Electronic Colloquium on Computational Complexity, Report No. 88 (2015)

Exponential Separation of Communication and External Information

Anat Ganor[*] Gillat Kol[†] Ran Raz[‡]

Abstract

We show an exponential gap between communication complexity and external information complexity, by analyzing a communication task suggested as a candidate by Braverman [Bra13]. Previously, only a separation of communication complexity and *internal* information complexity was known [GKR14, GKR15].

More precisely, we obtain an explicit example of a search problem with external information complexity $\leq O(k)$, with respect to any input distribution, and distributional communication complexity $\geq 2^k$, with respect to some input distribution. In particular, this shows that a communication protocol cannot always be compressed to its external information. By a result of Braverman [Bra12], our gap is the largest possible.

Moreover, since the upper bound of $O(k)$ on the external information complexity of the problem is obtained with respect to *any* input distribution, our result implies an exponential gap between communication complexity and information complexity (both internal and external) in the *non-distributional* setting of Braverman [Bra12]. In this setting, no gap was previously known, even for internal information complexity.

1 Introduction

Communication complexity is a central model in complexity theory that has been extensively studied in numerous works. In the two player distributional model, each player gets an input, where the inputs are sampled from a joint distribution that is known to both players. The players' goal is to solve a communication task that depends on both inputs. The players

[*] Weizmann Institute of Science, Israel. Research supported by an Israel Science Foundation grant and by the I-CORE Program of the Planning and Budgeting Committee and the Israel Science Foundation.

[†] Institute for Advanced Study, Princeton, NJ. Research supported by the National Science Foundation grant No. CCF-1412958 and the Weizmann Institute of Science National Postdoctoral Award Program for Advancing Women in Science.

[‡] Weizmann Institute of Science, Israel, and the Institute for Advanced Study, Princeton, NJ. Research supported by the Israel Science Foundation grant No. 1402/14, by the I-CORE Program of the Planning and Budgeting Committee and the Israel Science Foundation, by a grant from the Simons Foundation, by the Fund for Math at IAS, and by the National Science Foundation grant No. CCF-1412958.

ISSN 1433-8092

can use both common and private random strings and are allowed to err with some small probability. The players communicate in rounds, where in each round one of the players sends a message to the other player. The communication complexity of a protocol is the total number of bits communicated by the two players. The communication complexity of a communication task is the minimal number of bits that the players need to communicate in order to solve the task with high probability, where the minimum is taken over all protocols. For excellent surveys on communication complexity see [KN97, LS09].

The information complexity model, first introduced by [CSWY01, BYJKS04, BBCR10], studies the amount of information that the players need to reveal about their inputs in order to solve a communication task. The model was motivated by fundamental information theoretical questions of compressing communication, as well as by fascinating relations to communication complexity, and in particular to the direct sum problem in communication complexity.

In this paper, we will be interested in both internal and external information complexity (a.k.a, internal and external information cost). Roughly speaking, the *internal information complexity* of a protocol is the number of information bits that the players learn about each other's input, when running the protocol. The *external information complexity* of a protocol is the number of information bits that an external observer, who watches the execution of the protocol, learns about the players' inputs. The (internal or external) information complexity of a communication task is the infimum of the (internal or external) information complexity of a protocol, where the infimum is over all protocols that solve the task with high probability. It is well know that for any protocol (and thus also for any communication task), the internal information complexity of the protocol is at most its external information complexity, which is at most its communication complexity.

Many recent works focused on the problem of compressing interactive communication protocols. Given a communication protocol with small (internal or external) information complexity, can the protocol be compressed so that the total number of bits communicated by the protocol is also small? There are several beautiful known results, showing how to compress communication protocols in several cases [BBCR10, BR11, Bra12, BMY14, RR15]. Most relevant to our work are the result by Barak, Braverman, Chen and Rao that shows how to compress any protocol with external information complexity k and communication complexity c, to a protocol with communication complexity $O(k \cdot \text{polylog}(c))$ [BBCR10], and the result by Braverman that shows how to compress any protocol with internal (or external) information complexity k to a protocol with communication complexity $2^{O(k)}$ [Bra12].

The above compression results leave open the question of whether any protocol can be compressed all the way down to its (internal or external) information. Specifically, can a protocol with (internal or external) information complexity k be compressed to a protocol with communication complexity $O(k)$? In [GKR14, GKR15] (see also the simplification by [RS15]), we showed an exponential gap between *internal* information complexity and communication complexity. However, prior to the current work, no gap was known between *external* information complexity and communication complexity (recall that the external information complexity is always at least as large as the internal information complexity).

1.1 Our Results

1.1.1 Separation of Communication and External Information

We give the first gap between external information complexity and communication complexity. This is done by proving a tight lower bound for the communication complexity of a communication task, suggested by Braverman as a candidate for separating information complexity and communication complexity [Bra13]. We view this task as a search problem and refer to it as the *hidden layers game*, parameterized by $k \in \mathbb{N}$.

While both the internal and external information complexities of the hidden layers game are $O(k)$, with respect to *every* input distribution, we prove that for some input distribution, any communication protocol solving the problem, with communication complexity at most 2^k, has success probability at most 2^{-k}. By the above mentioned compression protocol of Braverman [Bra12], our result gives the largest possible gap between external information complexity and distributional communication complexity.

Theorem 1 (**Communication Lower Bound**)**.** *There exists an input distribution η, such that every randomized protocol (with shared randomness) for the hidden layers game with parameter k, that has communication complexity at most 2^k, errs with probability at least $1 - 2^{-k}$ (over the input distribution η).*

Theorem 2 (**External Information Upper Bound**, [Bra13])**.** *There exists a zero-error randomized protocol for the hidden layers game with parameter k, such that for any input distribution, the protocol has external information cost $O(k)$.*

We note that the inputs to the hidden layers game are very long, namely, quadruple exponential in k. The protocol that achieves information complexity $O(k)$ has communication complexity triple exponential in k.

1.1.2 Separation in the Non-Distributional Setting

In [Bra12], Braverman defined internal and external information complexity in a non-distributional (a.k.a prior-free) setting, where there is no underlying input distribution. Roughly speaking, the internal (or external) information complexity in this setting is defined as the maximum over all possible input distributions of the internal (or external) information complexity over that distribution.

Since the upper bound of $O(k)$ on the external information complexity of the hidden layers game is obtained with respect to *every* input distribution, our result implies an exponential gap between (randomized) communication complexity and information complexity (both internal and external) also in the non-distributional setting. We note that in this setting, no gap was previously known, even for internal information complexity.

1.2 Techniques

Our result is ultimately proved by a reduction from the communication complexity problem of set disjointness. However, the reduction is non-standard, in the sense that it is protocol-dependant. Given a communication protocol, that supposedly solves the hidden layers game, we use this protocol to solve set disjointness. This is done by embedding the inputs for the set disjointness problem in the inputs for the hidden layers game. However, the embedding is given by a different function for each protocol.

In this context, it may be interesting to note that the methods that were used to prove lower bounds for the communication complexity of set disjointness (as well as most other general methods for proving lower bounds for communication complexity) yield the same lower bounds for information complexity, and hence are not strong enough to establish gaps between communication complexity and information complexity [Bra12, BW12, KLL$^+$12, FJK$^+$15].

2 The Hidden Layers Game

The hidden layers game can be viewed as a communication game between two parties, called Alice and Bob. The game is specified by a parameter $k \in \mathbb{N}$ (we assume that k is larger than some large enough constant). We set $c = 2^{2^{8k}}$, $\ell = 2^{4k}$ and $h = 2^{c\ell}$. The game is played on the 2^{4k}-ary tree $\mathcal{T}$ with $h+1$ layers, where the root is in layer 0 and the leaves are in layer h, with edges directed from the root to the leaves. Denote the vertex set of $\mathcal{T}$ by V.

Alice gets an input $x = (a, f)$, where $a \in \{0, \ldots, h-1\}$ is an even index and f is a set of edges: The set f contains exactly one edge, x_v, going out of the vertex v, for every vertex v in layer a of $\mathcal{T}$. Bob gets an input $y = (b, g)$, where $b \in \{0, \ldots, h-1\}$ is an odd index and g is a set of edges: The set g contains exactly one edge, y_v, going out of the vertex v, for every vertex v in layer b of $\mathcal{T}$. We denote by Ω_x the set of all possible inputs for Alice, and by Ω_y the set of all possible inputs for Bob. We refer to a binary string in $\{0,1\}^\ell$ as a "block". We will often think of a and b as sequences of c blocks, i.e., $a, b \in \{0,1\}^{c\ell} = (\{0,1\}^\ell)^c$.

Let $v \in V$ be a leaf of $\mathcal{T}$. Let $x = (a, f) \in \Omega_x$ and $y = (b, g) \in \Omega_y$. We say that v is *consistent with* x if the path from the root to v contains an edge from the set f. We say that v is *consistent with* y if the path from the root to v contains an edge from the set g.

On inputs (x, y), the players' goal is to output the same leaf v, such that v is consistent with both x and y.

2.1 The Hard Distribution μ

The distributions $\rho, \rho^{i,z}$. For every $i \in [c]$ and $z \in \{0,1\}^{\ell(i-1)}$, we define the distribution $\rho^{i,z}$ over pairs of indices $(a, b) \in \{0,1\}^{c\ell} \times \{0,1\}^{c\ell}$ as follows: Select $w, w' \in_R \{0,1\}^{\ell(c-(i-1))}$ (independently), such that w is even and w' is odd (i.e., the least significant bit of w is 0, and the least significant bit of w' is 1). Define $a = (z, w)$ (that is, the string z concatenated with the string w), and $b = (z, w')$.

We define the distribution ρ over pairs of indices $(a, b) \in \{0,1\}^{c\ell} \times \{0,1\}^{c\ell}$ as:

$$\rho(a,b) = \mathop{\mathbf{E}}_{i \in_R [c]} \mathop{\mathbf{E}}_{z \in_R \{0,1\}^{\ell(i-1)}} \left[\rho^{i,z}(a,b)\right].$$

The distributions $\mu, \mu^{i,z}$. For every $i \in [c]$ and $z \in \{0,1\}^{\ell(i-1)}$, we define the distribution $\mu^{i,z}$ over pairs of inputs $(x, y) \in \Omega_x \times \Omega_y$ as follows: Select (a, b) according to $\rho^{i,z}$. For every vertex v in layer a of $\mathcal{T}$, we choose, independently at random, an edge x_v going out of v. We define f to be the set of all these edges. For every vertex v in layer b of $\mathcal{T}$, we choose, independently at random, an edge y_v going out of v. We define g to be the set of all these edges. Let $x = (a, f)$ and $y = (b, g)$.

We define the distribution μ over pairs of inputs $(x, y) \in \Omega_x \times \Omega_y$ as:

$$\mu(x,y) = \mathop{\mathbf{E}}_{i \in_R [c]} \mathop{\mathbf{E}}_{z \in_R \{0,1\}^{\ell(i-1)}} \left[\mu^{i,z}(x,y)\right].$$

3 Overview of the Lower Bound Proof

In this section, we overview the proof of the lower bound for the communication complexity of the hidden layers game with parameter k. We fix the random strings for the protocol so that we have a deterministic protocol. The main part of the proof is devoted to showing Theorem 8, that states that if the protocol communicates at most 8^k bits, it errs with probability at least 1/4 on inputs sampled according to μ. Given Theorem 8, we use Yao's Minimax principle and the fact that the correctness of the outputs can be verified, to show that there exists a (possibly different) input distribution η, such that any protocol errs with probability close to 1 on inputs sampled according to η.

We next sketch the proof of Theorem 8. Assume for contradiction that π is a deterministic communication protocol for the hidden layers game with parameter k, that has communication complexity at most 8^k and success probability greater than 3/4. Let $\{\mathcal{R}^1, \ldots, \mathcal{R}^m\}$ be the rectangle partition of the protocol π, where $m \le 2^{8^k}$.

Fixing i, z. Lemma 11 shows that for most selections of $i \in [c]$ and $z \in \{0,1\}^{\ell(i-1)}$, if $(X, Y) = ((A, F), (B, G))$ is distributed according to $\mu^{i,z}$, then the following information property holds: The transcript $\pi(X, Y)$ gives very little information on A_i, B_i, the i^{th} blocks of the indices A, B. Therefore, A_i, B_i are almost uniformly distributed in almost all the rectangles $\mathcal{R}^t$. The proof of the lemma is similar to the proof of Lemma 11 in [GKR14].

For the rest of the proof we fix i, z such that the above information property holds and, in addition, the success probability of π on inputs distributed according to $\mu^{i,z}$ remains close to 3/4. Note that $\mu^{i,z}$ is a product distribution, thus X, Y are independent.

Unique answers. Observe that the protocol π may give many different answers on a rectangle $\mathcal{R}^t$, as the answer given by a player may depend on his input. Nevertheless, since $\mu^{i,z}$ is a product distribution, by taking the most common answer in each rectangle, we are able to obtain a protocol π' where the answer on each rectangle is unique, without compromising the success probability by much (see Lemma 14).

Fixing s. We next want to fix both A_i, B_i to the same value $s \in \{0,1\}^\ell$. Once this is done, the distribution of the inputs becomes $\mu^{i,z}|_{A_i=s,B_i=s} = \mu^{i+1,(z,s)}$, where (z,s) is the concatenation of z and s.

Recall that by Lemma 11, A_i, B_i are almost uniformly distributed in almost all the rectangles $\mathcal{R}^t$. We show that this implies that for at least a constant fraction of the selections of a block assignment $s \in \{0,1\}^\ell$, the success probability of π' on inputs with both A_i, B_i set to s is still larger than some positive constant (see Lemma 16). This means that for at least a constant fraction of the s's, the success probability of π' on the input distribution $\mu^{i+1,(z,s)}$ is larger than some positive constant.

Let v^t be the output given by both Alice and Bob when reaching the rectangle $\mathcal{R}^t$ in the protocol π'. Let $d \in \{0,1\}^{c\ell}$ be a layer of $\mathcal{T}$. We say that layer d *splits paths* if there exist v^t and $v^{t'}$ whose lowest common ancestor in V is in layer d. That is, the path from the root to v^t and the path from the root to $v^{t'}$ are the same up-to layer d, and contain different vertices in layer $d+1$. Observe that there are at most m layers d that split paths, as there are only m outputs v^t.

We fix A_i, B_i to the same value $s \in \{0,1\}^\ell$ that satisfies the followings: Let $\mathcal{L}$ be the set of layers $d \in \{0,1\}^{c\ell}$, whose prefix is (z,s). The first requirement is that none of the layers $d \in \mathcal{L}$ split paths. The second requirement is that the success probability of π' on the input distribution $\mu^{i+1,(z,s)}$ is larger than some positive constant. Such a value s exists, as a constant fraction of the s's satisfy the second requirement, and since at most m layers split paths (and m is much smaller than 2^ℓ).

Reduction from disjointness. We consider the following unique disjointness problem: Denote $K = 2^{4k}$ (the arity of the tree $\mathcal{T}$). Alice gets an input $S \subset [K^2]$ of size K. Bob gets an input $T \subset [K^2]$ of size K. Given an input (S,T), the players' goal is to distinguish a yes instance, where $|S \cap T| = 1$, from a no instance, where $|S \cap T| = 0$. It is well known that even for a small constant $\epsilon > 0$, protocols with advantage ϵ for the above unique disjointness problem, have communication complexity at least $\Omega(K)$ [BFS86, KS92, Raz92].

Recall that $v^1, \ldots, v^m$ are the possible outputs given by the protocol π'. We consider the set $\{p^1, \ldots, p^{m'}\}$ of all sub-paths that are restrictions to the layers in $\mathcal{L}$ of the path from the root to v^t, for some $t \in [m]$.

Given inputs S, T for the unique disjointness problem, the players construct inputs x, y for the hidden layers game. First, the players select independently at random $a, b \in \mathcal{L}$, such that a is even and b is odd. Then, using their shared random string, for every $t \in [m']$, the players choose $j^t \in_R [K^2]$. Given S, Alice constructs f by selecting for every v in layer a, an edge x_v going out of v as follows: If v is not contained in any of the sub-paths p^t, then x_v is randomly chosen among the edges going out of v. Otherwise, v is contained in some sub-path p^t, for $t \in [m']$. Since, by the first requirement of s, none of the layers in $\mathcal{L}$ split paths, p^t is unique. If $j^t \in S$, then x_v is set to be the edge going out of v that is contained in sub-path p^t. Otherwise, $j^t \notin S$, and x_v is randomly selected among all edges that are going out of v and are not contained in p^t. Given T, Bob constructs g in a similar manner, independently.

After obtaining the inputs x, y, the players run the protocol π' on x, y, and get an output v^t. Let $p^{t'}$ be the sub-path associated with v^t. The players check whether or not $j^{t'} \in S \cap T$, and answer “yes” if and only if $j^{t'} \in S \cap T$.

If (S, T) is a no instance, the protocol always outputs “no”. We prove that if (S, T) is a yes instance, the protocol outputs “yes” with probability greater than some positive constant. This is done by observing that for every yes instance (S, T), the inputs (x, y) are distributed according to $\mu^{i+1,(z,s)}$, and using the second requirement of s. Therefore, we got a protocol for unique disjointness, with communication $o(K)$ and constant advantage, a contradiction.

4 Preliminaries

4.1 Notation

Let P, Q be two distributions. We denote by $\mathbb{D}(P\|Q)$ the relative entropy between P and Q. We denote by $|P - Q|_1$ the ℓ_1 distance between P and Q. Let X, Y be two random variables. We denote by $\mathbf{H}(X)$ the Shannon entropy of X. We denote by $\mathbf{I}(X; Y)$ the mutual information between X and Y.

Let X be a random variable and E be an event. We denote by $\mathbf{P}_X$ the distribution of the random variables X. We denote by $\mathbf{P}_{X|E}$ the distribution of the random variable X conditioned on the event E.

4.2 Definitions

Definition 1 (**Internal Information Cost**)**.** *The* internal information cost *of a protocol π over random inputs (X, Y) that are drawn according to a joint distribution μ, is defined as*

$$\mathsf{IC}_\mu(\pi) = \mathbf{I}(X; \pi(X, Y)|Y) + \mathbf{I}(Y; \pi(X, Y)|X),$$

where $\pi(X, Y)$ is a random variable which is the transcript of the protocol π with respect to μ. That is, $\pi(X, Y)$ is the concatenation of all the messages exchanged during the execution of π.

Definition 2 (**External Information Cost**)**.** *The* external information cost *of a protocol π over random inputs (X, Y) that are drawn according to a joint distribution μ, is defined as*

$$\mathsf{Ext}_\mu(\pi) = \mathbf{I}((X, Y); \pi(X, Y)),$$

where $\pi(X, Y)$ is a random variable which is the transcript of the protocol π with respect to μ. That is, $\pi(X, Y)$ is the concatenation of all the messages exchanged during the execution of π.

4.3 Propositions

Proposition 3. *For any two random variables A, B,*

$$\mathbf{I}(A; B) = \mathop{\mathbf{E}}_{b \leftarrow B} \left[\mathbb{D}(\mathbf{P}_{A|B=b} \| \mathbf{P}_A) \right].$$

Proposition 4. *For any two random variables A, B,*

$$\mathop{\mathbf{E}}_{b \leftarrow B} \left[|\mathbf{P}_{A|B=b} - \mathbf{P}_A|_1 \right] \leq 2\sqrt{\mathbf{I}(A; B)}.$$

Proof. It holds that

$$\begin{aligned} &\left(\mathop{\mathbf{E}}_{b \leftarrow B} \left[|\mathbf{P}_{A|B=b} - \mathbf{P}_A|_1 \right] \right)^2 \leq \mathop{\mathbf{E}}_{b \leftarrow B} \left[\left(|\mathbf{P}_{A|B=b} - \mathbf{P}_A|_1 \right)^2 \right] \\ &\leq 4 \mathop{\mathbf{E}}_{b \leftarrow B} \left[\mathbb{D} \left(\mathbf{P}_{A|B=b} \parallel \mathbf{P}_A \right) \right] && \text{(by Pinsker's inequlity)} \\ &= 4 \cdot \mathbf{I}(A; B). && \text{(by Proposition 3)} \end{aligned}$$

□

Proposition 5. *Let A, B, C be random variables, such that A, B are independent and A, B are also independent given C. Then,*

$$\mathbf{I}(A, B; C) = \mathbf{I}(A; C) + \mathbf{I}(B; C).$$

Proof. It holds that

$$\begin{aligned} \mathbf{I}(A, B; C) &= \mathbf{H}(A, B) - \mathbf{H}(A, B|C) \\ &= (\mathbf{H}(A) + \mathbf{H}(B)) - (\mathbf{H}(A|C) + \mathbf{H}(B|C)) = \mathbf{I}(A; C) + \mathbf{I}(B; C). \end{aligned}$$

□

Proposition 6. *Let $\Omega \neq \phi$ be a finite set. Let U be a random variable uniformly distributed over Ω. Let E be any event with $\Pr[E] > 0$. Then,*

$$\log(|\Omega|) - \mathbf{H}(U|E) \leq \log(1/\Pr[E]).$$

Proof. It holds that

$$\begin{aligned}
&\log(|\Omega|) - \mathbf{H}(U|E) \\
&= \log(|\Omega|) + \sum_{u\in\Omega} \Pr[U = u|E] \cdot \log(\Pr[U = u|E]) \\
&\leq \log(|\Omega|) + \sum_{u\in\Omega} \Pr[U = u|E] \cdot \log\left(\frac{\Pr[U = u]}{\Pr[E]}\right) \\
&= \log(|\Omega|) + \sum_{u\in\Omega} \Pr[U = u|E] \cdot (\log(\Pr[U = u]) - \log(\Pr[E])) \\
&= -\sum_{u\in\Omega} \Pr[U = u|E] \cdot \log(\Pr[E]) = \log(1/\Pr[E]).
\end{aligned}$$

□

Proposition 7. *Let $L \in \mathbb{N}$. For $i \in [L]$, let $\alpha_i, \beta_i \in [0,1]$. Then,*

$$\mathop{\mathbf{E}}_{i\in_R[L]} [\alpha_i \cdot \beta_i] \geq \mathop{\mathbf{E}}_{i\in_R[L]} [\alpha_i + \beta_i] - 1.$$

Proof. We assume, without loss of generality, that for every $i \in [L]$, it holds that $\alpha_i \geq \beta_i$, otherwise we switch their values. Assume, without loss of generality, that for every $i \in [L]$, either $\alpha_i = 1$ or $\beta_i = 0$. Otherwise, by increasing α_i by ϵ and decreasing β_i by ϵ, the left hand side may only decrease, while the right hand side is not effected. Therefore,

$$\mathop{\mathbf{E}}_{i\in_R[L]} [\alpha_i \cdot \beta_i] = \mathop{\mathbf{E}}_{i\in_R[L]} [\beta_i] = \mathop{\mathbf{E}}_{i\in_R[L]} [\beta_i + \alpha_i] - \mathop{\mathbf{E}}_{i\in_R[L]} [\alpha_i] \geq \mathop{\mathbf{E}}_{i\in_R[L]} [\beta_i + \alpha_i] - 1.$$

□

5 Communication Lower Bound

In this section, we prove Theorem 1. The proof follows easily from Theorem 8 below.

Theorem 8. *Every randomized protocol (with shared randomness) for the hidden layers game with parameter k, that has communication complexity at most 8^k, errs with probability at least $1/4$ over the input distribution μ.*

Proof of Theorem 1 given Theorem 8. Assume for contradiction that for every input distribution η, there exists a randomized protocol for the hidden layers game with parameter k, that has communication complexity at most 2^k, and has success probability at least 2^{-k} (over the input distribution η). Then, by Yao's Minimax principle, there exists a randomized protocol τ that has communication complexity at most 2^k, and has success probability at least 2^{-k} on every input pair.

Note that by exchanging $O(k)$ bits, the players can check whether their answers are correct, with error probability 2^{-100k}: They check that their answers are equal (by exchanging

$O(k)$ hash values), and exchange two additional bits to make sure that the output (that is now assumed to be the same for both players) is consistent with both x and y.

Consider the protocol τ' that on a given input (x, y), runs τ and checks the answer $100 \cdot 2^k$ times, and outputs a correct answer (if such an answer is found). The communication complexity of τ' is at most $(100 \cdot 2^k) \cdot (2^k + O(k)) \leq 8^k$ (for large enough k). The success probability of τ' is greater than $3/4$. In particular, τ' has communication complexity at most 8^k and success probability greater than $3/4$, over the input distribution μ. This contradicts Theorem 8. □

The rest of the section is devoted to proving Theorem 8. Fix $\lambda = 3/4$ (the proof works for any constant $1/\sqrt{2} < \lambda \leq 1$). Assume that π is a deterministic communication protocol for the hidden layers game with parameter k, that has communication complexity at most 8^k. The section is devoted to showing that π has success probability at most λ, when the inputs are selected according to the distribution μ. Observe that this also implies that every probabilistic protocol has success probability at most λ, since a probabilistic protocol is a distribution over deterministic protocols.

Assume for contradiction that π has success probability more than λ.

5.1 Notation

Let $\{\mathcal{R}^1, \ldots, \mathcal{R}^m\}$ be the rectangle partition induced by the protocol π, where $\mathcal{R}^t = \mathcal{X}^t \times \mathcal{Y}^t$ for $\mathcal{X}^t \subseteq \Omega_x$ and $\mathcal{Y}^t \subseteq \Omega_y$, and $m \leq 2^{8^k}$. We assume for simplicity and without loss of generality that $m = 2^{8^k}$ (as empty rectangles can always be added).

For every $i \in [c]$ and $z \in \{0,1\}^{\ell(i-1)}$, let $(X^{i,z}, Y^{i,z})$ be a pair of random variables distributed according to $\mu^{i,z}$, where $X^{i,z} = (A^{i,z}, F^{i,z})$ and $Y^{i,z} = (B^{i,z}, G^{i,z})$. Note that the pair $(A^{i,z}, B^{i,z})$ is distributed according to $\rho^{i,z}$.

Let $i \in [c]$. Let ψ be either a random variable taking values in $(\{0,1\}^\ell)^c$ or an element in $(\{0,1\}^\ell)^c$. Define $\psi_i \in \{0,1\}^\ell$ to be the i^{th} block of ψ. Define $\psi_{<i} = (\psi_1, \ldots, \psi_{i-1}) \in \{0,1\}^{\ell(i-1)}$ to be the first $i-1$ blocks of ψ.

The following two properties will be important for the rest of the proof:

Proposition 9. *For every $i \in [c]$ and $z \in \{0,1\}^{\ell(i-1)}$, it holds that $\rho^{i,z}$ and $\mu^{i,z}$ are product distributions. Thus, $X^{i,z}$ and $Y^{i,z}$ are independent.*

Proposition 10. *For every $i < i' \in [c]$, $z \in \{0,1\}^{\ell(i-1)}$, and $z' \in \{0,1\}^{\ell(i'-1)}$, such that z is a prefix of z', it holds that $\rho^{i',z'} = (\rho^{i,z} | A^{i,z}_{<i'} = z', B^{i,z}_{<i'} = z')$. That is, $\rho^{i',z'}$ is $\rho^{i,z}$ conditioned on the event $A^{i,z}_{<i'} = z', B^{i,z}_{<i'} = z'$. Similarly, $\mu^{i',z'} = (\mu^{i,z} | A^{i,z}_{<i'} = z', B^{i,z}_{<i'} = z')$.*

The proofs follow immediately from the definitions.

5.2 Bounding the Information on the i^{th} Block

Let $\pi(X^{i,z}, Y^{i,z})$ be a random variable representing the transcript of π when it is run on $(X^{i,z}, Y^{i,z})$. We associate a transcript with the rectangle $\mathcal{R}^t$ reached for this transcript. That is, we think of $\pi(X^{i,z}, Y^{i,z})$ as taking values in $[m]$.

Lemma 11. *It holds that*

$$\mathop{\mathbf{E}}_{i \in_R [c]} \mathop{\mathbf{E}}_{z \in_R \{0,1\}^{\ell(i-1)}} \left[\mathbf{I}(A_i^{i,z}; \pi(X^{i,z}, Y^{i,z})) \right] \leq \frac{m}{c}.$$

Similarly,

$$\mathop{\mathbf{E}}_{i \in_R [c]} \mathop{\mathbf{E}}_{z \in_R \{0,1\}^{\ell(i-1)}} \left[\mathbf{I}(B_i^{i,z}; \pi(X^{i,z}, Y^{i,z})) \right] \leq \frac{m}{c}.$$

Proof. We prove the first inequality, the second is similar. For every $i \in [c]$ and $z \in \{0,1\}^{\ell(i-1)}$, it holds that

$$\begin{aligned}
&\mathbf{I}(A_i^{i,z}; \pi(X^{i,z}, Y^{i,z})) && (1)\\
&= \mathbf{H}(A_i^{i,z}) - \mathbf{H}(A_i^{i,z} | \pi(X^{i,z}, Y^{i,z})) \\
&= \ell - \mathbf{H}(A_i^{i,z} | \pi(X^{i,z}, Y^{i,z})) && (A_i^{i,z} \text{ is uniformly distributed})\\
&= \sum_{t \in [m]} \Pr[\pi(X^{i,z}, Y^{i,z}) = t] \cdot (\ell - \mathbf{H}(A_i^{i,z} | \pi(X^{i,z}, Y^{i,z}) = t)) \\
&= \sum_{t \in [m]} \Pr[(X^{i,z}, Y^{i,z}) \in \mathcal{R}^t] \cdot (\ell - \mathbf{H}(A_i^{i,z} | (X^{i,z}, Y^{i,z}) \in \mathcal{R}^t)) \\
&= \sum_{t \in [m]} \Pr[X^{i,z} \in \mathcal{X}^t] \cdot \Pr[Y^{i,z} \in \mathcal{Y}^t] \cdot (\ell - \mathbf{H}(A_i^{i,z} | X^{i,z} \in \mathcal{X}^t)) && (X^{i,z}, Y^{i,z} \text{ are independent})\\
&\leq \sum_{t \in [m]} \Pr[X^{i,z} \in \mathcal{X}^t] \cdot (\ell - \mathbf{H}(A_i^{i,z} | X^{i,z} \in \mathcal{X}^t)). && (\text{as } \ell - \mathbf{H}(A_i^{i,z} | X^{i,z} \in \mathcal{X}^t) \geq 0)
\end{aligned}$$

Recall the distribution $\rho^{1,\phi}$, that is, the distribution $\rho^{i,z}$ for $i = 1$ and z that is the empty string (string of length 0). The distribution $\rho^{1,\phi}$ is uniform over $\{0,1\}^{c\ell} \times \{0,1\}^{c\ell}$. In the proof of this lemma we will denote $\rho^{\mathcal{U}} = \rho^{1,\phi}$, $\mu^{\mathcal{U}} = \mu^{1,\phi}$, $A^{\mathcal{U}} = A^{1,\phi}$, $B^{\mathcal{U}} = B^{1,\phi}$, $X^{\mathcal{U}} = X^{1,\phi}$, $Y^{\mathcal{U}} = Y^{1,\phi}$. By Proposition 10, $\rho^{i,z} = (\rho^{\mathcal{U}} | A_{<i}^{\mathcal{U}} = z, B_{<i}^{\mathcal{U}} = z)$. That is, $\rho^{i,z}$ is $\rho^{\mathcal{U}}$ conditioned on the event $A_{<i}^{\mathcal{U}} = z, B_{<i}^{\mathcal{U}} = z$. Similarly, $\mu^{i,z} = (\mu^{\mathcal{U}} | A_{<i}^{\mathcal{U}} = z, B_{<i}^{\mathcal{U}} = z)$.

We first consider the term $\Pr[X^{i,z} \in \mathcal{X}^t] \cdot \mathbf{H}(A_i^{i,z} | X^{i,z} \in \mathcal{X}^t)$ in Equation (1). For every $t \in [m]$,

$$\begin{aligned}
&\Pr[X^{i,z} \in \mathcal{X}^t] \cdot \mathbf{H}(A_i^{i,z} | X^{i,z} \in \mathcal{X}^t) \\
&= \Pr[X^{\mathcal{U}} \in \mathcal{X}^t | A_{<i}^{\mathcal{U}} = z, B_{<i}^{\mathcal{U}} = z] \cdot \mathbf{H}(A_i^{\mathcal{U}} | X^{\mathcal{U}} \in \mathcal{X}^t, A_{<i}^{\mathcal{U}} = z, B_{<i}^{\mathcal{U}} = z) \\
&= \Pr[X^{\mathcal{U}} \in \mathcal{X}^t | A_{<i}^{\mathcal{U}} = z] \cdot \mathbf{H}(A_i^{\mathcal{U}} | X^{\mathcal{U}} \in \mathcal{X}^t, A_{<i}^{\mathcal{U}} = z) && (\text{since } X^{\mathcal{U}}, Y^{\mathcal{U}} \text{ are independent})\\
&= \frac{\Pr[X^{\mathcal{U}} \in \mathcal{X}^t]}{\Pr[A_{<i}^{\mathcal{U}} = z]} \cdot \Pr[A_{<i}^{\mathcal{U}} = z | X^{\mathcal{U}} \in \mathcal{X}^t] \cdot \mathbf{H}(A_i^{\mathcal{U}} | X^{\mathcal{U}} \in \mathcal{X}^t, A_{<i}^{\mathcal{U}} = z) && (\text{Bayes' formula})
\end{aligned}$$

By taking expectation over z,

$$\begin{aligned}
&\mathop{\mathbf{E}}_{z\in_R\{0,1\}^{\ell(i-1)}}\left[\Pr[X^{i,z}\in\mathcal{X}^t]\cdot\mathbf{H}(A_i^{i,z}|X^{i,z}\in\mathcal{X}^t)\right]\\
&=\sum_{z\in\{0,1\}^{\ell(i-1)}}\Pr[X^{\mathcal{U}}\in\mathcal{X}^t]\cdot\Pr[A^{\mathcal{U}}_{<i}=z|X^{\mathcal{U}}\in\mathcal{X}^t]\cdot\mathbf{H}(A_i^{\mathcal{U}}|X^{\mathcal{U}}\in\mathcal{X}^t,A^{\mathcal{U}}_{<i}=z)\\
&=\Pr[X^{\mathcal{U}}\in\mathcal{X}^t]\cdot\sum_{z\in\{0,1\}^{\ell(i-1)}}\Pr[A^{\mathcal{U}}_{<i}=z|X^{\mathcal{U}}\in\mathcal{X}^t]\cdot\mathbf{H}(A_i^{\mathcal{U}}|A^{\mathcal{U}}_{<i}=z,X^{\mathcal{U}}\in\mathcal{X}^t)\\
&=\Pr[X^{\mathcal{U}}\in\mathcal{X}^t]\cdot\mathbf{H}(A_i^{\mathcal{U}}|A^{\mathcal{U}}_{<i},X^{\mathcal{U}}\in\mathcal{X}^t).
\end{aligned}$$

By taking expectation over i,

$$\begin{aligned}
&\mathop{\mathbf{E}}_{i\in_R[c]}\mathop{\mathbf{E}}_{z\in_R\{0,1\}^{\ell(i-1)}}\left[\Pr[X^{i,z}\in\mathcal{X}^t]\cdot\mathbf{H}(A_i^{i,z}|X^{i,z}\in\mathcal{X}^t)\right] && (2)\\
&=\Pr[X^{\mathcal{U}}\in\mathcal{X}^t]\cdot\mathop{\mathbf{E}}_{i\in_R[c]}\left[\mathbf{H}(A_i^{\mathcal{U}}|A^{\mathcal{U}}_{<i},X^{\mathcal{U}}\in\mathcal{X}^t)\right]\\
&=\frac{1}{c}\Pr[X^{\mathcal{U}}\in\mathcal{X}^t]\cdot\mathbf{H}(A^{\mathcal{U}}|X^{\mathcal{U}}\in\mathcal{X}^t). && \text{(by the chain rule for entropy)}
\end{aligned}$$

We next consider the term $\Pr[X^{i,z}\in\mathcal{X}^t]\cdot\ell$ in Equation (1). For every $i\in[c]$ and $t\in[m]$,

$$\mathop{\mathbf{E}}_{z\in_R\{0,1\}^{\ell(i-1)}}\left[\Pr[X^{i,z}\in\mathcal{X}^t]\cdot\ell\right]=\Pr\left[X^{\mathcal{U}}\in\mathcal{X}^t\right]\cdot\ell. \qquad (3)$$

Substituting (2) and (3) in (1), and taking expectation over i and z, we get

$$\begin{aligned}
&\mathop{\mathbf{E}}_{i\in_R[c]}\mathop{\mathbf{E}}_{z\in_R\{0,1\}^{\ell(i-1)}}\left[\mathbf{I}(A_i^{i,z};\pi(X^{i,z},Y^{i,z}))\right]\\
&\le\frac{1}{c}\sum_{t\in[m]}\Pr[X^{\mathcal{U}}\in\mathcal{X}^t]\cdot(c\ell-\mathbf{H}(A^{\mathcal{U}}|X^{\mathcal{U}}\in\mathcal{X}^t))\\
&\le\frac{1}{c}\sum_{t\in[m]}\Pr[X^{\mathcal{U}}\in\mathcal{X}^t]\cdot\log(1/\Pr[X^{\mathcal{U}}\in\mathcal{X}^t]) && \text{(by Proposition 6)}\\
&\le m/c. && \text{(as }-p\log(p)<1\text{ for }p\in[0,1])
\end{aligned}$$

□

5.3 Fixing i, z

Fix $i\in[c-1]$ and $z\in\{0,1\}^{\ell(i-1)}$ such that the following two properties hold:

1. The protocol π has success probability at least $\lambda-2/m$, when the inputs are selected according to the distribution $\mu^{i,z}$.

2. $\mathbf{I}(A_i^{i,z};\pi(X^{i,z},Y^{i,z}))\le\frac{m^2}{c}$ and $\mathbf{I}(B_i^{i,z};\pi(X^{i,z},Y^{i,z}))\le\frac{m^2}{c}$.

By Lemma 11 and Markov's inequality, the probability that $i\in_R[c]$ and $z\in_R\{0,1\}^{\ell(i-1)}$ do not satisfy the second item is at most $2/m$. The probability for $i=c$ is $1/c$. If there were no

$i \in [c-1]$ and $z \in \{0,1\}^{\ell(i-1)}$ satisfying both items, then the success probability of π would have been at most $(2/m + 1/c) \cdot 1 + (1 - (2/m + 1/c)) \cdot (\lambda - 2/m) < \lambda$, a contradiction. Therefore, there exist such i and z that satisfy both of the above items.

Let $\mathcal{U}_\ell$ be the uniform distribution over $\{0,1\}^\ell$.

Claim 12. *It holds that*

$$\mathop{\mathbf{E}}_{t \leftarrow \pi(X^{i,z}, Y^{i,z})} \left[|\mathbf{P}_{A_i^{i,z} | X^{i,z} \in \mathcal{X}^t} - \mathcal{U}_\ell|_1 \right] \leq 2m/\sqrt{c},$$

$$\mathop{\mathbf{E}}_{t \leftarrow \pi(X^{i,z}, Y^{i,z})} \left[|\mathbf{P}_{B_i^{i,z} | Y^{i,z} \in \mathcal{Y}^t} - \mathcal{U}_\ell|_1 \right] \leq 2m/\sqrt{c}.$$

Proof. We prove the first inequality, the second is similar.

$$\begin{aligned}
&\mathop{\mathbf{E}}_{t \leftarrow \pi(X^{i,z}, Y^{i,z})} \left[|\mathbf{P}_{A_i^{i,z} | X^{i,z} \in \mathcal{X}^t} - \mathcal{U}_\ell|_1 \right] \\
&= \mathop{\mathbf{E}}_{t \leftarrow \pi(X^{i,z}, Y^{i,z})} \left[|\mathbf{P}_{A_i^{i,z} | X^{i,z} \in \mathcal{X}^t} - \mathbf{P}_{A_i^{i,z}}|_1 \right] \\
&= \mathop{\mathbf{E}}_{t \leftarrow \pi(X^{i,z}, Y^{i,z})} \left[|\mathbf{P}_{A_i^{i,z} | \pi(X^{i,z}, Y^{i,z}) = t} - \mathbf{P}_{A_i^{i,z}}|_1 \right] && (X^{i,z}, Y^{i,z} \text{ are independent}) \\
&\leq 2\sqrt{\mathbf{I}(A_i^{i,z}; \pi(X^{i,z}, Y^{i,z}))} && \text{(by Proposition 4)} \\
&\leq 2m/\sqrt{c}. && \text{(by the second proprety of } i, z)
\end{aligned}$$

□

Claim 13. *It holds that*

$$\mathop{\mathbf{E}}_{s \leftarrow \mathcal{U}_\ell} \left[\left| \mathbf{P}_{\pi(X^{i,z}, Y^{i,z})} - \mathbf{P}_{\pi(X^{i+1,(z,s)}, Y^{i+1,(z,s)})} \right|_1 \right] \leq 2^{\ell+2} m/\sqrt{c}.$$

Proof. It holds that

$$\begin{aligned}
&\mathop{\mathbf{E}}_{s \leftarrow \mathcal{U}_\ell} \left[\left| \mathbf{P}_{\pi(X^{i,z}, Y^{i,z})} - \mathbf{P}_{\pi(X^{i+1,(z,s)}, Y^{i+1,(z,s)})} \right|_1 \right] \\
&= \mathop{\mathbf{E}}_{s \leftarrow \mathcal{U}_\ell} \left[\left| \mathbf{P}_{\pi(X^{i,z}, Y^{i,z})} - \mathbf{P}_{\pi(X^{i,z}, Y^{i,z}) | A_i^{i,z} = s, B_i^{i,z} = s} \right|_1 \right] && \text{(by Proposition 10)} \\
&\leq \sum_{s' \in \{0,1\}^\ell} \mathop{\mathbf{E}}_{s \leftarrow \mathcal{U}_\ell} \left[\left| \mathbf{P}_{\pi(X^{i,z}, Y^{i,z}) | A_i^{i,z} = s, B_i^{i,z} = s'} - \mathbf{P}_{\pi(X^{i,z}, Y^{i,z})} \right|_1 \right] \\
&= 2^\ell \cdot \mathop{\mathbf{E}}_{s, s' \leftarrow \mathcal{U}_\ell} \left[\left| \mathbf{P}_{\pi(X^{i,z}, Y^{i,z}) | A_i^{i,z} = s, B_i^{i,z} = s'} - \mathbf{P}_{\pi(X^{i,z}, Y^{i,z})} \right|_1 \right] \\
&\leq 2^{\ell+1} \sqrt{\mathbf{I}\left((A_i^{i,z}, B_i^{i,z}); \pi(X^{i,z}, Y^{i,z}) \right)} && \text{(by Proposition 4)} \\
&= 2^{\ell+1} \sqrt{\mathbf{I}\left(A_i^{i,z}; \pi(X^{i,z}, Y^{i,z}) \right) + \mathbf{I}\left(B_i^{i,z}; \pi(X^{i,z}, Y^{i,z}) \right)} && \text{(by Proposition 5)} \\
&\leq 2^{\ell+2} m/\sqrt{c}. && \text{(by the second proprety of } i, z)
\end{aligned}$$

□

5.4 The Unique Answer Protocol π'

5.4.1 A General Conversion to a Unique Answer Protocol

Lemma 14. *Let π be a deterministic communication protocol. At the beginning of the protocol, Alice holds an input in the set Ω_x and Bob holds an input in the set Ω_y. When the protocol ends, each player returns an output from the output set Θ. Let $\{\mathcal{R}^1, \ldots, \mathcal{R}^m\}$ be the rectangle partition induced by the protocol π, where $\mathcal{R}^t = \mathcal{X}^t \times \mathcal{Y}^t$ for $\mathcal{X}^t \subseteq \Omega_x$ and $\mathcal{Y}^t \subseteq \Omega_y$.*

Let γ be a product distribution over $\Omega_x \times \Omega_y$. Let $\lambda \geq 0$. Assume that with probability at least λ, over the selection of inputs according to γ, Alice and Bob output the same answer in the protocol π.

Then, there exists a deterministic communication protocol π', that on every pair of inputs exchanges the same messages as π (and, in particular, has the same communication complexity and rectangle partition as π), such that the followings hold:

1. *For every $t \in [m]$, there exists $\theta \in \Theta$ such that for every $(x, y) \in \mathcal{R}^t$, both Alice and Bob output θ in the protocol π'.*

2. *With probability at least λ^2, over the selection of inputs according to γ, the outputs given by both Alice and Bob in the protocol π' are the same as their outputs in the protocol π.*

Proof. Let $t \in [m]$. Consider the outputs given by Alice and Bob in the rectangle $\mathcal{R}^t$ by the protocol π. For $\theta \in \Theta$, let $p^t(\theta)$ be the probability that both Alice and Bob answer by the same answer θ, in the rectangle $\mathcal{R}^t$ by the protocol π, where the probability is over the distribution γ (conditioned on reaching the rectangle $\mathcal{R}^t$). Let θ^t be an output with maximal $p^t(\theta)$.

Let π' be the protocol that on every pair of inputs exchanges the same messages as π. For every $t \in [m]$, when the rectangle $\mathcal{R}^t$ is reached by π', both Alice and Bob output θ^t. The protocol π' clearly satisfies Property (1). We next show that π' satisfies Property (2).

Claim 15. *Let $t \in [m]$. Assume that in the protocol π, when reaching the rectangle $\mathcal{R}^t$, Alice and Bob give the same answer with probability δ, where the probability is over the distribution γ.*

Then, with probability at least δ^2, the outputs given by both Alice and Bob in the rectangle $\mathcal{R}^t$ by the protocol π' are the same as their outputs by the protocol π.

Proof. For $\theta \in \Theta$, let α_θ be the probability that Alice outputs θ in the rectangle $\mathcal{R}^t$ by the protocol π. Let β_θ be the probability that Bob outputs θ in the rectangle $\mathcal{R}^t$ by the protocol π. It holds that $\sum_{\theta\in\Theta} \alpha_\theta = \sum_{\theta\in\Theta} \beta_\theta = 1$. Since γ is a product distribution, we also have that $\sum_{\theta\in\Theta} \alpha_\theta \cdot \beta_\theta = \delta$. By the inequality of arithmetic and geometric means, for every $\alpha_\theta, \beta_\theta$ it holds that $\frac{\alpha_\theta+\beta_\theta}{2} \geq \sqrt{\alpha_\theta \cdot \beta_\theta}$. Summing over Θ, we get $1 \geq \sum_{\theta\in\Theta} \sqrt{\alpha_\theta \cdot \beta_\theta}$. It holds that

$$\delta = \sum_{\theta\in\Theta} \alpha_\theta \cdot \beta_\theta \leq \max_{\theta\in\Theta} \left\{\sqrt{\alpha_\theta \cdot \beta_\theta}\right\} \cdot \sum_{\theta\in\Theta} \sqrt{\alpha_\theta \cdot \beta_\theta} \leq \max_{\theta\in\Theta} \left\{\sqrt{\alpha_\theta \cdot \beta_\theta}\right\} = \sqrt{p^t(\theta^t)}.$$

Observe that the probability that the outputs given by both Alice and Bob in the rectangle $\mathcal{R}^t$ by the protocol π' are the same as their outputs by the protocol π is exactly $p^t(\theta^t)$. By the last equation, $p^t(\theta^t) \geq \delta^2$. □

For $t \in [m]$, let δ^t be the probability that Alice and Bob give the same answer in the rectangle $\mathcal{R}^t$, in the protocol π, as defined in Claim 15. Let γ^t be the probability that the protocol π reaches the rectangle $\mathcal{R}^t$, where the probability is over the distribution γ.

Since we assume that with probability at least λ, Alice and Bob output the same answer in the protocol π, it holds that $\sum_{t\in[m]} \gamma^t \cdot \delta^t \geq \lambda$. By Claim 15, the probability that the outputs given by both Alice and Bob by the protocol π' are the same as their outputs by the protocol π is at least $\sum_{t\in[m]} \gamma^t \cdot (\delta^t)^2$. It holds that $\sum_{t\in[m]} \gamma^t \cdot (\delta^t)^2 \geq (\sum_{t\in[m]} \gamma^t \cdot \delta^t)^2 \geq \lambda^2$. □

5.4.2 The Unique Answer Protocol π'

We assume, without loss of generality, that in the protocol π, on every input (x, y), Alice outputs a leaf $v \in V$ that is consistent with x, and Bob outputs a leaf $v \in V$ that is consistent with y. The reason is that if Alice outputs a leaf that is inconsistent with x, the players fail. Therefore, by outputting a leaf v that is consistent with x instead, the success probability can only increase. Since the players' goal is to output the same leaf v, such that v is consistent with both x and y, they succeed if and only if they output the same leaf.

Recall that $\mu^{i,z}$ is a product distribution. Therefore, we can apply Lemma 14 to the protocol π and the distribution $\mu^{i,z}$, and get a protocol π' satisfying the following properties:

1. The protocol π' has success probability at least $(\lambda - 2/m)^2 \geq \lambda^2 - 4/m$, when the inputs are selected according to the distribution $\mu^{i,z}$. This property is implied by Properties (1) and (2) of Lemma 14, and the fact that the players succeed in π if and only if they output the same leaf.

2. The rectangle partition induced by π' is $\{\mathcal{R}^1, \ldots, \mathcal{R}^m\}$ (the same as the partition induced by π). Furthermore, for every $t \in [m]$, there exists a leaf $v^t \in V$ such that for every $(x, y) \in \mathcal{R}^t$, both Alice and Bob output v^t in the protocol π'.

5.5 Fixing the i^{th} Block

Recall that we fixed i and z.

5.5.1 Fixing the i^{th} Block while Keeping π' Success Probability

Lemma 16. *There exist more than m values $s \in \{0,1\}^\ell$, such that the protocol π' has success probability at least $2\lambda^2 - 1 - o_k(1)$, when the inputs are selected according to the* *distribution $\mu^{i+1,(z,s)}$ (where (z,s) is the string with i blocks obtained by concatenating the string z with the string s).*

Proof. Recall that by the second property of the protocol π' in Section 5.4.2, the rectangle partition induced by π' is the same as the rectangle partition induced by π. In particular, the probability of reaching each rectangle when running π' is the same as the probability of reaching that rectangle when running π (under any input distribution). Fix $t \in [m]$. Denote by λ^t the success probability of the protocol π' when it reaches the rectangle $\mathcal{R}^t$, where the probability is over the distribution $\mu^{i,z}$. Let $\hat{\mathcal{X}}^t \subseteq \mathcal{X}^t$ be the set of inputs x for Alice, such that v^t is consistent with x. Let $\hat{\mathcal{Y}}^t \subseteq \mathcal{Y}^t$ be the set of inputs y for Bob, such that v^t is consistent with y.

Let $s \in \{0,1\}^\ell$. Let

$$\begin{aligned}\alpha^{t,s} &= \Pr\left[X^{i+1,(z,s)} \in \hat{\mathcal{X}}^t \mid X^{i+1,(z,s)} \in \mathcal{X}^t\right] \\ &= \Pr\left[X^{i,z} \in \hat{\mathcal{X}}^t \mid X^{i,z} \in \mathcal{X}^t, A_i^{i,z} = s, B_i^{i,z} = s\right] && \text{(by Proposition 10)} \\ &= \Pr\left[X^{i,z} \in \hat{\mathcal{X}}^t \mid X^{i,z} \in \mathcal{X}^t, A_i^{i,z} = s\right]. && (X^{i,z}, Y^{i,z} \text{ are independent})\end{aligned}$$

Similarly, let

$$\begin{aligned}\beta^{t,s} &= \Pr\left[Y^{i+1,(z,s)} \in \hat{\mathcal{Y}}^t \mid Y^{i+1,(z,s)} \in \mathcal{Y}^t\right] \\ &= \Pr\left[Y^{i,z} \in \hat{\mathcal{Y}}^t \mid Y^{i,z} \in \mathcal{Y}^t, B_i^{i,z} = s\right].\end{aligned}$$

Observe that for any s and t, the term $\alpha^{t,s} \cdot \beta^{t,s}$ is the success probability of π', on the rectangle $\mathcal{R}^t$, when the inputs are selected according to the distribution $\mu^{i+1,(z,s)}$ (as $\mu^{i+1,(z,s)}$ is a product distribution).

It holds that

$$\begin{aligned}&\Pr\left[X^{i,z} \in \hat{\mathcal{X}}^t \mid X^{i,z} \in \mathcal{X}^t\right] \\ &= \sum_{s \in \{0,1\}^\ell} \Pr[A_i^{i,z} = s \mid X^{i,z} \in \mathcal{X}^t] \cdot \Pr\left[X^{i,z} \in \hat{\mathcal{X}}^t \mid X^{i,z} \in \mathcal{X}^t, A_i^{i,z} = s\right] \\ &= \sum_{s \in \{0,1\}^\ell} \Pr[A_i^{i,z} = s \mid X^{i,z} \in \mathcal{X}^t] \cdot \alpha^{t,s} \\ &= \mathop{\mathbf{E}}_{s \leftarrow (A_i^{i,z} \mid X^{i,z} \in \mathcal{X}^t)} [\alpha^{t,s}].\end{aligned}$$

Similarly,

$$\Pr\left[Y^{i,z} \in \hat{\mathcal{Y}}^t \mid Y^{i,z} \in \mathcal{Y}^t\right] = \mathop{\mathbf{E}}_{s \leftarrow (B_i^{i,z} \mid Y^{i,z} \in \mathcal{Y}^t)} [\beta^{t,s}].$$

Therefore,

$$\mathop{\mathbf{E}}_{s \leftarrow (A_i^{i,z} \mid X^{i,z} \in \mathcal{X}^t)} \left[\alpha^{t,s}\right] \cdot \mathop{\mathbf{E}}_{s \leftarrow (B_i^{i,z} \mid Y^{i,z} \in \mathcal{Y}^t)} \left[\beta^{t,s}\right] = \lambda^t.$$

By the inequality of arithmetic and geometric means, it holds that

$$\mathop{\mathbf{E}}_{s\leftarrow(A_i^{i,z}|X^{i,z}\in\mathcal{X}^t)}\left[\alpha^{t,s}\right]+\mathop{\mathbf{E}}_{s\leftarrow(B_i^{i,z}|Y^{i,z}\in\mathcal{Y}^t)}\left[\beta^{t,s}\right]\geq 2\sqrt{\lambda^t}.$$

Let $\mathcal{U}_\ell$ be the uniform distribution over $\{0,1\}^\ell$. Denote $\Delta^t=|\mathbf{P}_{A_i^{i,z}|X^{i,z}\in\mathcal{X}^t}-\mathcal{U}_\ell|_1+|\mathbf{P}_{B_i^{i,z}|Y^{i,z}\in\mathcal{Y}^t}-\mathcal{U}_\ell|_1$. Then, by the triangle inequlity,

$$\mathop{\mathbf{E}}_{s\leftarrow\mathcal{U}_\ell}\left[\alpha^{t,s}\right]+\mathop{\mathbf{E}}_{s\leftarrow\mathcal{U}_\ell}\left[\beta^{t,s}\right]\geq 2\sqrt{\lambda^t}-\Delta^t.$$

By Proposition 7,

$$\mathop{\mathbf{E}}_{s\leftarrow\mathcal{U}_\ell}\left[\alpha^{t,s}\cdot\beta^{t,s}\right]\geq 2\sqrt{\lambda^t}-\Delta^t-1.$$

We take expectation over t and get

$$\mathop{\mathbf{E}}_{t\leftarrow\pi(X^{i,z},Y^{i,z})}\mathop{\mathbf{E}}_{s\leftarrow\mathcal{U}_\ell}\left[\alpha^{t,s}\cdot\beta^{t,s}\right]\geq\mathop{\mathbf{E}}_{t\leftarrow\pi(X^{i,z},Y^{i,z})}\left[2\sqrt{\lambda^t}-\Delta^t\right]-1. \tag{4}$$

We first consider the left hand side of Equation (4). By Claim 13,

$$\mathop{\mathbf{E}}_{s\leftarrow\mathcal{U}_\ell}\left[\left|\mathbf{P}_{\pi(X^{i,z},Y^{i,z})}-\mathbf{P}_{\pi(X^{i+1,(z,s)},Y^{i+1,(z,s)})}\right|_1\right]\leq 2^{\ell+2}m/\sqrt{c}.$$

Therefore, by the triangle inequality,

$$\begin{aligned}\mathop{\mathbf{E}}_{t\leftarrow\pi(X^{i,z},Y^{i,z})}\mathop{\mathbf{E}}_{s\leftarrow\mathcal{U}_\ell}\left[\alpha^{t,s}\cdot\beta^{t,s}\right]&=\mathop{\mathbf{E}}_{s\leftarrow\mathcal{U}_\ell}\mathop{\mathbf{E}}_{t\leftarrow\pi(X^{i,z},Y^{i,z})}\left[\alpha^{t,s}\cdot\beta^{t,s}\right]\\&\leq\mathop{\mathbf{E}}_{s\leftarrow\mathcal{U}_\ell}\mathop{\mathbf{E}}_{t\leftarrow\pi(X^{i+1,(z,s)},Y^{i+1,(z,s)})}\left[\alpha^{t,s}\cdot\beta^{t,s}\right]+2^{\ell+2}m/\sqrt{c}.\end{aligned}$$

We now consider the right hand side of Equation (4).

$$\begin{aligned}&\mathop{\mathbf{E}}_{t\leftarrow\pi(X^{i,z},Y^{i,z})}\left[2\sqrt{\lambda^t}-\Delta^t\right]-1\\&\geq\mathop{\mathbf{E}}_{t\leftarrow\pi(X^{i,z},Y^{i,z})}\left[2\lambda^t\right]-4m/\sqrt{c}-1 &&\text{(by Claim 12)}\\&\geq 2\left(\lambda^2-4/m\right)-4m/\sqrt{c}-1 &&\text{(by Property 1 of }\pi'\text{ in Section 5.4.2)}\end{aligned}$$

Therefore, Equation (4) implies that

$$\mathop{\mathbf{E}}_{s\leftarrow\mathcal{U}_\ell}\mathop{\mathbf{E}}_{t\leftarrow\pi(X^{i+1,(z,s)},Y^{i+1,(z,s)})}\left[\alpha^{t,s}\cdot\beta^{t,s}\right]\geq 2\lambda^2-1-o_k(1).$$

Since $m/2^\ell=o_k(1)$, it holds that for at least $m+1$ possible values $s\in\{0,1\}^\ell$,

$$\mathop{\mathbf{E}}_{t\leftarrow\pi(X^{i+1,(z,s)},Y^{i+1,(z,s)})}\left[\alpha^{t,s}\cdot\beta^{t,s}\right]\geq 2\lambda^2-1-o_k(1).$$

Recall that for any s and t, the term $\alpha^{t,s}\cdot\beta^{t,s}$ is the success probability of π', on the rectangle $\mathcal{R}^t$, when the inputs are selected according to the distribution $\mu^{i+1,(z,s)}$. Therefore, there exist at least $m+1$ values $s\in\{0,1\}^\ell$, such that the protocol π' has success probability at least $2\lambda^2-1-o_k(1)$, when the inputs are selected according to the distribution $\mu^{i+1,(z,s)}$. $\square$

5.5.2 Splitting Paths

Recall that v^t is the output given by both Alice and Bob when reaching the rectangle $\mathcal{R}^t$ in the protocol π'. Let $d \in \{0,1\}^{c\ell}$ be a layer of $\mathcal{T}$. We say that layer d *splits paths* if there exist v^t and $v^{t'}$ whose lowest common ancestor in V is in layer d. That is, the path from the root to v^t and the path from the root to $v^{t'}$ are the same up-to layer d, and contain different vertices in layer $d+1$.

5.5.3 Fixing s

For $s \in \{0,1\}^{\ell}$, we denote by $\mathcal{L}^{z,s}$ the set of layers $d \in \{0,1\}^{c\ell}$, whose prefix is (z,s) (that is, the most significant bits of d are (z,s)).

We fix a value $s \in \{0,1\}^{\ell}$ such that the followings hold:

1. None of the layers in $\mathcal{L}^{z,s}$ split paths.
2. The protocol π' has success probability at least $2\lambda^2 - 1 - o_k(1)$, when the inputs are selected according to the distribution $\mu^{i+1,(z,s)}$.

Observe that such a value s exists, as by Lemma 16 there are more than m possible values s that satisfy the second property, and because there are at most m layers d that split paths, as there are only m leaves v^t.

5.6 Reduction from Disjointness

5.6.1 Unique Disjointness

We consider the following unique disjointness problem: Recall that 2^{4k} is the arity of the tree $\mathcal{T}$. Denote $K = 2^{4k}$. Alice gets an input $S \subset [K^2]$ of size K. Bob gets an input $T \subset [K^2]$ of size K. We define the set of *yes* instances to be $\{(S,T) \mid |S \cap T| = 1\}$, and the set of *no* instances to be $\{(S,T) \mid |S \cap T| = 0\}$. Given an input (S,T), the players' goal is to distinguish a yes instance from a no instance.

It is well known that even for a small constant $\epsilon > 0$, protocols with advantage ϵ for the above unique disjointness problem, have communication complexity at least $\Omega(K)$ [BFS86, KS92, Raz92]. (A protocol has advantage ϵ, if on every yes instance it outputs "yes" with probability at least $p+\epsilon$, and on every no instance it outputs "yes" with probability at most p, for some $p \geq 0$).

5.6.2 The Reduction

Constructing inputs for the hidden layers game. Recall that $v^1, \dots, v^m$ are the possible outputs given by the protocol π'. We consider the set $\{p^1, \dots, p^{m'}\}$ of all sub-paths that are restrictions to the layers in $\mathcal{L}^{z,s}$ of the path from the root to v^t, for some $t \in [m]$.

Given inputs S, T for the unique disjointness problem, the players construct inputs x, y for the hidden layers game. First, the players select independently at random $a, b \in \mathcal{L}^{z,s}$, such that a is even and b is odd. Then, using their shared random string, for every $t \in [m']$, the players choose $j^t \in_R [K^2]$. Given S, Alice constructs f by selecting for every v in layer a, an edge x_v going out of v as follows: If v is not contained in any of the sub-paths p^t, then x_v is randomly chosen among the edges going out of v. Otherwise, v is contained in some sub-path p^t, for $t \in [m']$. Since, by the first property of s in Section 5.5.3, none of the layers in $\mathcal{L}^{z,s}$ split paths, p^t is unique. If $j^t \in S$, then x_v is set to be the edge going out of v that is contained in sub-path p^t. Otherwise, $j^t \notin S$, and x_v is randomly selected among all edges that are going out of v and are not contained in p^t. Given T, Bob constructs g in a similar manner, independently.

Running π' to solve unique disjointness. After obtaining the inputs x, y, the players run the protocol π' on x, y, and get an output v^t (recall that by the second property of the protocol π' in Section 5.4.2, both players get the same output). Let $p^{t'}$ be the sub-path associated with v^t. The players check if $j^{t'} \in S \cap T$, by having Alice check if $j^{t'} \in S$ and having Bob check if $j^{t'} \in T$. If indeed $j^{t'} \in S \cap T$, the players output "yes", otherwise, they output "no".

Reduction analysis. Clearly, if (S, T) is a no instance, the protocol always outputs "no". We claim that if (S, T) is a yes instance, the protocol outputs "yes" with probability at least $2\lambda^2 - 1 - o_k(1)$, where the probability is over the random selections the players make when constructing x, y from S, T. To see this, first observe that for every yes instance (S, T), the inputs (x, y) are distributed according to $\mu^{i+1,(z,s)}$: The indices a, b are clearly distributed according to $\rho^{i+1,(z,s)}$, by the definition of $\rho^{i+1,(z,s)}$. For vertices v in layer a, such that v is not on any of the sub-paths p^t, the element x_v is selected at random, independently of all other selections (as it should be selected by the distribution $\mu^{i+1,(z,s)}$). Similarly, for vertices v in layer b, such that v is not on any of the sub-paths p^t, the element y_v is selected at random, independently of all other selections. For every $t \in [m']$, the elements $x_v, y_{v'}$ for v, v' on the same sub-path p^t are selected independently of all other selections, as the coordinate j^t is selected independently for every $t \in [m']$. These elements $x_v, y_{v'}$ have the correct distribution: Both x_v and $y_{v'}$ are uniformly distributed because $\Pr_{j^t}[j^t \in S] = \Pr_{j^t}[j^t \in T] = 1/K$ (as $|S| = |T| = K$ and $j^t \in [K^2]$ is uniform). The pair $(x_v, y_{v'})$ has the correct joint distribution because $\Pr_{j^t}[j^t \in S \cap T] = 1/K^2$ (as $|S \cap T| = 1$ and $j^t \in [K^2]$ is uniform).

By the second property of s in Section 5.5.3, the protocol π' has success probability at least $2\lambda^2 - 1 - o_k(1)$, when the inputs are selected according to the distribution $\mu^{i+1,(z,s)}$. Whenever π' succeeds in returning a leaf v^t that is consistent with both x and y, the corresponding $j^{t'}$ satisfies $j^{t'} \in S \cap T$, and the players return "yes". Thus, for every yes instance (S, T), the players return "yes" with probability at least $2\lambda^2 - 1 - o_k(1)$.

Hence, π' must have communication complexity at least $\Omega(K) > 8^k$ (for large enough k), a contradiction.

6 External Information Upper Bound

In this section, we prove Theorem 2. Let (x, y) be an input pair to the hidden layers game, where $x = (a, f)$ and $y = (b, g)$. Consider the following protocol π for the hidden layers game. Starting from the root of the tree $\mathcal{T}$, until reaching a leaf, at every vertex v: If v is in layer a, Alice sends x_v. If v is in an even layer other than a, Alice samples and sends a random edge going out of v. If v is in layer b, Bob sends y_v. If v is in an odd layer other than b, Bob samples and sends a random edge going out of v. Both players then proceed to the vertex indicated by the communicated edge. The players output the leaf reached.

Clearly, the output of the protocol π is always correct. We show that the external information cost of π over any input distribution is $O(k)$. Recall that 2^{4k} is the arity of the tree $\mathcal{T}$. Denote $K = 2^{4k}$. Let η be an input distribution, and let (X, Y) be a pair of random variables distributed according to η. For every input (x, y), all edges sent in all the rounds but the rounds corresponding to layers a and b, are random edges. The edges sent in rounds corresponding to layers a and b are determined by x and y (and the messages communicated so far). Therefore, $\mathbf{H}(\pi(X, Y)|X = x, Y = y) = \log(K) \cdot (h - 2)$. Clearly, $\mathbf{H}(\pi(X, Y)) \leq \log(K) \cdot h$. We get that

$$\mathsf{Ext}_\eta(\pi) = \mathbf{I}((X, Y); \pi(X, Y)) = \mathbf{H}(\pi(X, Y)) - \mathbf{H}(\pi(X, Y)|X, Y) \leq 2\log(K) = O(k).$$

References

[BBCR10] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. In *STOC*, pages 67–76, 2010.

[BFS86] László Babai, Peter Frankl, and Janos Simon. Complexity classes in communication complexity theory (preliminary version). In *FOCS*, pages 337–347, 1986.

[BMY14] Balthazar Bauer, Shay Moran, and Amir Yehudayoff. Internal compression of protocols to entropy. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:101, 2014.

[BR11] Mark Braverman and Anup Rao. Information equals amortized communication. In *FOCS*, pages 748–757, 2011.

[Bra12] Mark Braverman. Interactive information complexity. In *STOC*, pages 505–524, 2012.

[Bra13] Mark Braverman. A hard-to-compress interactive task? *In 51th Annual Allerton Conference on Communication, Control, and Computing*, 2013.

[BW12] Mark Braverman and Omri Weinstein. A discrepancy lower bound for information complexity. In *APPROX-RANDOM*, pages 459–470, 2012.

[BYJKS04] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *J. Comput. Syst. Sci.*, 68(4):702–732, 2004.

[CSWY01] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Chi-Chih Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *FOCS*, pages 270–278, 2001.

[FJK$^+$15] Lila Fontes, Rahul Jain, Iordanis Kerenidis, Mathieu Lauri‘ere, Sophie Laplante, and Jeremie Roland. Relative discrepancy does not separate information and communication complexity. *Electronic Colloquium on Computational Complexity (ECCC)*, 2015.

[GKR14] Anat Ganor, Gillat Kol, and Ran Raz. Exponential separation of information and communication. In *FOCS*, pages 176–185, 2014.

[GKR15] Anat Ganor, Gillat Kol, and Ran Raz. Exponential separation of information and communication for boolean functions. In *STOC*, 2015.

[KLL$^+$12] Iordanis Kerenidis, Sophie Laplante, Virginie Lerays, Jérémie Roland, and David Xiao. Lower bounds on information complexity via zero-communication protocols and applications. In *FOCS*, pages 500–509, 2012.

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

[KS92] Bala Kalyanasundaram and Georg Schnitger. The probabilistic communication complexity of set intersection. *SIAM J. Discrete Math.*, 5(4):545–557, 1992.

[LS09] Troy Lee and Adi Shraibman. Lower bounds in communication complexity. *Foundations and Trends in Theoretical Computer Science*, 3(4):263–398, 2009.

[Raz92] Alexander A. Razborov. On the distributional complexity of disjointness. *Theor. Comput. Sci.*, 106(2):385–390, 1992.

[RR15] Sivaramakrishnan Natarajan Ramamoorthy and Anup Rao. How to compress asymmetric communication. *Electronic Colloquium on Computational Complexity (ECCC)*, 22:55, 2015.

[RS15] Anup Rao and Makrand Sinha. Simplified separation of information and communication. *Electronic Colloquium on Computational Complexity (ECCC)*, 22:57, 2015.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de