,|\, \boldsymbol{a}=a] - \mathbb{P}[\boldsymbol{b}=0]\big| \;=\; \left|\frac{\mathbb{P}[\boldsymbol{a}=a \,|\, \boldsymbol{b}=0] \cdot \mathbb{P}[\boldsymbol{b}=0]}{\mathbb{P}[\boldsymbol{a}=a]} - \mathbb{P}[\boldsymbol{b}=0]\right| \\
&=\; \mathbb{P}[\boldsymbol{b}=0] \cdot \left|\frac{\mathbb{P}[\boldsymbol{a}=a \,|\, \boldsymbol{b}=0] - \mathbb{P}[\boldsymbol{a}=a]}{\mathbb{P}[\boldsymbol{a}=a]}\right| \\
&=\; \mathbb{P}[\boldsymbol{b}=0] \cdot \mathbb{P}[\boldsymbol{b}=1] \cdot \frac{\big|\mathbb{P}[\boldsymbol{a}=a \,|\, \boldsymbol{b}=0] - \mathbb{P}[\boldsymbol{a}=a \,|\, \boldsymbol{b}=1]\big|}{\mathbb{P}[\boldsymbol{a}=a]}.
\end{aligned}$$

It follows that

$$\begin{aligned}
\mathbb{I}(\boldsymbol{a} \,;\, \boldsymbol{b}) \;&=\; \mathbb{E}_{a \sim \boldsymbol{a}} \mathbb{D}\big((\boldsymbol{b} \,|\, \boldsymbol{a}=a) \,\big\|\, \boldsymbol{b}\big) \\
&\geq\; \sum_a \mathbb{P}[\boldsymbol{a}=a] \cdot \tfrac{2}{\ln 2} \Delta\big((\boldsymbol{b} \,|\, \boldsymbol{a}=a), \boldsymbol{b}\big)^2 \\
&=\; \sum_a \mathbb{P}[\boldsymbol{a}=a] \cdot \tfrac{2}{\ln 2} \left(\mathbb{P}[\boldsymbol{b}=0] \cdot \mathbb{P}[\boldsymbol{b}=1] \cdot \frac{\big|\mathbb{P}[\boldsymbol{a}=a \,|\, \boldsymbol{b}=0] - \mathbb{P}[\boldsymbol{a}=a \,|\, \boldsymbol{b}=1]\big|}{\mathbb{P}[\boldsymbol{a}=a]}\right)^2 \\
&=\; \tfrac{2}{\ln 2} \cdot \mathbb{P}[\boldsymbol{b}=0]^2 \cdot \mathbb{P}[\boldsymbol{b}=1]^2 \cdot \sum_a \frac{\big|\mathbb{P}[\boldsymbol{a}=a \,|\, \boldsymbol{b}=0] - \mathbb{P}[\boldsymbol{a}=a \,|\, \boldsymbol{b}=1]\big|^2}{\mathbb{P}[\boldsymbol{a}=a]}
\end{aligned}$$

using the last two bullets of Fact 1 for the first two lines. Now we define vectors u and v indexed by the support of $\boldsymbol{a}$:

$$u_a \coloneqq \big|\mathbb{P}[\boldsymbol{a}=a \,|\, \boldsymbol{b}=0] - \mathbb{P}[\boldsymbol{a}=a \,|\, \boldsymbol{b}=1]\big| / \sqrt{\mathbb{P}[\boldsymbol{a}=a]} \qquad \text{and} \qquad v_a \coloneqq \sqrt{\mathbb{P}[\boldsymbol{a}=a]}.$$

By Cauchy–Schwarz,

$$u \cdot u \;\geq\; (u \cdot v)^2/(v \cdot v) \;=\; \big(2\,\Delta\big((\boldsymbol{a} \,|\, \boldsymbol{b}=0), (\boldsymbol{a} \,|\, \boldsymbol{b}=1)\big)\big)^2/1.$$

Combining, we get

$$\mathbb{I}(\boldsymbol{a} \,;\, \boldsymbol{b}) \;\geq\; \tfrac{2}{\ln 2} \cdot \mathbb{P}[\boldsymbol{b}=0]^2 \cdot \mathbb{P}[\boldsymbol{b}=1]^2 \cdot (u \cdot u) \;\geq\; \tfrac{8}{\ln 2} \cdot \mathbb{P}[\boldsymbol{b}=0]^2 \cdot \mathbb{P}[\boldsymbol{b}=1]^2 \cdot \Delta\big((\boldsymbol{a} \,|\, \boldsymbol{b}=0), (\boldsymbol{a} \,|\, \boldsymbol{b}=1)\big)^2.$$

Rearranging and using $\sqrt{8/\ln 2} \geq 3$ gives the lemma.

A Supplementary proofs

Restatement of Fact 1 (Direct sum). $\mathbb{I}(\boldsymbol{a}\,;\,\boldsymbol{b}_1\cdots\boldsymbol{b}_n) \geq \mathbb{I}(\boldsymbol{a}\,;\,\boldsymbol{b}_1)+\cdots+\mathbb{I}(\boldsymbol{a}\,;\,\boldsymbol{b}_n)$ *if* $\boldsymbol{b}_1\cdots\boldsymbol{b}_n$ *are fully independent.*

Proof. We have $\mathbb{H}(\boldsymbol{b}_1\cdots\boldsymbol{b}_n) = \mathbb{H}(\boldsymbol{b}_1)+\cdots+\mathbb{H}(\boldsymbol{b}_n)$ by full independence, and $\mathbb{H}(\boldsymbol{b}_1\cdots\boldsymbol{b}_n\,|\,\boldsymbol{a}) \leq \mathbb{H}(\boldsymbol{b}_1\,|\,\boldsymbol{a})+\cdots+\mathbb{H}(\boldsymbol{b}_n\,|\,\boldsymbol{a})$ by subadditivity of entropy. Thus

$$\mathbb{I}(\boldsymbol{a}\,;\,\boldsymbol{b}_1\cdots\boldsymbol{b}_n) \;=\; \mathbb{H}(\boldsymbol{b}_1\cdots\boldsymbol{b}_n)-\mathbb{H}(\boldsymbol{b}_1\cdots\boldsymbol{b}_n\,|\,\boldsymbol{a}) \;\geq\; \textstyle\sum_i\big(\mathbb{H}(\boldsymbol{b}_i)-\mathbb{H}(\boldsymbol{b}_i\,|\,\boldsymbol{a})\big) \;=\; \textstyle\sum_i \mathbb{I}(\boldsymbol{a}\,;\,\boldsymbol{b}_i). \quad \square$$

Restatement of Fact 2. *For every* $\tau\in A$, $q_{01}(\tau)\cdot q_{10}(\tau)\,/\,q_{00}(\tau) \geq q_{00}(\tau)-d_{01}(\tau)-d_{10}(\tau)$.

Proof. It suffices to show that

$$q_{01}(\tau)\cdot q_{10}(\tau) \;\geq\; q_{00}(\tau)^2 - q_{00}(\tau)\big(d_{01}(\tau)+d_{10}(\tau)\big). \tag{1}$$

(If $q_{00}(\tau)\neq 0$ then the desired inequality follows by dividing (1) by $q_{00}(\tau)$, and if $q_{00}(\tau)=0$ then it follows since its right side is ≤ 0 and its left side is 0; recall our convention that $0/0 \coloneqq 0$.) For some signs $\sigma_{x_1y_1}(\tau)\in\{1,-1\}$, the left side of (1) equals $\big(q_{00}(\tau)+\sigma_{01}(\tau)d_{01}(\tau)\big)\cdot\big(q_{00}(\tau)+\sigma_{10}(\tau)d_{10}(\tau)\big)$, which expands to

$$q_{00}(\tau)^2+\sigma_{01}(\tau)q_{00}(\tau)d_{01}(\tau)+\sigma_{10}(\tau)q_{00}(\tau)d_{10}(\tau)+\sigma_{01}(\tau)\sigma_{10}(\tau)d_{01}(\tau)d_{10}(\tau). \tag{2}$$

If $\sigma_{01}(\tau)=\sigma_{10}(\tau)$ then (2) is at least the right side of (1) since the last term of (2) is nonnegative. If $\sigma_{01}(\tau)\neq\sigma_{10}(\tau)$, say $\sigma_{01}(\tau)=-1$ and $\sigma_{10}(\tau)=1$, then (2) is at least the right side of (1) since the sum of the last two terms in (2) is $q_{00}(\tau)d_{10}(\tau)-d_{01}(\tau)d_{10}(\tau)=q_{01}(\tau)d_{10}(\tau)\geq 0$. $\square$

References

[AA05] Scott Aaronson and Andris Ambainis. Quantum search of spatial regions. *Theory of Computing*, 1(1):47–79, 2005. doi:10.4086/toc.2005.v001a004.

[Aar14] Scott Aaronson. The equivalence of sampling and searching. *Theory of Computing Systems*, 55(2):281–298, 2014. doi:10.1007/s00224-013-9527-3.

[ACK19] Sepehr Assadi, Yu Chen, and Sanjeev Khanna. Polynomial pass lower bounds for graph streaming algorithms. Technical Report abs/1904.04720, arXiv, 2019.

[ARW17] Amir Abboud, Aviad Rubinstein, and Ryan Williams. Distributed PCP theorems for hardness of approximation in P. In *Proceedings of the 58th Symposium on Foundations of Computer Science (FOCS)*, pages 25–36. IEEE, 2017. doi:10.1109/FOCS.2017.12.

[ASTS+03] Andris Ambainis, Leonard Schulman, Amnon Ta-Shma, Umesh Vazirani, and Avi Wigderson. The quantum communication complexity of sampling. *SIAM Journal on Computing*, 32(6):1570–1585, 2003. doi:10.1137/S009753979935476.

[AW09] Scott Aaronson and Avi Wigderson. Algebrization: A new barrier in complexity theory. *ACM Transactions on Computation Theory*, 1(1):2:1–2:54, 2009. doi:10.1145/1490270.1490272.

[AWY18] Josh Alman, Joshua Wang, and Huacheng Yu. Cell-probe lower bounds from online communication complexity. In *Proceedings of the 50th Symposium on Theory of Computing (STOC)*, pages 1003–1012. ACM, 2018. doi:10.1145/3188745.3188862.

[BCK+14] Joshua Brody, Amit Chakrabarti, Ranganath Kondapally, David Woodruff, and Grigory Yaroslavtsev. Beyond set disjointness: The communication complexity of finding the intersection. In *Proceedings of the 33rd Symposium on Principles of Distributed Computing (PODC)*, pages 106–113. ACM, 2014. doi:10.1145/2611462.2611501.

[BCS14] Itai Benjamini, Gil Cohen, and Igor Shinkar. Bi-lipschitz bijection between the boolean cube and the hamming ball. In *Proceedings of the 55th Symposium on Foundations of Computer Science (FOCS)*, pages 81–89. IEEE, 2014. doi:10.1109/FOCS.2014.17.

[BCW98] Harry Buhrman, Richard Cleve, and Avi Wigderson. Quantum vs. classical communication and computation. In *Proceedings of the 30th Symposium on Theory of Computing (STOC)*, pages 63–68. ACM, 1998. doi:10.1145/276698.276713.

[BEO+13] Mark Braverman, Faith Ellen, Rotem Oshman, Toniann Pitassi, and Vinod Vaikuntanathan. A tight bound for set disjointness in the message-passing model. In *Proceedings of the 54th Symposium on Foundations of Computer Science (FOCS)*, pages 668–677. IEEE, 2013. doi:10.1109/FOCS.2013.77.

[BFS86] László Babai, Peter Frankl, and Janos Simon. Complexity classes in communication complexity theory. In *Proceedings of the 27th Symposium on Foundations of Computer Science (FOCS)*, pages 337–347. IEEE, 1986. doi:10.1109/SFCS.1986.15.

[BGK15] Ralph Bottesch, Dmitry Gavinsky, and Hartmut Klauck. Correlation in hard distributions in communication complexity. In *Proceedings of the 19th International Workshop on Randomization and Computation (RANDOM)*, pages 544–572. Schloss Dagstuhl, 2015. doi:10.4230/LIPIcs.APPROX-RANDOM.2015.544.

[BGK+18] Mark Braverman, Ankit Garg, Young Kun-Ko, Jieming Mao, and Dave Touchette. Near-optimal bounds on the bounded-round quantum communication complexity of disjointness. *SIAM Journal on Computing*, 47(6):2277–2314, 2018. doi:10.1137/16M1061400.

[BGMdW13] Harry Buhrman, David Garcia-Soriano, Arie Matsliah, and Ronald de Wolf. The non-adaptive query complexity of testing k-parities. *Chicago Journal of Theoretical Computer Science*, 2013(6):1–11, 2013. doi:10.4086/cjtcs.2013.006.

[BGPW13] Mark Braverman, Ankit Garg, Denis Pankratov, and Omri Weinstein. From information to exact communication. In *Proceedings of the 45th Symposium on Theory of Computing (STOC)*, pages 151–160. ACM, 2013. doi:10.1145/2488608.2488628.

[BH09] Paul Beame and Dang-Trinh Huynh-Ngoc. Multiparty communication complexity and threshold circuit size of AC^0. In *Proceedings of the 50th Symposium on Foundations of Computer Science (FOCS)*, pages 53–62. IEEE, 2009. doi:10.1109/FOCS.2009.12.

[BIL12] Christopher Beck, Russell Impagliazzo, and Shachar Lovett. Large deviation bounds for decision trees and sampling lower bounds for AC^0-circuits. In *Proceedings of the 53rd Symposium on Foundations of Computer Science (FOCS)*, pages 101–110. IEEE, 2012. doi:10.1109/FOCS.2012.82.

[BKM18] Lucas Boczkowski, Iordanis Kerenidis, and Frédéric Magniez. Streaming communication protocols. *ACM Transactions on Computation Theory*, 10(4):19:1–19:21, 2018. doi:10.1145/3276748.

[BM13] Mark Braverman and Ankur Moitra. An information complexity approach to extended formulations. In *Proceedings of the 45th Symposium on Theory of Computing (STOC)*, pages 161–170. ACM, 2013. doi:10.1145/2488608.2488629.

[BO15] Mark Braverman and Rotem Oshman. On information complexity in the broadcast model. In *Proceedings of the 34th Symposium on Principles of Distributed Computing (PODC)*, pages 355–364. ACM, 2015. doi:10.1145/2767386.2767425.

[BO17] Mark Braverman and Rotem Oshman. A rounds vs. communication tradeoff for multi-party set disjointness. In *Proceedings of the 58th Symposium on Foundations of Computer Science (FOCS)*, pages 144–155. IEEE, 2017. doi:10.1109/FOCS.2017.22.

[BPSW06] Paul Beame, Toniann Pitassi, Nathan Segerlind, and Avi Wigderson. A strong direct product theorem for corruption and the multiparty communication complexity of disjointness. *Computational Complexity*, 15(4):391–432, 2006. doi:10.1007/s00037-007-0220-2.

[BRdW08] Avraham Ben-Aroya, Oded Regev, and Ronald de Wolf. A hypercontractive inequality for matrix-valued functions with applications to quantum computing and LDCs. In *Proceedings of the 49th Symposium on Foundations of Computer Science (FOCS)*, pages 477–486. IEEE, 2008. doi:10.1109/FOCS.2008.45.

[BYJKS04] Ziv Bar-Yossef, T.S. Jayram, Ravi 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. doi:10.1016/j.jcss.2003.11.006.

[CA08] Arkadev Chattopadhyay and Anil Ada. Multiparty communication complexity of disjointness. Technical Report TR08-002, Electronic Colloquium on Computational Complexity (ECCC), 2008. URL: https://eccc.weizmann.ac.il//eccc-reports/2008/TR08-002/.

[Che18] Lijie Chen. On the hardness of approximate and exact (bichromatic) maximum inner product. In *Proceedings of the 33rd Computational Complexity Conference (CCC)*, pages 14:1–14:45. Schloss Dagstuhl, 2018. doi:10.4230/LIPIcs.CCC.2018.14.

[CKS03] Amit Chakrabarti, Subhash Khot, and Xiaodong Sun. Near-optimal lower bounds on the multi-party communication complexity of set disjointness. In *Proceedings of the 18th Conference on Computational Complexity*, pages 107–117. IEEE, 2003. doi:10.1109/CCC.2003.1214414.

[CSWY01] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *Proceedings of the 42nd Symposium on Foundations of Computer Science (FOCS)*, pages 270–278. IEEE, 2001. doi:10.1109/SFCS.2001.959901.

[CT06] Thomas Cover and Joy Thomas. *Elements of Information Theory*. Wiley, 2006.

[DFHL18] Yuval Dagan, Yuval Filmus, Hamed Hatami, and Yaqiao Li. Trading information complexity for error. *Theory of Computing*, 14(1):1–73, 2018. doi:10.4086/toc.2018.v014a006.

[DKS12] Anirban Dasgupta, Ravi Kumar, and D. Sivakumar. Sparse and lopsided set disjointness via information theory. In *Proceedings of the 16th International Workshop on Randomization and Computation (RANDOM)*, pages 517–528. Springer, 2012. doi:10.1007/978-3-642-32512-0_44.

[DP09] Devdatt Dubhashi and Alessandro Panconesi. *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge University Press, 2009.

[DW12] Anindya De and Thomas Watson. Extractors and lower bounds for locally samplable sources. *ACM Transactions on Computation Theory*, 4(1):3:1–3:21, 2012. doi:10.1145/2141938.2141941.

[FHLY17] Yuval Filmus, Hamed Hatami, Yaqiao Li, and Suzin You. Information complexity of the AND function in the two-party and multi-party settings. In *Proceedings of the 23rd International Computing and Combinatorics Conference (COCOON)*, pages 200–211. Springer, 2017. doi:10.1007/978-3-319-62389-4_17.

[Gav16] Dmitry Gavinsky. Communication complexity of inevitable intersection. Technical Report abs/1611.08842, arXiv, 2016.

[GGN10] Oded Goldreich, Shafi Goldwasser, and Asaf Nussboim. On the implementation of huge random objects. *SIAM Journal on Computing*, 39(7):2761–2822, 2010. doi:10.1137/080722771.

[GJW18] Mika Göös, Rahul Jain, and Thomas Watson. Extension complexity of independent set polytopes. *SIAM Journal on Computing*, 47(1):241–269, 2018. doi:10.1137/16M109884X.

[GLM+16] Mika Göös, Shachar Lovett, Raghu Meka, Thomas Watson, and David Zuckerman. Rectangles are nonnegative juntas. *SIAM Journal on Computing*, 45(5):1835–1869, 2016. doi:10.1137/15M103145X.

[GPW16] Mika Göös, Toniann Pitassi, and Thomas Watson. Zero-information protocols and unambiguity in Arthur–Merlin communication. *Algorithmica*, 76(3):684–719, 2016. doi:10.1007/s00453-015-0104-9.

[Gro94] Vince Grolmusz. The BNS lower bound for multi-party protocols is nearly optimal. *Information and Computation*, 112(1):51–54, 1994. doi:10.1006/inco.1994.1051.

[Gro09] André Gronemeier. Asymptotically optimal lower bounds on the NIH-multi-party information complexity of the AND-function and disjointness. In *Proceedings of the 26th International Symposium on Theoretical Aspects of Computer Science (STACS)*, pages 505–516. Schloss Dagstuhl, 2009. doi:10.4230/LIPIcs.STACS.2009.1846.

[GS10] Dmitry Gavinsky and Alexander Sherstov. A separation of NP and coNP in multiparty communication complexity. *Theory of Computing*, 6(1):227–245, 2010. doi:10.4086/toc.2010.v006a010.

[GW16] Mika Göös and Thomas Watson. Communication complexity of set-disjointness for all probabilities. *Theory of Computing*, 12(9):1–23, 2016. doi:10.4086/toc.2016.v012a009.

[HdW02] Peter Høyer and Ronald de Wolf. Improved quantum communication complexity bounds for disjointness and equality. In *Proceedings of the 19th Symposium on Theoretical Aspects of Computer Science (STACS)*, pages 299–310. Springer, 2002. doi:10.1007/3-540-45841-7_24.

[HW07] Johan Håstad and Avi Wigderson. The randomized communication complexity of set disjointness. *Theory of Computing*, 3(1):211–219, 2007. doi:10.4086/toc.2007.v003a011.

[Jay09] T.S. Jayram. Hellinger strikes back: A note on the multi-party information complexity of AND. In *Proceedings of the 13th International Workshop on Randomization and Computation (RANDOM)*, pages 562–573. Springer, 2009. doi:10.1007/978-3-642-03685-9_42.

[JK10] Rahul Jain and Hartmut Klauck. The partition bound for classical communication complexity and query complexity. In *Proceedings of the 25th Conference on Computational Complexity (CCC)*, pages 247–258. IEEE, 2010. doi:10.1109/CCC.2010.31.

[JKN08] Rahul Jain, Hartmut Klauck, and Ashwin Nayak. Direct product theorems for classical communication complexity via subdistribution bounds. In *Proceedings of the 40th Symposium on Theory of Computing (STOC)*, pages 599–608. ACM, 2008. doi:10.1145/1374376.1374462.

[JRS03] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A lower bound for the bounded round quantum communication complexity of set disjointness. In *Proceedings of the 44th Symposium on Foundations of Computer Science (FOCS)*, pages 220–229. IEEE, 2003. doi:10.1109/SFCS.2003.1238196.

[JSWZ13] Rahul Jain, Yaoyun Shi, Zhaohui Wei, and Shengyu Zhang. Efficient protocols for generating bipartite classical distributions and quantum states. *IEEE Transactions on Information Theory*, 59(8):5171–5178, 2013. doi:10.1109/TIT.2013.2258372.

[Kla03] Hartmut Klauck. Rectangle size bounds and threshold covers in communication complexity. In *Proceedings of the 18th Conference on Computational Complexity (CCC)*, pages 118–134. IEEE, 2003. doi:10.1109/CCC.2003.1214415.

[Kla10] Hartmut Klauck. A strong direct product theorem for disjointness. In *Proceedings of the 42nd Symposium on Theory of Computing (STOC)*, pages 77–86. ACM, 2010. doi:10.1145/1806689.1806702.

[KMSY19] Gillat Kol, Shay Moran, Amir Shpilka, and Amir Yehudayoff. Approximate nonnegative rank is equivalent to the smooth rectangle bound. *Computational Complexity*, 28(1):1–25, 2019. doi:10.1007/s00037-018-0176-4.

[KN97] Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, 1997.

[KNTZ07] Hartmut Klauck, Ashwin Nayak, Amnon Ta-Shma, and David Zuckerman. Interaction in quantum communication. *IEEE Transactions on Information Theory*, 53(6):1970–1982, 2007. doi:10.1109/TIT.2007.896888.

[KP14] Hartmut Klauck and Supartha Podder. New bounds for the garden-hose model. In *Proceedings of the 34th International Conference on Foundation of Software Technology and Theoretical Computer Science (FSTTCS)*, pages 481–492. Schloss Dagstuhl, 2014. doi:10.4230/LIPIcs.FSTTCS.2014.481.

[KS92] Bala Kalyanasundaram and Georg Schnitger. The probabilistic communication complexity of set intersection. *SIAM Journal on Discrete Mathematics*, 5(4):545–557, 1992. doi:10.1137/0405044.

[KSdW07] Hartmut Klauck, Robert Spalek, and Ronald de Wolf. Quantum and classical strong direct product theorems and optimal time-space tradeoffs. *SIAM Journal on Computing*, 36(5):1472–1493, 2007. doi:10.1137/05063235X.

[KW09] Eyal Kushilevitz and Enav Weinreb. The communication complexity of set-disjointness with small sets and 0-1 intersection. In *Proceedings of the 50th Symposium on Foundations of Computer Science (FOCS)*, pages 63–72. IEEE, 2009. doi:10.1109/FOCS.2009.15.

[Lin91] Jianhua Lin. Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*, 37(1):145–151, 1991. doi:10.1109/18.61115.

[LS09] Troy Lee and Adi Shraibman. Disjointness is hard in the multiparty number-on-the-forehead model. *Computational Complexity*, 18(2):309–336, 2009. doi:10.1007/s00037-009-0276-2.

[LV12] Shachar Lovett and Emanuele Viola. Bounded-depth circuits cannot sample good codes. *Computational Complexity*, 21(2):245–266, 2012. doi:10.1007/s00037-012-0039-3.

[Pat11] Mihai Patrascu. Unifying the landscape of cell-probe lower bounds. *SIAM Journal on Computing*, 40(3):827–847, 2011. doi:10.1137/09075336X.

[PS17] Vladimir Podolskii and Alexander Sherstov. Inner product and set disjointness: Beyond logarithmically many parties. Technical Report abs/1711.10661, arXiv, 2017.

[Raz92] Alexander Razborov. On the distributional complexity of disjointness. *Theoretical Computer Science*, 106(2):385–390, 1992. doi:10.1016/0304-3975(92)90260-M.

[Raz03] Alexander Razborov. Quantum communication complexity of symmetric predicates. *Izvestiya: Mathematics*, 67(1):145–159, 2003. doi:10.1070/IM2003v067n01ABEH000422.

[Rub18] Aviad Rubinstein. Hardness of approximate nearest neighbor search. In *Proceedings of the 50th Symposium on Theory of Computing (STOC)*, pages 1260–1268. ACM, 2018. doi:10.1145/3188745.3188916.

[RY15] Anup Rao and Amir Yehudayoff. Simplified lower bounds on the multiparty communication complexity of disjointness. In *Proceedings of the 30th Computational Complexity Conference (CCC)*, pages 88–101. Schloss Dagstuhl, 2015. doi:10.4230/LIPIcs.CCC.2015.88.

[She11] Alexander Sherstov. The pattern matrix method. *SIAM Journal on Computing*, 40(6):1969–2000, 2011. doi:10.1137/080733644.

[She12] Alexander Sherstov. Strong direct product theorems for quantum communication and query complexity. *SIAM Journal on Computing*, 41(5):1122–1165, 2012. doi:10.1137/110842661.

[She14] Alexander Sherstov. Communication lower bounds using directional derivatives. *Journal of the ACM*, 61(6):1–71, 2014. doi:10.1145/2629334.

[She16] Alexander Sherstov. The multiparty communication complexity of set disjointness. *SIAM Journal on Computing*, 45(4):1450–1489, 2016. doi:10.1137/120891587.

[ST13] Mert Saglam and Gábor Tardos. On the communication complexity of sparse set disjointness and exists-equal problems. In *Proceedings of the 54th Symposium on Foundations of Computer Science (FOCS)*, pages 678–687. IEEE, 2013. doi:10.1109/FOCS.2013.78.

[SZ09] Yaoyun Shi and Yufan Zhu. Quantum communication complexity of block-composed functions. *Quantum Information and Computation*, 9(5–6):444–460, 2009.

[Tes03] Pascal Tesson. *Computational Complexity Questions Related to Finite Monoids and Semigroups*. PhD thesis, McGill University, 2003.

[Vio12a] Emanuele Viola. The complexity of distributions. *SIAM Journal on Computing*, 41(1):191–218, 2012. doi:10.1137/100814998.

[Vio12b] Emanuele Viola. Extractors for Turing-machine sources. In *Proceedings of the 16th International Workshop on Randomization and Computation (RANDOM)*, pages 663–671. Springer, 2012. doi:10.1007/978-3-642-32512-0_56.

[Vio14] Emanuele Viola. Extractors for circuit sources. *SIAM Journal on Computing*, 43(2):655–672, 2014. doi:10.1137/11085983X.

[Vio16] Emanuele Viola. Quadratic maps are hard to sample. *ACM Transactions on Computation Theory*, 8(4):18:1–18:4, 2016. doi:10.1145/2934308.

[Vio18] Emanuele Viola. Sampling lower bounds: Boolean average-case and permutations. Technical Report TR18-060, Electronic Colloquium on Computational Complexity (ECCC), 2018. URL: https://eccc.weizmann.ac.il/report/2018/060.

[Wat14] Thomas Watson. Time hierarchies for sampling distributions. *SIAM Journal on Computing*, 43(5):1709–1727, 2014. doi:10.1137/120898553.

[Wat16] Thomas Watson. Nonnegative rank vs. binary rank. *Chicago Journal of Theoretical Computer Science*, 2016(2):1–13, 2016. doi:10.4086/cjtcs.2016.002.

[Wat18] Thomas Watson. Communication complexity with small advantage. In *Proceedings of the 33rd Computational Complexity Conference (CCC)*, pages 9:1–9:17. Schloss Dagstuhl, 2018. doi:10.4230/LIPIcs.CCC.2018.9.

[WW15] Omri Weinstein and David Woodruff. The simultaneous communication of disjointness with applications to data streams. In *Proceedings of the 42nd International Colloquium on Automata, Languages, and Programming (ICALP)*, pages 1082–1093. Springer, 2015. doi:10.1007/978-3-662-47672-7_88.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il