Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 148 (2016)

Communication Complexity with Small Advantage

Thomas Watson*

April 5, 2020

Abstract

We study problems in randomized communication complexity when the protocol is only required to attain some small advantage over purely random guessing, i.e., it produces the correct output with probability at least ϵ greater than one over the codomain size of the function. Previously, Braverman and Moitra (STOC 2013) showed that the set-intersection function requires $\Theta(\epsilon n)$ communication to achieve advantage ϵ. Building on this, we prove the same bound for several variants of set-intersection: (1) the classic "tribes" function obtained by composing with AND (provided $1/\epsilon$ is at most the width of the AND), and (2) the variant where the sets are uniquely intersecting and the goal is to determine partial information about (say, certain bits of the index of) the intersecting coordinate.

1 Introduction

In randomized communication complexity, protocols are commonly required to succeed with probability at least some constant less than 1, such as 3/4. Achieving success probability one over the codomain size of the function is trivial by outputting a uniformly random guess. There is a spectrum of complexities between these extremes, where we require a protocol to achieve success probability ϵ greater than one over the codomain size, i.e., *advantage* ϵ. We study the fine-grained question "How does the communication complexity of achieving advantage ϵ depend on ϵ?"

Formally, for a two-party function F, let $\mathrm{R}_p(F)$ denote the minimum worst-case communication cost of any randomized protocol (with both public and private coins) that is *p-correct* in the sense that for each input (X, Y) in the domain of F, it outputs $F(X, Y)$ with probability at least p.

First let us consider functions with codomain size 2. One observation is that running an advantage-ϵ protocol $O(1/\epsilon^2)$ times independently and taking the majority outcome yields an advantage-1/4 protocol (we call this "majority-amplification"); i.e., $\mathrm{R}_{1/2+\epsilon}(F) \geq \Omega(\epsilon^2 \mathrm{R}_{3/4}(F))$. However, this does not tell the whole story; achieving advantage ϵ may be harder than this bound suggests, depending on the function. For example, consider the well-studied functions INNER-PROD (inner product mod 2), SET-INTER (set-intersection, where 1-inputs are intersecting), and GAP-HAMMING (determining whether the Hamming distance is $\geq n/2 + \sqrt{n}$ or $\leq n/2 - \sqrt{n}$). Each of these three functions F satisfies $\mathrm{R}_{3/4}(F) = \Theta(n)$, and yet

- $\mathrm{R}_{1/2+\epsilon}(\text{INNER-PROD}) = \Theta(n)$ provided $\epsilon \geq 2^{-o(n)}$ [CG88];
- $\mathrm{R}_{1/2+\epsilon}(\text{SET-INTER}) = \Theta(\epsilon n)$ provided $\epsilon n \geq 1$ [BM13, GW16];
- $\mathrm{R}_{1/2+\epsilon}(\text{GAP-HAMMING}) = \Theta(\epsilon^2 n)$ provided $\epsilon^2 n \geq 1$ [CR12, Vid12, She12].

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

ISSN 1433-8092

(We provide a proof of the Gap-Hamming upper bound in Appendix A since it does not appear in the above references.)

Hence it is naturally interesting to study the dependence of the complexity on ϵ for different important functions, in order to build a more complete understanding of randomized communication. For functions with codomain size greater than 2, small-advantage protocols are not even amenable to amplification, so no lower bounds for them follow a priori from lower bounds for higher-advantage protocols.

The functions we study are defined using composition. Letting $g\colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a two-party total function (usually called a *gadget*), and $f\colon \{0,1\}^n \to \{0,1\}$ be a (possibly partial) function, the two-party composed (possibly partial) function $f \circ g^n\colon \mathcal{X}^n \times \mathcal{Y}^n \to \{0,1\}$ is defined by $(f \circ g^n)(X,Y) \coloneqq f\big(g(X_1,Y_1),\ldots,g(X_n,Y_n)\big)$ where $X = (X_1,\ldots,X_n)$ and $Y = (Y_1,\ldots,Y_n)$ with $X_i \in \mathcal{X}$ and $Y_i \in \mathcal{Y}$ for each i. Sometimes, the outer function f itself will be defined using standard function composition.

In the functions $\textsc{And}_m$ and $\textsc{Or}_m$, the subscript indicates the number of input bits.

1.1 Tribes

Just as Set-Inter is the canonical NP-complete communication problem, so-called Tribes is the canonical $\Pi_2\mathsf{P}$-complete communication problem. A linear randomized lower bound for Tribes (with constant advantage) was shown in [JKS03] using information complexity (thereby giving a nearly optimal (quadratic) separation between the $(\mathsf{NP} \cap \mathsf{coNP})$-type and BPP-type communication complexity measures for a total function). An alternative proof of the lower bound for Tribes was given in [HJ13] using the smooth rectangle bound technique introduced by [JK10, CR12].

Analogously to $\textsc{Set-Inter}_m \coloneqq \textsc{Or}_m \circ \textsc{And}_2^m$, we have the definition

$$\textsc{Tribes}_{\ell,m} \;\coloneqq\; \textsc{And}_\ell \circ \textsc{Or}_m^\ell \circ \textsc{And}_2^{\ell \times m} \;=\; \textsc{And}_\ell \circ \textsc{Set-Inter}_m^\ell.$$

We always assume $m \geq 2$ (since if $m = 1$ then $\textsc{Tribes}_{\ell,m}$ is trivially computable with constant communication). Note that the outer function $\textsc{And}_\ell \circ \textsc{Or}_m^\ell$ takes a boolean $\ell \times m$ matrix and indicates whether every row has at least one 1. For $\textsc{Tribes}_{\ell,m}$, Alice and Bob each get such a matrix, and the above function is applied to the bitwise And of the two matrices.

Theorem 1. $\mathrm{R}_{1/2+\epsilon}(\textsc{Tribes}_{\ell,m}) = \Theta(\epsilon\ell m)$ *provided* $\epsilon\ell \geq 1$.

The upper bound is shown as follows. Let M denote the boolean $\ell \times m$ matrix that is fed into $\textsc{And}_\ell \circ \textsc{Or}_m^\ell$. Consider the protocol in which Alice and Bob publicly sample a uniformly random set of $4\epsilon\ell$ rows, evaluate all the bits of M in those rows (using $O(\epsilon\ell m)$ communication), and accept iff each of those rows of M contains at least one 1. For a 1-input, this rejects with probability 0, and for a 0-input it finds an all-0 row (and hence rejects) with probability at least 4ϵ. Now if we modify the above protocol so it rejects automatically with probability $1/2 - \epsilon$ and otherwise proceeds as before, then it rejects 1-inputs with probability $1/2 - \epsilon$ and 0-inputs with probability at least $(1/2 - \epsilon) + (1/2 + \epsilon) \cdot 4\epsilon \geq 1/2 + \epsilon$. The provision $\epsilon\ell \geq 1$ was stated cleanly to ensure that we can round $4\epsilon\ell$ up to an integer without affecting the asymptotic complexity. (If $\epsilon\ell \leq o(1)$ then just evaluating a single row of M takes $\omega(\epsilon\ell m)$ communication.) The lower bound, which we prove in Section 2, does not require this provision.

Our basic approach to prove the lower bound in Theorem 1 is to combine the information complexity techniques of [BM13] (developed for the ϵ-advantage lower bound for Set-Inter) with the

information complexity techniques of [JKS03] (developed for the constant-advantage lower bound for TRIBES). However, in trying to combine these techniques, there are a variety of technical hurdles, which require several new ideas to overcome. In Section 1.4 we discuss why other approaches fail to prove Theorem 1.

1.2 What if $\epsilon\ell \leq o(1)$?

As mentioned above, when $\epsilon\ell \leq o(1)$, our proof of the $O(\epsilon\ell m)$ upper bound for $\text{TRIBES}_{\ell,m}$ breaks down. So what upper bound can we give in this case? Let us restrict our attention to $\ell = 2$ (and let $\epsilon > 0$ be arbitrary).

First of all, notice that the communication protocol in Section 1.1 is actually a *query complexity* (a.k.a. *decision tree complexity*) upper bound for the outer function. A communication protocol for any composed function (with constant-size gadget) can simulate a decision tree for the outer function, using constant communication to evaluate the output of each gadget when queried by the decision tree. In the next paragraph, we describe an $O(\sqrt{\epsilon}m)$-query ϵ-advantage randomized decision tree for $\text{AND}_2 \circ \text{OR}_m^2$ (thus showing that $\text{R}_{1/2+\epsilon}(\text{TRIBES}_{2,m}) \leq O(\sqrt{\epsilon}m)$ provided $\sqrt{\epsilon}m \geq 1$).

Say the input is $z = (z_1, z_2) \in \{0,1\}^m \times \{0,1\}^m$. Consider the following randomized decision tree: Pick $S_1, S_2 \subseteq [m]$ both of size $2\sqrt{\epsilon}m$, independently uniformly at random, and accept iff $z_1|_{S_1}$ and $z_2|_{S_2}$ each contain at least one 1. For a 1-input, each of these two events happens with probability at least $2\sqrt{\epsilon}$, so they happen simultaneously with probability at least 4ϵ. For a 0-input, one of the two events never happens, and hence this accepts with probability 0. Now if we modify the above randomized decision tree so it accepts automatically with probability $1/2 - \epsilon$ and otherwise proceeds as before, then it accepts 0-inputs with probability $1/2 - \epsilon$ and 1-inputs with probability at least $(1/2 - \epsilon) + (1/2 + \epsilon) \cdot 4\epsilon \geq 1/2 + \epsilon$, and queries at most $O(\sqrt{\epsilon}m)$ bits.

We conjecture that this communication upper bound is tight, i.e., $\text{R}_{1/2+\epsilon}(\text{TRIBES}_{2,m}) \geq \Omega(\sqrt{\epsilon}m)$. This remains open, but we at least prove the query complexity version of this conjecture, which can be construed as evidence for the communication version. (The query complexity measure $\text{R}_p^{\text{dt}}(f)$ is defined in the natural way.)

Theorem 2. $\text{R}^{\text{dt}}_{1/2+\epsilon}(\text{AND}_2 \circ \text{OR}_m^2) = \Theta(\sqrt{\epsilon}m)$ *provided* $\sqrt{\epsilon}m \geq 1$.

We prove the lower bound of Theorem 2 in Section 3. There are some known powerful "simulation theorems" (e.g., [GLM+16, GPW17]) for converting query lower bounds for an outer function into matching communication lower bounds for a composed function; however, we lack a simulation theorem powerful enough to convert Theorem 2 into a communication lower bound. Furthermore, we have not found a way to emulate the query lower bound proof with information complexity tools to get a communication lower bound.

1.3 Which part contains the intersecting coordinate?

We now turn our attention away from TRIBES.

Suppose Alice and Bob are given uniquely intersecting subsets X and Y from a universe of size n that is partitioned into $\ell \geq 2$ equal-size parts, and they wish to identify which part contains the intersection. Of course, they can succeed with probability $1/\ell$ by random guessing without communicating about their sets. To do better they can use the following protocol.

Alice and Bob publicly sample a uniformly random subset S of size $2\epsilon n$ They exchange $X \cap S$ and $Y \cap S$ using $4\epsilon n$ bits of communication If $S \cap X \cap Y \neq \emptyset$ they output the label of the part containing the known point of intersection Otherwise they publicly sample and output a uniformly random part label

This protocol succeeds with probability $2\epsilon + (1-2\epsilon)/\ell = 1/\ell + (1-1/\ell)\cdot 2\epsilon \geq 1/\ell + \epsilon$. We prove that this is optimal: $\Omega(\epsilon n)$ communication is necessary to achieve advantage ϵ.

We state this using the following notation. Define the partial function $\textsc{Which}_\ell \colon \{0,1\}^\ell \to [\ell]$ that takes a string of Hamming weight 1 and outputs the coordinate of the only 1. Define the "unambiguous-or" function $\textsc{Unambig-Or}_m$ as $\textsc{Or}_m$ restricted to the domain of strings of Hamming weight 0 or 1. Define the "unambiguous-set-intersection" function[1] $\textsc{Unambig-Inter}_m := \textsc{Unambig-Or}_m \circ \textsc{And}_2^m$.

Theorem 3. $\mathrm{R}_{1/\ell+\epsilon}(\textsc{Which}_\ell \circ \textsc{Unambig-Inter}_m^\ell) = \Theta(\epsilon\ell m)$ *provided* $\epsilon\ell m \geq 1$.

We prove the lower bound in Section 4, where we also describe some ways to reinterpret Theorem 3. This problem is motivated partly as a simple variant of set-intersection whose communication complexity was not fully understood before, and partly because of the corollaries presented in Section 4, which show the maximum possible gap between randomized small-advantage complexity and so-called SV-nondeterminism, and between sampling from distributions with/without the ability to condition on an event.

The key to the proof is in relating the complexity of $\textsc{Which}_\ell \circ F^\ell$ to the complexity of F (for an arbitrary two-party F with boolean output). It is natural to conjecture that the complexity goes up by roughly a factor of ℓ after composition with $\textsc{Which}_\ell$; this is an alternative form of direct sum problem. In the standard direct sum setting, the goal is to evaluate F on each of ℓ independent inputs; our form is equivalent but under the promise that one of the inputs evaluates to 1 and the rest to 0. Thus proving the direct sum conjecture (factor ℓ increase in complexity) appears qualitatively harder in our setting than in the standard setting. We show an information complexity version of the conjecture, and we combine this with [BM13] to derive Theorem 3.

For worst-case communication, we at least show that the complexity does not go down after composition with $\textsc{Which}_\ell$. In particular, this yields a simple proof of a communication lower bound due to [Kla03] which implies the communication complexity class separation $\mathsf{UP} \cap \mathsf{coUP} \not\subseteq \mathsf{BPP}$. The proof in [Kla03] is technically somewhat involved, exploiting a "fine-tuned" version of Razborov's corruption lemma [Raz92]; our simple proof of the same lower bound is by a black-box reduction to the standard (constant-advantage) lower bound for $\textsc{Unambig-Inter}$.

1.4 Related work

We now describe why the $\Omega(\epsilon\ell m)$ lower bound in Theorem 1 does not follow straightforwardly from known results. First of all, applying standard majority-amplification to the known $\Omega(\ell m)$ lower bound for constant advantage only yields an $\Omega(\epsilon^2\ell m)$ lower bound. What about the technique used by [GW16] to give a simplified proof of the tight ϵ-advantage lower bound for $\textsc{Set-Inter}$? Let us summarize this technique (known as "and-amplification") as applied to the complement function $\textsc{Set-Disj}$: Running an ϵ-advantage protocol $O(1/\epsilon)$ times, and accepting iff all runs

[1] Sometimes this is called "unique-set-intersection", but our terminology is more consistent with classical complexity; see [GPW18].

accept, yields a so-called SBP-type protocol, for which the complexity is characterized by the corruption bound. Hence the ϵ-advantage complexity is always at least $\Omega(\epsilon)$ times the corruption bound (which is $\Omega(n)$ for Set-Disj$_n$ by [Raz92]). Applied to Tribes$_{\ell,m}$ (or its complement), the and-amplification technique can only yield an essentially $\Omega(\epsilon \cdot \max(\ell, m))$ lower bound, since Tribes$_{\ell,m}$ has an $O(\ell \log m)$-communication nondeterministic (in particular, SBP-type) protocol and an $O(m + \log \ell)$-communication conondeterministic (in particular, coSBP-type) protocol.

Can we leverage the known smooth rectangle lower bound for Tribes$_{\sqrt{n},\sqrt{n}}$ [HJ13]? The smooth rectangle bound in general characterizes the complexity of so-called WAPP-type protocols [JK10, GLM+16]. Thus if we could "amplify" an ϵ-advantage protocol into a (sufficiently-large-constant-advantage) WAPP-type protocol with $o(1/\epsilon^2)$ factor overhead, we would get a nontrivial ϵ-advantage lower bound for Tribes$_{\sqrt{n},\sqrt{n}}$. However, the smooth rectangle lower bound for Gap-Hamming [CR12] shows that this cannot always be done, i.e., an $\Omega(1/\epsilon^2)$ overhead is sometimes necessary (at least for general partial functions). In summary, the known "rectangle-based" lower bound techniques fail to yield Theorem 1, so we use an information complexity approach instead.

There is some other work related to the Tribes lower bound. The paper [RS10] proved that if the definition of R is changed to allow only private coins (no public coins) then $\mathrm{R}_{1/2+\epsilon}(\text{Tribes}_{\ell,\ell^2}) \geq \Omega(\ell)$ for all $\epsilon > 0$ (no matter how small). The original constant-advantage lower bound for Tribes [JKS03] spawned a line of research on the communication complexity of read-once formulas [JKR09, LS10, JKZ10, GJ16]. A multi-party version of Tribes has also been studied in the message-passing model [CM15].

Regarding the problem of finding which part contains the unique intersecting coordinate, we mention that there is some prior work studying a peripherally related topic: the randomized complexity of "finding the exact intersection" [BGPW13, BCK+14, BCK+16], albeit not restricting the size of the intersection.

One of our corollaries of Theorem 3 concerns the communication complexity of problems where the goal is to sample from a distribution. Not much is known about this topic, and the existing works focus on problems where there is no input and the goal is to sample random input-output pairs of a function [ASTS+03, JSWZ13, Wat16]. The complexity of "small-advantage sampling" was studied in the context of time-bounded computation in [Wat14].

1.5 Preliminaries

We first note that it suffices to prove our lower bounds for $\text{And}_\ell \circ \text{Or}_m^\ell \circ \text{And}_2^{\ell \times m}$ (Theorem 1) and $\text{Which}_\ell \circ \text{Unambig-Or}_m^\ell \circ \text{And}_2^{\ell \times m}$ (Theorem 3) with And_2 replaced by a different two-party gadget, namely the equality function on trits $\text{3Eq}\colon \{0,1,2\} \times \{0,1,2\} \to \{0,1\}$ ($\text{3Eq}(X,Y) = 1$ iff $X = Y$). This is because 3Eq reduces to $\text{Unambig-Or}_3 \circ \text{And}_2^3$ (with Alice and Bob both mapping their trit to its characteristic bit vector of Hamming weight 1), and thus $\text{Unambig-Or}_m \circ \text{3Eq}^m$ reduces to $\text{Unambig-Or}_{3m} \circ \text{And}_2^{3m}$, and $\text{Or}_m \circ \text{3Eq}^m$ reduces to $\text{Or}_{3m} \circ \text{And}_2^{3m}$.

We now mention some notational conventions. 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 (total variation) distance. We use bold letters to denote random variables, and non-bold letters for particular outcomes. We use $\in_{\mathrm{u}}$ to denote that a random variable is distributed uniformly over some set.

Fact 1. *Mutual information and relative entropy satisfy the following standard properties [CT06]:*

(i) 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.*

(ii) Alternative definition: $\mathbb{I}(\boldsymbol{A}\,;\,\boldsymbol{B}) = \mathbb{E}_{A\sim\boldsymbol{A}}\mathbb{D}\big((\boldsymbol{B}\,|\,\boldsymbol{A}=A)\,\big\|\,\boldsymbol{B}\big)$.
(iii) Pinsker's inequality: $\mathbb{D}(\boldsymbol{A}\,\|\,\boldsymbol{B}) \geq 2\Delta(\boldsymbol{A},\boldsymbol{B})^2$.

All protocols Π are randomized and have both public and private coins, unless otherwise stated, and we use $CC(\Pi)$ to denote the worst-case communication cost. When we speak of an arbitrary F, by default it is assumed to be a two-party partial function. Also, complexity class names (such as BPP) refer to classes of (families of) two-party partial functions with polylogarithmic communication protocols of the relevant type.

2 Communication Lower Bound for Tribes

The upper bound for Theorem 1 was shown in Section 1.1. In this section we give the proof of the lower bound, which is broken into four steps corresponding to the four subsections. In Section 2.1 we use known techniques [BYJKS04, JKS03, BM13] to show that it suffices to prove a certain information complexity lower bound for a constant-size function; there are no substantially new ideas in this step. Then in Section 2.2 we further reduce to a problem involving just nine inputs at a time. In Section 2.3 and Section 2.4 we finish the proof of the lower bound by showing how to tightly relate probabilities (coming from the protocol's correctness) to the corresponding contributions to the protocol's information cost.

2.1 Step 1: Conditioning and direct sum

As noted in Section 1.5, it suffices to prove the lower bound for $\textsc{Tribes}'_{\ell,m} \coloneqq \textsc{And}_\ell \circ \textsc{Or}_m^\ell \circ 3\textsc{Eq}^{\ell\times m}$ instead of $\textsc{Tribes}_{\ell,m}$. Suppose for contradiction there is a $(1/2+\epsilon)$-correct protocol Π for $\textsc{Tribes}'_{\ell,m}$ with $CC(\Pi) \leq o(\epsilon\ell m)$. As a technicality, we assume Π has been converted into a private-coin-only protocol, where Alice first privately samples the public coins (if any) and sends them to Bob. (This could blow up the communication, but we will only use the fact that the "original communication" part of the transcript has bounded length, not the "public coins" part.)

We can think of the input to $\textsc{Tribes}'_{\ell,m}$ as an $\ell \times m$ table where each cell has two trits, one for Alice and one for Bob. As is standard in information complexity lower bounds, we define a distribution over inputs, equipped with a "conditioning scheme" that decomposes the distribution into a mixture of product distributions (where Alice's and Bob's parts of the input are independent of each other). We do this by placing a uniformly random 1-input to 3Eq at a uniformly random cell in each row, and for each of the remaining cells choosing at random a rectangular "window" of 0-inputs to 3Eq, from which the input to that cell is drawn.

Formally, let us define $\mathscr{W}_1 \coloneqq \big\{\{00\},\{11\},\{22\}\big\}$ as the set of "1-windows" of 3Eq, and define $\mathscr{W}_0 \coloneqq \big\{\{01,02\},\{10,12\},\{20,21\},\{10,20\},\{01,21\},\{02,12\}\big\}$ as the set of "0-windows" of 3Eq. We define a probability space with the following random variables: $\boldsymbol{X} \in \{0,1,2\}^{\ell\times m}$, $\boldsymbol{Y} \in \{0,1,2\}^{\ell\times m}$, $\boldsymbol{\tau} \in \{0,1\}^*$, $\boldsymbol{J} \in [m]^\ell$, and $\boldsymbol{W} \in (2^{\{0,1,2\}\times\{0,1,2\}})^{\ell\times m}$. Choose $\boldsymbol{J}$ uniformly, and for each $(i,j) \in [\ell]\times[m]$ independently, let

$$\boldsymbol{W}_{i,j} \;\in_{\mathrm{u}}\; \begin{cases} \mathscr{W}_1 & \text{if } j = \boldsymbol{J}_i \\ \mathscr{W}_0 & \text{if } j \neq \boldsymbol{J}_i \end{cases}$$

and let $(\boldsymbol{X}_{i,j}\boldsymbol{Y}_{i,j}) \in_{\mathrm{u}} \boldsymbol{W}_{i,j}$. Note that $\boldsymbol{XY}$ is supported on 1-inputs of $\textsc{Tribes}'_{\ell,m}$, and that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent conditioned on $\boldsymbol{W}$. Finally, let $\boldsymbol{\tau}$ be the random transcript on input $(\boldsymbol{X},\boldsymbol{Y})$.

Define $\boldsymbol{X}_{-\boldsymbol{J}} \coloneqq (\boldsymbol{X}_{i,j})_{j \neq \boldsymbol{J}_i}$ (and $\boldsymbol{Y}_{-\boldsymbol{J}}$ similarly), and let $\boldsymbol{\tau}^{\mathrm{C}}$ denote the "original communication" part of $\boldsymbol{\tau}$, and $\boldsymbol{\tau}^{\mathrm{R}}$ denote the "public coins" part of $\boldsymbol{\tau}$. We have

$$\mathbb{I}\big(\boldsymbol{\tau}\,;\,\boldsymbol{X}_{-\boldsymbol{J}}\boldsymbol{Y}_{-\boldsymbol{J}} \,\big|\, \boldsymbol{W}\big) \;=\; \mathbb{I}\big(\boldsymbol{\tau}^{\mathrm{C}}\,;\,\boldsymbol{X}_{-\boldsymbol{J}}\boldsymbol{Y}_{-\boldsymbol{J}} \,\big|\, \boldsymbol{W}\boldsymbol{\tau}^{\mathrm{R}}\big) \;\leq\; \mathbb{H}\big(\boldsymbol{\tau}^{\mathrm{C}} \,\big|\, \boldsymbol{W}\boldsymbol{\tau}^{\mathrm{R}}\big) \;\leq\; CC(\Pi) \;\leq\; o(\epsilon\ell m)$$

where the equality holds by the chain rule and independence of $\boldsymbol{\tau}^{\mathrm{R}}$ and $\boldsymbol{W}\boldsymbol{X}\boldsymbol{Y}$. If we augment the probability space with random variables $(\boldsymbol{i}, \boldsymbol{k})$ sampled uniformly from $([\ell]\times[m]) \setminus \{(i, \boldsymbol{J}_i) \,:\, i \in [\ell]\}$ (independent of the other random variables, conditioned on $\boldsymbol{J}$), then by Fact 1.*(i)* we have

$$\mathbb{I}\big(\boldsymbol{\tau}\,;\,\boldsymbol{X}_{\boldsymbol{i},\boldsymbol{k}}\boldsymbol{Y}_{\boldsymbol{i},\boldsymbol{k}} \,\big|\, \boldsymbol{W}\boldsymbol{i}\boldsymbol{k}\big) \;\leq\; \tfrac{1}{\ell(m-1)} \cdot \mathbb{I}\big(\boldsymbol{\tau}\,;\,\boldsymbol{X}_{-\boldsymbol{J}}\boldsymbol{Y}_{-\boldsymbol{J}} \,\big|\, \boldsymbol{W}\big) \;\leq\; o(\epsilon)$$

(using $(\boldsymbol{\tau} \,|\, \boldsymbol{W} = W)$ as $\boldsymbol{A}$ and $(\boldsymbol{X}_{i,k}\boldsymbol{Y}_{i,k} \,|\, \boldsymbol{W} = W)$ as a component of $\boldsymbol{B}$, for any particular W). For convenience let $\boldsymbol{j} \coloneqq \boldsymbol{J}_{\boldsymbol{i}}$, let $\boldsymbol{h} \coloneqq \{\boldsymbol{j}, \boldsymbol{k}\}$, let $\boldsymbol{W}_{\boldsymbol{i},\boldsymbol{h}}$ be the restriction of $\boldsymbol{W}$ to the 2 coordinates in $\{\boldsymbol{i}\} \times \boldsymbol{h}$, and let $\boldsymbol{W}_{-\boldsymbol{i},\boldsymbol{h}}$ be the restriction of $\boldsymbol{W}$ to the remaining $\ell \times m - 2$ coordinates. There must exist outcomes i^*, h^*, $W^*_{-i^*,h^*}$ such that

$$\mathbb{I}\big(\boldsymbol{\tau}\,;\,\boldsymbol{X}_{\boldsymbol{i},\boldsymbol{k}}\boldsymbol{Y}_{\boldsymbol{i},\boldsymbol{k}} \,\big|\, \boldsymbol{W}_{\boldsymbol{i},\boldsymbol{h}}\boldsymbol{k},\; \boldsymbol{i} = i^*,\; \boldsymbol{h} = h^*,\; \boldsymbol{W}_{-\boldsymbol{i},\boldsymbol{h}} = W^*_{-i^*,h^*}\big) \;\leq\; o(\epsilon). \tag{1}$$

Note that given this $i^*, h^*, W^*_{-i^*,h^*}$, the remaining conditioning variables $\boldsymbol{W}_{\boldsymbol{i},\boldsymbol{h}}\boldsymbol{k}$ have 36 possible outcomes: 2 choices for $\boldsymbol{k}$ (it could be either element of h^*, and $\boldsymbol{j}$ is the other), 3 choices for $\boldsymbol{W}_{\boldsymbol{i},\boldsymbol{j}}$, and 6 choices for $\boldsymbol{W}_{\boldsymbol{i},\boldsymbol{k}}$.

We rephrase the situation by considering a protocol Π^* that interprets its input as X_{i^*,h^*}, Y_{i^*,h^*}, uses private coins to sample $X_{-i^*,h^*}, Y_{-i^*,h^*}$ uniformly from $W^*_{-i^*,h^*}$, then runs the private-coin protocol Π on the combined input X, Y. Observe that Π^* is a $(1/2+\epsilon)$-correct protocol for $\textsc{Or}_2 \circ \textsc{3Eq}^2$ since with probability 1, $(\textsc{Or}_2 \circ \textsc{3Eq}^2)(X_{i^*,h^*}, Y_{i^*,h^*}) = \textsc{Tribes}'_{\ell,m}(X, Y)$ (as the evaluation of the 3Eq functions on $X_{-i^*,h^*}, Y_{-i^*,h^*}$ is guaranteed to have a 1 in each of the non-i^* rows, and 0's in the non-h^* columns of the i^* row). Here, we now think of the two coordinates in $\{i^*\} \times h^*$ as being labeled 1 and 2.

For convenience, we henceforth recycle notation by letting Π denote the new protocol Π^* and letting $(\boldsymbol{j}, \boldsymbol{k}) \in_{\mathrm{u}} \{(1,2),(2,1)\}$, $\boldsymbol{W}_{\boldsymbol{j}} \in_{\mathrm{u}} \mathscr{W}_1$, $\boldsymbol{W}_{\boldsymbol{k}} \in_{\mathrm{u}} \mathscr{W}_0$, $(\boldsymbol{X}_1\boldsymbol{Y}_1) \in_{\mathrm{u}} \boldsymbol{W}_1$, $(\boldsymbol{X}_2\boldsymbol{Y}_2) \in_{\mathrm{u}} \boldsymbol{W}_2$. With respect to this recycled notation, the inequality (1) becomes

$$\mathbb{I}\big(\boldsymbol{\tau}\,;\,\boldsymbol{X}_{\boldsymbol{k}}\boldsymbol{Y}_{\boldsymbol{k}} \,\big|\, \boldsymbol{W}\boldsymbol{k}\big) \;\leq\; o(\epsilon). \tag{2}$$

The following lemma, whose proof occupies the remaining three subsections, provides the contradiction, completing the proof of Theorem 1.

Lemma 1. *If (2) holds then Π is not a $(1/2+\epsilon)$-correct protocol for $\textsc{Or}_2 \circ \textsc{3Eq}^2$.*

2.2 Step 2: Uniformly covering a pair of gadgets

Let us set up some notation (all in reference to the private-coin protocol Π). If X is an Alice input and Y is a Bob input, let $\pi_{\mathrm{X},\mathrm{Y}}$ denote the probability Π accepts on input (X, Y). For a 1×2 rectangle of inputs $\{\mathrm{U}\} \times \{\mathrm{V}, \mathrm{W}\}$ let $\iota_{\mathrm{U},\mathrm{VW}}$ denote the mutual information between the random transcript of Π and a uniformly random input from $\{(\mathrm{U}, \mathrm{V}), (\mathrm{U}, \mathrm{W})\}$. Similarly, for a 2×1 rectangle of inputs $\{\mathrm{V}, \mathrm{W}\} \times \{\mathrm{U}\}$ let $\iota_{\mathrm{VW},\mathrm{U}}$ denote the mutual information between the random transcript of Π and a uniformly random input from $\{(\mathrm{V}, \mathrm{U}), (\mathrm{W}, \mathrm{U})\}$. We write $\mathrm{U} = \mathrm{U}_1\mathrm{U}_2 \in \{0,1,2\}^2$ and similarly for V and W.

Since in the inequality (2) there are only a constant number of possible outcomes for $\boldsymbol{Wk}$, the $o(\epsilon)$ bound holds conditioned on each of those outcomes. Thus, (2) can be further rephrased as

$$\begin{aligned}\iota_{\mathrm{U,VW}} \le o(\epsilon) \text{ and } \iota_{\mathrm{VW,U}} \le o(\epsilon) \text{ if } \mathrm{U}_1, \mathrm{V}_1, \mathrm{W}_1 \text{ are all equal and } \mathrm{U}_2, \mathrm{V}_2, \mathrm{W}_2 \text{ are all distinct,}\\ \text{or } \mathrm{U}_2, \mathrm{V}_2, \mathrm{W}_2 \text{ are all equal and } \mathrm{U}_1, \mathrm{V}_1, \mathrm{W}_1 \text{ are all distinct.}\end{aligned} \tag{3}$$

The following lemma (illustrated in Figure 1) is proved in the remaining two subsections.

Lemma 2. *For any Alice inputs* $\mathrm{A}, \mathrm{B}, \mathrm{C}$ *and Bob inputs* $\mathrm{D}, \mathrm{E}, \mathrm{F}$*, we have*

$$\pi_{\mathrm{A,D}} - \pi_{\mathrm{A,F}} - \pi_{\mathrm{C,D}} + \pi_{\mathrm{C,F}} \;\le\; 128\big(\iota_{\mathrm{A,DE}} + \iota_{\mathrm{AB,D}} + \iota_{\mathrm{C,FE}} + \iota_{\mathrm{CB,F}}\big).$$

We now show how to use Lemma 2 to prove Lemma 1.

Proof of Lemma 1. First we define a map from $\{0,1,2\}^2 \times \{\pm 1\}^2$ to $(\{0,1,2\}^2)^6$ that takes "data" consisting of $t_1, t_2 \in \{0,1,2\}$ and $\delta_1, \delta_2 \in \{\pm 1\}$ and maps it to a tuple of Alice inputs $\mathrm{A}, \mathrm{B}, \mathrm{C}$ and Bob inputs $\mathrm{D}, \mathrm{E}, \mathrm{F}$ defined by

$$\mathrm{A} \coloneqq t_1, (t_2 + \delta_2) \quad \mathrm{B} \coloneqq t_1, t_2 \quad \mathrm{C} \coloneqq (t_1 + \delta_1), t_2 \quad \mathrm{D} \coloneqq t_1, (t_2 - \delta_2) \quad \mathrm{E} \coloneqq t_1, t_2 \quad \mathrm{F} \coloneqq (t_1 - \delta_1), t_2$$

(where the addition is mod 3). For any choice of the data, we have $(\mathrm{B}, \mathrm{E}) \in (3\mathrm{EQ}^2)^{-1}(11)$ (hence the dark gray shading in Figure 1), $(\mathrm{A}, \mathrm{D}), (\mathrm{B}, \mathrm{D}), (\mathrm{A}, \mathrm{E}) \in (3\mathrm{EQ}^2)^{-1}(10)$ and $(\mathrm{C}, \mathrm{F}), (\mathrm{C}, \mathrm{E}), (\mathrm{B}, \mathrm{F}) \in (3\mathrm{EQ}^2)^{-1}(01)$ (hence the light gray shading), and $(\mathrm{A}, \mathrm{F}), (\mathrm{C}, \mathrm{D}) \in (3\mathrm{EQ}^2)^{-1}(00)$.

Note that there are 36 possible choices of the data, and that $\big|(3\mathrm{EQ}^2)^{-1}(10)\big| = \big|(3\mathrm{EQ}^2)^{-1}(01)\big| = 18$ and $\big|(3\mathrm{EQ}^2)^{-1}(00)\big| = 36$. It is straightforward to verify the following key properties of our map.

- The A, D coordinates form a 2-to-1 map onto $(3\mathrm{EQ}^2)^{-1}(10)$ (since δ_1 is irrelevant).
- The C, F coordinates form a 2-to-1 map onto $(3\mathrm{EQ}^2)^{-1}(01)$ (since δ_2 is irrelevant).
- The A, F coordinates form a 1-to-1 map onto $(3\mathrm{EQ}^2)^{-1}(00)$.
- The C, D coordinates form a 1-to-1 map onto $(3\mathrm{EQ}^2)^{-1}(00)$.
- The quantities $\iota_{\mathrm{A,DE}},\ \iota_{\mathrm{AB,D}},\ \iota_{\mathrm{C,FE}},\ \iota_{\mathrm{CB,F}}$ are always $\le o(\epsilon)$ by (3).

Now we have (letting the dependence of $\mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{D}, \mathrm{E}, \mathrm{F}$ on $t_1, t_2, \delta_1, \delta_2$ be implicit)

$$\begin{aligned}&\textstyle\sum_{(\mathrm{X,Y}) \in (3\mathrm{EQ}^2)^{-1}(10) \cup (3\mathrm{EQ}^2)^{-1}(01)} \pi_{\mathrm{X,Y}} - \sum_{(\mathrm{X,Y}) \in (3\mathrm{EQ}^2)^{-1}(00)} \pi_{\mathrm{X,Y}}\\ =\;& \tfrac{1}{2} \textstyle\sum_{t_1,t_2,\delta_1,\delta_2} \big(\pi_{\mathrm{A,D}} - \pi_{\mathrm{A,F}} - \pi_{\mathrm{C,D}} + \pi_{\mathrm{C,F}}\big)\\ \le\;& \tfrac{1}{2} \textstyle\sum_{t_1,t_2,\delta_1,\delta_2} 128\big(\iota_{\mathrm{A,DE}} + \iota_{\mathrm{AB,D}} + \iota_{\mathrm{C,FE}} + \iota_{\mathrm{CB,F}}\big)\\ \le\;& \tfrac{1}{2} \cdot 36 \cdot 128 \cdot 4 \cdot o(\epsilon)\\ =\;& o(\epsilon)\end{aligned}$$

where the second line is by the first four key properties of our map, the third line is by Lemma 2, and the fourth line is by the last key property. Hence Π cannot be $(1/2+\epsilon)$-correct for $\mathrm{OR}_2 \circ 3\mathrm{EQ}^2$ since otherwise the first line would be at least $36 \cdot (1/2 + \epsilon) - 36 \cdot (1/2 - \epsilon) = 72\epsilon$. ∎

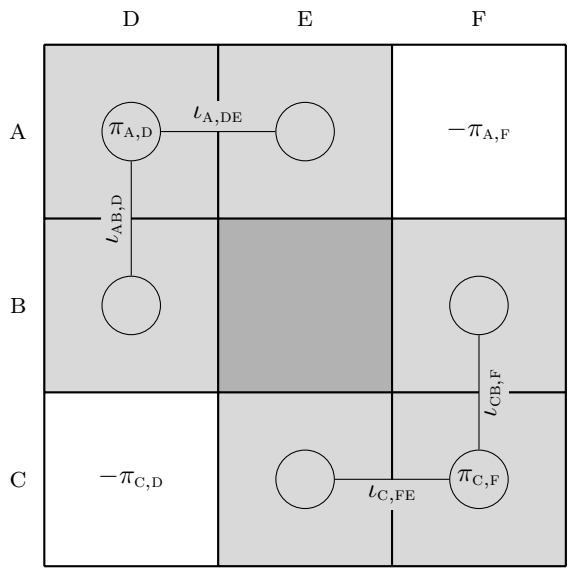

Figure 1: Illustration for Lemma 2.

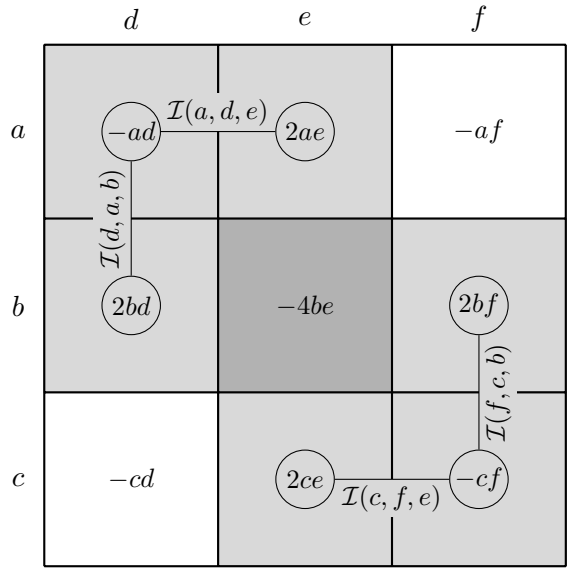

Figure 2: Illustration for Lemma 5.

2.3 Step 3: Relating information and probabilities for inputs

We first set up some notation. For numbers $u, v, w \in [0,1]$, define $\mathcal{I}(u,v,w) \coloneqq u(v-w)^2/(v+w)$ (with the convention that $0/0 = 0$). For an input (X, Y) and a transcript τ, we let the numbers $\tau_{\mathrm{X}}, \tau_{\mathrm{Y}} \in [0,1]$ be such that $\mathbb{P}[\Pi(\mathrm{X},\mathrm{Y}) \text{ has transcript } \tau] = \tau_{\mathrm{X}} \cdot \tau_{\mathrm{Y}}$ (where τ_{X} does not depend on Y, and τ_{Y} does not depend on X). Note that $\pi_{\mathrm{X},\mathrm{Y}} = \sum_{\text{accepting } \tau} \tau_{\mathrm{X}} \cdot \tau_{\mathrm{Y}}$.

The following fact was also used in [BM13]; we provide a proof for completeness.

Lemma 3. *For any rectangle $\{\mathrm{U}\} \times \{\mathrm{V},\mathrm{W}\}$ we have $\iota_{\mathrm{U},\mathrm{VW}} \geq \frac{1}{4}\sum_\tau \mathcal{I}(\tau_{\mathrm{U}}, \tau_{\mathrm{V}}, \tau_{\mathrm{W}})$. Symmetrically, for any rectangle $\{\mathrm{V},\mathrm{W}\} \times \{\mathrm{U}\}$ we have $\iota_{\mathrm{VW},\mathrm{U}} \geq \frac{1}{4}\sum_\tau \mathcal{I}(\tau_{\mathrm{U}}, \tau_{\mathrm{V}}, \tau_{\mathrm{W}})$.*

Proof. Assume the random variable $\boldsymbol{Y} \in_{\mathrm{u}} \{\mathrm{V},\mathrm{W}\}$ is jointly distributed with $\boldsymbol{\tau}$ (the random variable representing the transcript). Note that $\mathbb{P}[\boldsymbol{\tau} = \tau] = \frac{1}{2}\tau_{\mathrm{U}}(\tau_{\mathrm{V}} + \tau_{\mathrm{W}})$ and that $\Delta\big((\boldsymbol{Y} \mid \boldsymbol{\tau} = \tau), \boldsymbol{Y}\big) = \frac{1}{2} - \min(\tau_{\mathrm{V}}, \tau_{\mathrm{W}})/(\tau_{\mathrm{V}} + \tau_{\mathrm{W}}) = \frac{1}{2}|\tau_{\mathrm{V}} - \tau_{\mathrm{W}}|/(\tau_{\mathrm{V}} + \tau_{\mathrm{W}})$. Then we have

$$\begin{aligned}
\iota_{\mathrm{U},\mathrm{VW}} &\coloneqq \mathbb{I}(\boldsymbol{\tau}\,;\,\boldsymbol{Y}) \\
&= \mathbb{E}_{\tau\sim\boldsymbol{\tau}}\mathbb{D}\big((\boldsymbol{Y} \mid \boldsymbol{\tau} = \tau) \,\big\|\, \boldsymbol{Y}\big) \\
&\geq \textstyle\sum_\tau \mathbb{P}[\boldsymbol{\tau} = \tau] \cdot 2\Delta\big((\boldsymbol{Y} \mid \boldsymbol{\tau} = \tau), \boldsymbol{Y}\big)^2 \\
&= \textstyle\sum_\tau \big(\frac{1}{2}\tau_{\mathrm{U}}(\tau_{\mathrm{V}} + \tau_{\mathrm{W}})\big) \cdot 2\big(\frac{1}{2}(\tau_{\mathrm{V}} - \tau_{\mathrm{W}})/(\tau_{\mathrm{V}} + \tau_{\mathrm{W}})\big)^2 \\
&= \textstyle\frac{1}{4}\sum_\tau \tau_{\mathrm{U}}(\tau_{\mathrm{V}} - \tau_{\mathrm{W}})^2/(\tau_{\mathrm{V}} + \tau_{\mathrm{W}})
\end{aligned}$$

where the second line is by Fact 1.*(ii)*, and the third line is by Fact 1.*(iii)*. ∎

Intuitively, Lemma 3 means $\mathcal{I}(\tau_{\mathrm{U}}, \tau_{\mathrm{V}}, \tau_{\mathrm{W}})$ lower bounds the "contribution" of τ to the information cost. Now that we have related the information costs to the contributions, we need to relate the contributions to the probabilities of observing individual transcripts. The following two lemmas allow us to do this.

Lemma 4. *For any four numbers $q, r, s, t \in [0,1]$, we have*

$$-qs + qt + rs - rt \;\le\; 2\big(\mathcal{I}(q,s,t) + \mathcal{I}(s,q,r)\big).$$

Lemma 5. *For any six numbers $a, b, c, d, e, f \in [0,1]$, we have*

$$-ad + 2ae - af + 2bd - 4be + 2bf - cd + 2ce - cf \;\le\; 32\big(\mathcal{I}(a,d,e) + \mathcal{I}(d,a,b) + \mathcal{I}(c,f,e) + \mathcal{I}(f,c,b)\big).$$

Lemma 4 is from [BM13]. Lemma 5 (illustrated in Figure 2) is more involved and constitutes one of the key technical novelties in our proof of Theorem 1. For example, one insight is in finding the proper list of coefficients on the left side of the inequality in Lemma 5, to simultaneously make the lemma true and enable it to be used in our proof approach for Lemma 2.

The proof of Lemma 4 in [BM13] proceeds by clearing denominators and then decomposing the difference between the right and left sides into a sum of parts, such that the (weighted) AM–GM inequality implies each part is nonnegative. A priori, it is conceivable the same approach could work for Lemma 5; however, the problem of finding an appropriate decomposition can be expressed as a linear program feasibility question, and with the help of an LP solver we found that this approach actually does not work for Lemma 5 (even with 32 replaced by other constants). To get around this, we begin by giving a significantly different proof of Lemma 4,[2] which we *are* able to generalize to prove Lemma 5. We provide our proofs of both lemmas in the remaining subsection, where we also give some intuition.

For now we complete the proof of Lemma 2. Here we employ another key idea (beyond the proof structure of [BM13]): The corresponding part of the argument in [BM13] finishes by simply summing Lemma 4 over accepting transcripts, but this approach does not work in our context. We also need to take into account the rejecting transcripts and the fact that the acceptance and rejection probabilities sum to 1, in order to orchestrate all the necessary cancellations.

Proof of Lemma 2. We have

$$\begin{aligned}
&-\pi_{\mathrm{A,D}} + 2\pi_{\mathrm{A,E}} - \pi_{\mathrm{A,F}} + 2\pi_{\mathrm{B,D}} - 4\pi_{\mathrm{B,E}} + 2\pi_{\mathrm{B,F}} - \pi_{\mathrm{C,D}} + 2\pi_{\mathrm{C,E}} - \pi_{\mathrm{C,F}} \\
&= \sum_{\text{accepting } \tau} \big(-\tau_{\mathrm{A}}\tau_{\mathrm{D}} + 2\tau_{\mathrm{A}}\tau_{\mathrm{E}} - \tau_{\mathrm{A}}\tau_{\mathrm{F}} + 2\tau_{\mathrm{B}}\tau_{\mathrm{D}} - 4\tau_{\mathrm{B}}\tau_{\mathrm{E}} + 2\tau_{\mathrm{B}}\tau_{\mathrm{F}} - \tau_{\mathrm{C}}\tau_{\mathrm{D}} + 2\tau_{\mathrm{C}}\tau_{\mathrm{E}} - \tau_{\mathrm{C}}\tau_{\mathrm{F}}\big) \\
&\le 32 \sum_{\text{accepting } \tau} \big(\mathcal{I}(\tau_{\mathrm{A}}, \tau_{\mathrm{D}}, \tau_{\mathrm{E}}) + \mathcal{I}(\tau_{\mathrm{D}}, \tau_{\mathrm{A}}, \tau_{\mathrm{B}}) + \mathcal{I}(\tau_{\mathrm{C}}, \tau_{\mathrm{F}}, \tau_{\mathrm{E}}) + \mathcal{I}(\tau_{\mathrm{F}}, \tau_{\mathrm{C}}, \tau_{\mathrm{B}})\big). \qquad (4)
\end{aligned}$$

by Lemma 5 with $(a,b,c,d,e,f) = (\tau_{\mathrm{A}}, \tau_{\mathrm{B}}, \tau_{\mathrm{C}}, \tau_{\mathrm{D}}, \tau_{\mathrm{E}}, \tau_{\mathrm{F}})$. We also have

$$\begin{aligned}
2\big(\pi_{\mathrm{A,D}} - \pi_{\mathrm{A,E}} - \pi_{\mathrm{B,D}} + \pi_{\mathrm{B,E}}\big) &= 2\big(-(1-\pi_{\mathrm{A,D}}) + (1-\pi_{\mathrm{A,E}}) + (1-\pi_{\mathrm{B,D}}) - (1-\pi_{\mathrm{B,E}})\big) \\
&= 2\sum_{\text{rejecting } \tau} \big(-\tau_{\mathrm{A}}\tau_{\mathrm{D}} + \tau_{\mathrm{A}}\tau_{\mathrm{E}} + \tau_{\mathrm{B}}\tau_{\mathrm{D}} - \tau_{\mathrm{B}}\tau_{\mathrm{E}}\big) \\
&\le 4\sum_{\text{rejecting } \tau} \big(\mathcal{I}(\tau_{\mathrm{A}}, \tau_{\mathrm{D}}, \tau_{\mathrm{E}}) + \mathcal{I}(\tau_{\mathrm{D}}, \tau_{\mathrm{A}}, \tau_{\mathrm{B}})\big) \qquad (5)
\end{aligned}$$

by Lemma 4 with $(q,r,s,t) = (\tau_{\mathrm{A}}, \tau_{\mathrm{B}}, \tau_{\mathrm{D}}, \tau_{\mathrm{E}})$. Similarly,

$$2\big(\pi_{\mathrm{C,F}} - \pi_{\mathrm{C,E}} - \pi_{\mathrm{B,F}} + \pi_{\mathrm{B,E}}\big) \;\le\; 4\sum_{\text{rejecting } \tau} \big(\mathcal{I}(\tau_{\mathrm{C}}, \tau_{\mathrm{F}}, \tau_{\mathrm{E}}) + \mathcal{I}(\tau_{\mathrm{F}}, \tau_{\mathrm{C}}, \tau_{\mathrm{B}})\big) \qquad (6)$$

by Lemma 4 with $(q,r,s,t) = (\tau_{\mathrm{C}}, \tau_{\mathrm{B}}, \tau_{\mathrm{F}}, \tau_{\mathrm{E}})$. Summing the inequalities (4), (5), (6) yields

$$\begin{aligned}
\pi_{\mathrm{A,D}} - \pi_{\mathrm{A,F}} - \pi_{\mathrm{C,D}} + \pi_{\mathrm{C,F}} &\le 32\sum_{\tau} \big(\mathcal{I}(\tau_{\mathrm{A}}, \tau_{\mathrm{D}}, \tau_{\mathrm{E}}) + \mathcal{I}(\tau_{\mathrm{D}}, \tau_{\mathrm{A}}, \tau_{\mathrm{B}}) + \mathcal{I}(\tau_{\mathrm{C}}, \tau_{\mathrm{F}}, \tau_{\mathrm{E}}) + \mathcal{I}(\tau_{\mathrm{F}}, \tau_{\mathrm{C}}, \tau_{\mathrm{B}})\big) \\
&\le 128\big(\iota_{\mathrm{A,DE}} + \iota_{\mathrm{AB,D}} + \iota_{\mathrm{C,FE}} + \iota_{\mathrm{CB,F}}\big)
\end{aligned}$$

by Lemma 3. ∎

[2] In fact, properly balancing the calculations in our proof of Lemma 4 shows that the factor of 2 can be improved to the golden ratio $\phi \approx 1.618$, which does not seem to follow from the proof in [BM13].

2.4 Step 4: Relating information and probabilities for transcripts

We first give some intuition for why the inequality in Lemma 5 is true. Suppose for some small $\delta, \epsilon > 0$ we have $a = 1/2 + \delta$, $e = 1/2 + \epsilon$, and $b = c = d = f = 1/2$, as illustrated in Figure 3. (Although this is just a specific example, the phenomenon it illustrates turns out to hold in general.)

The left side of the inequality is the linear combination of the areas of the 9 rectangles, with coefficients as indicated in the figure. The purple regions are congruent and hence cancel out since the coefficients sum to 0. The red regions are congruent and hence cancel out since the coefficients in the top row sum to 0. The blue regions are congruent and hence cancel out since the coefficients in the middle column sum to 0. Thus the left side is $2\delta\epsilon$ since only the green region contributes.

Regarding the four terms on the right side of the inequality, the first and third are $\Theta(\epsilon^2)$, the second is $\Theta(\delta^2)$, and the fourth is 0. Hence left side $= \Theta(\delta\epsilon) \le \Theta(\epsilon^2 + \delta^2) =$ right side. The point is that the right side only has terms that are quadratic in δ, ϵ, while the left side has "higher-order" terms (at least linear in δ, ϵ) but those higher-order terms miraculously cancel out leaving only quadratic terms. The key property for the cancellation is that in every row and every column, the coefficients sum to 0.[3]

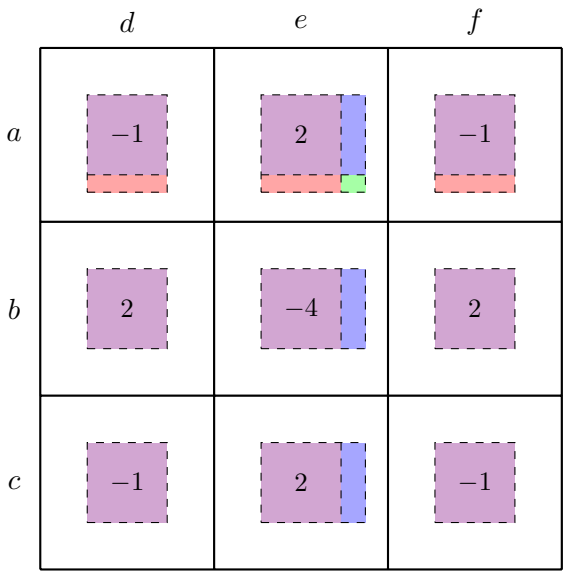

Figure 3: Intuition for Lemma 5.

We proceed to our formal proofs of Lemma 4 and Lemma 5. To avoid division-by-0 technicalities, we assume the relevant quantities are infinitesimally perturbed so none are 0.

Proof of Lemma 4. Define

$$\mathcal{L} \coloneqq -qs + qt + rs - rt = (q - r)(t - s)$$

to be the left side of the inequality in the statement of Lemma 4, and define

$$\mathcal{R} \coloneqq \mathcal{I}(q, s, t) + \mathcal{I}(s, q, r) = \frac{q}{t + s}(t - s)^2 + \frac{s}{q + r}(q - r)^2$$

to be the right side except for the factor of 2. The goal is to show that $\mathcal{R} \ge \mathcal{L}/2$. If $q \ge r$ and $s \ge t$, or if $r \ge q$ and $t \ge s$, then $\mathcal{L} \le 0 \le \mathcal{R}$, so we are done in these cases. Now consider the case that $q \ge r$ and $t \ge s$. (The remaining case, that $r \ge q$ and $s \ge t$, is symmetric.) If $t \le 3s$ (so $s/(t+s) \ge 1/4$) then since $q/(q+r) \ge 1/2$, the product of the two terms of $\mathcal{R}$ is $\ge (q-r)^2(t-s)^2/8$, so by AM–GM, $\mathcal{R} \ge 2(q - r)(t - s)/\sqrt{8} \ge \mathcal{L}/2$. If $t \ge 3s$ then $t + s \le 2(t - s)$ so the first term of $\mathcal{R}$ is $\ge (q/2(t - s))(t - s)^2 = q(t - s)/2 \ge \mathcal{L}/2$. □

Proof of Lemma 5. Define

$$\mathcal{L} \coloneqq -ad + 2ae - af + 2bd - 4be + 2bf - cd + 2ce - cf = (a - 2b + c)(-d + 2e - f)$$

[3] We have not attempted to verify whether an analogue of Lemma 5 holds for every such list of coefficients.

to be the left side of the inequality in the statement of Lemma 5, and define

$$\begin{aligned}\mathcal{R} &:= \mathcal{I}(a,d,e)+\mathcal{I}(d,a,b)+\mathcal{I}(c,f,e)+\mathcal{I}(f,c,b)\\ &= \frac{a}{e+d}(e-d)^2+\frac{d}{a+b}(a-b)^2+\frac{c}{e+f}(e-f)^2+\frac{f}{c+b}(c-b)^2\end{aligned}$$

to be the right side except for the factor of 32. The goal is to show that $\mathcal{R} \geq \mathcal{L}/32$. If $a+c \geq 2b$ and $d+f \geq 2e$, or if $a+c \leq 2b$ and $d+f \leq 2e$, then $\mathcal{L} \leq 0 \leq \mathcal{R}$, so we are done in these cases. Now consider the case that $a+c \geq 2b$ and $d+f \leq 2e$. (The remaining case, that $a+c \leq 2b$ and $d+f \geq 2e$, is symmetric.) We consider four subcases; the first two are just like our argument for Lemma 4, but the other two are a bit more complicated.

$c \leq a$ and $d \leq f$: Then $\mathcal{L} \leq 4(a-b)(e-d)$. If $e \leq 3d$ (so $d/(e+d) \geq 1/4$) then since $a/(a+b) \geq 1/2$ (because $b \leq a$ follows from $a+c \geq 2b$ and $c \leq a$), the product of the first two terms of $\mathcal{R}$ is $\geq (a-b)^2(e-d)^2/8$, so by AM–GM, the sum of these two terms is $\geq 2(a-b)(e-d)/\sqrt{8} \geq \mathcal{L}/6$. If $e \geq 3d$ then $e+d \leq 2(e-d)$ so the first term of $\mathcal{R}$ is $\geq (a/2(e-d))(e-d)^2 = a(e-d)/2 \geq (a-b)(e-d)/2 \geq \mathcal{L}/8$.

$a \leq c$ and $f \leq d$: Then $\mathcal{L} \leq 4(c-b)(e-f)$. If $e \leq 3f$ (so $f/(e+f) \geq 1/4$) then since $c/(c+b) \geq 1/2$ (because $b \leq c$ follows from $a+c \geq 2b$ and $a \leq c$), the product of the last two terms of $\mathcal{R}$ is $\geq (c-b)^2(e-f)^2/8$, so by AM–GM, the sum of these two terms is $\geq 2(c-b)(e-f)/\sqrt{8} \geq \mathcal{L}/6$. If $e \geq 3f$ then $e+f \leq 2(e-f)$ so the third term of $\mathcal{R}$ is $\geq (c/2(e-f))(e-f)^2 = c(e-f)/2 \geq (c-b)(e-f)/2 \geq \mathcal{L}/8$.

$a \leq c$ and $d \leq f$: Then $\mathcal{L} \leq 4(c-b)(e-d)$. If $e \leq 2f$ (so $f/(e+d) \geq 1/3$) and $c \leq 5a$ (so $a/(c+b) \geq 1/10$) then the product of the first and last terms of $\mathcal{R}$ is $\geq (c-b)^2(e-d)^2/30$, so by AM–GM, the sum of these two terms is $\geq 2(c-b)(e-d)/\sqrt{30} \geq \mathcal{L}/12$. If $e \leq 2f$ and $c \geq 5a$ then $f \geq (e-d)/2$ and $c+b \leq 4(c-b)$ (because $6c \geq 5c+5a \geq 10b$) so the last term of $\mathcal{R}$ is $\geq (f/4(c-b))(c-b)^2 = f(c-b)/4 \geq (c-b)(e-d)/8 \geq \mathcal{L}/32$. If $e \geq 2f$ then $e+f \leq 3(e-f)$ and $e-f \geq e/2 \geq (e-d)/2$ so the third term of $\mathcal{R}$ is $\geq (c/3(e-f))(e-f)^2 = c(e-f)/3 \geq c(e-d)/6 \geq (c-b)(e-d)/6 \geq \mathcal{L}/24$.

$c \leq a$ and $f \leq d$: Then $\mathcal{L} \leq 4(a-b)(e-f)$. If $e \leq 2d$ (so $d/(e+f) \geq 1/3$) and $a \leq 5c$ (so $c/(a+b) \geq 1/10$) then the product of the middle two terms of $\mathcal{R}$ is $\geq (a-b)^2(e-f)^2/30$, so by AM–GM, the sum of these two terms is $\geq 2(a-b)(e-f)/\sqrt{30} \geq \mathcal{L}/12$. If $e \leq 2d$ and $a \geq 5c$ then $d \geq (e-f)/2$ and $a+b \leq 4(a-b)$ (because $6a \geq 5a+5c \geq 10b$) so the second term of $\mathcal{R}$ is $\geq (d/4(a-b))(a-b)^2 = d(a-b)/4 \geq (a-b)(e-f)/8 \geq \mathcal{L}/32$. If $e \geq 2d$ then $e+d \leq 3(e-d)$ and $e-d \geq e/2 \geq (e-f)/2$ so the first term of $\mathcal{R}$ is $\geq (a/3(e-d))(e-d)^2 = a(e-d)/3 \geq a(e-f)/6 \geq (a-b)(e-f)/6 \geq \mathcal{L}/24$. □

3 Query Lower Bound for Tribes

The upper bound for Theorem 2 was shown in Section 1.2; we now prove the matching lower bound.

Suppose for contradiction there is a randomized decision tree, which is a distribution $\mathcal{T}$ over deterministic decision trees that always make at most $\sqrt{\epsilon}m/2$ queries, and which accepts 0-inputs with probability at most $1/2-\epsilon$ and 1-inputs with probability at least $1/2+\epsilon$. Consider the

following pair of distributions (D_0, D_1) over 0-inputs and 1-inputs respectively: To sample from D_0, pick $\boldsymbol{i} \in_{\mathrm{u}} \{1,2\}, \boldsymbol{j} \in_{\mathrm{u}} [m], \boldsymbol{k} \in_{\mathrm{u}} [m]$ independently and set $\boldsymbol{z}_{\boldsymbol{i},\boldsymbol{j}} = \boldsymbol{z}_{\boldsymbol{i},\boldsymbol{k}} = 1$ (and the rest of the bits to 0). To sample from D_1, pick $\boldsymbol{j} \in_{\mathrm{u}} [m], \boldsymbol{k} \in_{\mathrm{u}} [m]$ independently and set $\boldsymbol{z}_{1,\boldsymbol{j}} = \boldsymbol{z}_{2,\boldsymbol{k}} = 1$ (and the rest of the bits to 0).

We claim that for an arbitrary T in the support of $\mathcal{T}$, for each $r \in \{0,1,2\}$, letting A_r be the set of z's such that $T(z)$ accepts after having read exactly r 1's, we have $\mathbb{P}_{D_1}[A_r] - \mathbb{P}_{D_0}[A_r] \leq \epsilon/4$. This yields the following contradiction:

$$\begin{aligned}
2\epsilon &= (1/2+\epsilon) - (1/2-\epsilon) \\
&\leq \mathbb{E}_{\boldsymbol{z}\sim D_1}\big[\mathbb{P}_{\boldsymbol{T}\sim\mathcal{T}}[\boldsymbol{T}(\boldsymbol{z}) \text{ accepts}]\big] - \mathbb{E}_{\boldsymbol{z}\sim D_0}\big[\mathbb{P}_{\boldsymbol{T}\sim\mathcal{T}}[\boldsymbol{T}(\boldsymbol{z}) \text{ accepts}]\big] \\
&= \mathbb{E}_{\boldsymbol{T}\sim\mathcal{T}}\Big[\mathbb{P}_{\boldsymbol{z}\sim D_1}[\boldsymbol{T}(\boldsymbol{z}) \text{ accepts}] - \mathbb{P}_{\boldsymbol{z}\sim D_0}[\boldsymbol{T}(\boldsymbol{z}) \text{ accepts}]\Big] \\
&= \mathbb{E}_{\boldsymbol{T}\sim\mathcal{T}}\Big[\textstyle\sum_r \big(\mathbb{P}_{D_1}[A_r] - \mathbb{P}_{D_0}[A_r]\big)\Big] \\
&\leq \epsilon/4 + \epsilon/4 + \epsilon/4
\end{aligned}$$

(where the dependence of A_r on $\boldsymbol{T}$ is implicit on the fourth line). To prove the claim, we first set up some notation. Consider the execution of T when it reads only 0's until it halts. Let $S_i \subseteq [m]$ ($i \in \{1,2\}$) be the coordinates of z_i queried on this execution, and let $\delta_i \coloneqq |S_i|/m$; note that $\delta_1 + \delta_2 \leq \sqrt{\epsilon}/2$. For each $q \in \big[|S_1| + |S_2|\big]$, let

- B^q be the set of z's that cause T to read $q-1$ 0's then a 1,
- $i^q \in \{1,2\}$, $h^q \in [m]$ be such that z_{i^q,h^q} is the location of that 1,
- $C^q \subseteq B^q$ be the set of z's that cause T to read $q-1$ 0's, then a 1, then only 0's until it halts,
- $S_i^q \subseteq [m]$ ($i \in \{1,2\}$) be the coordinates of z_i queried on the execution corresponding to C^q,
- $\delta_i^q \coloneqq |S_i^q|/m$ ($i \in \{1,2\}$); note that $\delta_1^q + \delta_2^q \leq \sqrt{\epsilon}/2$.

Case $\boldsymbol{r=0}$: If the execution that reads only 0's rejects then $\mathbb{P}_{D_1}[A_0] = \mathbb{P}_{D_0}[A_0] = 0$; otherwise

$$\mathbb{P}_{D_1}[A_0] - \mathbb{P}_{D_0}[A_0] \;=\; (1-\delta_1)(1-\delta_2) - \tfrac{1}{2}(1-\delta_1)^2 - \tfrac{1}{2}(1-\delta_2)^2 \;=\; \delta_1\delta_2 - \tfrac{1}{2}(\delta_1^2+\delta_2^2) \;\leq\; \epsilon/4.$$

Case $\boldsymbol{r=1}$: For each q, assuming for convenience that $i^q = 1$, we have

$$\mathbb{P}_{D_1}[C^q] \;=\; \mathbb{P}\big[\boldsymbol{j} = h^q \text{ and } \boldsymbol{k} \notin S_2^q\big] \;=\; (1-\delta_2^q)/m \;\leq\; 1/m$$

and

$$\begin{aligned}
\mathbb{P}_{D_0}[C^q] &\geq \mathbb{P}[\boldsymbol{i}=1] \cdot \mathbb{P}\big[(\boldsymbol{j} = h^q \text{ and } \boldsymbol{k} \notin S_1^q) \text{ or } (\boldsymbol{k} = h^q \text{ and } \boldsymbol{j} \notin S_1^q)\big] \\
&= \tfrac{1}{2} \cdot 2 \cdot (1-\delta_1^q)/m \\
&\geq (1-\sqrt{\epsilon}/2)/m
\end{aligned}$$

and so $\mathbb{P}_{D_1}[C^q] - \mathbb{P}_{D_0}[C^q] \leq \sqrt{\epsilon}/(2m)$. Letting $Q \subseteq \big[|S_1|+|S_2|\big]$ be those q's for which the execution corresponding to C^q accepts, and noting that $A_1 = \bigcup_{q\in Q} C^q$, we have

$$\mathbb{P}_{D_1}[A_1] - \mathbb{P}_{D_0}[A_1] \;=\; \textstyle\sum_{q\in Q}\big(\mathbb{P}_{D_1}[C^q] - \mathbb{P}_{D_0}[C^q]\big) \;\leq\; (\sqrt{\epsilon}m/2) \cdot \sqrt{\epsilon}/(2m) \;=\; \epsilon/4.$$

Case $\boldsymbol{r=2}$: We have

$$\mathbb{P}_{\boldsymbol{z}\sim D_1}\big[T(\boldsymbol{z}) \text{ reads at least one 1}\big] \;=\; \mathbb{P}\big[\boldsymbol{j} \in S_1 \text{ or } \boldsymbol{k} \in S_2\big] \;\leq\; \delta_1 + \delta_2 \;\leq\; \sqrt{\epsilon}/2.$$

For each q, assuming for convenience that $i^q = 1$, we have

$$\mathbb{P}_{\boldsymbol{z}\sim D_1}\big[T(\boldsymbol{z}) \text{ reads two 1's} \,\big|\, \boldsymbol{z} \in B^q\big] \;=\; \mathbb{P}_{\boldsymbol{z}\sim D_1}\big[\boldsymbol{k} \in S_2^q \,\big|\, \boldsymbol{z} \in B^q\big] \;\le\; \delta_2^q \;\le\; \sqrt{\epsilon}/2$$

(the middle inequality may not be an equality, since prior to reading the first 1, T may have read some 0's in $\boldsymbol{z}_2$). Hence

$$\begin{aligned}
&\mathbb{P}_{D_1}[A_2] - \mathbb{P}_{D_0}[A_2] \\
\le\;& \mathbb{P}_{\boldsymbol{z}\sim D_1}\big[T(\boldsymbol{z}) \text{ reads two 1's}\big] \\
=\;& \mathbb{P}_{\boldsymbol{z}\sim D_1}\big[T(\boldsymbol{z}) \text{ reads at least one 1}\big] \cdot \mathbb{P}_{\boldsymbol{z}\sim D_1}\big[T(\boldsymbol{z}) \text{ reads two 1's} \,\big|\, T(\boldsymbol{z}) \text{ reads at least one 1}\big] \\
\le\;& (\sqrt{\epsilon}/2) \cdot (\sqrt{\epsilon}/2) \\
=\;& \epsilon/4.
\end{aligned}$$

4 Which One Is the 1-Input?

We prove Theorem 3 and related results in this section. We state and apply the key lemmas in Section 4.1, and we prove them in Section 4.2. We describe some ways to reinterpret Theorem 3 in Section 4.3. We discuss some related questions in Section 4.4.

4.1 Overview

Let us first review some definitions.

Correctness: We say Π is p-correct if for each (X, Y) in the domain of F, we have $\mathbb{P}[\Pi(X, Y) = F(X, Y)] \ge p$ over the randomness of Π. For a distribution D over the domain of F, we say Π is (p, D)-correct if $\mathbb{P}[\Pi(\boldsymbol{X}, \boldsymbol{Y}) = F(\boldsymbol{X}, \boldsymbol{Y})] \ge p$ over both the randomness of Π and $\boldsymbol{XY} \sim D$.

Efficiency: We let $CC(\Pi)$ denote the worst-case communication cost of Π. Letting D' be a distribution over the set of all possible inputs to Π (which is a superset of the domain of F), define $IC^{D'}(\Pi) \coloneqq \mathbb{I}\big(\boldsymbol{\tau}\,;\,\boldsymbol{X} \,\big|\, \boldsymbol{Y}\boldsymbol{R}^{\text{pub}}\big) + \mathbb{I}\big(\boldsymbol{\tau}\,;\,\boldsymbol{Y} \,\big|\, \boldsymbol{X}\boldsymbol{R}^{\text{pub}}\big)$ to be the internal information cost with respect to $\boldsymbol{XY} \sim D'$ (where $\boldsymbol{\tau}$ denotes the random transcript and $\boldsymbol{R}^{\text{pub}}$ denotes the public coins)[4].

Complexity: We can define the following complexity measures. (Note that in this notation, the subscripts are related to correctness and the superscripts are related to efficiency.)

$$\begin{aligned}
\mathrm{R}_p(F) &\coloneqq \min_{p\text{-correct }\Pi} CC(\Pi) \\
\mathrm{R}_{p,D}(F) &\coloneqq \min_{(p,\,D)\text{-correct }\Pi} CC(\Pi) \\
\mathrm{I}_p^{D'}(F) &\coloneqq \inf_{p\text{-correct }\Pi} IC^{D'}(\Pi) \\
\mathrm{I}_{p,D}^{D'}(F) &\coloneqq \inf_{(p,\,D)\text{-correct }\Pi} IC^{D'}(\Pi)
\end{aligned}$$

Lemma 6. *For every F and balanced distribution $D = \frac{1}{2}D_0 + \frac{1}{2}D_1$ on the domain of F, we have $\mathrm{I}_{1/2+\epsilon/2,D}^{D_0}(F) \le \mathrm{R}_{1/\ell+\epsilon}(\textsc{Which}_\ell \circ F^\ell)/\ell$.*

Lemma 7. *For every F we have $\mathrm{R}_{1/2+\epsilon/4}(F) \le \mathrm{R}_{1/\ell+\epsilon}(\textsc{Which}_\ell \circ F^\ell)$.*

[4]This notation is somewhat different than in Section 2.1, where we found it more convenient to let $\boldsymbol{\tau}$ denote the concatenation of the communication transcript and the public coins.

We provide the (very similar) proofs of these two lemmas in Section 4.2. The key idea is that if we embed a random 1-input of F into a random coordinate and fill the other $\ell - 1$ coordinates with random 0-inputs of F, then the protocol for $\textsc{Which}_\ell \circ F^\ell$ will find the embedded 1-input with advantage ϵ, whereas if we embed a random 0-input in the same way then the protocol cannot achieve any advantage since the coordinate of the embedding becomes independent of the ℓ-tuple of 0-inputs given to the protocol. For Lemma 6 we use a direct sum property for information to get the factor ℓ decrease in cost; for Lemma 7 we do not get a decrease since there is no available analogous direct sum property for communication.

Proof of Theorem 3. The upper bound was shown in Section 1.3. Let $F \coloneqq \textsc{Unambig-Or}_m \circ 3\textsc{Eq}^m$. As noted in Section 1.5, it suffices to prove the lower bound for $\textsc{Which}_\ell \circ F^\ell$ instead of $\textsc{Which}_\ell \circ \textsc{Unambig-Inter}_m^\ell$. For $b \in \{0,1\}$ let D_b be the uniform distribution over $F^{-1}(b)$, and let $D \coloneqq \frac{1}{2}D_0 + \frac{1}{2}D_1$. It was shown in [BM13] that $\mathrm{I}^{D_0}_{1/2+\epsilon,D}(F) \geq \Omega(\epsilon m)$;[5] the result was not stated in this way in that paper, but careful inspection of the proof yields it.[6] Then $\mathrm{R}_{1/\ell+\epsilon}(\textsc{Which}_\ell \circ F^\ell) \geq \Omega(\epsilon \ell m)$ follows immediately from this and Lemma 6. □

Note that for any communication complexity class $\mathcal{C}$, if $F \in \mathcal{C}$ then $\textsc{Which}_2 \circ F^2 \in \mathcal{C} \cap \mathsf{co}\mathcal{C}$. Hence for $\ell = 2$ and ϵ a positive constant, Lemma 7 implies that if $\mathcal{C} \not\subseteq \mathsf{BPP}$ then $\mathcal{C} \cap \mathsf{co}\mathcal{C} \not\subseteq \mathsf{BPP}$. In particular, taking $F = \textsc{Unambig-Inter}$ (and $\mathcal{C} = \mathsf{UP}$), we have a simple proof of a result of [Kla03, Theorem 2 of the arXiv version], using as a black box the fact that $F \notin \mathsf{BPP}$.

4.2 Proofs

Proof of Lemma 6. Consider an arbitrary $(1/\ell + \epsilon)$-correct protocol Π for $\textsc{Which}_\ell \circ F^\ell$. Define a probability space with the following random variables: $\boldsymbol{i} \in_{\mathrm{u}} [\ell]$, $\boldsymbol{XY}$ is an input to Π such that $\boldsymbol{X_i Y_i} \sim D$ and $\boldsymbol{X_j Y_j} \sim D_0$ for $j \in [\ell] \setminus \{\boldsymbol{i}\}$ (with the ℓ coordinates independent conditioned on $\boldsymbol{i}$), $\boldsymbol{\tau}$ is the communication transcript of Π, and $\boldsymbol{R}^{\mathrm{pub}}, \boldsymbol{R}^{\mathrm{priv}}_{\mathrm{A}}, \boldsymbol{R}^{\mathrm{priv}}_{\mathrm{B}}$ are the public, Alice's private, and Bob's private coins, respectively. Let Π' be the following protocol with input interpreted as $\boldsymbol{X_i Y_i}$.

> Publicly sample $\boldsymbol{i}$, $\boldsymbol{X}_{1,\ldots,\boldsymbol{i}-1}$, $\boldsymbol{Y}_{\boldsymbol{i}+1,\ldots,\ell}$, and $\boldsymbol{R}^{\mathrm{pub}}$
> Alice privately samples $\boldsymbol{X}_{\boldsymbol{i}+1,\ldots,\ell}$ (conditioned on the outcome of $\boldsymbol{Y}_{\boldsymbol{i}+1,\ldots,\ell}$) and $\boldsymbol{R}^{\mathrm{priv}}_{\mathrm{A}}$
> Bob privately samples $\boldsymbol{Y}_{1,\ldots,\boldsymbol{i}-1}$ (conditioned on the outcome of $\boldsymbol{X}_{1,\ldots,\boldsymbol{i}-1}$) and $\boldsymbol{R}^{\mathrm{priv}}_{\mathrm{B}}$
> Run Π on the combined input $\boldsymbol{XY}$ with coins $\boldsymbol{R}^{\mathrm{pub}}, \boldsymbol{R}^{\mathrm{priv}}_{\mathrm{A}}, \boldsymbol{R}^{\mathrm{priv}}_{\mathrm{B}}$
> If Π outputs $\boldsymbol{i}$ then output 1, otherwise output 0

For a bit b, let E_b denote the event that $F(\boldsymbol{X_i}, \boldsymbol{Y_i}) = b$. We have

$$\begin{aligned}
& IC^{D_0}(\Pi') \\
\coloneqq\ & \mathbb{I}\big(\boldsymbol{\tau}\ ;\ \boldsymbol{X_i} \,\big|\, \boldsymbol{Y_i}, \boldsymbol{i}, \boldsymbol{X}_{1,\ldots,\boldsymbol{i}-1}, \boldsymbol{Y}_{\boldsymbol{i}+1,\ldots,\ell}, \boldsymbol{R}^{\mathrm{pub}}, E_0\big) + \mathbb{I}\big(\boldsymbol{\tau}\ ;\ \boldsymbol{Y_i} \,\big|\, \boldsymbol{X_i}, \boldsymbol{i}, \boldsymbol{X}_{1,\ldots,\boldsymbol{i}-1}, \boldsymbol{Y}_{\boldsymbol{i}+1,\ldots,\ell}, \boldsymbol{R}^{\mathrm{pub}}, E_0\big) \\
=\ & \tfrac{1}{\ell} \cdot \textstyle\sum_{i=1}^{\ell} \Big(\mathbb{I}\big(\boldsymbol{\tau}\ ;\ \boldsymbol{X}_i \,\big|\, \boldsymbol{X}_{1,\ldots,i-1}, \boldsymbol{Y}_{i,\ldots,\ell}, \boldsymbol{R}^{\mathrm{pub}}, E_0\big) + \mathbb{I}\big(\boldsymbol{\tau}\ ;\ \boldsymbol{Y}_i \,\big|\, \boldsymbol{X}_{1,\ldots,i}, \boldsymbol{Y}_{i+1,\ldots,\ell}, \boldsymbol{R}^{\mathrm{pub}}, E_0\big) \Big) \\
\leq\ & \tfrac{1}{\ell} \cdot IC^{D_0^\ell}(\Pi)
\end{aligned}$$

[5] The simplified proof of the main conclusion $\mathrm{R}_{1/2+\epsilon}(\textsc{Unambig-Inter}_m) \geq \Omega(\epsilon m)$ given in [GW16] does not yield the needed information complexity lower bound.

[6] For one thing, the write-up in [BM13] indicates that the information lower bound argument only works for protocols that have been "smoothed" in some sense, but actually this assumption is not necessary.

$\leq \frac{1}{\ell} \cdot CC(\Pi)$

where the inequalities follow by known facts (see [BM13, Fact 2.3 of the ECCC Revision #1 version] and [BR14, Lemma 3.14 of the ECCC Revision #1 version]). We also have $\mathbb{P}\big[\Pi' \text{ outputs } 1 \,\big|\, E_1\big] = \mathbb{P}\big[\Pi \text{ outputs } \boldsymbol{i} \,\big|\, E_1\big] \geq 1/\ell + \epsilon$ by the correctness of Π (since $\boldsymbol{i} = \big(\textsc{Which}_\ell \circ F^\ell\big)(\boldsymbol{X}, \boldsymbol{Y})$ assuming E_1). We also have $\mathbb{P}\big[\Pi' \text{ outputs } 1 \,\big|\, E_0\big] = \mathbb{P}\big[\Pi \text{ outputs } \boldsymbol{i} \,\big|\, E_0\big] = 1/\ell$ since conditioned on E_0, $\boldsymbol{i}$ is independent of $\boldsymbol{XY}$. Hence over the randomness of the whole experiment, the probability Π' is correct is at least $(1/2) \cdot (1/\ell + \epsilon) + (1/2) \cdot (1 - 1/\ell) = 1/2 + \epsilon/2$. □

Proof of Lemma 7. By the minimax theorem, it suffices to show that for every distribution D over the domain of F, $\mathrm{R}_{1/2+\epsilon/4,D}(F) \leq \mathrm{R}_{1/\ell+\epsilon}(\textsc{Which}_\ell \circ F^\ell)$. If either $F^{-1}(0)$ or $F^{-1}(1)$ has probability at least $1/2 + \epsilon/4$ under D, then a protocol that outputs a constant witnesses $\mathrm{R}_{1/2+\epsilon/4,D}(F) = 0$, so we may assume otherwise. For a bit b, let D_b be the distribution D conditioned on $F^{-1}(b)$.

Consider an arbitrary $(1/\ell + \epsilon)$-correct protocol Π for $\textsc{Which}_\ell \circ F^\ell$. Define a probability space with the following random variables: $\boldsymbol{i} \in_{\mathrm{u}} [\ell]$, $\boldsymbol{XY}$ is an input to Π such that $\boldsymbol{X_i Y_i} \sim D$ and $\boldsymbol{X}_j \boldsymbol{Y}_j \sim D_0$ for $j \in [\ell] \setminus \{\boldsymbol{i}\}$ (with the ℓ coordinates independent conditioned on $\boldsymbol{i}$), and $\boldsymbol{R}^{\mathrm{pub}}, \boldsymbol{R}^{\mathrm{priv}}_{\mathrm{A}}, \boldsymbol{R}^{\mathrm{priv}}_{\mathrm{B}}$ are the public, Alice's private, and Bob's private coins, respectively. Let $\boldsymbol{X}_{-\boldsymbol{i}} \boldsymbol{Y}_{-\boldsymbol{i}}$ denote $\boldsymbol{XY}$ restricted to coordinates in $[\ell] \setminus \{\boldsymbol{i}\}$. Let Π' be the following protocol with input interpreted as $\boldsymbol{X_i Y_i}$.

> Publicly sample $\boldsymbol{i}$, $\boldsymbol{X}_{-\boldsymbol{i}}$, $\boldsymbol{Y}_{-\boldsymbol{i}}$, and $\boldsymbol{R}^{\mathrm{pub}}$
> Alice and Bob privately sample $\boldsymbol{R}^{\mathrm{priv}}_{\mathrm{A}}$ and $\boldsymbol{R}^{\mathrm{priv}}_{\mathrm{B}}$, respectively
> Run Π on the combined input $\boldsymbol{XY}$ with coins $\boldsymbol{R}^{\mathrm{pub}}, \boldsymbol{R}^{\mathrm{priv}}_{\mathrm{A}}, \boldsymbol{R}^{\mathrm{priv}}_{\mathrm{B}}$
> If Π outputs $\boldsymbol{i}$ then output 1, otherwise output 0

Note that $CC(\Pi') \leq CC(\Pi)$. For a bit b, let E_b denote the event that $F(\boldsymbol{X_i}, \boldsymbol{Y_i}) = b$. We have $\mathbb{P}\big[\Pi' \text{ outputs } 1 \,\big|\, E_1\big] = \mathbb{P}\big[\Pi \text{ outputs } \boldsymbol{i} \,\big|\, E_1\big] \geq 1/\ell + \epsilon$ by the correctness of Π (since $\boldsymbol{i} = \big(\textsc{Which}_\ell \circ F^\ell\big)(\boldsymbol{X}, \boldsymbol{Y})$ assuming E_1). We also have $\mathbb{P}\big[\Pi' \text{ outputs } 1 \,\big|\, E_0\big] = \mathbb{P}\big[\Pi \text{ outputs } \boldsymbol{i} \,\big|\, E_0\big] = 1/\ell$ since conditioned on E_0, $\boldsymbol{i}$ is independent of $\boldsymbol{XY}$. Hence over the randomness of the whole experiment, the probability Π' is correct is at least the minimum of $(1/2 + \epsilon/4) \cdot (1/\ell + \epsilon) + (1/2 - \epsilon/4) \cdot (1 - 1/\ell)$ and $(1/2 - \epsilon/4) \cdot (1/\ell + \epsilon) + (1/2 + \epsilon/4) \cdot (1 - 1/\ell)$, both of which are at least $1/2 + \epsilon/4$. □

4.3 Corollaries

We now describe how Theorem 3 implies two other results, which give different perspectives. One result concerns so-called SV-nondeterminism, and the other concerns protocols whose output is a sample from a distribution.

Generally speaking, for a function with codomain $[\ell]$, an SV-nondeterministic algorithm can make a nondeterministic guess and output a value from $[\ell] \cup \{\perp\}$, and on every input it must (1) output the correct value for at least one guess, and (2) for each guess output either the correct value or $\perp$.[7] (For $\ell = 2$, this corresponds to an $\mathsf{NP} \cap \mathsf{coNP}$ type of computation.) This definition makes sense for communication complexity, where it turns out an SV-nondeterministic protocol can be equivalently defined as follows: There is a collection of rectangles each labeled with a value from $[\ell]$, such that the union of the rectangles labeled $v \in [\ell]$ exactly covers the set of all v-inputs. We let

[7] SV stands for "single-valued", which historically comes from the fact that the set of non-$\perp$ values that are output on a given input (over the possible guesses) must be a singleton.

$\mathrm{SV}(F)$ denote the minimum cost, i.e., log of the number of rectangles, of an SV-nondeterministic protocol for F.

Corollary 1. *There exists an F with codomain $[\ell]$ such that $\mathrm{R}_{1/\ell+\epsilon}(F) \geq \Omega(\epsilon 2^{\mathrm{SV}(F)})$. Moreover, this is tight: For every F with codomain $[\ell]$ we have $\mathrm{R}_{1/\ell+\epsilon}(F) \leq O(\epsilon 2^{\mathrm{SV}(F)})$ provided $\epsilon 2^{\mathrm{SV}(F)} \geq 1$.*

Proof. By Theorem 3, the first part is witnessed by $F \coloneqq \textsc{Which}_\ell \circ \textsc{Unambig-Inter}_m^\ell$ (for any ℓ and m with $\epsilon\ell m \geq 1$) since $\mathrm{SV}(F) \leq \log(\ell m)$. As for the second part, given a cost-c SV-nondeterministic protocol for F, Alice and Bob can publicly sample a subset of $2\epsilon 2^c$ of the 2^c rectangles, and if the input lies in any of them (which can be checked with $O(\epsilon 2^c)$ bits of communication) then they output the label of that rectangle, otherwise they output a uniformly random value from $[\ell]$. □

Let $\mathcal{D}_\ell$ denote the set of all probability distributions over $[\ell]$. A function F with codomain $\mathcal{D}_\ell$ can be viewed as a *sampling problem*, where given input (X, Y) the goal is to output a sample from (or close to) the distribution $F(X, Y)$. We define $\mathrm{S}_p(F)$ as the minimum worst-case communication cost of any protocol Π that, for each input (X, Y), outputs a sample from a distribution $\Pi(X, Y) \in \mathcal{D}_\ell$ such that $\Delta\big(\Pi(X, Y), F(X, Y)\big) \leq 1 - p$. Note that the uniform distribution over $[\ell]$ is within distance $1 - 1/\ell$ of every distribution in $\mathcal{D}_\ell$, so $\mathrm{S}_{1/\ell}(F) = 0$ for all F. Thus it makes sense to consider the complexity of achieving advantage ϵ, i.e., $\mathrm{S}_{1/\ell+\epsilon}(F)$.

A natural nondeterministic analogue of sampling is sampling with *postselection*: A protocol may output $\perp$ with probability < 1, and *conditioned* on not outputting $\perp$, the output should be a sample from (or close to) $F(X, Y)$. An issue is that if we do not restrict the probability of outputting $\perp$, then every F can be sampled with postselection with constant communication (by using public coins to guess what the joint input is). Hence we define $\mathrm{PS}_p(F)$ as the minimum $CC(\Pi) + \log(1/\alpha)$ of any protocol Π that, for each input (X, Y), conditioned on not outputting $\perp$, outputs a sample from a distribution $\Pi(X, Y) \in \mathcal{D}_\ell$ such that $\Delta\big(\Pi(X, Y), F(X, Y)\big) \leq 1 - p$, and where $\alpha > 0$ is defined as the minimum over inputs of the probability of not outputting $\perp$. (Such logarithmic terms appear in the cost measures for several other communication models; see [GLM+16] for more details.) We note that a protocol with communication cost c and associated α can be modified to have communication cost 2 and associated $\alpha' \coloneqq \alpha/2^c$: Assuming w.l.o.g. that for each outcome of the public coins, the corresponding deterministic protocol has exactly 2^c possible transcripts, Alice and Bob can sample all the coins as usual as well as publicly sample a uniformly random transcript; they can then check whether the guessed transcript would have been the real one, and if so output the same value and if not output $\perp$.

Corollary 2. *There exists an F with codomain $\mathcal{D}_\ell$ such that $\mathrm{S}_{1/\ell+\epsilon}(F) \geq \Omega(\epsilon 2^{\mathrm{PS}_1(F)})$. Moreover, this is tight: For every F with codomain $\mathcal{D}_\ell$ we have $\mathrm{S}_{1/\ell+\epsilon}(F) \leq O(\epsilon 2^{\mathrm{PS}_1(F)})$ provided $\epsilon 2^{\mathrm{PS}_1(F)} \geq 1$.*

Proof. By Theorem 3, the first part is witnessed by $F \coloneqq \textsc{Which}_\ell \circ \textsc{Unambig-Inter}_m^\ell$ (for any ℓ and m with $\epsilon\ell m \geq 1$), where we identify the output of F (a value from $[\ell]$) with the distribution completely concentrated on that value, in which case we have $\mathrm{S}_{1/\ell+\epsilon}(F) = \mathrm{R}_{1/\ell+\epsilon}(F) \geq \Omega(\epsilon\ell m)$ and $\mathrm{PS}_1(F) \leq O(1) + \log(\ell m)$. As for the second part, given a PS_1 protocol for F with communication cost 2 (which is w.l.o.g. as noted above) and associated α, Alice and Bob can run that protocol $O(\epsilon/\alpha)$ times; if it ever produces a non-$\perp$ output (which happens with probability $\geq 2\epsilon$) then they output the same value, otherwise they output a uniformly random value from $[\ell]$. The statistical distance of this distribution to $F(X, Y)$ is $\leq 2\epsilon \cdot 0 + (1 - 2\epsilon) \cdot (1 - 1/\ell) \leq 1 - 1/\ell - \epsilon$. □

4.4 Related questions

One question is how strong of a converse there is to Lemma 7, i.e., how well $\mathrm{R}_{1/\ell+\epsilon}(\textsc{Which}_\ell \circ F^\ell)$ can be upper bounded in terms of $\mathrm{R}_{1/2+\delta}(F)$. Doing so as a black-box reduction (which would also work for query complexity) can be phrased as the following problem: Supposing there are ℓ coins, one of which is good (having heads probability $\geq 1/2+\delta$) and the rest of which are bad (having heads probability $\leq 1/2-\delta$), identify the good coin with probability $\geq 1/\ell+\epsilon$ (over the randomness of both the algorithm and the coin flips). This has somewhat of a multi-armed bandit flavor and fits in the framework of "noisy decision trees". As far as we know, it is open to determine an optimal strategy for arbitrary ℓ, ϵ, δ, but here are some observations. (In conjunction with Lemma 7, these show that F and $\textsc{Which}_\ell \circ F^\ell$ are at least qualitatively equivalent in complexity for small ℓ.)

- $\mathrm{R}_{3/4}(\textsc{Which}_\ell \circ F^\ell) \leq \ell \cdot \mathrm{R}_{1-1/(4\ell)}(F)$ since we can just flip each coin once, and by a union bound, with probability 3/4 all the coins will have the "right" outcomes. (This does not exploit any properties of $\textsc{Which}_\ell$.) Of course, $\mathrm{R}_{1-1/(4\ell)}(F)$ can be further upper bounded in terms of smaller-advantage complexities by majority-amplification.
- $\mathrm{R}_{1/\ell+\epsilon}(\textsc{Which}_\ell \circ F^\ell) \leq \mathrm{R}_{1/2+\epsilon\ell/2}(F)$ (provided $\epsilon\ell \leq 1$) since we can pick a coin uniformly at random and flip it; if it comes up heads then output the index of that coin; otherwise output a uniformly random one of the other $\ell-1$ indices. This implies that $\mathrm{R}_{1/\ell+\epsilon}(\textsc{Which}_\ell \circ F^\ell) \leq O\big(\ell^2 \cdot \mathrm{R}_{1/2+\epsilon}(F)\big)$ (provided $\epsilon\ell \leq 1$) since by Lemma 8 we can boost ϵ advantage to $\epsilon\ell/2$ advantage with $O(\ell^2)$-repetition majority-amplification.

We also remark that $\mathrm{R}^{\mathrm{dt}}_{1/2+\epsilon}(f) \leq O\big(\mathrm{R}^{\mathrm{dt}}_{1/\ell+\epsilon}(\textsc{Which}_\ell \circ f^\ell)/(\epsilon\ell)\big)$ follows by combining the idea behind Lemma 7 with the idea behind the "AND-composition lemma" in [GJPW18] (namely, halting and outputting 1 if the number of queries exceeds $O(1/(\epsilon\ell))$ times the height of the randomized decision tree for $\textsc{Which}_\ell \circ f^\ell$). We omit the details of the simple analysis.

Finally, we remark that by combining Lemma 6 with the "one-sided vs. two-sided information complexity" equivalence of [GJPW18] and the "worst-case vs. average-case information complexity" equivalence of [Bra15], it is possible to derive a version of Lemma 6 with worst-case information complexity on the left side of the inequality.

A A Delicate Concentration Bound

Lemma 8. *Suppose a coin with heads probability $\frac{1}{2}+\delta$ is tossed N times (where $\delta \geq 0$ and N is odd). Then the probability of getting a majority of heads is at least $\frac{1}{2}+\Omega(\delta\sqrt{N})$ provided $\delta\sqrt{N} \leq 1$.*

As mentioned at the beginning of this paper, Lemma 8 implies the following bound.

Corollary 3. $\mathrm{R}_{1/2+\epsilon}(\textsc{Gap-Hamming}) \leq O(\epsilon^2 n)$ *provided* $\epsilon^2 n \geq 1$.

Proof of Corollary 3. Suppose Alice and Bob publicly sample a uniformly random one of the n coordinates and accept iff their bits are unequal there. This can be viewed as a coin toss with heads probability at least $\frac{1}{2}+1/\sqrt{n}$ (where heads represents the output being correct). Repeating the experiment $\Theta(\epsilon^2 n)$ times and taking the majority outcome boosts the success probability to $\frac{1}{2}+\epsilon$. ∎

Lemma 8 follows without difficulty from a Chernoff bound when $N = \Theta(1/\delta^2)$, and from the Berry–Esseen theorem when $\omega(1/\delta) \le N \le O(1/\delta^2)$, but the general case seems to require a direct proof, and we provide one below. We could not find a proof in the literature. After this paper was written, other proofs of Lemma 8 were discovered and presented at [Vio].

Proof of Lemma 8. We think of N as a fixed, sufficiently large number, and δ as varying in the range $[0, 1/\sqrt{N}]$. In fact, since the probability in question is a monotonically increasing function of δ, it suffices to consider $\delta \in [0, 0.01/\sqrt{N}]$.

Letting $p_{i,\delta} \coloneqq \binom{N}{i} \cdot (\frac{1}{2}+\delta)^i \cdot (\frac{1}{2}-\delta)^{N-i}$, the probability is $\sum_{i=\lceil N/2 \rceil}^{N} p_{i,\delta}$. When $\delta = 0$ this equals $\frac{1}{2}$ since N is odd, so it suffices to show that the derivative of the probability with respect to δ is $\Omega(\sqrt{N})$ for all $\delta \in \left[0, 0.01/\sqrt{N}\right]$. We introduce the shorthand $\gamma \coloneqq \frac{i}{N} - \frac{1}{2}$ (and hence $i = N \cdot (\frac{1}{2} + \gamma)$), keeping in mind that γ is a function of i even though we suppress this dependence in the notation.

The key claim is that $\frac{d}{d\delta}[p_{i,\delta}] = c_{i,\delta} \cdot \sqrt{N} \cdot (\gamma - \delta)$ for some $c_{i,\delta}$ that is nonnegative for all $\lceil N/2 \rceil \le i \le N$ and is in $[2.5, 3.7]$ for all $\lceil N/2 \rceil \le i \le N/2 + 0.03\sqrt{N}$ (so $\gamma \in [0, 0.03/\sqrt{N}]$). Then

$$\begin{aligned}
\tfrac{d}{d\delta}\Big[\textstyle\sum_{i=\lceil N/2 \rceil}^{N} p_{i,\delta}\Big] &= \textstyle\sum_{i=\lceil N/2 \rceil}^{N} c_{i,\delta} \cdot \sqrt{N} \cdot (\gamma - \delta) \\
&\ge \Big(\textstyle\sum_{i\,:\,\gamma \in [0,\delta)} 3.7 \cdot \sqrt{N} \cdot (\gamma - \delta)\Big) + \Big(\textstyle\sum_{i\,:\,\gamma \in (\delta, 0.03/\sqrt{N}]} 2.5 \cdot \sqrt{N} \cdot (\gamma - \delta)\Big) \\
&\ge \Big(-3.7 \cdot \sqrt{N} \cdot \textstyle\sum_{j=1}^{\lfloor 0.01\sqrt{N} \rfloor} \frac{j}{N}\Big) + \Big(2.5 \cdot \sqrt{N} \cdot \textstyle\sum_{j=1}^{\lfloor 0.02\sqrt{N} \rfloor} \frac{j}{N}\Big) \\
&\ge \Big(-\tfrac{3.7}{\sqrt{N}} \cdot \big(0.01\sqrt{N} + 1\big)^2/2\Big) + \Big(\tfrac{2.5}{\sqrt{N}} \cdot \big(0.02\sqrt{N} - 1\big)^2/2\Big) \\
&\ge 0.0003\sqrt{N}.
\end{aligned}$$

It remains to prove the key claim. We have

$$\begin{aligned}
\tfrac{d}{d\delta}[p_{i,\delta}] &= \tbinom{N}{i} \cdot \Big(i \cdot (\tfrac{1}{2}+\delta)^{i-1} \cdot (\tfrac{1}{2}-\delta)^{N-i} - (N-i) \cdot (\tfrac{1}{2}+\delta)^i \cdot (\tfrac{1}{2}-\delta)^{N-i-1}\Big) \\
&= \tbinom{N}{i} \cdot (\tfrac{1}{2}+\delta)^i \cdot (\tfrac{1}{2}-\delta)^{N-i} \cdot \Big(\tfrac{i}{\frac{1}{2}+\delta} - \tfrac{N-i}{\frac{1}{2}-\delta}\Big).
\end{aligned}$$

As a special case, this is nonnegative when $i = N$, so henceforth assume $i < N$ (i.e., $\gamma < \frac{1}{2}$). By Stirling approximations,

$$\tbinom{N}{i} = c_i' \cdot \tfrac{N^N}{i^i \cdot (N-i)^{N-i}} \cdot \sqrt{\tfrac{N}{i \cdot (N-i)}} = c_i' \cdot \tfrac{1}{(\frac{1}{2}+\gamma)^i \cdot (\frac{1}{2}-\gamma)^{N-i}} \cdot \tfrac{1}{\sqrt{(\frac{1}{2}+\gamma) \cdot (\frac{1}{2}-\gamma)}} \cdot \tfrac{1}{\sqrt{N}}$$

for some $c_i' \in [0.33, 0.44]$. We also have

$$\tfrac{i}{\frac{1}{2}+\delta} - \tfrac{N-i}{\frac{1}{2}-\delta} = N \cdot \Big(\tfrac{\frac{1}{2}+\gamma}{\frac{1}{2}+\delta} - \tfrac{\frac{1}{2}-\gamma}{\frac{1}{2}-\delta}\Big) = N \cdot \Big(\tfrac{(1+2\delta)+(2\gamma-2\delta)}{1+2\delta} - \tfrac{(1-2\delta)-(2\gamma-2\delta)}{1-2\delta}\Big) = 2N \cdot \big(\tfrac{1}{1+2\delta} + \tfrac{1}{1-2\delta}\big) \cdot (\gamma - \delta).$$

Putting these things together, we have

$$\tfrac{d}{d\delta}[p_{i,\delta}] = 2c_i' \cdot \underbrace{\big(\tfrac{1+2\delta}{1+2\gamma}\big)^i \cdot \big(\tfrac{1-2\delta}{1-2\gamma}\big)^{N-i}}_{c_{i,\delta}''} \cdot \underbrace{\tfrac{1}{\sqrt{(\frac{1}{2}+\gamma) \cdot (\frac{1}{2}-\gamma)}}}_{c_i'''} \cdot \underbrace{\big(\tfrac{1}{1+2\delta} + \tfrac{1}{1-2\delta}\big)}_{c_\delta''''} \cdot \sqrt{N} \cdot (\gamma - \delta).$$

Thus $c_{i,\delta} = 2c'_i \cdot c''_{i,\delta} \cdot c'''_i \cdot c''''_\delta$, which is certainly nonnegative for $\gamma \in [0, \frac{1}{2})$. Henceforth assume $\gamma \in [0, 0.03/\sqrt{N}]$; then in particular, $c'''_i \in [1.99, 2.01]$ since N is sufficiently large. Similarly, $c''''_\delta \in [1.99, 2.01]$. Note that

$$c''_{i,\delta} = c^*_{i,\delta} \cdot c^{**}_{i,\delta} \qquad \text{where} \qquad c^*_{i,\delta} \coloneqq \left(\tfrac{1+2\delta}{1+2\gamma}\right)^{2i-N} \qquad \text{and} \qquad c^{**}_{i,\delta} \coloneqq \left(\tfrac{1-4\delta^2}{1-4\gamma^2}\right)^{N-i}.$$

We have the following calculations.

- $c^*_{i,\delta} \geq \left(\frac{1}{1+2\gamma}\right)^{2i-N} \geq \left(\frac{1}{1+0.06/\sqrt{N}}\right)^{0.06\sqrt{N}} \geq (e^{-0.06/\sqrt{N}})^{0.06\sqrt{N}} = e^{-0.0036} \geq 0.99$
- $c^*_{i,\delta} \leq (1+2\delta)^{2i-N} \leq (1+0.02/\sqrt{N})^{0.06\sqrt{N}} \leq (e^{0.03/\sqrt{N}})^{0.06\sqrt{N}} = e^{0.0018} \leq 1.01$
- $c^{**}_{i,\delta} \geq (1-4\delta^2)^{N-i} \geq (1-0.0004/N)^{N/2} \geq (e^{-0.01/N})^{N/2} = e^{-0.005} \geq 0.99$
- $c^{**}_{i,\delta} \leq \left(\frac{1}{1-4\gamma^2}\right)^{N-i} \leq \left(\frac{1}{1-0.0036/N}\right)^{N/2} \leq (e^{0.01/N})^{N/2} = e^{0.005} \leq 1.01$

It follows that $c''_{i,\delta} \in [0.99^2, 1.01^2]$. In conclusion, we have

$$c_{i,\delta} \in \left[2 \cdot 0.33 \cdot 0.99^2 \cdot 1.99 \cdot 1.99,\ 2 \cdot 0.44 \cdot 1.01^2 \cdot 2.01 \cdot 2.01\right] \subseteq [2.5, 3.7]$$

which proves the key claim. □

References

[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.

[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.

[BCK+16] Joshua Brody, Amit Chakrabarti, Ranganath Kondapally, David Woodruff, and Grigory Yaroslavtsev. Certifying equality with limited interaction. *Algorithmica*, 76(3):796–845, 2016. doi:10.1007/s00453-016-0163-6.

[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, 2013. doi:10.1145/2488608.2488628.

[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.

[BR14] Mark Braverman and Anup Rao. Information equals amortized communication. *IEEE Transactions on Information Theory*, 60(10):6058–6069, 2014. doi:10.1109/TIT.2014.2347282.

[Bra15] Mark Braverman. Interactive information complexity. *SIAM Journal on Computing*, 44(6):1698–1739, 2015. doi:10.1137/130938517.

[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.

[CG88] Benny Chor and Oded Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM Journal on Computing*, 17(2):230–261, 1988. doi:10.1137/0217015.

[CM15] Arkadev Chattopadhyay and Sagnik Mukhopadhyay. Tribes is hard in the message passing model. In *Proceedings of the 32nd International Symposium on Theoretical Aspects of Computer Science (STACS)*, pages 224–237. Schloss Dagstuhl, 2015. doi:10.4230/LIPIcs.STACS.2015.224.

[CR12] Amit Chakrabarti and Oded Regev. An optimal lower bound on the communication complexity of Gap-Hamming-Distance. *SIAM Journal on Computing*, 41(5):1299–1317, 2012. doi:10.1137/120861072.

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

[GJ16] Mika Göös and T. S. Jayram. A composition theorem for conical juntas. In *Proceedings of the 31st Computational Complexity Conference (CCC)*, pages 5:1–5:16. Schloss Dagstuhl, 2016. doi:10.4230/LIPIcs.CCC.2016.5.

[GJPW18] Mika Göös, T. S. Jayram, Toniann Pitassi, and Thomas Watson. Randomized communication vs. partition number. *ACM Transactions on Computation Theory*, 10(1):4:1–4:20, 2018. doi:10.1145/3170711.

[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.

[GPW17] Mika Göös, Toniann Pitassi, and Thomas Watson. Query-to-communication lifting for BPP. In *Proceedings of the 58th Symposium on Foundations of Computer Science (FOCS)*, pages 132–143. IEEE, 2017. doi:10.1109/FOCS.2017.21.

[GPW18] Mika Göös, Toniann Pitassi, and Thomas Watson. The landscape of communication complexity classes. *Computational Complexity*, 27(2):245–304, 2018. doi:10.1007/s00037-018-0166-6.

[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.

[HJ13] Prahladh Harsha and Rahul Jain. A strong direct product theorem for the tribes function via the smooth-rectangle bound. In *Proceedings of the 33rd Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS)*, pages 141–152. Schloss Dagstuhl, 2013. doi:10.4230/LIPIcs.FSTTCS.2013.141.

[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.

[JKR09] T.S. Jayram, Swastik Kopparty, and Prasad Raghavendra. On the communication complexity of read-once AC^0 formulae. In *Proceedings of the 24th Conference on Computational Complexity (CCC)*, pages 329–340. IEEE, 2009. doi:10.1109/CCC.2009.39.

[JKS03] T.S. Jayram, Ravi Kumar, and D. Sivakumar. Two applications of information complexity. In *Proceedings of the 35th Symposium on Theory of Computing (STOC)*, pages 673–682. ACM, 2003. doi:10.1145/780542.780640.

[JKZ10] Rahul Jain, Hartmut Klauck, and Shengyu Zhang. Depth-independent lower bounds on the communication complexity of read-once boolean formulas. In *Proceedings of the 16th International Computing and Combinatorics Conference (COCOON)*, pages 54–59. Springer, 2010. doi:10.1007/978-3-642-14031-0_8.

[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.

[LS10] Nikos Leonardos and Michael Saks. Lower bounds on the randomized communication complexity of read-once functions. *Computational Complexity*, 19(2):153–181, 2010. doi:10.1007/s00037-010-0292-2.

[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.

[RS10] Alexander Razborov and Alexander Sherstov. The sign-rank of AC^0. *SIAM Journal on Computing*, 39(5):1833–1855, 2010. doi:10.1137/080744037.

[She12] Alexander Sherstov. The communication complexity of Gap Hamming Distance. *Theory of Computing*, 8(1):197–208, 2012. doi:10.4086/toc.2012.v008a008.

[Vid12] Thomas Vidick. A concentration inequality for the overlap of a vector on a large set, with application to the communication complexity of the Gap-Hamming-Distance problem. *Chicago Journal of Theoretical Computer Science*, 2012(1):1–12, 2012. doi:10.4086/cjtcs.2012.001.

[Vio] Statistical distance between uniform and biased coin. URL: https://cstheory.stackexchange.com/questions/36517/statistical-distance-between-uniform-and-biased-coin.

[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.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il