Electronic Colloquium on Computational Complexity, Report No. 49 (2014)

Exponential Separation of Information and Communication

Anat Ganor* Gillat Kol† Ran Raz‡

Abstract

We show an exponential gap between communication complexity and information complexity, by giving an explicit example for a communication task (relation), with information complexity $\leq O(k)$, and distributional communication complexity $\geq 2^k$. This shows that a communication protocol cannot always be compressed to its internal information. By a result of Braverman [Bra12b], our gap is the largest possible. By a result of Braverman and Rao [BR11], our example shows a gap between communication complexity and amortized communication complexity, implying that a tight direct sum result for distributional communication complexity cannot hold.

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 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].

*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 at the IAS supported by The Fund For Math and the Weizmann Institute of Science National Postdoctoral Award Program for Advancing Women in Science.

‡Weizmann Institute of Science, Israel and Institute for Advanced Study, Princeton, NJ. Research supported by an Israel Science Foundation grant, by the I-CORE Program of the Planning and Budgeting Committee and the Israel Science Foundation. Supported at the IAS by the The Fund For Math and The Simonyi Fund, and by NSF grants number CCF-0832797, DMS-0835373.

ISSN 1433-8092

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, a problem that has a rich history, and has been studied in many works and various settings [FKNN95, CSWY01, JRS03, HJMR07, BBCR10, Kla10, Jai11, JPY12, BRWY12, BRWY13] (and many other works). In this paper we will mainly be interested in internal information complexity (a.k.a, information complexity and 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 information complexity of a communication task is the minimal number of information bits that the players learn about each other's input when solving the task, where the minimum is taken over all protocols.

Many recent works focused on the problem of compressing interactive communication protocols. Given a communication protocol with small 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. Barak, Braverman, Chen and Rao showed how to compress any protocol with information complexity k and communication complexity c, to a protocol with communication complexity $\tilde{O}(\sqrt{ck})$ in the general case, and $\tilde{O}(k)$ in the case where the underlying distribution is a product distribution [BBCR10]. Braverman and Rao showed how to compress any one round (or small number of rounds) protocol with information complexity k to a protocol with communication complexity $O(k)$ [BR11]. Braverman showed how to compress any protocol with information complexity k to a protocol with communication complexity $2^{O(k)}$ [Bra12b] (see also [BW12, KLL+12]). This last protocol is the most related to our work, as it gives a compression result that works in the general case and doesn't depend at all on the communication complexity of the original protocol. Braverman also described a communication complexity task that has information complexity $O(k)$ and no known communication protocol with communication complexity smaller than 2^k [Bra13]. However, there is no known lower bound on the communication complexity of that problem.

Another line of works shows that many of the known general techniques for proving lower bounds for randomized communication complexity also give lower bounds for information complexity [Bra12b, BW12, KLL+12].

In this work we show the first gap between information complexity and communication complexity of a communication task. We give an explicit example for a communication task (a relation), called the *bursting noise game*, parameterized by $k \in \mathbb{N}$ and played with an input distribution μ. We prove that the information complexity of the game is $O(k)$, while any communication protocol for solving this game, with communication complexity at most 2^k, almost always errs. By the above mentioned compression protocol of Braverman [Bra12b], our result gives the largest possible gap between information complexity and communication complexity.

Theorem 1 (**Communication Lower Bound**). *Every randomized protocol (with shared randomness) for the bursting noise game with parameter k, that has communication complexity at most 2^k, errs with probability $\epsilon \geq 1 - o(1)$ (over the input distribution μ, where the $o(1)$ is sub-constant in k).*

Theorem 2 (**Information Upper Bound**). *There exists a randomized protocol for the bursting noise game with parameter k, that has information cost $O(k)$ and errs with probability $\epsilon \leq 2^{-\Omega(k)}$ (over the input distribution μ).*

We note that both the inputs and the outputs in our bursting noise game example are very long. Namely, the input length is triple exponential in k, and the output length is double exponential. The protocol that achieves information complexity $O(k)$ has communication complexity double exponential in k.

As mentioned above, information complexity is also related to the direct sum problem in communication complexity. Braverman and Rao showed that information complexity is equal to the amortized communication complexity, that is, the limit of the communication complexity needed to solve n tasks of the same type, divided by n [BR11] (see also [Bra12a, Bra12b, Bra13]). Our result therefore shows a gap between distributional communication complexity and amortized distributional communication complexity, proving that tight direct sum results for the communication complexity of relations cannot hold.

Organization. The paper is organized as follows. In Section 2 we define the bursting noise game. Section 3 gives an overview of our main result, the lower bound for the communication complexity of the bursting noise game (Theorem 1). In Section 4 we give general definitions and preliminaries. In Section 5 we prove the graph correlation lemma, a central tool that we will use in the lower bound proof. In Section 6 we prove the communication complexity lower bound (Theorem 1). In Section 7 we show that the straightforward protocol for the bursting noise game has low information cost, thus proving the upper bound required by Theorem 2. The appendix contains proofs of information theoretic lemmas that are used by the lower bound proof.

2 Bursting Noise Games

The *bursting noise game* is a communication game between two parties, called the *first player* and the *second player*. The game is specified by a parameter $k \in \mathbb{N}$, where $k > 2^{100}$. We set $c = 2^{4^k}$, $w_0 = 2^{100}k$, $w_1 = 2^{100k}$, $w = w_0 + w_1$.

The game is played on the binary tree $\mathcal{T}$ with $c \cdot w$ layers (the root is in layer 1 and the leaves are in layer $c \cdot w$), with edges directed from the root to the leaves. Denote the vertex set of $\mathcal{T}$ by V. Each player gets as input a bit for every vertex in the tree. Let x be the input given to the first player, and y be the input given to the second player, where $x, y \in \{0,1\}^V$. For a vertex $v \in V$, we denote by x_v and y_v the bits in x and y associated with v. The input pair (x, y) is selected according to a joint distribution μ on $\{0,1\}^V \times \{0,1\}^V$, defined below.

Denote by $\mathrm{Even}(\mathcal{T}) \subseteq V$ the set of non-leaf vertices in an even layer of $\mathcal{T}$ and by $\mathrm{Odd}(\mathcal{T}) \subseteq V$ the set of non-leaf vertices in an odd layer of $\mathcal{T}$. We think of the vertices in $\mathrm{Odd}(\mathcal{T})$ as "owned" by the first player and the vertices in $\mathrm{Even}(\mathcal{T})$ as "owned" by the second player. Let $v \in V$ be a non-leaf vertex. Let v_0 be the left child of v and v_1 be the right child of v. Let $b \in \{0,1\}$. We say that v_b is the *correct child* of v with respect to x, y, if either the first player owns v and $x_v = b$, or the second player owns v and $y_v = b$.

For $s \leq t \in \mathbb{N}$, denote by $[s,t]$ the set $\{s, \ldots, t\}$ and by $[t]$ the set $\{1, \ldots, t\}$. Let $s \leq t < t' \in [c \cdot w]$ be layers of $\mathcal{T}$ and let $v \in V$ be a vertex in layer t'. For $j \in [s, t+1]$, let v_j be v's ancestor in layer j. Let $\Delta = t - s + 1$. We say that v is *typical* with respect to s, Δ, x, y, if the followings hold:

1. For at least 0.8-fraction of the indices $j \in [s,t] \cap \mathrm{Odd}(\mathcal{T})$, the vertex v_{j+1} is the correct child of v_j with respect to x, y.

2. For at least 0.8-fraction of the indices $j \in [s,t] \cap \mathrm{Even}(\mathcal{T})$, the vertex v_{j+1} is the correct child of v_j with respect to x, y.

Observe that in order to decide whether v is typical with respect to s, Δ, x, y, it suffices to know the bits that x, y assign to the vertices $v_s, \ldots, v_t$.

We think of the $c \cdot w$ layers of the tree $\mathcal{T}$ as partitioned into c multi-layers, each consisting of w consecutive layers (e.g., the first multi-layer consists of layers 1 to w). We denote by i^* the first layer of the i^{th} multi-layer, that is, $i^* = (i-1)w + 1$.

We next define the distribution μ on $\{0,1\}^V \times \{0,1\}^V$ by an algorithm for sampling an input pair (x, y) (Algorithm 1 below). In the algorithm, when we say "set v to be non-noisy", we mean "select $x_v \in \{0,1\}$ uniformly at random and set $y_v = x_v$". By "set v to be noisy", we mean "select $x_v \in \{0,1\}$ and $y_v \in \{0,1\}$ independently and uniformly at random". Figure 1 illustrates Algorithm 1.

The players' mutual goal is to output the same leaf $v \in V$, where v is typical with respect to i^*, w_0, x, y (that is, v is typical with respect to the main noise layers).

For $i \in [c]$, we denote by μ_i the distribution μ conditioned on the event that the noisy multi-layer selected by Step 1 of the algorithm defining μ, is i. Note that $\mu = \frac{1}{c}\sum_{i \in [c]} \mu_i$.

Remark. *Observe that it is not always possible to deduce i (i.e., the index of the noisy multi-layer used to construct the pair (x, y)) from the pair (x, y). Therefore, the bursting noise game does not induce a relation. Nevertheless, with extremely high probability, the first multi-layer on which x and y disagree is i. Thus, the game can be easily converted to a relation, by omitting the rare inputs (x, y) that agree on multi-layer i. Note that since the statistical distance between the two distributions is negligible, both our upper bound and lower bound trivially apply to the new game as well. For that reason, it will be helpful to think of the supports of the different μ_i's as if they were pairwise disjoint.*

Algorithm 1 Sample (x, y) according to μ

1. Randomly select $i \in [c]$ (the noisy multi-layer).
2. Set every vertex in layers $[i^*, i^* + w_0 - 1]$ to be noisy (main noise).
3. Let L_1 be the set of all non-typical vertices in layer $i^* + w_0$ with respect to i^*, w_0, x, y (note that x, y were already defined on layers $[i^*, i^* + w_0 - 1]$, and therefore the typical vertices are defined). For every $v \in L_1$, set all the vertices in the subtree with w_1 layers rooted at v to be noisy.
4. If $i < c$: Let L_2 be the set of all vertices v in layer $i^* + w_0 + w_1 = (i+1)^*$ with ancestors in L_1, such that v is non-typical with respect to $i^* + w_0, w_1, x, y$ (note that x, y were already defined on the required vertices, and therefore the typical vertices are defined). For every $v \in L_2$, set all the vertices in the subtree rooted at v to be noisy.
5. Set all unset vertices in V to be non-noisy.

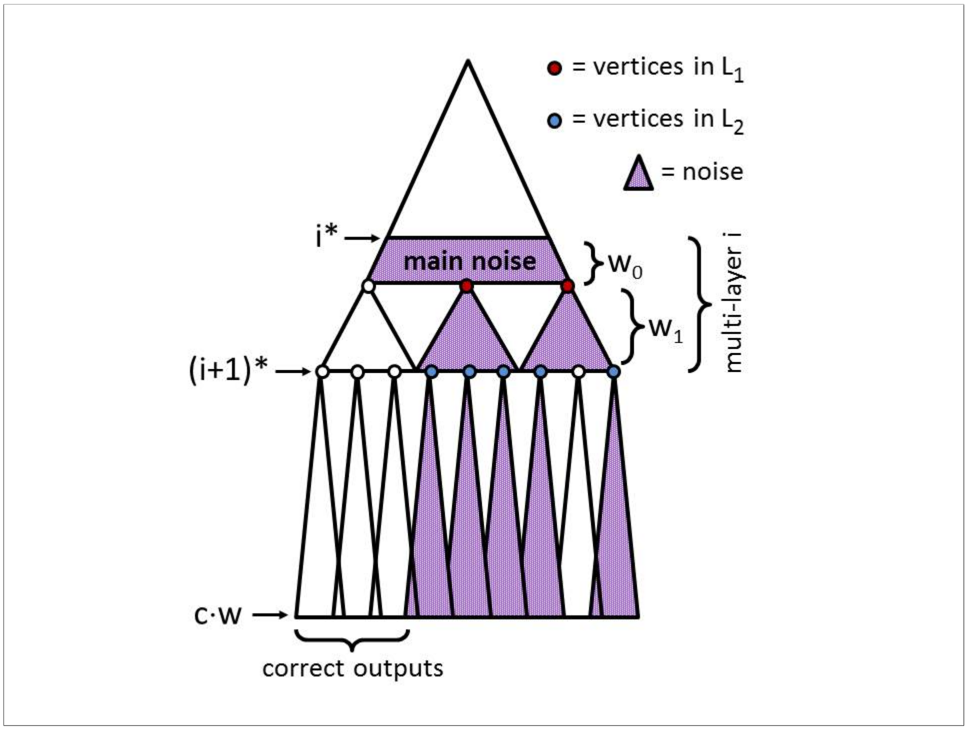

Figure 1: Illustration of Algorithm 1

The straightforward protocol. Consider the following straightforward protocol for the bursting noise game. Starting from the root until reaching a leaf, at every vertex v, if the first player owns v, she sends the bit x_v with probability 0.9, and the bit $1-x_v$ with probability 0.1. Similarly, if the second player owns v, she sends the bit y_v with probability 0.9, and the bit $1-y_v$ with probability 0.1. Both players continue to the child of v that is indicated by the communicated bit. When they reach a leaf they output that leaf.

By the Chernoff bound, the probability that the players output a leaf that is not typical with respect to the main noise layers is at most $2^{-\Omega(w_0)}$. That is, the error probability is exponentially small in k. In Section 7, we show that the information cost of this protocol is $O(k)$. Intuitively, this will follow since the expected number of vertices reached by the protocol, on which the players' inputs disagree, is $O(k)$ (with high probability the disagreement is only on vertices in the main noise layers).

Remark. *Observe that c is set to be double exponential in k. If c were set to be just exponential in k, a simple binary search algorithm would have been able to find the location of the main noise layers, and thus solve the bursting noise game with communication complexity polynomial in k.*

3 Overview of the Lower Bound Proof

Rectangle Partition

We will describe the proof of the lower bound for the communication complexity of the bursting noise game. We fix the random strings for the protocol so that we have a deterministic protocol. We show that if the protocol communicates at most 2^k bits, it errs with probability $1-o(1)$ on inputs sampled according to μ. We will show that for almost all $i \in [c]$, the protocol errs with probability $1-o(1)$ on inputs sampled according to μ_i, that is, the distribution μ conditioned on the event that the noisy multi-layer selected by Step 1 of Algorithm 1 defining μ, is i. Note that the distribution μ_i is uniformly distributed over $\mathrm{supp}(\mu_i)$, and that for every pair of inputs $(x, y) \in \mathrm{supp}(\mu_i)$, the projection of x and y on the first $i-1$ multi-layers is the same.

As mentioned above, it will be helpful to think of the supports of the different μ_i's as if they were pairwise disjoint (this property holds if we remove a μ_i-negligible set of inputs from the support of each μ_i).

Let $\{R^1, \ldots, R^m\}$ be the rectangle partition induced by the protocol, where $R^t = A^t \times B^t$, and $m \leq 2^{2^k}$. For $i \in [c]$ and an assignment z to the first $i-1$ multi-layers, we denote by $R^{t,z} = A^{t,z} \times B^{t,z}$, the rectangle of all pairs of inputs $(x, y) \in R^t$, such that the projection of both x, y on the first $i-1$ multi-layers is equal to z. Let $X^{t,z}$ be a random variable uniformly distributed over $A^{t,z}$. Let $Y^{t,z}$ be a random variable uniformly distributed over $B^{t,z}$. We denote by $X_i^{t,z}, Y_i^{t,z}$ the projections of $X^{t,z}, Y^{t,z}$, respectively, on multi-layer i.

For fixed i, z, we define $\rho^{i,z}$ to be a probability distribution that selects a rectangle in

$\{R^{1,z},\ldots,R^{m,z}\}$ according to its relative size. That is, $\rho^{i,z}$ is defined as follows: Randomly select x, y, such that the projection of both x and y on the first $i-1$ multi-layers is z. Select t to be the index of the unique rectangle $R^{t,z}$ containing (x, y).

Bounding the Information on the Noisy Multi-Layer

The main intuition of the proof is that since c is significantly larger than 2^k, the protocol cannot make progress on all multi-layers $i \in [c]$ simultaneously. We first show that for a random $i \in [c]$, a random z, and a random rectangle $R^{t,z}$, chosen according to $\rho^{i,z}$, very little information is known about $X_i^{t,z}$ and $Y_i^{t,z}$.

Formally, we prove in Lemma 11 that

$$\mathop{\mathbf{E}}_{i} \mathop{\mathbf{E}}_{z} \mathop{\mathbf{E}}_{t \leftarrow \rho^{i,z}} \left[\mathbf{I}\left(X_i^{t,z}\right)\right] \leq \frac{m}{c}, \tag{1}$$

and similarly,

$$\mathop{\mathbf{E}}_{i} \mathop{\mathbf{E}}_{z} \mathop{\mathbf{E}}_{t \leftarrow \rho^{i,z}} \left[\mathbf{I}\left(Y_i^{t,z}\right)\right] \leq \frac{m}{c}, \tag{2}$$

where we denote by $\mathbf{I}(Z) := \log(|\Omega|) - \mathbf{H}(Z)$ the information known about a random variable Z, where Ω is the space that Z is defined over.

The proof of Lemma 11 doesn't follow by a trivial application of super-additivity of information. That's because choosing i, z at random and t according to $\rho^{i,z}$ and then choosing a random variable X to be uniformly distributed on $A^{t,z}$, gives a random variable X with distribution that may be very far from uniform. Moreover, the probability that X is in the set A^t, associated with a rectangle R^t, may be very far from the probability that a uniformly distributed input is in A^t. Nevertheless, we are still able to prove Lemma 11, using the fact that we have a bound of m on the total number of times that an input x appears in the cover $\{A^1,\ldots,A^m\}$.

We fix $\frac{1}{k^{o(1)}} \leq \gamma \leq o(1)$ to be sub-constant, and we fix i, z, t, such that,

1. $\mathbf{I}\left(X_i^{t,z}\right) \leq \frac{1}{\gamma} \cdot \frac{m}{c}$
2. $\mathbf{I}\left(Y_i^{t,z}\right) \leq \frac{1}{\gamma} \cdot \frac{m}{c}$
3. The rectangle $R^{t,z}$ is not too small.

By Equation (1) and Equation (2), and by Markov's inequality, we know that when we choose i, z uniformly at random, and t according to $\rho^{i,z}$, the triplet (i, z, t) satisfies all three conditions with high probability. Therefore, we ignore triplets (i, z, t) that do not satisfy all three conditions.

How the Proof Works

Let Λ^i be the set of input pairs $(x, y) \in \text{supp}(\mu_i)$, such that the protocol errs on (x, y). Let P_i be the probability for a uniformly distributed pair of inputs (x, y), that have the same projection on the first $i-1$ multi-layers, to be in $\mathrm{supp}(\mu_i)$. We prove that

$$\Pr_{X^{t,z},Y^{t,z}}\left[(X^{t,z},Y^{t,z})\in\Lambda^i\right]\geq P_i\cdot(1-o(1)).$$

Summing over all possibilities for z,t, we obtain (for almost all $i\in[c]$) that the protocol errs on μ_i with probability $1-o(1)$.

Unique Answer Rectangles

In the rectangle $R^{t,z}$, the answer of each of the two players in the protocol may not be unique, as the answer of each player may also depend on the input that she gets. Nevertheless, using the fact that if the two players answer differently then the protocol errs, we are able to subdivide the rectangle $R^{t,z}$ into $\mathrm{poly}(1/\gamma)$ sub-rectangles $R^{t,s,z}$, such that in each rectangle $R^{t,s,z}$ the answer is unique, except for a bad set of rectangles whose total size is negligible compared to the size of $R^{t,z}$. When subdividing $R^{t,z}$, we also need to change the answers given by the two players on each rectangle, but we are able to do that without adding errors to the protocol.

We ignore rectangles $R^{t,s,z}$ where the answer of the protocol is not unique, as their total size is small, and only consider rectangles $R^{t,s,z}=A^{t,s,z}\times B^{t,s,z}$ where the answer is unique. Let $X^{t,s,z}$ be a random variable uniformly distributed over $A^{t,s,z}$. Let $Y^{t,s,z}$ be a random variable uniformly distributed over $B^{t,s,z}$. For the rectangles $R^{t,s,z}$ we no longer have the strong bounds $\mathbf{I}\left(X_i^{t,z}\right)\leq\frac{1}{\gamma}\cdot\frac{m}{c}$, and $\mathbf{I}\left(Y_i^{t,z}\right)\leq\frac{1}{\gamma}\cdot\frac{m}{c}$, but rather the weaker bounds

$$\mathbf{I}\left(X_i^{t,s,z}\right)\leq O\left(\log\left(1/\gamma\right)\right),$$

and

$$\mathbf{I}\left(Y_i^{t,s,z}\right)\leq O\left(\log\left(1/\gamma\right)\right).$$

The Main Lemma

Fix i,z,t,s. In the rectangle $R^{t,s,z}$ the answer is unique, denote that answer by $\omega^{t,s,z}$. We define $\Lambda^{t,s,z}$ to be the set of input pairs $(x,y)\in\mathrm{supp}(\mu_i)$, such that $\omega^{t,s,z}$ is not a correct answer for inputs (x,y). In Lemma 14, we prove that

$$\Pr_{X^{t,s,z},Y^{t,s,z}}\left[(X^{t,s,z},Y^{t,s,z})\in\Lambda^{t,s,z}\right]\geq P_i\cdot(1-o(1)),\tag{3}$$

and as before, summing over all possibilities for t,s,z, this implies, for almost all $i\in[c]$, that the protocol errs on μ_i with probability $1-o(1)$.

In what follows, we outline the proof of Equation (3).

The Super Graph

We define the bipartite graph $G=(U\cup W,E)$ with sets of vertices U,W and set of edges E as follows: Let $U=W$ be the set of all possible assignments for multi-layer i (for one player).

For $u \in U, w \in W$, we have $(u, w) \in E$ if there exists $(x, y) \in \text{supp}(\mu_i)$, such that $x_i = u$ and $y_i = w$.

Let M be the number of vertices in layer $(i+1)^*$ of the tree $\mathcal{T}$. We identify the set $[M]$ with the set of vertices in layer $(i+1)^*$. Let $u \in U, w \in W$. We define $T(u, w) \subset [M]$ to be the set of all vertices in layer $(i+1)^*$ that are set to be non-noisy for inputs u, w, by Algorithm 1 defining μ, when the noisy multi-layer is i. Observe that u and w determine for every vertex in layer $(i+1)^*$ if it is noisy or not. Note that by a symmetry argument, G is bi-regular, and $T(u, w)$ is of the same size T for every u, w.

Let $\mathcal{E}^{t,s,z} \subseteq E$ be the set of all $(u, w) \in E$ for which the output $\omega^{t,s,z}$ is correct for some input $(x, y) \in \text{supp}(\mu_i)$, where $x_i = u$ and $y_i = w$. Note that if the noise is taken on the i^{th} multi-layer, then u and w determine the correctness of $\omega^{t,s,z}$. It holds that

$$|\mathcal{E}^{t,s,z}| \leq 2^{-20k}|E|,$$

as for any fixed u and every $v \in [M]$, at most a fraction of 2^{-20k} of the sets $\{T(u, w)\}_{(u,w)\in E}$ contain v, and the output $\omega^{t,s,z}$ is correct only if it has an ancestor in $T(u, w)$.

Let Σ be the set of all possible boolean assignments to the vertices of a subtree of $\mathcal{T}$ rooted at layer $(i+1)^*$.

For $u \in U$, we define the random variable X^u, over the domain $\Sigma^{[M]}$, to be the conditional variable $(X^{t,s,z}_{\geq i} | X^{t,s,z}_i = u)$, that is, X^u has the distribution of $X^{t,s,z}_{\geq i}$ conditioned on the event $X^{t,s,z}_i = u$, where $X^{t,s,z}_{\geq i}$ denotes the projection of $X^{t,s,z}$ to all multi-layers after multi-layer i. Similarly, for $w \in W$, we define the random variable Y^w, over the domain $\Sigma^{[M]}$, to be $(Y^{t,s,z}_{\geq i} | Y^{t,s,z}_i = w)$, that is, Y^w has the distribution of $Y^{t,s,z}_{\geq i}$ conditioned on the event $Y^{t,s,z}_i = w$.

First Application of the Graph Correlation Lemma

By the definition of the distribution μ_i, the left hand side of Equation (3) is equal to

$$\sum_{(u,w)\in E\setminus\mathcal{E}^{t,s,z}} \Pr\left[X^{t,s,z}_i = u\right] \cdot \Pr\left[Y^{t,s,z}_i = w\right] \cdot \Pr\left[X^u_{T(u,w)} = Y^w_{T(u,w)}\right], \tag{4}$$

where $X^u_{T(u,w)}$ and $Y^w_{T(u,w)}$ are the projections of X^u, Y^w, respectively, to coordinates in $T(u, w)$. This is true because a pair (x, y) is in $\text{supp}(\mu_i)$ if and only if $(x_i, y_i) \in E$ and x, y agree on all the subtrees rooted at vertices in layer $(i+1)^*$ that are set to be non-noisy for inputs x_i, y_i, by Algorithm 1 defining μ, when the noisy multi-layer is i.

Our graph correlation lemma (Lemma 9), that may be interesting in its own right, gives a general way to bound such expressions by

$$\geq (1 - o(1))\, |\Sigma|^{-T} \sum_{(u,w)\in E\setminus(\mathcal{E}^{t,s,z}\cup\mathcal{D}^{t,s,z})} \Pr\left[X^{t,s,z}_i = u\right] \cdot \Pr\left[Y^{t,s,z}_i = w\right], \tag{5}$$

where $\mathcal{D}^{t,s,z} \subset E$ is a small set, compared to the size of E, and $|\Sigma|^{-T}$ is a normalization factor that would be equal to $\Pr[X^u_{T(u,w)} = Y^w_{T(u,w)}]$ if X^u, Y^w were uniformly distributed (independent) random variables.

Thus, using Lemma 9, we are able to bound the left hand side of Equation (3), which is an expression that depends on the variables $X^{t,s,z}, Y^{t,s,z}$, by the expression in Equation (5) that depends only on the projections of these variables to multi-layer i.

We still need to bound from below the expression

$$\sum_{(u,w)\in E\setminus(\mathcal{E}^{t,s,z}\cup\mathcal{D}^{t,s,z})} \Pr\left[X_i^{t,s,z}=u\right]\cdot\Pr\left[Y_i^{t,s,z}=w\right]. \tag{6}$$

Since $\mathcal{E}^{t,s,z}\cup\mathcal{D}^{t,s,z}$ is a small set (compared to the size of E), we will first ignore the set $\mathcal{E}^{t,s,z}\cup\mathcal{D}^{t,s,z}$, and bound from below the expression

$$\sum_{(u,w)\in E} \Pr\left[X_i^{t,s,z}=u\right]\cdot\Pr\left[Y_i^{t,s,z}=w\right]. \tag{7}$$

In Lemma 18 we give a general bound on such expressions: Let P be a random variable over the domain U, and Q be a random variable over the domain W, such that P and Q are independent. Denote $I=\max\{\mathbf{I}(P),\mathbf{I}(Q),1\}$. It holds that

$$\Pr_{P,Q}\left[(P,Q)\in E\right]\geq\frac{|E|}{|U|\cdot|W|}\left(1-\frac{I}{k}\right).$$

Note that the expression in Equation (7) is of this type.

The proof of the lemma is by applying Lemma 9 once again, as described next.

The Mini Graph

Let $\hat{G}=(\hat{U}\cup\hat{W},\hat{E})$ be the complete bipartite graph with sets of vertices $\hat{U},\hat{W}$ and set of edges $\hat{E}$, defined as follows: Let $\hat{U}=\hat{W}$ be the set of all boolean assignments (for one player) to the vertices in layers i^* to i^*+w_0-1 of the tree $\mathcal{T}$ (that is, the first w_0 layers in multi-layer i; see Section 2). Set $\hat{E}=\hat{U}\times\hat{W}$.

Let $\hat{P}$ and $\hat{Q}$ be the projections of P and Q on layers i^* to i^*+w_0-1 (respectively). Let $\tilde{P}$ and $\tilde{Q}$ be the projections of P and Q on layers i^*+w_0 to $(i+1)^*-1$ (respectively).

Let $\hat{M}$ be the number of vertices in layer i^*+w_0 of the tree $\mathcal{T}$. We identify the set $[\hat{M}]$ with the set of vertices in layer i^*+w_0. Let $\hat{u}\in\hat{U},\hat{w}\in\hat{W}$. We define $\hat{T}(\hat{u},\hat{w})\subset[\hat{M}]$ to be the set of all vertices in layer i^*+w_0 that are set to be non-noisy for inputs $\hat{u},\hat{w}$, by Algorithm 1 defining μ, when the noisy multi-layer is i. That is, $\hat{T}(\hat{u},\hat{w})$ is the set of all typical vertices in layer i^*+w_0 with respect to $i^*,w_0,\hat{u},\hat{w}$. Observe that $\hat{u}$ and $\hat{w}$ determine for every vertex in layer i^*+w_0 if it is noisy or not.

Note that by a symmetry argument, $\hat{T}(\hat{u},\hat{w})$ is of the same size $\hat{T}$ for every $\hat{u},\hat{w}$.

Define the random variable $X^{\hat{u}}$ to be $(\tilde{P}|\hat{P}=\hat{u})$, that is, $X^{\hat{u}}$ has the distribution of $\tilde{P}$ conditioned on the event $\hat{P}=\hat{u}$. Define the random variable $Y^{\hat{w}}$ to be $(\tilde{Q}|\hat{Q}=\hat{w})$, that is, $Y^{\hat{w}}$ has the distribution of $\tilde{Q}$ conditioned on the event $\hat{Q}=\hat{w}$.

Let $\hat{\Sigma}$ be the set of all possible boolean assignments to the vertices of a subtree with w_1

layers, rooted at layer $i^* + w_0$.

Second Application of the Graph Correlation Lemma

By the definitions of the graphs $G, \hat{G}$,

$$\Pr\left[(P,Q) \in E\right] = \sum_{(\hat{u},\hat{w})\in\hat{E}} \Pr_P\left[\hat{P} = \hat{u}\right] \cdot \Pr_Q\left[\hat{Q} = \hat{w}\right] \cdot \Pr_{X^{\hat{u}}, Y^{\hat{w}}}\left[X^{\hat{u}}_{\hat{T}(\hat{u},\hat{w})} = Y^{\hat{w}}_{\hat{T}(\hat{u},\hat{w})}\right],$$

where $X^{\hat{u}}_{\hat{T}(\hat{u},\hat{w})}$ and $Y^{\hat{w}}_{\hat{T}(\hat{u},\hat{w})}$ are the projections of $X^{\hat{u}}, Y^{\hat{w}}$, respectively, to coordinates in $\hat{T}(\hat{u},\hat{w})$. This is true because a pair (u,w) is in E if and only if $(\hat{u},\hat{w}) \in \hat{E}$ and u,w agree on all the subtrees rooted at vertices in layer $i^* + \omega_0$ that are set to be non-noisy for inputs $\hat{u},\hat{w}$, by Algorithm 1 defining μ, when the noisy multi-layer is i.

Lemma 9 gives a general way to bound such expressions by

$$\geq (1 - o(1)) \, |\hat{\Sigma}|^{-\hat{T}} \sum_{(\hat{u},\hat{w})\in\hat{E}\setminus\hat{D}} \Pr_P\left[\hat{P} = \hat{u}\right] \cdot \Pr_Q\left[\hat{Q} = \hat{w}\right], \tag{8}$$

where $\hat{D} \subset \hat{E}$ is a small set, compared to the size of $\hat{E}$, and $|\hat{\Sigma}|^{-\hat{T}}$ is a normalization factor that would be equal to $\Pr[X^{\hat{u}}_{\hat{T}(\hat{u},\hat{w})} = Y^{\hat{w}}_{\hat{T}(\hat{u},\hat{w})}]$ if $X^{\hat{u}}_{\hat{T}(\hat{u},\hat{w})}, Y^{\hat{w}}_{\hat{T}(\hat{u},\hat{w})}$ were uniformly distributed (independent) random variables.

Thus, using Lemma 9, we are able to bound the expression in Equation (7), which is an expression that depends on the variables P, Q, by an expression that depends only on the projections of these variables to the first ω_0 layers of multi-layer i.

Since $\hat{D}$ is small, we will first ignore it, and note that since $\hat{G}$ is the complete graph,

$$\sum_{(\hat{u},\hat{w})\in\hat{E}} \Pr_P\left[\hat{P} = \hat{u}\right] \cdot \Pr_Q\left[\hat{Q} = \hat{w}\right] = 1,$$

and we use this to bound the right hand side of Equation (8).

Completing the Proof

In both applications of Lemma 9, we ignored the sum on small sets, $\mathcal{E}^{t,s,z} \cup \mathcal{D}^{t,s,z}$ in the first application, and $\hat{D}$ in the second application. To complete the proof, we need to show that

$$\sum_{(u,w)\in\mathcal{E}^{t,s,z}\cup\mathcal{D}^{t,s,z}} \Pr\left[X^{t,s,z}_i = u\right] \cdot \Pr\left[Y^{t,s,z}_i = w\right],$$

in the first application (see Equation (6) and Equation (7)), and

$$\sum_{(\hat{u},\hat{w})\in\hat{D}} \Pr_P\left[\hat{P} = \hat{u}\right] \cdot \Pr_Q\left[\hat{Q} = \hat{w}\right],$$

in the second application, are both negligible.

In the second application, we are able to prove that, using the fact that $\hat{D}$ is small and $\hat{G}$ is the complete graph, and using the bound that we have on $\mathbf{I}(P)$ and $\mathbf{I}(Q)$.

In the first application, we use the fact that $R^{t,s,z} \subseteq R^{t,z}$, to bound the sum by

$$\frac{|R^{t,z}|}{|R^{t,s,z}|} \sum_{(u,w) \in \mathcal{E}^{t,s,z} \cup \mathcal{D}^{t,s,z}} \Pr\left[X_i^{t,z} = u\right] \cdot \Pr\left[Y_i^{t,z} = w\right].$$

Since $\mathbf{I}\left(X_i^{t,z}\right) \leq \frac{1}{\gamma} \cdot \frac{m}{c}$, and $\mathbf{I}\left(Y_i^{t,z}\right) \leq \frac{1}{\gamma} \cdot \frac{m}{c}$, we know that the distributions of $X_i^{t,z}$ and $Y_i^{t,z}$ are extremely close to uniform, and hence the sum in the last expression is negligible. Using also the fact that $\frac{|R^{t,z}|}{|R^{t,s,z}|}$ is relatively small, we obtain that the entire expression is negligible.

Another difficulty that we ignored in the discussion so far, is that Lemma 9 requires random variables X^u, Y^w with bounded information for all u, w, while we have variables with bounded information for almost all u, w. In the two applications of the lemma, we fix this by two different simple manipulations. In the first application, we change the variables $X^{t,s,z}$ and $Y^{t,s,z}$ by taking them to be uniformly distributed with very small probability. This works because $X^{t,s,z}$ and $Y^{t,s,z}$ are uniformly distributed over a subset. In the second application, we just replace every $X^{\hat{u}}$ or $Y^{\hat{w}}$ that has large information, with a uniformly distributed random variable. This works since $\hat{G}$ is the complete graph.

Proof of the Graph Correlation Lemma and Shearer's Inequality

To bound expressions such as the expression in Equation (4), we show that if $\Pr[X^u_{T(u,w)} = Y^w_{T(u,w)}]$ is significantly larger than what is obtained by uniformly distributed variables, then either $\mathbf{I}(X^u_{T(u,w)})$ or $\mathbf{I}(Y^w_{T(u,w)})$ are non negligible (or both). We use this to show that for some u (or some w) we have that $\mathbf{I}(X^u)$ (or $\mathbf{I}(Y^w)$) are large, deriving a contradiction.

Our proof relies on a variant of Shearer's inequality [CGFS86, Kah01] that follows easily by Radhakrishnan's beautiful information theoretical proof [Rad03] (see Lemmas 7 and 8 and [MT10]).

4 Definitions and Preliminaries

4.1 General Notation

Throughout the paper, all logarithms are taken with base 2, and we define $0 \log(0) = 0$. For a set S, when we write "$x \in_R S$" we mean that x is selected uniformly at random from the set S. For a distribution τ, when we write "$x \leftarrow \tau$" we mean that x is selected according to the distribution τ. For Z that is either a random variable taking values in $\{0,1\}^V$ or an element in $\{0,1\}^V$, and a set $T \subseteq V$, we define Z_T to be the projection of Z to T.

4.2 Information Cost

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

$$IC_\mu(\pi) = \mathbf{I}(\Pi; X|Y) + \mathbf{I}(\Pi; Y|X),$$

where Π is a random variable which is the transcript of the protocol π with respect to μ. That is, Π is the concatenation of all the messages exchanged during the execution of π. The ϵ information cost of a computational task f with respect to a distribution μ is defined as

$$IC_\mu(f, \epsilon) = \inf_\pi IC_\mu(\pi),$$

where the infimum ranges over all protocols π that solve f with error at most ϵ on inputs that are sampled according to μ.

4.3 Relative Entropy

Definition 2 (Relative Entropy). *Let $\mu_1, \mu_2 : \Omega \to [0, 1]$ be two distributions, where Ω is discrete (but not necessarily finite). The* relative entropy *between μ_1 and μ_2, denoted $\mathbf{D}(\mu_1 \| \mu_2)$, is defined as*

$$\mathbf{D}(\mu_1 \| \mu_2) = \sum_{x \in \Omega} \mu_1(x) \log \left(\tfrac{\mu_1(x)}{\mu_2(x)} \right).$$

Proposition 3. *Let $\mu_1, \mu_2 : \Omega \to [0, 1]$ be two distributions. Then,*

$$\mathbf{D}(\mu_1 \| \mu_2) \geq 0.$$

The following relation is called Pinsker's inequality, and it relates the relative entropy to the ℓ_1 distance.

Proposition 4 (Pinsker's Inequality). *Let $\mu_1, \mu_2 : \Omega \to [0, 1]$ be two distributions. Then,*

$$2 \ln(2) \cdot \mathbf{D}(\mu_1 \| \mu_2) \geq \|\mu_1 - \mu_2\|^2,$$

where

$$\|\mu_1 - \mu_2\| = \sum_{x \in \Omega} |\mu_1(x) - \mu_2(x)| = 2 \max_{E \subseteq \Omega} \{\mu_1(E) - \mu_2(E)\}.$$

4.4 Information

Definition 3 (Information). *Let $\mu : \Omega \to [0, 1]$ be a distribution and let $\mathcal{U}$ be the uniform distribution over Ω. The* information *of μ, denoted $\mathbf{I}(\mu)$, is defined by*

$$\mathbf{I}(\mu) = \mathbf{D}(\mu \parallel \mathcal{U}) = \sum_{x \in \mathrm{supp}(\mu)} \mu(x) \log \left(\frac{\mu(x)}{\frac{1}{|\Omega|}} \right) = \sum_{x \in \mathrm{supp}(\mu)} \mu(x) \log \left(|\Omega| \mu(x) \right).$$

Equivalently,

$$\mathbf{I}(\mu) = \log(|\Omega|) - \mathbf{H}(\mu),$$

where $\mathbf{H}(\mu)$ *denotes the Shannon entropy of* μ.

For a random variable X *taking values in* Ω*, with distribution* $P_X : \Omega \to [0,1]$*, we define* $\mathbf{I}(X) = \mathbf{I}(P_X)$.

Proposition 5 (**Supper-Additivity of Information**). *Let* $X_1, \ldots, X_m$ *be* m *random variables, taking values in* $\Omega_1, \ldots, \Omega_m$*, respectively. Consider the random variable* $(X_1, \ldots, X_m)$*, taking values in* $\Omega_1 \times \ldots \times \Omega_m$*. Then,*

$$\mathbf{I}\left((X_1, \ldots, X_m)\right) \geq \sum_{i \in [m]} \mathbf{I}(X_i).$$

Proof. Using the sub-additivity of the Shannon entropy function, we have

$$\begin{aligned}
\mathbf{I}\left((X_1, \ldots, X_m)\right) &= \log(|\Omega_1 \times \ldots \times \Omega_m|) - \mathbf{H}(X_1, \ldots, X_m) \\
&\geq \sum_{i \in [m]} \log(|\Omega_i|) - \sum_{i \in [m]} \mathbf{H}(X_i) \\
&= \sum_{i \in [m]} (\log(|\Omega_i|) - \mathbf{H}(X_i)) = \sum_{i \in [m]} \mathbf{I}(X_i).
\end{aligned}$$

□

4.5 Shearer-Like Inequality for Information

The following version of Shearer's inequality [CGFS86, Kah01] is due to [Rad03].

Lemma 6 (**Shearer's Inequality**). *Let* $X_1, \ldots, X_M$ *be* M *random variables. Let* $X = (X_1, \ldots, X_M)$*. Let* $T = \{T_i\}_{i \in I}$ *be a collection of subsets of* $[M]$*, such that each element of* $[M]$ *appears in at least* K *members of* T*. For* $A \subseteq [M]$*, let* $X_A = \{X_j : \ j \in A\}$*. Then,*

$$\sum_{i \in I} \mathbf{H}[X_{T_i}] \geq K \cdot \mathbf{H}[X].$$

We state and prove here the following "Shearer-like" inequality for information. A variant of this lemma was proved in [MT10].

Lemma 7 (**Shearer-Like Inequality for Information**). *Let* $X_1, \ldots, X_M$ *be* M *random variables, taking values in* $\Omega_1, \ldots, \Omega_M$*, respectively. Let* $X = (X_1, \ldots, X_M)$ *be a random variable, taking values in* $\Omega_1 \times \cdots \times \Omega_M$*. Let* $T = \{T_i\}_{i \in I}$ *be a collection of subsets of* $[M]$*, such that each element of* $[M]$ *appears in at most* $\frac{1}{K}$ *fraction of the members of* T*. For* $A \subseteq [M]$*, let* $X_A = \{X_j : \ j \in A\}$*. Then,*

$$K \cdot \mathop{\mathbf{E}}_{i \in_R I}[\mathbf{I}(X_{T_i})] \leq \mathbf{I}(X).$$

Proof. Fix $i \in I$. By the definition of information,

$$\mathbf{I}(X_{T_i}) = \sum_{j \in T_i} \log(|\Omega_j|) - \mathbf{H}[X_{T_i}].$$

For every $j \in [M]$, define $\mathbf{H}[X_j | X_{<j}] = \mathbf{H}[X_j | (X_\ell \ : \ \ell < j)]$. By the chain rule for the entropy function,

$$\begin{aligned} \mathbf{I}(X) &= \sum_{j \in [M]} \Big(\log(|\Omega_j|) - \mathbf{H}[X_j | X_{<j}] \Big), \\ \mathbf{I}(X_{T_i}) &= \sum_{j \in T_i} \Big(\log(|\Omega_j|) - \mathbf{H}[X_j | (X_\ell \ : \ \ell \in T_i, \ell < j)] \Big). \end{aligned}$$

For every $j \in T_i$ it holds that $\mathbf{H}[X_j | (X_\ell \ : \ \ell \in T_i, \ell < j)] \geq \mathbf{H}[X_j | X_{<j}]$. Therefore,

$$\mathbf{I}(X_{T_i}) \leq \sum_{j \in T_i} \Big(\log(|\Omega_j|) - \mathbf{H}[X_j | X_{<j}] \Big).$$

Summing over all $i \in I$ we get that

$$\sum_{i \in I} \mathbf{I}(X_{T_i}) \leq \sum_{i \in I} \sum_{j \in T_i} \Big(\log(|\Omega_j|) - \mathbf{H}[X_j | X_{<j}] \Big). \tag{9}$$

For every $j \in [M]$, the term $\log(|\Omega_j|) - \mathbf{H}[X_j | X_{<j}]$ appears on the right-hand side of Equation (9) at most $\frac{|I|}{K}$ times. Therefore,

$$\begin{aligned} \sum_{i \in I} \mathbf{I}(X_{T_i}) &\leq \frac{|I|}{K} \cdot \sum_{j \in [M]} \Big(\log(|\Omega_j|) - \mathbf{H}[X_j | X_{<j}] \Big) \\ &= \frac{|I|}{K} \cdot \mathbf{I}(X). \end{aligned}$$

Dividing by $\frac{|I|}{K}$ we get that the claim holds. ☐

The next lemma generalizes Lemma 7, and gives a Shearer-like inequality for relative entropy. A variant of this lemma was proved in [MT10]. The lemma will not be used in the paper, but we include it here as it may be useful in this context. The proof is given in Appendix A.

Lemma 8 (Shearer-Like Inequality for Relative Entropy). *Let $P, Q : \Omega_1 \times \cdots \times \Omega_M \to [0,1]$ be two distributions, such that Q is a product distribution, i.e., for every $j \in [M]$, there exists $Q_j : \Omega_j \to [0,1]$, such that $Q(x_1, \ldots, x_M) = \prod_{j \in [M]} Q_j(x_j)$. Let $T = \{T_i\}_{i \in I}$ be a collection of subsets of $[M]$, such that each element of $[M]$ appears in at most $\frac{1}{K}$ fraction of the members of T. For $A \subseteq [M]$, let P_A and Q_A be the marginal distributions of A in the distributions P and Q (respectively). Then,*

$$K \cdot \mathop{\mathbf{E}}_{i \in_R I} [\mathbf{D}(P_{T_i} \| Q_{T_i})] \leq \mathbf{D}(P \| Q).$$

5 The Graph Correlation Lemma

Lemma 9 (**Graph Correlation Lemma**). [1] *Let $G = (U \cup W, E)$ be a bipartite (multi)-graph with sets of vertices U, W and (multi)-set of edges E, such that, G is bi-regular and $|U| = |W|$. Let $M > T > k \in \mathbb{N}$ be such that, $T \leq 2^{-20k}M$, and $k \geq 4$. For every $(u, w) \in E$, let $T(u, w) \subset [M]$ be a set of size T, such that, for every $u \in U$, each element of $[M]$ appears in at most 2^{-20k} fraction of the sets in $\{T(u, w)\}_{(u,w)\in E}$, and for every $w \in W$, each element of $[M]$ appears in at most 2^{-20k} fraction of the sets in $\{T(u, w)\}_{(u,w)\in E}$.*

Let Σ be a finite set. For every $u \in U$, let $X^u \in \Sigma^M$ be a random variable, such that, $\mathbf{I}(X^u) \leq 2^{4k}$, and for every $w \in W$, let $Y^w \in \Sigma^M$ be a random variable, such that, $\mathbf{I}(Y^w) \leq 2^{4k}$, and such that, for every $u \in U$ and $w \in W$, the random variables X^u and Y^w are mutually independent.

For $(u, w) \in E$, denote

$$\mu(u, w) = \frac{\Pr_{X^u, Y^w}[X^u_{T(u,w)} = Y^w_{T(u,w)}]}{|\Sigma|^{-T}}.$$

Let

$$\mathcal{D} = \{(u, w) \in E \ : \ \mu(u, w) \leq 1 - 2^{-4k}\}.$$

Then,

$$\frac{|\mathcal{D}|}{|E|} \leq 2^{-4k}.$$

Proof. We will start by proving the following claim.

Claim 10. *If $(u, w) \in \mathcal{D}$ then at least one of the following two inequalities holds,*

$$\mathbf{I}\left(X^u_{T(u,w)}\right) \geq 2^{-8k-4},$$

$$\mathbf{I}\left(Y^w_{T(u,w)}\right) \geq 2^{-8k-4}.$$

Proof. Assume $(u, w) \in \mathcal{D}$. Thus,

$$-2^{-4k} \geq \mu(u, w) - 1 = |\Sigma|^T \cdot \left(\Pr_{X^u, Y^w}[X^u_{T(u,w)} = Y^w_{T(u,w)}] - |\Sigma|^{-T} \right) =$$

$$|\Sigma|^T \cdot \left(\left(\sum_{z \in \Sigma^{T(u,w)}} \Pr_{X^u}[X^u_{T(u,w)} = z] \cdot \Pr_{Y^w}[Y^w_{T(u,w)} = z] \right) - |\Sigma|^{-T} \right) =$$

$$|\Sigma|^T \cdot \sum_{z \in \Sigma^{T(u,w)}} \left(\Pr_{X^u}[X^u_{T(u,w)} = z] - |\Sigma|^{-T} \right) \cdot \left(\Pr_{Y^w}[Y^w_{T(u,w)} = z] - |\Sigma|^{-T} \right). \tag{10}$$

[1] Many variants of this lemma can be proven. In particular, a similar argument can be used to prove a similar statement with sets $T(u, w)$ that are not of the same size. We state the lemma here for sets $T(u, w)$ of the same size T, for convenience of notation.

In the last sum, we can omit the positive summands (and the inequality still holds). As for the negative summands, we split them into summands where $(\Pr[X^u_{T(u,w)} = z] - |\Sigma|^{-T})$ is negative and $(\Pr[Y^w_{T(u,w)} = z] - |\Sigma|^{-T})$ is positive, and summands where it's the other way around. In the first case, we bound the first term by

$$\left(\Pr_{X^u}[X^u_{T(u,w)} = z] - |\Sigma|^{-T}\right) \geq -|\Sigma|^{-T},$$

and for the second term, we use

$$\left(\Pr_{Y^w}[Y^w_{T(u,w)} = z] - |\Sigma|^{-T}\right) = \left|\Pr_{Y^w}[Y^w_{T(u,w)} = z] - |\Sigma|^{-T}\right|.$$

Similarly, in the second case, we bound the terms the other way around. Note also that we can add to the sum arbitrary negative summands (and the inequality still holds). Thus, Equation (10) implies

$$\begin{aligned}
-2^{-4k} \geq \quad & |\Sigma|^T \cdot \sum_{z \in \Sigma^{T(u,w)}} \left(-|\Sigma|^{-T}\right) \cdot \left|\Pr_{Y^w}[Y^w_{T(u,w)} = z] - |\Sigma|^{-T}\right| + \\
& |\Sigma|^T \cdot \sum_{z \in \Sigma^{T(u,w)}} \left|\Pr_{X^u}[X^u_{T(u,w)} = z] - |\Sigma|^{-T}\right| \cdot \left(-|\Sigma|^{-T}\right) = \\
& - \sum_{z \in \Sigma^{T(u,w)}} \left|\Pr_{Y^w}[Y^w_{T(u,w)} = z] - |\Sigma|^{-T}\right| - \sum_{z \in \Sigma^{T(u,w)}} \left|\Pr_{X^u}[X^u_{T(u,w)} = z] - |\Sigma|^{-T}\right|,
\end{aligned}$$

that is,

$$\sum_{z \in \Sigma^{T(u,w)}} \left|\Pr_{Y^w}[Y^w_{T(u,w)} = z] - |\Sigma|^{-T}\right| + \sum_{z \in \Sigma^{T(u,w)}} \left|\Pr_{X^u}[X^u_{T(u,w)} = z] - |\Sigma|^{-T}\right| \geq 2^{-4k}.$$

Hence, for every $(u, w) \in \mathcal{D}$, at least one of the following two inequalities holds,

$$\sum_{z \in \Sigma^{T(u,w)}} \left|\Pr_{X^u}[X^u_{T(u,w)} = z] - |\Sigma|^{-T}\right| \geq 2^{-4k-1},$$

$$\sum_{z \in \Sigma^{T(u,w)}} \left|\Pr_{Y^w}[Y^w_{T(u,w)} = z] - |\Sigma|^{-T}\right| \geq 2^{-4k-1}.$$

The claim follows by Pinsker's inequality. □

We will now proceed with the proof of Lemma 9. By Claim 10, we know that one of the following two statements must hold:

1. For at least half of the edges $(u, w) \in \mathcal{D}$, we have $\mathbf{I}\left(X^u_{T(u,w)}\right) \geq 2^{-8k-4}$.

2. For at least half of the edges $(u, w) \in \mathcal{D}$, we have $\mathbf{I}\left(Y^w_{T(u,w)}\right) \geq 2^{-8k-4}$.

Without loss of generality, assume that the first statement holds.

Assume for a contradiction that

$$\frac{|\mathcal{D}|}{|E|} > 2^{-4k}.$$

Thus, by an averaging argument, there exists $u \in U$, such that, for at least 2^{-4k-1} fraction of the edges $(u, w) \in E$, we have $\mathbf{I}\left(X^u_{T(u,w)}\right) \geq 2^{-8k-4}$. Fix $u \in U$ that has this property. Denote by $E(u)$ the (multi)-set of edges in E that contain u, that is, $E(u) = \{(u, w) : (u, w) \in E\}$. Thus,

$$\mathop{\mathbf{E}}_{(u,w)\in_R E(u)} \left[\mathbf{I}\left(X^u_{T(u,w)}\right)\right] \geq 2^{-4k-1} \cdot 2^{-8k-4} = 2^{-12k-5}.$$

Since each element of $[M]$ appears in at most 2^{-20k} fraction of the sets in $\{T(u, w)\}_{(u,w)\in E(u)}$, we have by Lemma 7,

$$\mathbf{I}(X^u) \geq 2^{-12k-5} \cdot 2^{20k} = 2^{8k-5},$$

in contradiction to the assumption of the lemma. □

6 Communication Lower Bound

In this section we prove Theorem 1. Assume that π is a deterministic communication protocol for the bursting noise game with parameter k, that has communication complexity at most 2^k. The rest of this section is devoted to showing that π has error $\epsilon \geq 1-o(1)$ (when the inputs are selected according to the distribution μ). That is, the protocol almost always errs. Observe that this also implies that every probabilistic protocol errs with probability $\epsilon \geq 1 - o(1)$, as it is a distribution over deterministic protocols.

6.1 Notation

Let $\{R^1, \ldots, R^m\}$ be the rectangle partition induced by the protocol π, where $R^t = A^t \times B^t$ for $A^t, B^t \subseteq \{0, 1\}^V$ and $m \leq 2^{2^k}$. Let $t \in [m]$. Let X^t be a random variable taking values in $\{0, 1\}^V$, that is uniformly distributed over A^t. Let Y^t be a random variable taking values in $\{0, 1\}^V$, that is uniformly distributed over B^t.

Let $i \in [c]$ be a multi-layer. Define $V_{<i} \subseteq V$ to be the set of vertices in multi-layers 1 to $i - 1$. Define $V_i \subseteq V$ to be the set of vertices in multi-layer i. Define $V_{\geq i} \subseteq V$ to be the set of vertices in multi-layers i to c. For Z that is either a random variable taking values in $\{0, 1\}^V$ or an element in $\{0, 1\}^V$, we define $Z_{<i}$, Z_i, $Z_{\geq i}$ to be the projections of Z to $V_{<i}$, V_i, $V_{\geq i}$ (respectively).

Let $i \in [c]$ and $z \in \{0, 1\}^{V_{<i}}$. Define Ψ^z to be the set of all elements $\psi \in \{0, 1\}^V$ with $\psi_{<i} = z$. It holds that $|\Psi^z| = |\{0, 1\}^{V_{\geq i}}|$.

Let $i \in [c]$, $z \in \{0, 1\}^{V_{<i}}$ and $t \in [m]$. Define $A^{t,z} = A^t \cap \Psi^z$ and $B^{t,z} = B^t \cap \Psi^z$. Define $R^{t,z} = A^{t,z} \times B^{t,z}$. Let $X^{t,z}$ be a random variable taking values in Ψ^z, that is uniformly

distributed over $A^{t,z}$. Let $Y^{t,z}$ be a random variable taking values in Ψ^z, that is uniformly distributed over $B^{t,z}$.

6.2 Bounding the Information on the Noisy Multi-Layer

Let $i \in [c]$ and $z \in \{0,1\}^{V_{<i}}$. We define $\rho^{i,z} : [m] \to [0,1]$ to be the distribution that selects a rectangle index $t \in [m]$ as follows: Randomly select an input pair $(x,y) \in \Psi^z \times \Psi^z$. Select t to be the index of the unique rectangle R^t containing (x,y). That is, $\rho^{i,z}(t)$ is the density of the rectangle $R^{t,z}$ with respect to input pairs that agree with z,

$$\rho^{i,z}(t) = \frac{|R^{t,z}|}{|\Psi^z \times \Psi^z|}.$$

The following lemma shows that, in expectation, the distribution of the projections of inputs in $R^{t,z}$ to multi-layer i is close to uniform.

Lemma 11. *It holds that*

$$\mathop{\mathbf{E}}_{i \in_R [c]} \; \mathop{\mathbf{E}}_{z \in_R \{0,1\}^{V_{<i}}} \; \mathop{\mathbf{E}}_{t \leftarrow \rho^{i,z}} \left[\mathbf{I}\left(X_i^{t,z}\right) \right] \le \frac{m}{c},$$

and similarly,

$$\mathop{\mathbf{E}}_{i \in_R [c]} \; \mathop{\mathbf{E}}_{z \in_R \{0,1\}^{V_{<i}}} \; \mathop{\mathbf{E}}_{t \leftarrow \rho^{i,z}} \left[\mathbf{I}\left(Y_i^{t,z}\right) \right] \le \frac{m}{c}.$$

Proof. Fix $i \in [c]$. It holds that

$$\begin{aligned}
&\mathop{\mathbf{E}}_{z \in_R \{0,1\}^{V_{<i}}} \; \mathop{\mathbf{E}}_{t \leftarrow \rho^{i,z}} \left[\mathbf{I}\left(X_i^{t,z}\right) \right] \\
&= \sum_{z \in \{0,1\}^{V_{<i}}} \frac{1}{|\{0,1\}^{V_{<i}}|} \sum_{t \in [m]} \frac{|R^{t,z}|}{|\{0,1\}^{V_{\ge i}}|^2} \cdot \mathbf{I}\left(X_i^{t,z}\right) \\
&= \sum_{z \in \{0,1\}^{V_{<i}}} \frac{1}{|\{0,1\}^{V}|} \sum_{t \in [m]} \frac{|A^{t,z}| \cdot |B^{t,z}|}{|\{0,1\}^{V_{\ge i}}|} \cdot \mathbf{I}\left(X_i^{t,z}\right) \\
&\le \sum_{z \in \{0,1\}^{V_{<i}}} \frac{1}{|\{0,1\}^{V}|} \sum_{t \in [m]} \frac{|A^{t,z}| \cdot |\{0,1\}^{V_{\ge i}}|}{|\{0,1\}^{V_{\ge i}}|} \cdot \mathbf{I}\left(X_i^{t,z}\right) \\
&= \sum_{t \in [m]} \frac{1}{|\{0,1\}^{V}|} \sum_{z \in \{0,1\}^{V_{<i}}} |A^{t,z}| \cdot \mathbf{I}\left(X_i^{t,z}\right) \\
&= \sum_{t \in [m]} \frac{|A^t|}{|\{0,1\}^{V}|} \sum_{z \in \{0,1\}^{V_{<i}}} \frac{|A^{t,z}|}{|A^t|} \cdot \mathbf{I}\left(X_i^{t,z}\right) \\
&= \sum_{t \in [m]} \frac{|A^t|}{|\{0,1\}^{V}|} \sum_{z \in \{0,1\}^{V_{<i}}} \frac{|A^{t,z}|}{|A^t|} \left(|V_i| - \mathbf{H}\left(X_i^{t,z}\right)\right).
\end{aligned}$$

Denote

$$s := \sum_{t\in[m]} \frac{|A^t|}{|\{0,1\}^V|} = \sum_{t\in[m]} \sum_{z\in\{0,1\}^{V_{<i}}} \frac{|A^{t,z}|}{|\{0,1\}^V|}.$$

Observe that $1 \leq s \leq m$. We have that

$$\begin{aligned}
&\mathop{\mathbf{E}}_{z\in_R\{0,1\}^{V_{<i}}} \mathop{\mathbf{E}}_{t\leftarrow\rho^{i,z}} \left[\mathbf{I}\left(X_i^{t,z}\right)\right] \\
&\leq s|V_i| - \sum_{t\in[m]} \frac{|A^t|}{|\{0,1\}^V|} \sum_{z\in\{0,1\}^{V_{<i}}} \frac{|A^{t,z}|}{|A^t|} \cdot \mathbf{H}\left(X_i^{t,z}\right) \\
&= s|V_i| - \sum_{t\in[m]} \frac{|A^t|}{|\{0,1\}^V|} \sum_{z\in\{0,1\}^{V_{<i}}} \frac{|A^{t,z}|}{|A^t|} \cdot \mathbf{H}\left(X_i^t|X_{<i}^t = z\right) \\
&= s|V_i| - \sum_{t\in[m]} \frac{|A^t|}{|\{0,1\}^V|} \mathop{\mathbf{E}}_{z\leftarrow X_{<i}^t} \left[\mathbf{H}\left(X_i^t|X_{<i}^t = z\right)\right] \\
&= s|V_i| - \sum_{t\in[m]} \frac{|A^t|}{|\{0,1\}^V|} \cdot \mathbf{H}\left(X_i^t|X_{<i}^t\right).
\end{aligned}$$

By the chain rule for the entropy function,

$$\begin{aligned}
&\mathop{\mathbf{E}}_{i\in_R[c]} \mathop{\mathbf{E}}_{z\in_R\{0,1\}^{V_{<i}}} \mathop{\mathbf{E}}_{t\leftarrow\rho^{i,z}} \left[\mathbf{I}\left(X_i^{t,z}\right)\right] \\
&\leq \mathop{\mathbf{E}}_{i\in_R[c]} \left[s|V_i| - \sum_{t\in[m]} \frac{|A^t|}{|\{0,1\}^V|} \cdot \mathbf{H}\left(X_i^t|X_{<i}^t\right)\right] \\
&= \mathop{\mathbf{E}}_{i\in_R[c]} [s|V_i|] - \sum_{t\in[m]} \frac{|A^t|}{|\{0,1\}^V|} \cdot \mathop{\mathbf{E}}_{i\in_R[c]} \left[\mathbf{H}\left(X_i^t|X_{<i}^t\right)\right] \\
&= \frac{s|V|}{c} - \frac{1}{c} \sum_{t\in[m]} \frac{|A^t|}{|\{0,1\}^V|} \sum_{i\in[c]} \left[\mathbf{H}\left(X_i^t|X_{<i}^t\right)\right] \\
&= \frac{s}{c}\left(|V| - \sum_{t\in[m]} \frac{|A^t|}{s|\{0,1\}^V|} \cdot \mathbf{H}(X^t)\right) \\
&= \frac{s}{c}\left(|V| + \sum_{t\in[m]} \frac{|A^t|}{s|\{0,1\}^V|} \cdot \log\left(\frac{1}{|A^t|}\right)\right) \\
&\leq \frac{s}{c}\left(|V| + \log\left(\sum_{t\in[m]} \frac{|A^t|}{s|\{0,1\}^V|} \cdot \frac{1}{|A^t|}\right)\right) \\
&\leq \frac{s}{c} \log\left(\frac{m}{s}\right) \\
&\leq \frac{m}{c},
\end{aligned}$$

where the third to last inequality is by the concavity of the log function. The last inequality holds as $-x\log(x) < 1$ for $x \in [0,1]$. $\square$

6.3 Unique Answer Rectangles

Lemma 12 (**Unique Answer Lemma**). *Let A be a set of inputs for the first player, B be a set of inputs for the second player, and Ω be a set of possible outputs. Let $\pi_1 : A \to \Omega$, $\pi_2 : B \to \Omega$ be any functions determining the players' outputs.*

Let $\gamma > 0$. There exist a partition of A into a disjoint union $A = A^1 \cup \cdots \cup A^\ell$, a partition of B into a disjoint union $B = B^1 \cup \cdots \cup B^\ell$, where $\ell = O(1/\gamma^4)$, and for every $s_1, s_2 \in [\ell]$ there exist functions $\pi_1^{s_1,s_2} : A^{s_1} \to \Omega$, $\pi_2^{s_1,s_2} : B^{s_2} \to \Omega$, such that the followings hold: Denote $R^{s_1,s_2} = A^{s_1} \times B^{s_2}$.

1. *Let $s_1, s_2 \in [\ell]$ and let $(x,y) \in R^{s_1,s_2}$. If $\pi_1(x) = \pi_2(y)$ then*
$$\pi_1^{s_1,s_2}(x) = \pi_2^{s_1,s_2}(y) = \pi_1(x) = \pi_2(y).$$

2. *For $s_1, s_2 \in [\ell]$, we say that the rectangle R^{s_1,s_2} is a* unique answer rectangle *if there exists $\omega \in \Omega$ such that for every $(x,y) \in R^{s_1,s_2}$, it holds that $\pi_1^{s_1,s_2}(x) = \pi_2^{s_1,s_2}(y) = \omega$. Let S be the union of all unique answer rectangles R^{s_1,s_2}, where $s_1, s_2 \in [\ell]$. Then,*
$$\Pr_{(x,y)\in_R A\times B}[(x,y) \notin S] \leq \gamma.$$

3. *For $s_1, s_2 \in [\ell]$, we say that the rectangle R^{s_1,s_2} is a γ-*large rectangle *if $|A^{s_1}||B^{s_2}| \geq \frac{\gamma^4}{10^4}|A||B|$. Let L be the union of all γ-large rectangles R^{s_1,s_2}, where $s_1, s_2 \in [\ell]$. Then,*
$$\Pr_{(x,y)\in_R A\times B}[(x,y) \notin L] \leq \gamma.$$

Proof. For $\Omega' \subseteq \Omega$, define

$$\begin{aligned}
p_1(\Omega') &= \Pr_{x\in_R A}[\pi_1(x) \in \Omega'],\\
p_2(\Omega') &= \Pr_{y\in_R B}[\pi_2(y) \in \Omega'],\\
A(\Omega') &= \{x \in A \ : \ \pi_1(x) \in \Omega'\},\\
B(\Omega') &= \{y \in B \ : \ \pi_2(y) \in \Omega'\}.
\end{aligned}$$

We define the partitions of A and B as follows: For every $\omega \in \Omega$ such that either $p_1(\{\omega\}) \geq \frac{\gamma}{10}$ or $p_2(\{\omega\}) \geq \frac{\gamma}{10}$, add $A(\{\omega\})$ to the partition of A and $B(\{\omega\})$ to the partition of B. So far, we added at most $\frac{20}{\gamma}$ sets to each partition. Let

$$T = \left\{\omega \in \Omega \ : \ p_1(\{\omega\}), p_2(\{\omega\}) < \frac{\gamma}{10}\right\}.$$

Let $T_1, \ldots, T_t$ be a minimal partition of T, such that for every $j \in [t]$, we have

$$p_1(T_j), p_2(T_j) < \frac{\gamma}{10}.$$

Since T is minimal, there is at most one set T_j in T with both $p_1(T_j), p_2(T_j) < \frac{\gamma}{20}$ (as if there were two such sets we could have merged them). Therefore, $t \leq 2 \cdot \frac{20}{\gamma} + 1$. For every $i \in [t]$, add $A(T_i)$ to the partition of A and $B(T_i)$ to the partition of B. This concludes the definition of the partitions of A and B, where the number of sets in each partition, denoted by ℓ, is at most $\frac{20}{\gamma} + \frac{40}{\gamma} + 1 \leq \frac{70}{\gamma}$.

Fix $s_1, s_2 \in [\ell]$. Let $\Omega_A = \{\pi_1(x) \ : \ x \in A^{s_1}\}$ and $\Omega_B = \{\pi_2(y) \ : \ y \in B^{s_2}\}$. For every $(x, y) \in A^{s_1} \times B^{s_2}$, we define $\pi_1^{s_1,s_2}(x)$ and $\pi_2^{s_1,s_2}(y)$ by the following steps (once the conditions of a step are fulfilled and the outputs of the functions are defined we do not continue to the next step):

1. If $p_1(\Omega_A) \geq \frac{\gamma}{10}$, then Ω_A contains a single value ω. Define $\pi_1^{s_1,s_2}(x) = \pi_2^{s_1,s_2}(y) = \omega$.
2. If $p_2(\Omega_B) \geq \frac{\gamma}{10}$, then Ω_B contains a single value ω. Define $\pi_1^{s_1,s_2}(x) = \pi_2^{s_1,s_2}(y) = \omega$.
3. If $\Omega_A \cap \Omega_B \neq \emptyset$, define $\pi_1^{s_1,s_2}(x) = \pi_1(x)$ and $\pi_2^{s_1,s_2}(y) = \pi_2(y)$.
4. Define $\pi_1^{s_1,s_2}(x) = \pi_2^{s_1,s_2}(y) = w$ for some fixed $w \in \Omega$.

We prove that the three requirements of the lemma are met:

1. Assume that $\pi_1(x) = \pi_2(y) = \omega$. Then, $\omega \in \Omega_A \cap \Omega_B$. Thus, the outputs $\pi_1^{s_1,s_2}(x)$ and $\pi_2^{s_1,s_2}(y)$ are defined in one of the first three steps, and therefore, $\pi_1^{s_1,s_2}(x) = \pi_2^{s_1,s_2}(y) = \omega$.
2. Assume that R^{s_1,s_2} is not a unique answer rectangle. Then, there exists $(x, y) \in R^{s_1,s_2}$ such that $\pi_1^{s_1,s_2}(x) \neq \pi_2^{s_1,s_2}(y)$. Then, the outputs are defined in Step 3, and it holds that $p_1(\Omega_A), p_2(\Omega_B) < \frac{\gamma}{10}$ and $\Omega_A \cap \Omega_B \neq \emptyset$. By the definition of the partitions of A and B, there exists $i \in [t]$ such that $\Omega_A, \Omega_B \subseteq T_i$ (we mention that if there is no $\omega \in T_i$ with $p_1(\omega) = 0$ or $p_2(\omega) = 0$ then $\Omega_A = \Omega_B = T_i$). Therefore,
$$\Pr_{(x,y) \in_R A \times B}[(x, y) \notin S] \leq \sum_{i \in [t]} p_1(T_i) \cdot p_2(T_i) \leq \frac{\gamma}{10} \sum_{i \in [t]} p_1(T_i) \leq \frac{\gamma}{10}.$$
3. Since the number of rectangles R^{s_1,s_2}, where $s_1, s_2 \in [\ell]$, is $\ell^2 \leq \frac{70^2}{\gamma^2}$,
$$\Pr_{(x,y) \in_R A \times B}[(x, y) \notin L] \leq \frac{70^2}{\gamma^2} \cdot \frac{\gamma^4}{10^4} \leq \gamma.$$

□

6.4 Good Rectangles

Fix $\gamma = \gamma(k) > 0$ to be sub-constant (i.e., $\gamma = o(1)$), and such that $\gamma > \frac{1}{k^{o(1)}}$. Let $i \in [c]$. We say that i is *good* if

$$\mathop{\mathbf{E}}_{z \in_R \{0,1\}^{V_{<i}}} \mathop{\mathbf{E}}_{t \leftarrow \rho^{i,z}} \left[\mathbf{I}\left(X_i^{t,z}\right) \right] \leq \frac{m}{\gamma c}. \tag{11}$$

$$\mathop{\mathbf{E}}_{z \in_R \{0,1\}^{V_{<i}}} \mathop{\mathbf{E}}_{t \leftarrow \rho^{i,z}} \left[\mathbf{I}\left(Y_i^{t,z}\right) \right] \leq \frac{m}{\gamma c}. \tag{12}$$

By Markov's inequality and Lemma 11,

$$\Pr_{i \in_R [c]} [\ i \text{ is good }] \geq 1 - 2\gamma. \tag{13}$$

For the rest of the lower bound proof, fix a good $i \in [c]$.

For every $z \in \{0,1\}^{V_{<i}}$ and $t \in [m]$, consider the rectangle $R^{t,z} = A^{t,z} \times B^{t,z}$. Apply Lemma 12 to the sets $A^{t,z}$, $B^{t,z}$, where the functions π_1, π_2 are the outputs of the two players for the rectangle R^t, in the protocol π. Lemma 12 partitions each rectangle $R^{t,z}$ into $\ell = O(1/\gamma^8)$ new rectangles, denoted $R^{t,1,z}, \ldots, R^{t,\ell,z}$. Observe that $\{R^{t,s,z}\}_{t \in [m], s \in [\ell]}$ is a cover of $\Psi^z \times \Psi^z$,

$$\bigcup_{\substack{t \in [m] \\ s \in [\ell]}} R^{t,s,z} = \Psi^z \times \Psi^z. \tag{14}$$

For $t \in [m]$ and $s \in [\ell]$, denote $R^{t,s,z} = A^{t,s,z} \times B^{t,s,z}$, where $A^{t,s,z}, B^{t,s,z} \subseteq \Psi^z$. Let $X^{t,s,z}$ be a random variable taking values in Ψ^z, that is uniformly distributed over $A^{t,s,z}$. Let $Y^{t,s,z}$ be a random variable taking values in Ψ^z, that is uniformly distributed over $B^{t,s,z}$.

Let $z \in \{0,1\}^{V_{<i}}$. We define $\eta^{i,z} : [m] \times [\ell] \to [0,1]$ to be the distribution that selects rectangle indices $(t,s) \in [m] \times [\ell]$ as follows: Randomly select an input pair $(x,y) \in \Psi^z \times \Psi^z$. Select (t,s) to be the indices of the unique rectangle $R^{t,s,z}$ containing (x,y). That is, $\eta^{i,z}(t,s)$ is the density of the rectangle $R^{t,s,z}$ with respect to input pairs that agree with z,

$$\eta^{i,z}(t,s) = \frac{|R^{t,s,z}|}{|\Psi^z \times \Psi^z|}.$$

Observe that for every $t \in [m]$ it holds that

$$\rho^{i,z}(t) = \sum_{s \in [\ell]} \eta^{i,z}(t,s).$$

We say that (i,z,t,s) is *good* if all the followings holds:

1. $R^{t,s,z}$ is a unique answer rectangle (as in Lemma 12). Recall that for each unique answer rectangle $R^{t,s,z}$, there is a unique leaf in V, denoted $\omega^{t,s,z}$, returned as an output by both players on all input pairs in the rectangle $R^{t,s,z}$.

2. $R^{t,s,z}$ is γ-large (as in Lemma 12). That is, $|R^{t,s,z}| \geq \frac{\gamma^4}{10^4}|R^{t,z}|$.

3. $\mathbf{I}\left(X^{t,s,z}\right) \leq 2\log(m)$, and therefore also, $\mathbf{I}\left(X^{t,z}\right) \leq 2\log(m)$.

4. $\mathbf{I}\left(Y^{t,s,z}\right) \leq 2\log(m)$, and therefore also, $\mathbf{I}\left(Y^{t,z}\right) \leq 2\log(m)$.

5. $\mathbf{I}\left(X_i^{t,z}\right) \leq \frac{m}{\gamma^2 c}$.

6. $\mathbf{I}\left(Y_i^{t,z}\right) \leq \frac{m}{\gamma^2 c}$.

7. $\mathbf{I}\left(X_i^{t,s,z}\right) \leq O\left(\log(1/\gamma)\right)$.

8. $\mathbf{I}\left(Y_i^{t,s,z}\right) \leq O\left(\log(1/\gamma)\right)$.

Lemma 13. *It holds that*

$$\Pr_{\substack{z \in_R \{0,1\}^{V_{<i}},\\ (t,s) \leftarrow \eta^{i,z}}} \left[\ (i,z,t,s) \text{ is good}\ \right] \geq 1 - O(\gamma).$$

Proof. We claim that each of the eight requirements in the definition of a good tuple (i,z,t,x) is violated with probability $O(\gamma)$:

1. By the second item of Lemma 12.

2. By the third item of Lemma 12.

3. If $\mathbf{I}\left(X^{t,s,z}\right) > 2\log(m)$, then $\eta^{i,z}(t,s) = \frac{|R^{t,s,z}|}{|\Psi^z \times \Psi^z|} \leq 1/m^2$. Since for every z there are at most $m \cdot \ell$ rectangles $R^{t,s,z}$, the $\eta^{i,z}$-measure of all such rectangles is at most $\ell/m < O(\gamma)$.

4. Same.

5. Since i is good, follows from Equation (11) and Markov's inequality.

6. Since i is good, follows from Equation (12) and Markov's inequality.

7. Since the the second and fifth properties of a good (i,z,t,s) imply this (seventh) property as follows. Let τ and τ' be the distributions of the random variables $X_i^{t,s,z}$ and $X_i^{t,z}$ (respectively). By the second property of a good tuple (i,z,t,s) (i.e., γ-largeness), for every $\omega \in \{0,1\}^{V_i}$ it holds that $\tau(\omega) \leq \frac{10^4}{\gamma^4} \cdot \tau'(\omega)$. We denote $\Psi = \{0,1\}^{V_i}$,

$\Psi^{pos} = \{\omega \in \Psi : \log(|\Psi| \cdot \tau'(\omega)) \geq 0\}$ and $\Psi^{neg} = \Psi \setminus \Psi^{pos}$. Therefore,

$$\begin{aligned}
\mathbf{I}\left(X_i^{t,s,z}\right) = \mathbf{I}(\tau) &= \sum_{\omega \in \Psi} \tau(\omega) \log(|\Psi| \cdot \tau(\omega)) \\
&\leq \sum_{\omega \in \Psi} \tau(\omega) \log\left(|\Psi| \cdot \frac{10^4}{\gamma^4} \cdot \tau'(\omega)\right) \\
&\leq O(\log(1/\gamma)) + \sum_{\omega \in \Psi} \tau(\omega) \log(|\Psi| \cdot \tau'(\omega)) \\
&\leq O(\log(1/\gamma)) + \sum_{\omega \in \Psi^{pos}} \tau(\omega) \log(|\Psi| \cdot \tau'(\omega)) \\
&\leq O(\log(1/\gamma)) + \frac{10^4}{\gamma^4} \sum_{\omega \in \Psi^{pos}} \tau'(\omega) \log(|\Psi| \cdot \tau'(\omega)).
\end{aligned}$$

By the fifth property of a good tuple (i, z, t, s), it holds that

$$\mathbf{I}(\tau') = \mathbf{I}\left(X_i^{t,z}\right) \leq \frac{m}{\gamma^2 c} < 0.01,$$

and thus by Lemma 5.11 in [KR13] (stated for convenience in Appendix A, Lemma 24)

$$-\sum_{\omega \in \Psi^{neg}} \tau'(\omega) \log(|\Psi| \cdot \tau'(\omega)) < 4\mathbf{I}(\tau')^{0.1}.$$

Therefore,

$$\begin{aligned}
\mathbf{I}\left(X_i^{t,s,z}\right) &< O(\log(1/\gamma)) + \frac{10^4}{\gamma^4}\left(\sum_{\omega \in \Psi} \tau'(\omega) \log(|\Psi| \cdot \tau'(\omega)) + 4\mathbf{I}(\tau')^{0.1}\right) \\
&= O(\log(1/\gamma)) + \frac{10^4}{\gamma^4}\left(\mathbf{I}(\tau') + 4\mathbf{I}(\tau')^{0.1}\right) \\
&\leq O(\log(1/\gamma)).
\end{aligned}$$

8. Since the the second and sixth properties of a good (i, z, t, s) imply this (eighth) property, as above.

□

We recall that for every unique answer rectangle $R^{t,s,z}$ and for every $(x, y) \in R^{t,s,z}$, if the protocol π is correct on (x, y), then in the protocol π both players output $\omega^{t,s,z}$ on the input (x, y). The reason is that if the protocol π is correct on (x, y), then, in particular, both players return the same output ω on (x, y) in the protocol π. In this case, by the first item in Lemma 12, the output on the rectangle $R^{t,s,z}$ is the same as the original output, i.e., $\omega = \omega^{t,s,z}$.

6.5 The Super Graph

Let $G = (U \cup W, E)$ be the bipartite graph with sets of vertices U, W and set of edges E, defined as follows: Recall that we fixed a good i. Let $U = W = \{0,1\}^{V_i}$. For $u \in U, w \in W$, we have $(u,w) \in E$ if there exists $(x,y) \in \text{supp}(\mu_i)$, such that $x_i = u$ and $y_i = w$.

Let M be the number of vertices in layer $(i+1)^*$ of the tree $\mathcal{T}$. We identify the set $[M]$ with the set of vertices in layer $(i+1)^*$. Let $u \in U, w \in W$, such that $(u,w) \in E$. We define $T(u,w) \subset [M]$ to be the set of all vertices in layer $(i+1)^*$ that are set to be non-noisy for inputs u, w, by Algorithm 1 defining μ, when the noisy multi-layer is i. Observe that u and w determine for every vertex in layer $(i+1)^*$ if it is noisy or not. By a symmetry argument, G is bi-regular and $T(u,w)$ is of the same size T for every u, w. Let Σ be the set of all possible boolean assignments to the vertices of a subtree of $\mathcal{T}$ rooted at layer $(i+1)^*$.

We observe that an input pair (x,y) with $x_{<i} = y_{<i}$ and $(x_i, y_i) \in E$, is in $\text{supp}(\mu_i)$, if and only if x and y agree on the subtrees rooted at each of the vertices in $T(x_i, y_i)$ (these are non-noisy subtrees). Therefore,

$$\Pr_{(x,y) \in_R \Psi_z \times \Psi_z} [(x,y) \in \text{supp}(\mu_i)] = |\Sigma|^{-T} \frac{|E|}{|U| \cdot |W|}. \tag{15}$$

Let $\mathcal{G}_i$ be the set of all $(t,s,z) \in [m] \times [\ell] \times \{0,1\}^{V_{<i}}$ such that (i,z,t,s) is good (see Section 6.4). Let $(t,s,z) \in \mathcal{G}_i$. Let $\mathcal{E}^{t,s,z} \subseteq E$ be the set of all $(u,w) \in E$ for which the output $\omega^{t,s,z}$ is correct for inputs $(x,y) \in \text{supp}(\mu_i)$, with $x_i = u$ and $y_i = w$. Note that if the noise is taken on the i^{th} multi-layer, then u and w determine the correctness of $\omega^{t,s,z}$. Let

$$\Lambda^{t,s,z} = \left\{ (x,y) \in \text{supp}(\mu_i) \ : \ (x_i, y_i) \in E \setminus \mathcal{E}^{t,s,z} \right\}.$$

Let S_i be the set of inputs $(x,y) \in \text{supp}(\mu_i)$ that the protocol π errs on, when the noisy multi-layer is i. Our goal is to lower bound the size of S_i. Observe that the protocol π errs on all the inputs in $\Lambda^{t,s,z} \cap R^{t,s,z}$, when the noisy multi-layer is i. The reason is that, as remarked at the end of Section 6.4, by the first property of a good tuple (i,z,t,s) (i.e., unique answer), if the protocol π is correct on $(x,y) \in R^{t,s,z}$, then in the protocol π both players output $\omega^{t,s,z}$ on the input (x,y). Therefore,

$$\Lambda^{t,s,z} \cap R^{t,s,z} \subseteq S_i.$$

Define

$$P^{t,s,z} := \Pr_{(x,y) \in_R R^{t,s,z}} \left[(x,y) \in \Lambda^{t,s,z} \right] = \Pr \left[(X^{t,s,z}, Y^{t,s,z}) \in \Lambda^{t,s,z} \right],$$

and get that

$$|S_i| \geq \sum_{(t,s,z) \in \mathcal{G}_i} |R^{t,s,z}| \cdot P^{t,s,z}.$$

The following lemma, proved in Section 6.5.1, lower bounds $P^{t,s,z}$.

Lemma 14. *For $(t,s,z) \in \mathcal{G}_i$, it holds that*

$$P^{t,s,z} \geq \left(1 - \frac{\mathrm{poly}(1/\gamma)}{k}\right) \Pr_{(x,y)\in_R \Psi_z \times \Psi_z}\left[(x,y) \in \mathrm{supp}(\mu_i)\right].$$

We continue to bound the size of S_i using Lemma 14,

$$|S_i| \geq \left(1 - \frac{\mathrm{poly}(1/\gamma)}{k}\right) \Pr_{(x,y)\in_R \Psi_z \times \Psi_z}\left[(x,y) \in \mathrm{supp}(\mu_i)\right] \cdot \sum_{(t,s,z)\in \mathcal{G}_i} |R^{t,s,z}|.$$

By Lemma 13,

$$\sum_{(t,s,z)\in \mathcal{G}_i} |R^{t,s,z}| \geq (1 - O(\gamma)) \cdot |\{0,1\}^{V^{<i}}| \cdot |\{0,1\}^{V^{\geq i}}|^2.$$

Therefore,

$$|S_i| \geq \left(1 - \frac{\mathrm{poly}(1/\gamma)}{k} - O(\gamma)\right) \cdot |\{0,1\}^{V^{<i}}| \cdot |\{0,1\}^{V^{\geq i}}|^2 \Pr_{(x,y)\in_R \Psi_z \times \Psi_z}\left[(x,y) \in \mathrm{supp}(\mu_i)\right],$$

which implies

$$\frac{|S_i|}{|\mathrm{supp}(\mu_i)|} \geq 1 - \frac{\mathrm{poly}(1/\gamma)}{k} - O(\gamma).$$

We conclude that the protocol π errs on $1 - o(1)$ fraction of the inputs in $\mathrm{supp}(\mu_i)$, when the noisy multi-layer is i. The lower bound follows.

6.5.1 Proof of Lemma 14

In this section we prove Lemma 14. By Equation (15), it suffices to show that

$$P^{t,s,z} := \Pr\left[(X^{t,s,z}, Y^{t,s,z}) \in \Lambda^{t,s,z}\right] \geq \left(1 - \frac{\mathrm{poly}(1/\gamma)}{k}\right) |\Sigma|^{-T} \frac{|E|}{|U| \cdot |W|}.$$

The proof uses several claims and lemmas. All the claims used by the proof are proved in Section 6.5.2. Lemma 18 is stated and proved in Section 6.6. Lemma 22 is stated and proved in Appendix A.

Adding randomness. We start by "adding" some more randomness to the variables $X^{t,s,z}$ and $Y^{t,s,z}$, to obtain new variables $\breve{X}^{t,s,z}$ and $\breve{Y}^{t,s,z}$, for a reason that will be explained shortly. Let $\delta = 2^{-2^{2k}}$. Define $\breve{X}^{t,s,z}$ to be $X^{t,s,z}$ with probability $1-\delta$, and uniformly distributed over Ψ^z with probability δ. Define $\breve{Y}^{t,s,z}$ to be $Y^{t,s,z}$ with probability $1-\delta$, and uniformly distributed over Ψ^z with probability δ.

For $u \in U$, we define the random variable X^u to be $(\breve{X}^{t,s,z}_{\geq i} | \breve{X}^{t,s,z}_i = u)$, that is, X^u has the distribution of $\breve{X}^{t,s,z}_{\geq i}$ conditioned on the event $\breve{X}^{t,s,z}_i = u$. For $w \in W$, we define the random variable Y^w to be $(\breve{Y}^{t,s,z}_{\geq i} | \breve{Y}^{t,s,z}_i = w)$, that is, Y^w has the distribution of $\breve{Y}^{t,s,z}_{\geq i}$ conditioned on the event $\breve{Y}^{t,s,z}_i = w$. By applying Lemma 22 twice (see Appendix A), once with $\Theta_1 = X^{t,s,z}_i$ and $\Theta_2 = X^{t,s,z}_{\geq i}$ and then with $\Theta_1 = Y^{t,s,z}_i$ and $\Theta_2 = Y^{t,s,z}_{\geq i}$, and using the third and fourth properties of a good tuple (i,z,t,s), we get $\mathbf{I}(X^u), \mathbf{I}(Y^w) \leq 2^{4k}$ for every $u \in U, w \in W$. We mention that Lemma 9, that will be applied next, requires that $\mathbf{I}(X^u), \mathbf{I}(Y^w) \leq 2^{4k}$, and that without the addition of randomness to the variables $X^{t,s,z}$ and $Y^{t,s,z}$, this requirement may not by satisfied.

Recall that we want to lower bound the expression

$$P^{t,s,z} := \Pr\left[(X^{t,s,z}, Y^{t,s,z}) \in \Lambda^{t,s,z}\right].$$

Consider the expression

$$\breve{P}^{t,s,z} := \Pr\left[(\breve{X}^{t,s,z}, \breve{Y}^{t,s,z}) \in \Lambda^{t,s,z}\right].$$

The following claim shows that $P^{t,s,z}$ cannot be much smaller than $\breve{P}^{t,s,z}$, thus it suffices to bound $\breve{P}^{t,s,z}$.

Claim 15. *It holds that*

$$P^{t,s,z} \geq \breve{P}^{t,s,z} - 2\delta|\Sigma|^{-T}\frac{|E|}{|U|\cdot|W|}.$$

Applying Lemma 9 to bound $\breve{P}_{t,s,z}$. We turn to bound $\breve{P}^{t,s,z}$. Lemma 9 can be applied to the graph G, as by a symmetry argument, for any fixed u or w and every $v \in [M]$, it holds that at most a fraction of 2^{-20k} of the sets $\{T(u,w)\}_{(u,w)\in E}$ contain v (by definition of the bursting noise game and by the Chernoff bound). This argument also shows that $T \leq 2^{-20k}M$, and that

$$|\mathcal{E}^{t,s,z}| \leq 2^{-20k}|E|, \tag{16}$$

as the output $\omega^{t,s,z}$ is correct only if it has an ancestor in $T(u,w)$.

By applying Lemma 9 to the graph G, there exists a set $\mathcal{D}^{t,s,z} \subseteq E$ such that

$$\frac{|\mathcal{D}^{t,s,z}|}{|E|} \leq 2^{-4k}, \tag{17}$$

and for every $(u,w) \in E \setminus \mathcal{D}^{t,s,z}$ it holds that

$$\Pr_{X^u,Y^w}\left[X^u_{T(u,w)} = Y^w_{T(u,w)}\right] \geq \left(1-2^{-4k}\right)|\Sigma|^{-T}.$$

Recall that an input pair (x,y) with $x_{<i} = y_{<i}$ and $(x_i,y_i) \in E$, is in $\mathrm{supp}(\mu_i)$, if and only if x and y agree on the subtrees rooted at each of the vertices in $T(x_i,y_i)$ (these are non-noisy subtrees). Therefore, by the definition of X^u,Y^w and the last equation,

$$\begin{aligned}\breve{P}^{t,s,z} &= \sum_{(u,w)\in E\setminus\mathcal{E}^{t,s,z}} \Pr\left[\breve{X}^{t,s,z}_i = u\right]\cdot\Pr\left[\breve{Y}^{t,s,z}_i = w\right]\cdot\Pr_{X^u,Y^w}\left[X^u_{T(u,w)} = Y^w_{T(u,w)}\right]\\ &\geq \left(1-2^{-4k}\right)|\Sigma|^{-T}\sum_{(u,w)\in E\setminus(\mathcal{E}^{t,s,z}\cup\mathcal{D}^{t,s,z})}\Pr\left[\breve{X}^{t,s,z}_i = u\right]\cdot\Pr\left[\breve{Y}^{t,s,z}_i = w\right].\end{aligned} \tag{18}$$

We write the sum in the last expression as the sum on pairs $(u,w) \in E$, minus the sum on pairs $(u,w) \in \mathcal{E}^{t,s,z} \cup \mathcal{D}^{t,s,z}$ (sum over the bad sets), and bound each partial sum separately.

Applying Lemma 18 to lower bound the sum on E. We apply Lemma 18 (see Section 6.6) with $P = \breve{X}_i^{t,s,z}$, $Q = \breve{Y}_i^{t,s,z}$ and $I \leq O(\log(1/\gamma))$, as by the seventh and eighth properties of a good tuple (i,z,t,s),

$$\mathbf{I}(\breve{X}_i^{t,s,z}) \leq \mathbf{I}(X_i^{t,s,z}) \leq O(\log(1/\gamma)),$$

$$\mathbf{I}(\breve{Y}_i^{t,s,z}) \leq \mathbf{I}(Y_i^{t,s,z}) \leq O(\log(1/\gamma)).$$

We get

$$\sum_{(u,w)\in E} \Pr\left[\breve{X}_i^{t,s,z} = u\right] \cdot \Pr\left[\breve{Y}_i^{t,s,z} = w\right] \geq \frac{|E|}{|U|\cdot|W|}\left(1 - \frac{O(\log(1/\gamma))}{k}\right). \tag{19}$$

Upper bounding the sum over the bad sets. In order to lower bound $\breve{P}^{t,s,z}$, it remains to upper bound the expression

$$\breve{C}^{t,s,z} := \sum_{(u,w)\in\mathcal{E}^{t,s,z}\cup\mathcal{D}^{t,s,z}} \Pr\left[\breve{X}_i^{t,s,z} = u\right] \cdot \Pr\left[\breve{Y}_i^{t,s,z} = w\right].$$

Consider the expression

$$C^{t,s,z} := \sum_{(u,w)\in\mathcal{E}^{t,s,z}\cup\mathcal{D}^{t,s,z}} \Pr\left[X_i^{t,s,z} = u\right] \cdot \Pr\left[Y_i^{t,s,z} = w\right].$$

The following claim shows that $\breve{C}^{t,s,z}$ cannot be much larger than $C^{t,s,z}$, thus it suffices to bound $C^{t,s,z}$.

Claim 16. *It holds that*

$$\breve{C}^{t,s,z} \leq C^{t,s,z} + 2\delta \cdot \frac{|E|}{|U|\cdot|W|}.$$

The following claim upper bounds $C^{t,s,z}$.

Claim 17. *It holds that*

$$C^{t,s,z} \leq O(1/\gamma^4) \cdot 2^{-2k} \cdot \frac{|E|}{|U|\cdot|W|}.$$

By combining Claims 16 and 17, we get

$$\breve{C}^{t,s,z} \leq \mathrm{poly}(1/\gamma) \cdot 2^{-k} \cdot \frac{|E|}{|U|\cdot|W|}. \tag{20}$$

By Equations (18), (19), (20) and Claim 15,

$$P^{t,s,z} \geq \left(1 - \frac{\mathrm{poly}(1/\gamma)}{k}\right)|\Sigma|^{-T}\frac{|E|}{|U|\cdot|W|},$$

and the assertion holds.

6.5.2 Proof of Claims

Proof of Claim 15. Let $\mathcal{U}_1, \mathcal{U}_2$ be two independent random variables, each uniformly distributed over Ψ^z. It holds that

$$\begin{aligned}
\breve{P}^{t,s,z} &= \Pr\left[(\breve{X}^{t,s,z}, \breve{Y}^{t,s,z}) \in \Lambda^{t,s,z}\right] \\
&= (1-\delta)^2 \cdot \Pr\left[(X^{t,s,z}, Y^{t,s,z}) \in \Lambda^{t,s,z}\right] + \delta^2 \cdot \Pr\left[(\mathcal{U}_1, \mathcal{U}_2) \in \Lambda^{t,s,z}\right] \\
&\quad + (1-\delta)\delta \cdot \Pr\left[(X^{t,s,z}, \mathcal{U}_2) \in \Lambda^{t,s,z}\right] + (1-\delta)\delta \cdot \Pr\left[(\mathcal{U}_1, Y^{t,s,z}) \in \Lambda^{t,s,z}\right] \\
&\leq (1-\delta)^2 \cdot \Pr\left[(X^{t,s,z}, Y^{t,s,z}) \in \Lambda^{t,s,z}\right] + \delta^2 \cdot \Pr\left[(\mathcal{U}_1, \mathcal{U}_2) \in \operatorname{supp}(\mu_i)\right] \\
&\quad + (1-\delta)\delta \cdot \Pr\left[(X^{t,s,z}, \mathcal{U}_2) \in \operatorname{supp}(\mu_i)\right] + (1-\delta)\delta \cdot \Pr\left[(\mathcal{U}_1, Y^{t,s,z}) \in \operatorname{supp}(\mu_i)\right] \\
&\leq \Pr\left[(X^{t,s,z}, Y^{t,s,z}) \in \Lambda^{t,s,z}\right] + 2\delta|\Sigma|^{-T}\frac{|E|}{|U| \cdot |W|} \\
&= P^{t,s,z} + 2\delta|\Sigma|^{-T}\frac{|E|}{|U| \cdot |W|}.
\end{aligned}$$

The last inequality uses the fact that for every $x, y \in \Psi^z$,

$$\Pr\left[(x, \mathcal{U}_2) \in \operatorname{supp}(\mu_i)\right] = \Pr\left[(\mathcal{U}_1, y) \in \operatorname{supp}(\mu_i)\right] = \Pr\left[(\mathcal{U}_1, \mathcal{U}_2) \in \operatorname{supp}(\mu_i)\right].$$

By Equation (15), this equals $|\Sigma|^{-T}\frac{|E|}{|U| \cdot |W|}$. □

Proof of Claim 16. Let $\mathcal{U}_1, \mathcal{U}_2$ be two independent random variables, each uniformly distributed over $\{0,1\}^{V_i}$. It holds that

$$\begin{aligned}
\breve{C}^{t,s,z} &= \sum_{(u,w) \in \mathcal{E}^{t,s,z} \cup \mathcal{D}^{t,s,z}} \Pr\left[\breve{X}_i^{t,s,z} = u\right] \cdot \Pr\left[\breve{Y}_i^{t,s,z} = w\right] \\
&= (1-\delta)^2 \sum_{(u,w) \in \mathcal{E}^{t,s,z} \cup \mathcal{D}^{t,s,z}} \Pr\left[X_i^{t,s,z} = u\right] \cdot \Pr\left[Y_i^{t,s,z} = w\right] \\
&\quad + \delta^2 \sum_{(u,w) \in \mathcal{E}^{t,s,z} \cup \mathcal{D}^{t,s,z}} \Pr\left[\mathcal{U}_1 = u\right] \cdot \Pr\left[\mathcal{U}_2 = w\right] \\
&\quad + (1-\delta)\delta \sum_{(u,w) \in \mathcal{E}^{t,s,z} \cup \mathcal{D}^{t,s,z}} \Pr\left[X_i^{t,s,z} = u\right] \cdot \Pr\left[\mathcal{U}_2 = w\right] \\
&\quad + (1-\delta)\delta \sum_{(u,w) \in \mathcal{E}^{t,s,z} \cup \mathcal{D}^{t,s,z}} \Pr\left[\mathcal{U}_1 = u\right] \cdot \Pr\left[Y_i^{t,s,z} = w\right].
\end{aligned}$$

The first term (out of the four) in the last sum is $(1-\delta)^2 \cdot C^{t,s,z}$. We bound each of the last three terms in the same way, and get the desired inequality. For example,

$$\begin{aligned}
&\sum_{(u,w)\in\mathcal{E}^{t,s,z}\cup\mathcal{D}^{t,s,z}} \Pr\left[\mathcal{U}_1 = u\right]\cdot\Pr\left[Y_i^{t,s,z} = w\right] \\
&\le \sum_{(u,w)\in E} \Pr\left[\mathcal{U}_1 = u\right]\cdot\Pr\left[Y_i^{t,s,z} = w\right] \\
&\le \frac{1}{|U|}\sum_{(u,w)\in E} \Pr\left[Y_i^{t,s,z} = w\right] \\
&= \frac{1}{|U|}\cdot\frac{|E|}{|W|}\sum_{w\in W} \Pr\left[Y_i^{t,s,z} = w\right] = \frac{|E|}{|U|\cdot|W|},
\end{aligned}$$

where second to last equality is due to the bi-regularity of G. □

Proof of Claim 17. We want to upper bound the expression

$$C^{t,s,z} := \sum_{(u,w)\in\mathcal{E}^{t,s,z}\cup\mathcal{D}^{t,s,z}} \Pr\left[X_i^{t,s,z} = u\right]\cdot\Pr\left[Y_i^{t,s,z} = w\right].$$

First, observe that $C^{t,s,z}\cdot|R^{t,s,z}|$ is exactly the number of input pairs $(x,y)\in R^{t,s,z}$ with $(x_i,y_i)\in\mathcal{E}^{t,s,z}\cup\mathcal{D}^{t,s,z}$. Consider the expression

$$C^{t,z} := \sum_{(u,w)\in\mathcal{E}^{t,s,z}\cup\mathcal{D}^{t,s,z}} \Pr\left[X_i^{t,z} = u\right]\cdot\Pr\left[Y_i^{t,z} = w\right].$$

Again, $C^{t,z}\cdot|R^{t,z}|$ is exactly the number of input pairs $(x,y)\in R^{t,z}$ with $(x_i,y_i)\in\mathcal{E}^{t,s,z}\cup\mathcal{D}^{t,s,z}$. Since $R^{t,z}$ contains $R^{t,s,z}$, and by the second property of a good tuple (i,z,t,s) (i.e., γ-largeness), it holds that

$$C^{t,s,z} \le \frac{|R^{t,z}|}{|R^{t,s,z}|}C^{t,z} \le O(1/\gamma^4)\cdot C^{t,z}. \tag{21}$$

Therefore, in order to bound $C^{t,s,z}$, it suffices to bound $C^{t,z}$.

By the third and fourth properties of a good tuple (i,z,t,s), we have $\mathbf{I}\left(X^{t,z}\right), \mathbf{I}\left(Y^{t,z}\right) \le 2\log(m)$, which means that $|R^{t,z}| \ge \frac{1}{m^4}|\Psi^z\times\Psi^z|$. This implies that for every set $L\subseteq\Psi^z$, the probability that $X^{t,z}$ is in L is at most m^4 the probability that a uniformly distributed variable over Ψ^z obtains a value in L. In particular, for every $u\in U$,

$$\Pr\left[X_i^{t,z} = u\right] \le \frac{m^4}{|U|}. \tag{22}$$

Similarly, for $w\in W$,

$$\Pr\left[Y_i^{t,z} = w\right] \le \frac{m^4}{|W|}. \tag{23}$$

Define

$$U' = \left\{u\in U \ : \ \Pr\left[X_i^{t,z} = u\right] \ge \frac{2}{|U|}\right\},$$

$$W' = \left\{ w \in W \ : \ \Pr\left[Y_i^{t,z} = w\right] \geq \frac{2}{|W|} \right\}.$$

By the fifth and sixth properties of a good tuple (i, z, t, s), we have

$$\mathbf{I}\left(X_i^{t,z}\right) \leq \frac{m}{\gamma^2 c},$$

$$\mathbf{I}\left(Y_i^{t,z}\right) \leq \frac{m}{\gamma^2 c}.$$

Using Lemma 5.12 in [KR13] (stated for convenience in Appendix A, Lemma 25) it holds that

$$\Pr\left[X_i^{t,z} \in U'\right] < 5 \cdot \left(\frac{m}{\gamma^2 c}\right)^{0.1}, \tag{24}$$

Similarly,

$$\Pr\left[Y_i^{t,z} \in W'\right] < 5 \cdot \left(\frac{m}{\gamma^2 c}\right)^{0.1}, \tag{25}$$

The expression $C^{t,z}$ is a sum over pairs $(u, w) \in \mathcal{E}^{t,s,z} \cup \mathcal{D}^{t,s,z}$. We bound $C^{t,z}$ by a sum of three partial sums, and work on each partial sum separately. The first partial sum is over pairs $(u, w) \in E$ with $u \in U'$, the second is over pairs $(u, w) \in E$ with $w \in W'$, the third is over pairs $(u, w) \in \mathcal{E}^{t,s,z} \cup \mathcal{D}^{t,s,z}$ with $u \notin U'$ and $w \notin W'$.

We bound the first partial sum as follows. We use Equation (23) for the first inequality, the bi-regularity of G for the second, and Equation (24) for the third.

$$\begin{aligned}
&\sum_{\substack{(u,w)\in E \\ u \in U'}} \Pr\left[X_i^{t,z} = u\right] \cdot \Pr\left[Y_i^{t,z} = w\right] \\
&\leq \frac{m^4}{|W|} \sum_{\substack{(u,w)\in E \\ u \in U'}} \Pr\left[X_i^{t,z} = u\right] \\
&= \frac{m^4}{|W|} \cdot \frac{|E|}{|U|} \sum_{u \in U'} \Pr\left[X_i^{t,z} = u\right] \\
&\leq \frac{5m^4}{|W|} \cdot \frac{|E|}{|U|} \cdot \left(\frac{m}{\gamma^2 c}\right)^{0.1} \leq \frac{|E|}{|U| \cdot |W|} \cdot c^{-0.05}.
\end{aligned}$$

The second partial sum is bounded in a similar way. We bound the third partial sum using Equations (16) and (17),

$$\sum_{\substack{(u,w)\in \mathcal{E}^{t,s,z} \cup \mathcal{D}^{t,s,z} \\ u \notin U', w \notin W'}} \Pr\left[X_i^{t,z} = u\right] \cdot \Pr\left[Y_i^{t,z} = w\right] \leq \left|\mathcal{E}^{t,s,z} \cup \mathcal{D}^{t,s,z}\right| \cdot \frac{2}{|U|} \cdot \frac{2}{|W|} < \frac{|E|}{|U| \cdot |W|} \cdot 2^{-3k}.$$

We conclude that $C^{t,z} \leq \frac{|E|}{|U|\cdot|W|} \cdot 2^{-2k}$. Using Equation (21),

$$C^{t,s,z} \leq O(1/\gamma^4) \cdot \frac{|E|}{|U| \cdot |W|} \cdot 2^{-2k}.$$

$\square$

6.6 The Mini Graph

Lemma 18. *Let P be a random variable over the domain U, and Q be a random variable over the domain W, such that P and Q are independent. Denote $I = \max\{\mathbf{I}(P), \mathbf{I}(Q), 1\}$. It holds that*

$$\Pr_{P,Q}[(P,Q) \in E] \geq \frac{|E|}{|U| \cdot |W|}\left(1 - \frac{I}{k}\right).$$

Proof. Let $\hat{G} = (\hat{U} \cup \hat{W}, \hat{E})$ be the complete bipartite graph with sets of vertices $\hat{U}, \hat{W}$ and set of edges $\hat{E}$, defined as follows: Let $\hat{U} = \hat{W}$ be the set of all boolean assignments to the vertices in layers i^* to $i^* + w_0 - 1$ of the tree $\mathcal{T}$ (that is, the first w_0 layers in multi-layer i; see Section 2). Let $\tilde{U} = \tilde{W}$ be the set of all boolean assignments to the vertices in layers $i^* + w_0$ to $i^* + w_0 + w_1 - 1 = (i+1)^* - 1$ of the tree $\mathcal{T}$ (that is, all other layers in multi-layer i). Set $\hat{E} = \hat{U} \times \hat{W}$.

Let $\hat{M}$ be the number of vertices in layer $i^* + w_0$ of the tree $\mathcal{T}$. We identify the set $[\hat{M}]$ with the set of vertices in layer $i^* + w_0$. Let $\hat{u} \in \hat{U}, \hat{w} \in \hat{W}$. We define $\hat{T}(\hat{u}, \hat{w}) \subset [\hat{M}]$ to be the set of all vertices in layer $i^* + w_0$ that are set to be non-noisy for inputs $\hat{u}, \hat{w}$, by Algorithm 1 defining μ, when the noisy multi-layer is i. That is, $\hat{T}(\hat{u}, \hat{w})$ is the set of all typical vertices in layer $i^* + w_0$ with respect to $i^*, w_0, \hat{u}, \hat{w}$. Observe that $\hat{u}$ and $\hat{w}$ determine for every vertex in layer $i^* + w_0$ if it is noisy or not.

Note that by a symmetry argument, $\hat{G}$ is bi-regular and $\hat{T}(\hat{u}, \hat{w})$ is of the same size $\hat{T}$ for every $\hat{u}, \hat{w}$. By the definition of the bursting noise game and by the Chernoff bound, for any fixed $\hat{u}$ or $\hat{w}$ and every $v \in [\hat{M}]$, it holds that at most a fraction of 2^{-20k} of the sets $\{\hat{T}(\hat{u}, \hat{w})\}_{(\hat{u},\hat{w})\in\hat{E}}$ contain v.

Let P be a random variable over the domain U, and Q be a random variable over the domain W, such that P and Q are independent. Let $\hat{P}$ and $\hat{Q}$ be the projections of P and Q on layers i^* to $i^* + w_0 - 1$ (respectively). Let $\tilde{P}$ and $\tilde{Q}$ be the projections of P and Q on layers $i^* + w_0$ to $(i+1)^* - 1$ (respectively).

Define the bad sets

$$\mathcal{D}_1 = \left\{\hat{u} \in \hat{U} \; : \; \Pr_P[\hat{P} = \hat{u}] = 0 \;\text{ or }\; \mathbf{I}(\tilde{P}|\hat{P} = \hat{u}) > 2^{4k}\right\},$$

$$\mathcal{D}_2 = \left\{\hat{w} \in \hat{W} \; : \; \Pr_Q[\hat{Q} = \hat{w}] = 0 \;\text{ or }\; \mathbf{I}(\tilde{Q}|\hat{Q} = \hat{w}) > 2^{4k}\right\}.$$

By the chain rule for the entropy function,

$$
\begin{aligned}
I \geq \mathbf{I}(P) &= \log(|U|) - \mathbf{H}(P) \\
&= \log(|\hat{U}| \cdot |\tilde{U}|) - \mathbf{H}(\hat{P}, \tilde{P}) \\
&= \log(|\hat{U}|) + \log(|\tilde{U}|) - \mathbf{H}(\hat{P}) - \mathbf{H}(\tilde{P}|\hat{P}) \\
&= \mathbf{I}(\hat{P}) + \log(|\tilde{U}|) - \mathop{\mathbf{E}}_{\hat{u} \leftarrow \hat{P}} \left[\mathbf{H}(\tilde{P}|\hat{P} = \hat{u}) \right] \\
&\geq \mathop{\mathbf{E}}_{\hat{u} \leftarrow \hat{P}} \left[\mathbf{I}(\tilde{P}|\hat{P} = \hat{u}) \right].
\end{aligned}
$$

By Markov's inequality,

$$
\Pr_P \left[\hat{P} \in \mathcal{D}_1 \right] \leq \frac{I}{2^{4k}}. \tag{26}
$$

By a similar argument,

$$
\Pr_Q \left[\hat{Q} \in \mathcal{D}_2 \right] \leq \frac{I}{2^{4k}}. \tag{27}
$$

For $\hat{u} \notin \mathcal{D}_1$, we define the random variable $X^{\hat{u}}$ to be $(\tilde{P}|\hat{P} = \hat{u})$, that is, $X^{\hat{u}}$ has the distribution of $\tilde{P}$ conditioned on the event $\hat{P} = \hat{u}$. For $\hat{u} \in \mathcal{D}_1$, we define the random variable $X^{\hat{u}}$ to be uniformly distributed over $\tilde{U}$. Similarly, for $\hat{w} \notin \mathcal{D}_2$, we define the random variable $Y^{\hat{w}}$ to be $(\tilde{Q}|\hat{Q} = \hat{w})$, that is, $Y^{\hat{w}}$ has the distribution of $\tilde{Q}$ conditioned on the event $\hat{Q} = \hat{w}$. For $\hat{w} \in \mathcal{D}_2$, we define the random variable $Y^{\hat{w}}$ to be uniformly distributed over $\tilde{W}$. Let $\hat{\Sigma}$ be the set of all possible boolean assignments to the vertices of a subtree with w_1 layers, rooted at layer $i^* + w_0$.

By Lemma 9 applied to the graph $\hat{G}$, there exists a set $\hat{\mathcal{D}} \subset \hat{E}$ such that $\frac{|\hat{\mathcal{D}}|}{|\hat{E}|} \leq 2^{-4k}$, and for every $(\hat{u}, \hat{w}) \notin \hat{\mathcal{D}}$ it holds that

$$
\Pr_{X^{\hat{u}}, Y^{\hat{w}}} \left[X^{\hat{u}}_{\hat{T}(\hat{u},\hat{w})} = Y^{\hat{w}}_{\hat{T}(\hat{u},\hat{w})} \right] \geq \left(1 - 2^{-4k}\right) |\hat{\Sigma}|^{-\hat{T}}. \tag{28}
$$

It holds that

$$
\begin{aligned}
&\Pr\left[(P,Q) \in E\right] \\
&= \sum_{\hat{u} \in \hat{U}, \hat{w} \in \hat{W}} \Pr_P \left[\hat{P} = \hat{u} \right] \cdot \Pr_Q \left[\hat{Q} = \hat{w} \right] \cdot \Pr_{P,Q} \left[(P,Q) \in E \;\middle|\; \hat{P} = \hat{u}, \hat{Q} = \hat{w} \right] \\
&\geq \sum_{\substack{\hat{u} \in \hat{U} \setminus \mathcal{D}_1, \\ \hat{w} \in \hat{W} \setminus \mathcal{D}_2, \\ (\hat{u},\hat{w}) \notin \hat{\mathcal{D}}}} \Pr_P \left[\hat{P} = \hat{u} \right] \cdot \Pr_Q \left[\hat{Q} = \hat{w} \right] \cdot \Pr_{P,Q} \left[(P,Q) \in E \;\middle|\; \hat{P} = \hat{u}, \hat{Q} = \hat{w} \right].
\end{aligned}
$$

By the definition of the bursting noise game (when the main noise is on multi-layer i), for every $\hat{u}, \hat{w}$, the following holds: Conditioned on $\hat{P} = \hat{u}$ and $\hat{Q} = \hat{w}$, we have $(P,Q) \in E$ if and only if $\tilde{P}$ and $\tilde{Q}$ agree on the subtrees rooted at vertices in $\hat{T}(\hat{u}, \hat{w})$. Therefore, using

Equation (28) and the fact that $\hat{E}$ contains all pairs $(\hat{u}, \hat{w})$,

$$
\begin{aligned}
&\Pr\left[(P,Q) \in E\right] \\
&\geq \sum_{\substack{\hat{u} \in \hat{U} \setminus \mathcal{D}_1, \\ \hat{w} \in \hat{W} \setminus \mathcal{D}_2, \\ (\hat{u},\hat{w}) \notin \hat{\mathcal{D}}}} \Pr_P\left[\hat{P} = \hat{u}\right] \cdot \Pr_Q\left[\hat{Q} = \hat{w}\right] \cdot \Pr_{X^{\hat{u}}, Y^{\hat{w}}}\left[X^{\hat{u}}_{\hat{T}(\hat{u},\hat{w})} = Y^{\hat{w}}_{\hat{T}(\hat{u},\hat{w})}\right] \\
&\geq \left(1 - 2^{-4k}\right) |\hat{\Sigma}|^{-\hat{T}} \sum_{\substack{\hat{u} \in \hat{U} \setminus \mathcal{D}_1, \\ \hat{w} \in \hat{W} \setminus \mathcal{D}_2, \\ (\hat{u},\hat{w}) \notin \hat{\mathcal{D}}}} \Pr_P\left[\hat{P} = \hat{u}\right] \cdot \Pr_Q\left[\hat{Q} = \hat{w}\right].
\end{aligned}
$$

To bound the last term, we consider four partial sums. Clearly,

$$\sum_{(\hat{u},\hat{w}) \in \hat{U} \times \hat{W}} \Pr_P\left[\hat{P} = \hat{u}\right] \cdot \Pr_Q\left[\hat{Q} = \hat{w}\right] = 1.$$

By Equation (26),

$$\sum_{\hat{u} \in \mathcal{D}_1} \Pr_P\left[\hat{P} = \hat{u}\right] \leq \frac{I}{2^{4k}},$$

and by Equation (27),

$$\sum_{\hat{w} \in \mathcal{D}_2} \Pr_Q\left[\hat{Q} = \hat{w}\right] \leq \frac{I}{2^{4k}}.$$

We next apply Lemma 23 (stated and proved in Appendix A) with a distribution τ over the set $\hat{U} \times \hat{W}$, defined by $\tau(\hat{u}, \hat{w}) = \Pr[\hat{P} = \hat{u}] \cdot \Pr[\hat{Q} = \hat{w}]$. Observe that by the independence of P and Q and by the super additivity of information, $\mathbf{I}(\tau) = \mathbf{I}(\hat{P}) + \mathbf{I}(\hat{Q}) \leq \mathbf{I}(P) + \mathbf{I}(Q) \leq 2I$. Lemma 23 implies that

$$\sum_{(\hat{u},\hat{w}) \in \hat{\mathcal{D}}} \Pr_P\left[\hat{P} = \hat{u}\right] \cdot \Pr_Q\left[\hat{Q} = \hat{w}\right] \leq \frac{2I+1}{4k}.$$

Therefore,

$$\Pr\left[(P,Q) \in E\right] \geq \left(1 - 2^{-4k}\right) |\hat{\Sigma}|^{-\hat{T}} \left(1 - \frac{I}{2^{4k}} - \frac{I}{2^{4k}} - \frac{2I+1}{4k}\right) \geq |\hat{\Sigma}|^{-\hat{T}} \left(1 - \frac{I}{k}\right).$$

Finally, note that for every $u \in U$ and $w \in W$, such that the projection of u and w to layers i^* to $i^* + w_0 - 1$ is $\hat{u}$ and $\hat{w}$ (respectively), the following holds: $(u, w) \in E$ if and only if u and w agree on the subtrees rooted at vertices in $\hat{T}(\hat{u}, \hat{w})$. Therefore, $|E| = |U| \cdot |W| \cdot |\hat{\Sigma}|^{-\hat{T}}$, and the assertion follows. □

Remark. *We remark that Lemma 18 is only used in the proof of Lemma 14, for P and Q such that the sets $\mathcal{D}_1, \mathcal{D}_2$ defined in the proof of Lemma 18 are empty. Hence, we could have used a restricted version of Lemma 18 with a simpler proof. We included here the more general version of Lemma 18, as the general claim and proof method may be of independent* *interest.*

7 Information Upper Bound

In this section we prove Theorem 2. Let $(x, y) \in \{0,1\}^V \times \{0,1\}^V$ be an input pair to the bursting noise game. For every vertex $v \in V$, we define two distributions $P_v = (p_v, 1 - p_v)$ and $Q_v = (q_v, 1 - q_v)$, both over $\{0,1\}$. If the player who owns v gets as an input the bit 0 for v, then $P_v = (0.9, 0.1)$, else $P_v = (0.1, 0.9)$. Similarly, if the player who doesn't own v gets as an input the bit 0 for v, then $Q_v = (0.9, 0.1)$, else $Q_v = (0.1, 0.9)$. We think of every P_v as the "correct" distribution over the two children of v, and of every Q_v as an estimation of P_v, based on the knowledge of the player who doesn't own v.

Consider the following protocol for the bursting noise game. Starting from the root until reaching a leaf, at every vertex v, the player who owns v samples a bit according to P_v and sends this bit to the other player. Both players continue to the child of v that is indicated by the communicated bit. When they reach a leaf they output that leaf. We denote this protocol by π. By the Chernoff bound, the probability that the players output a leaf that is not typical with respect to the main noise layers is at most $2^{-\Omega(w_0)}$. That is, the error probability of π is exponentially small in k.

To upper bound the information cost of the protocol π it is convenient to use the notion of divergence cost of a tree. This notion is implicit in [BBCR10] and was formally defined in [BR11].

Definition 4 (**Divergence Cost**)**.** *Given a binary tree $\mathcal{T}$, whose root is r, and distributions $P_v = (p_v, 1 - p_v), Q_v = (q_v, 1 - q_v)$ for every vertex v in the tree, we define the* divergence cost *of the tree, recursively, as follows.* $\mathbf{D}(\mathcal{T}) = 0$ *if the tree has depth 0, otherwise,*

$$\mathbf{D}(\mathcal{T}) = \mathbf{D}(P_r \| Q_r) + \mathop{\mathbf{E}}_{v \sim P_r}[\mathbf{D}(\mathcal{T}_v)],$$

where for every vertex v, $\mathcal{T}_v$ is the subtree of $\mathcal{T}$ whose root is v.

Let $\mathcal{T}$ be the binary tree on which the bursting noise game is played, and let P_v, Q_v, for every vertex $v \in V$, be the distributions defined according to the input pair $(x, y) \in \{0,1\}^V \times \{0,1\}^V$. Following the recursion in the definition of the divergence cost we get the following equation:

$$\mathbf{D}(\mathcal{T}) = \sum_{v \in V} \tilde{p_v} \cdot \mathbf{D}(P_v \| Q_v), \tag{29}$$

where $\tilde{p}_v$, for a vertex $v \in V$, is the probability to reach v during the execution of the protocol π, defined above. That is, if v is the root of the tree, $\tilde{p}_v = 1$, otherwise,

$$\tilde{p}_v = \begin{cases} \tilde{p}_u \cdot p_u & \text{if } v \text{ is the left-hand child of } u \\ \tilde{p}_u \cdot (1 - p_u) & \text{if } v \text{ is the right-hand child of } u. \end{cases}$$

Using Equation (29), we give an upper bound on the divergence cost of $\mathcal{T}$.

Proposition 19.
$$\mathbf{D}(\mathcal{T}) = O(k).$$

Proof. For every non-noisy vertex v we have $\mathbf{D}(P_v \| Q_v) = 0$, and for every noisy vertex v we have $\mathbf{D}(P_v \| Q_v) \leq 4$. Let i be the noisy multi-layer. Then,

1. The vertices above the i^{th} multi-layer add nothing to the divergence cost.

2. The main noise in layers $[i^*, i^* + w_0 - 1]$ adds $O(w_0)$ to the divergence cost.

3. Let v be the vertex that the players reach during the execution of the protocol π, in layer $i^* + w_0$, that is, at the end of the main noise. If v is typical with respect to i^*, w_0, x, y, the vertices below v add nothing to the divergence cost. The probability that v is a non-typical vertex with respect to i^*, w_0, x, y, that is, $v \in L_1$, is at most $2^{-\Omega(w_0)}$. In such a case, the vertices in the subtree with w_1 layers rooted at v add $O(2^{-\Omega(w_0)} \cdot w_1) = O(1)$ to the divergence cost (where w_0 was chosen to be sufficiently larger than $\log(w_1)$).

4. If $i < c$: Let v be the vertex that the players reach during the execution of the protocol in layer $i^* + w_0 + w_1$. If v is a typical vertex with respect to $i^* + w_0, w_1, x, y$ or v has no ancestor in L_1, the vertices below v add nothing to the divergence cost. The probability that v has an ancestor in L_1 and is a non-typical vertex with respect to $i^* + w_0, w_1, x, y$, that is, $v \in L_2$, is at most $2^{-\Omega(w_1)}$. In such a case, the vertices in the subtree rooted at v add at most $O(2^{-\Omega(w_1)} \cdot w \cdot c) = O(1)$ to the divergence cost (where w_1 was chosen to be sufficiently larger than $\log(w \cdot c)$).

Together, the total divergence cost is $O(w_0) = O(k)$ as claimed. □

Let X be the input to the first player and Y be the input to the second player. In the protocol π, the players use two private random strings and no public randomness. Denote the private random string of the first player by R_1, and the private random string of the second player by R_2. For a layer $d \in [c \cdot w]$, let Π_d be the vertex in layer d that the players reach during the execution of the protocol π, when the inputs are (X, Y) and the private random strings are R_1 and R_2. Let the tree $\mathcal{T}'$ be the same as $\mathcal{T}$, except that every distribution Q_v, for every vertex $v \in V$, is replaced with the distribution $Q'_v = (q'_v, 1 - q'_v)$, where q'_v is defined as follows: Let d be the layer of v. If v is owned by the first player, q'_v is the function of v, y and r_2, defined as
$$q'_v = \mathop{\mathbf{E}}_{X,R_1}[p_v | Y = y, R_2 = r_2, \Pi_d = v].$$
If v is owned by the second player, q'_v is the function of v, x and r_1, defined as
$$q'_v = \mathop{\mathbf{E}}_{Y,R_2}[p_v | X = x, R_1 = r_1, \Pi_d = v].$$

We think of Q'_v as the best estimation of the correct distribution P_v, based on the knowledge of the player who doesn't own v, whereas Q_v is some estimation. Intuitively,

$\mathbf{D}(P_v \| Q_v)$ is the information that the player who doesn't own v learns on P_v from the bit sent during the protocol at the vertex v, assuming that she expects this bit to be distributed according to Q_v, whereas $\mathbf{D}(P_v \| Q'_v)$ is the information that she learns based on the best possible estimation of P_v. Therefore, intuitively, the divergence cost of $\mathcal{T}'$ is at most the divergence cost of $\mathcal{T}$, in expectation. This is formulated in the following proposition.

Proposition 20.

$$\mathbf{E}[\mathbf{D}(\mathcal{T}')] \leq \mathbf{E}[\mathbf{D}(\mathcal{T})],$$

where the expectation is over the sampling of the inputs according to μ and over the randomness.

Proof. By Equation (29),

$$\mathop{\mathbf{E}}_{X,Y,R_1,R_2}[\mathbf{D}(\mathcal{T}) - \mathbf{D}(\mathcal{T}')] = \mathop{\mathbf{E}}_{X,Y,R_1,R_2}\left[\sum_{v \in V} \tilde{p_v}\left(\mathbf{D}(P_v \| Q_v) - \mathbf{D}(P_v \| Q'_v)\right)\right].$$

We separate the sum on the vertices to layers and work on each layer separately. Fix a layer d in the tree. Let L_d be the set of vertices in layer d. To simplify notation, let A denote (X, R_1), let B denote (Y, R_2), and let V denote Π_d. Then,

$$\mathop{\mathbf{E}}_{X,Y,R_1,R_2}\left[\sum_{v \in L_d} \tilde{p_v}\left(\mathbf{D}(P_v \| Q_v) - \mathbf{D}(P_v \| Q'_v)\right)\right] = \mathop{\mathbf{E}}_{A,B,V}\left[\mathbf{D}(P_V \| Q_V) - \mathbf{D}(P_V \| Q'_V)\right].$$

By the definition of relative entropy,

$$\begin{aligned}
&\mathop{\mathbf{E}}_{A,B,V}\left[\mathbf{D}(P_V \| Q_V) - \mathbf{D}(P_V \| Q'_V)\right] \\
&= \mathop{\mathbf{E}}_{A,B,V}\left[p_V\left(\log\left(\frac{p_V}{q_V}\right) - \log\left(\frac{p_V}{q'_V}\right)\right) + (1-p_V)\left(\log\left(\frac{1-p_V}{1-q_V}\right) - \log\left(\frac{1-p_V}{1-q'_V}\right)\right)\right] \\
&= \mathop{\mathbf{E}}_{A,B,V}\left[p_V \log\left(\frac{q'_V}{q_V}\right) + (1-p_V)\log\left(\frac{1-q'_V}{1-q_V}\right)\right]. \qquad (30)
\end{aligned}$$

Assume that the vertices in layer d are owned by the first player. The case that the vertices in layer d are owned by the second player is analogous. Consider the first summand in Equation (30). It holds that,

$$\mathop{\mathbf{E}}_{A,B,V}\left[p_V \log\left(\frac{q'_V}{q_V}\right)\right] = \mathop{\mathbf{E}}_{B,V}\left[\mathop{\mathbf{E}}_{A}\left[\left(p_V \log\left(\frac{q'_V}{q_V}\right)\right)\middle| B, V\right]\right].$$

By the definition of q'_V, for fixed B, V, it holds that $q'_V = \mathbf{E}_A[p_V | B, V]$. Since q'_V and q_V are functions of B and V, when we condition on B and V, q'_V and q_V are fixed. Therefore, conditioned on B and V, the term $\log\left(\frac{q'_V}{q_V}\right)$ is independent of A. We get that,

$$\begin{aligned}\mathop{\mathbf{E}}_{B,V}\left[\mathop{\mathbf{E}}_{A}\left[\left(p_V \log\left(\frac{q'_V}{q_V}\right)\right)\middle| B,V\right]\right] &= \mathop{\mathbf{E}}_{B,V}\left[\mathop{\mathbf{E}}_{A}\left[p_V|B,V\right]\log\left(\frac{q'_V}{q_V}\right)\right] \\ &= \mathop{\mathbf{E}}_{B,V}\left[q'_V \log\left(\frac{q'_V}{q_V}\right)\right].\end{aligned}$$

In the same way, we get that the second summand in Equation (30) is

$$\mathop{\mathbf{E}}_{A,B,V}\left[(1-p_V)\log\left(\frac{1-q'_V}{1-q_V}\right)\right] = \mathop{\mathbf{E}}_{B,V}\left[(1-q'_V)\log\left(\frac{1-q'_V}{1-q_V}\right)\right].$$

Put together it holds that,

$$\mathop{\mathbf{E}}_{A,B,V}\left[\mathbf{D}(P_V\|Q_V) - \mathbf{D}(P_V\|Q'_V)\right] = \mathop{\mathbf{E}}_{B,V}\left[\mathbf{D}(Q'_V\|Q_V)\right].$$

Since the divergence is non-negative, $\mathbf{E}_{A,B,V}\left[\mathbf{D}(P_V\|Q_V) - \mathbf{D}(P_V\|Q'_V)\right] \geq 0$. This is true for every layer d in the tree, therefore, summing over all layers, we get that $\mathbf{E}_{A,B}[\mathbf{D}(\mathcal{T}')] \leq \mathbf{E}_{A,B}[\mathbf{D}(\mathcal{T})]$, as stated. $\square$

The following lemma relates the information cost of π to the expected divergence cost of $\mathcal{T}'$. The proof appears in [BR11] (see Lemma 5.3 therein). Together with Propositions 19 and 20 we get that $IC_\mu(\pi) = O(k)$.

Lemma 21.

$$\mathbf{E}[\mathbf{D}(\mathcal{T}')] = IC_\mu(\pi),$$

where the expectation is over the sampling of the inputs according to μ and over the randomness.

References

[BBCR10] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. In *STOC*, pages 67–76, 2010. 2, 36

[BR11] Mark Braverman and Anup Rao. Information equals amortized communication. In *FOCS*, pages 748–757, 2011. 1, 2, 3, 36, 39

[Bra12a] Mark Braverman. Coding for interactive computation: progress and challenges. *In 50th Annual Allerton Conference on Communication, Control, and Computing*, 2012. 3

[Bra12b] Mark Braverman. Interactive information complexity. In *STOC*, pages 505–524, 2012. 1, 2, 3

[Bra13] Mark Braverman. A hard-to-compress interactive task? *In 51th Annual Allerton Conference on Communication, Control, and Computing*, 2013. 2, 3

[BRWY12] Mark Braverman, Anup Rao, Omri Weinstein, and Amir Yehudayoff. Direct products in communication complexity. *Electronic Colloquium on Computational Complexity (ECCC)*, 19:143, 2012. 2

[BRWY13] Mark Braverman, Anup Rao, Omri Weinstein, and Amir Yehudayoff. Direct product via round-preserving compression. *Electronic Colloquium on Computational Complexity (ECCC)*, 20:35, 2013. 2

[BW12] Mark Braverman and Omri Weinstein. A discrepancy lower bound for information complexity. In *APPROX-RANDOM*, pages 459–470, 2012. 2

[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. 2

[CGFS86] Fan R. K. Chung, Ronald L. Graham, Peter Frankl, and James B. Shearer. Some intersection theorems for ordered sets and graphs. *J. Comb. Theory, Ser. A*, 43(1):23–37, 1986. 12, 14

[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. 2

[CT06] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006. 41, 42

[FKNN95] Tomás Feder, Eyal Kushilevitz, Moni Naor, and Noam Nisan. Amortized communication complexity. *SIAM J. Comput.*, 24(4):736–750, 1995. 2

[Gra90] Robert M. Gray. *Entropy and Information Theory*. Springer-Verlag New York, Inc., New York, NY, USA, 1990. 42

[HJMR07] Prahladh Harsha, Rahul Jain, David A. McAllester, and Jaikumar Radhakrishnan. The communication complexity of correlation. In *IEEE Conference on Computational Complexity*, pages 10–23, 2007. 2

[Jai11] Rahul Jain. New strong direct product results in communication complexity. *Electronic Colloquium on Computational Complexity (ECCC)*, 18:24, 2011. 2

[JPY12] Rahul Jain, Attila Pereszlényi, and Penghui Yao. A direct product theorem for the two-party bounded-round public-coin communication complexity. In *FOCS*, pages 167–176, 2012. 2

[JRS03] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A direct sum theorem in communication complexity via message compression. In *ICALP*, pages 300–315, 2003. 2

[Kah01] Jeff Kahn. An entropy approach to the hard-core model on bipartite graphs. *Combinatorics, Probability and Computing*, 10:219–237, 5 2001. 12, 14

[Kla10] Hartmut Klauck. A strong direct product theorem for disjointness. In *STOC*, pages 77–86, 2010. 2

[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. 2

[KN97] Eyal Kushilevitz and Noam Nisan. Communication complexity. *Cambridge University Press*, 1997. 1

[KR13] Gillat Kol and Ran Raz. Interactive channel capacity. In *STOC*, pages 715–724, 2013. 25, 32, 44

[LS09] Troy Lee and Adi Shraibman. Lower bounds in communication complexity. *Foundations and Trends in Theoretical Computer Science*, 3(4):263–398, 2009. 1

[MT10] Mokshay M. Madiman and Prasad Tetali. Information inequalities for joint distributions, with interpretations and applications. *IEEE Transactions on Information Theory*, 56(6):2699–2713, 2010. 12, 14, 15

[Rad03] Jaikumar Radhakrishnan. Entropy and counting. *IIT Kharagpur Golden Jubilee Volume*, page 125, 2003. 12, 14

A Information Theoretic Lemmas

Lemma (Lemma 8 restated, Shearer-Like Inequality for Relative Entropy). *Let $P, Q : \Omega_1 \times \cdots \times \Omega_M \to [0,1]$ be two distributions, such that Q is a product distribution, i.e., for every $j \in [M]$, there exists $Q_j : \Omega_j \to [0,1]$, such that $Q(x_1, \ldots, x_M) = \prod_{j \in [M]} Q_j(x_j)$. Let $T = \{T_i\}_{i \in I}$ be a collection of subsets of $[M]$, such that each element of $[M]$ appears in at most $\frac{1}{K}$ fraction of the members of T. For $A \subseteq [M]$, let P_A and Q_A be the marginal distributions of A in the distributions P and Q (respectively). Then,*

$$K \cdot \mathop{\mathbf{E}}_{i \in_R I}[\mathbf{D}(P_{T_i} \| Q_{T_i})] \leq \mathbf{D}(P \| Q).$$

Proof. For the proof of this lemma, it will be convenient to use the notion of conditional relative entropy using the notation of [CT06] (see Section 2.5 of [CT06]). For $A \subseteq [M]$, and

$(x_1, \ldots, x_M) \in \Omega_1 \times \ldots \times \Omega_M$, we denote $x_A = \{x_j : j \in A\}$. In this notation, we need to prove that

$$K \cdot \mathop{\mathbf{E}}_{i \in_R I} \left[\mathbf{D}\left(P(x_{T_i}) \| Q(x_{T_i}) \right) \right] \le \mathbf{D}\left(P(x_{[M]}) \| Q(x_{[M]}) \right).$$

Define $x_{<j} = \{x_\ell : \ell < j\}$ and $x_{T_i,<j} = \{x_\ell : \ell \in T_i, \ell < j\}$. By the chain rule for relative entropy (see Section 2.5 in [CT06]),

$$\mathbf{D}\left(P(x_{[M]}) \| Q(x_{[M]}) \right) = \sum_{j \in [M]} \mathbf{D}\left(P(x_j | x_{<j}) \| Q(x_j | x_{<j}) \right),$$

$$\mathbf{D}\left(P(x_{T_i}) \| Q(x_{T_i}) \right) = \sum_{j \in T_i} \mathbf{D}\left(P(x_j | x_{T_i,<j}) \| Q(x_j | x_{T_i,<j}) \right).$$

Since Q is a product distribution, conditioning can only increase the relative entropy (see, for example, Lemma 2.5.3 in [Gra90]). In particular, for every $j \in T_i$ it holds that

$$\mathbf{D}\left(P(x_j | x_{T_i,<j}) \| Q(x_j | x_{T_i,<j}) \right) \le \mathbf{D}\left(P(x_j | x_{<j}) \| Q(x_j | x_{<j}) \right).$$

Therefore,

$$\mathbf{D}\left(P(x_{T_i}) \| Q(x_{T_i}) \right) \le \sum_{j \in T_i} \mathbf{D}\left(P(x_j | x_{<j}) \| Q(x_j | x_{<j}) \right).$$

Summing over all $i \in I$ we get that

$$\sum_{i \in I} \mathbf{D}\left(P(x_{T_i}) \| Q(x_{T_i}) \right) \le \sum_{i \in I} \sum_{j \in T_i} \mathbf{D}\left(P(x_j | x_{<j}) \| Q(x_j | x_{<j}) \right). \tag{31}$$

For every $j \in [M]$, the term $\mathbf{D}\left(P(x_j | x_{<j}) \| Q(x_j | x_{<j}) \right)$ appears on the right-hand side of Equation (31) at most $\frac{|I|}{K}$ times. Therefore,

$$\begin{aligned} \sum_{i \in I} \mathbf{D}\left(P(x_{T_i}) \| Q(x_{T_i}) \right) &\le \frac{|I|}{K} \cdot \sum_{j \in [M]} \mathbf{D}\left(P(x_j | x_{<j}) \| Q(x_j | x_{<j}) \right) \\ &= \frac{|I|}{K} \cdot \mathbf{D}\left(P(x_{[M]}) \| Q(x_{[M]}) \right). \end{aligned}$$

Dividing by $\frac{|I|}{K}$ we get that the claim holds. □

Lemma 22. *Let $\Theta = (\Theta_1, \Theta_2)$ be a random variable uniformly distributed over a subset Γ of a domain $\Omega = \Omega_1 \times \Omega_2$, such that $\beta = \frac{|\Gamma|}{|\Omega|} \ge 2^{-2^{2k}}$. Let $\delta = 2^{-2^{2k}}$. Let Z be a boolean random variable, obtaining the value 1 with probability $1 - \delta$ and 0 with probability δ. Define $\Theta' = (\Theta'_1, \Theta'_2)$ to be Θ whenever $Z = 1$ and uniformly distributed over Ω whenever $Z = 0$.*

Then, for every $\theta_1 \in \Omega_1$, it holds that

$$\mathbf{I}(\Theta'_2 | \Theta'_1 = \theta_1) \le 2^{4k}.$$

Proof. For $\theta_1 \in \Omega_1$, define

$$\Gamma(\theta_1) = \{\theta_2 \in \Omega_2 \mid (\theta_1, \theta_2) \in \Gamma\}.$$

Let $\beta_1 = \frac{|\Gamma(\theta_1)|}{|\Omega_2|}$. It holds that

$$\begin{aligned}
&\mathbf{H}(\Theta_2'|\Theta_1' = \theta_1) \geq \mathbf{H}(\Theta_2'|Z, \Theta_1' = \theta_1) \\
&= \Pr_{Z,\Theta}[Z = 1|\Theta_1' = \theta_1] \cdot \log(|\Gamma(\theta_1)|) + \Pr_{Z,\Theta}[Z = 0|\Theta_1' = \theta_1] \cdot \log(|\Omega_2|) \\
&= \log(|\Omega_2|) - \Pr_{Z,\Theta}[Z = 1|\Theta_1' = \theta_1] \cdot \log\left(1/\beta_1\right).
\end{aligned}$$

Therefore,

$$\mathbf{I}(\Theta_2'|\Theta_1' = \theta_1) \leq \Pr_{Z,\Theta}[Z = 1|\Theta_1' = \theta_1] \cdot \log\left(1/\beta_1\right).$$

If $\beta_1 \geq 2^{-2^{4k}}$, then the assertion follows. Otherwise, $\beta_1 < 2^{-2^{4k}}$.

$$\begin{aligned}
&\Pr_{Z,\Theta}[Z = 1|\Theta_1' = \theta_1] \\
&= \frac{(1-\delta) \cdot \Pr[\Theta_1' = \theta_1|Z = 1]}{\delta \cdot \Pr[\Theta_1' = \theta_1|Z = 0] + (1-\delta) \cdot \Pr[\Theta_1' = \theta_1|Z = 1]} \\
&= \frac{(1-\delta) \cdot \frac{|\Gamma(\theta_1)|}{|\Gamma|}}{\delta \cdot \frac{1}{|\Omega_1|} + (1-\delta) \cdot \frac{|\Gamma(\theta_1)|}{|\Gamma|}} \\
&= \frac{(1-\delta) \cdot \frac{\beta_1}{\beta}}{\delta + (1-\delta) \cdot \frac{\beta_1}{\beta}} < \frac{\beta_1}{\delta \cdot \beta}.
\end{aligned}$$

This implies,

$$\mathbf{I}(\Theta_2'|\Theta_1' = \theta_1) \leq \frac{\beta_1 \log\left(1/\beta_1\right)}{\delta \cdot \beta} < \frac{2^{-2^{4k}} \cdot 2^{4k}}{2^{-2^{2k}} \cdot 2^{-2^{2k}}} < 1,$$

where the second to last inequality is since the function $x \log(1/x)$ is monotone increasing for $0 \leq x \leq \frac{1}{e}$. $\square$

Lemma 23. *Let τ be a probability distribution over a domain Ω. Assume $\mathbf{I}(\tau) \leq I$ for $I \in \mathbb{R}^+$. Let $D \subset \Omega$ be such that $\frac{|D|}{|\Omega|} \leq \alpha$. Then,*

$$\tau(D) \leq \frac{I+1}{\log\left(\frac{1}{\alpha}\right)}.$$

Proof. Let $\delta := \tau(D)$. Consider the distribution τ' that assigns each of the elements in D a probability of $\frac{\delta}{|D|} \geq \frac{\delta}{\alpha|\Omega|}$, and each of the other elements a probability of $\frac{1-\delta}{|\Omega|-|D|} \geq \frac{1-\delta}{|\Omega|}$. That is, τ' is obtained by re-distributing the weight of τ, so it will be uniform on D and on $\Omega \setminus D$. Observe that $\mathbf{H}(\tau') \geq \mathbf{H}(\tau)$, and hence $\mathbf{I}(\tau') \leq \mathbf{I}(\tau)$. Therefore,

$$\begin{aligned}
I \geq \mathbf{I}(\tau) \geq \mathbf{I}(\tau') &\geq \delta \log\left(\tfrac{\delta}{\alpha}\right) + (1-\delta)\log(1-\delta) \\
&= \delta \log\left(\tfrac{1}{\alpha}\right) - \mathbf{H}(\delta, 1-\delta) \geq \delta \log\left(\tfrac{1}{\alpha}\right) - 1 = \tau(D) \log\left(\tfrac{1}{\alpha}\right) - 1.
\end{aligned}$$

$\square$

Lemma 24 (**Lemma 5.11 in** [KR13])**.** *Let* $\mu : \Omega \to [0,1]$ *be a distribution satisfying* $I = \mathbf{I}(\mu) \leq 0.01$*. Let* $\mathcal{A} \subseteq \Omega$ *be the set of elements with* $\mu(x) < \frac{1}{|\Omega|}$*. Denote*

$$I^{neg}(\mu) = -\sum_{x \in \mathcal{A}} \mu(x) \log(|\Omega|\mu(x)).$$

Then,

$$I^{neg}(\mu) < 4I^{0.25} \log\left(\tfrac{1}{I^{0.25}}\right) < 4I^{0.1}.$$

Lemma 25 (**Lemma 5.12 in** [KR13])**.** *Let* $\mu : \Omega \to [0,1]$ *be a distribution satisfying* $I = \mathbf{I}(\mu) \leq 0.01$*. Let* $\mathcal{A} \subseteq \Omega$ *be the set of elements with* $\mu(x) \geq \frac{2}{|\Omega|}$*. Then,*

$$\mu(\mathcal{A}) < 4I^{0.25} \log\left(\tfrac{1}{I^{0.25}}\right) + I < 5I^{0.1}.$$

ECCC ISSN 1433-8092

http://eccc.hpi-web.de