Electronic Colloquium on Computational Complexity, Report No. 113 (2014)

Exponential Separation of Information and Communication for Boolean Functions

Anat Ganor* Gillat Kol† Ran Raz‡

Abstract

We show an exponential gap between communication complexity and information complexity for boolean functions, by giving an explicit example of a partial function with information complexity $\leq O(k)$, and distributional communication complexity $\geq 2^k$. This shows that a communication protocol for a partial boolean function cannot always be compressed to its internal information. By a result of Braverman [Bra12], 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 of boolean functions cannot hold, answering a long standing open problem.

Our techniques build on [GKR14], that proved a similar result for relations with very long outputs (double exponentially long in k). In addition to the stronger result, the current work gives a simpler proof, benefiting from the short output length of boolean functions.

Another (conceptual) contribution of our work is the *relative discrepancy* method, a new rectangle-based method for proving communication complexity lower bounds for boolean functions, powerful enough to separate information complexity and communication complexity.

1 Introduction

The classical works of Shannon, Fano and Huffman show that if Alice wants to send a message x to Bob, it's sufficient for her to send $\lceil \mathbf{H}(x) \rceil$ bits, in expectation, where $\mathbf{H}$ denotes

*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

Shannon's entropy function [Sha48, Fan49, Huf52]. In other words, the length of the message x can be compressed to roughly $\mathbf{H}(x)$, the information content of the message. Can one prove analogous results in the interactive setting, where Alice and Bob engage in an interactive communication protocol? The standard way to formalize this question is as whether or not there exist gaps between the *information complexity* and *communication complexity* of communication tasks.

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, such as, computing a function of 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].

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 [KRW95, FKNN95, CSWY01, JRS03, Sha03, 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] (where $\tilde{O}$ hides logarithmic factors in both k and c). 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 $2^{O(k)}$ [Bra12] (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.

Another line of works shows that many of the known general techniques for proving lower bounds for randomized communication complexity also give similar lower bounds for information complexity [Bra12, BW12, KLL+12], and hence cannot be used to separate information complexity and communication complexity.

In [GKR14] we showed an exponential gap between information complexity and communication complexity for relations with very long outputs (double exponentially long in k). The current work builds on the techniques of [GKR14] and gives the first gap between information complexity and communication complexity for boolean functions. We give an explicit example for a partial[1] boolean function, called the *bursting noise function*, parameterized by $k \in \mathbb{N}$ and applied on inputs distributed according to an input distribution μ. We prove that the information complexity of the function is $O(k)$, while any communication protocol for computing this function, with communication complexity at most 2^k, has error very close to $1/2$. By the above mentioned compression protocol of Braverman [Bra12], 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 function with parameter k, that has communication complexity at most 2^k, errs with probability $\epsilon \geq \frac{1}{2} - 2^{-k}$ (over the input distribution μ).*

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

We note that the inputs to the bursting noise function are very long, namely, triple exponential in k. The protocol that achieves information complexity $O(k)$ has communication complexity double exponential in k.

The Direct Sum Problem

As mentioned above, information complexity is also related to the direct sum problem in communication complexity.

Let μ be a distribution on $\{0,1\}^n \times \{0,1\}^n$, and let $f : \text{supp}(\mu) \to \{0,1\}$ be a boolean function. Let $D^\mu(f)$ be the communication complexity of the best protocol that computes f

[1] We note that since information complexity and distributional communication complexity are always defined with respect to an input distribution μ, there is no difference in this context between a partial boolean function and a total boolean function.

with probability at least $2/3$, where the probability is over inputs distributed according to μ and over the random bits of the protocol. Let $D^{\mu,N}(f)$ be the communication complexity of the best protocol that computes f on N independent pairs sampled according to μ, getting the answer correct with probability at least $2/3$ in each coordinate (where the probability is over the inputs and over the random bits of the protocol). The amortized communication complexity of f is defined to be $\lim_{N\to\infty} \frac{D^{\mu,N}(f)}{N}$, that is, the limit of the communication complexity needed to solve N tasks of the same type, divided by N.

Braverman and Rao showed that information complexity is equal to the amortized communication complexity [BR11]. Our result therefore shows an exponential gap between distributional communication complexity and amortized distributional communication complexity for a boolean function, proving that tight direct sum results cannot hold. This gives the first gap between distributional communication complexity and amortized distributional communication complexity for boolean functions.

Techniques

Underlying our lower bound proof is the *relative discrepancy* method, a new rectangle-based method for proving communication complexity lower bounds for boolean functions, powerful enough to separate information complexity and communication complexity. We describe this new method in Section 3.

Our techniques build on [GKR14]. We improve the lower bound technique so that it can be applied for boolean functions, rather than for relations with extremely long outputs. In addition to the stronger result, the current work gives a simpler proof, benefiting from the short output length of boolean functions. Roughly speaking, since the output length of a boolean function is 1, it is easy to ensure that in each rectangle induced by the communication protocol, the answer is unique and does not depend on the inputs.

Organization

The paper is organized as follows. In Section 2 we define the bursting noise function. Section 3 describes the relative discrepancy bound. In Section 4, we give an overview of the lower bound proof. In Section 5 we give general definitions and preliminaries. In Section 6 we prove the graph correlation lemma, a central tool that we will use in the lower bound proof. In Section 7 we prove the communication complexity lower bound (Theorem 1). Section 8 gives a general tool that can be used to upper bound the information cost of a protocol, using the notion of a divergence cost of a tree. In Section 9 we give a protocol for the bursting noise function with low information cost, thus proving the upper bound required by Theorem 2. Section 6 and Section 8 are similar to the corresponding sections in [GKR14], and are included here for completeness.

2 The Bursting Noise Function

The *bursting noise function* can be viewed as 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 $c = 2^{4^k}$ and $w = 2^{100}k$.

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 $\text{Even}(\mathcal{T}) \subseteq V$ the set of non-leaf vertices in an even layer of $\mathcal{T}$ and by $\text{Odd}(\mathcal{T}) \subseteq V$ the set of non-leaf vertices in an odd layer of $\mathcal{T}$. We think of the vertices in $\text{Odd}(\mathcal{T})$ as "owned" by the first player and the vertices in $\text{Even}(\mathcal{T})$ 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 $y_v = b$.

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, $i^* = (i-1)w + 1$.

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 $i \in [c]$ be a multi-layer. Denote $s = i^*$ and $t = s + w - 1 = (i+1)^* - 1$. Let $t' \in [(i+1)^*, cw]$, and let $v \in V$ be a vertex in layer t' of $\mathcal{T}$. For $j \in [s, t+1]$, let v_j be v's ancestor in layer j. We say that v is *typical* with respect to i, x, y, if the followings hold:

1. For at least 0.8-fraction of the indices $j \in [s,t] \cap \text{Odd}(\mathcal{T})$, 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 $j \in [s,t] \cap \text{Even}(\mathcal{T})$, 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 i, x, y, it suffices to know the bits that x, y assign to the vertices $v_s, \ldots, v_t$. When x, y are clear from the context, we omit x, y and say that v is typical with respect to multi-layer i.

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 bit b defined by Step (5) of Algorithm 1. Note that for any leaf $v \in V$, where v is typical with respect to i, x, y (that is, v is typical with respect to the noisy multi-layer; see Algorithm 1), we have that $x_v \oplus y_v = b$, by Step (4)

Algorithm 1 Sample (x, y) according to μ

1. Randomly select $i \in [c]$ (the noisy multi-layer).
2. Set every vertex in multi-layer i (layers $[i^*, i^* + w - 1]$) to be noisy.
3. If $i < c$: Let L be the set of all non-typical vertices in layer $i^* + w = (i + 1)^*$ with respect to i, x, y (note that x, y were already defined on layers $[i^*, i^* + w - 1]$, and therefore the typical vertices are defined). For every $v \in L$, set all the vertices in the subtree rooted at v to be noisy.
4. Set all unset vertices in V to be non-noisy.
5. Randomly select a bit $b \in \{0, 1\}$.
   For every leaf $v \in V$, add b to y_v, that is, $y_v \leftarrow y_v \oplus b$.

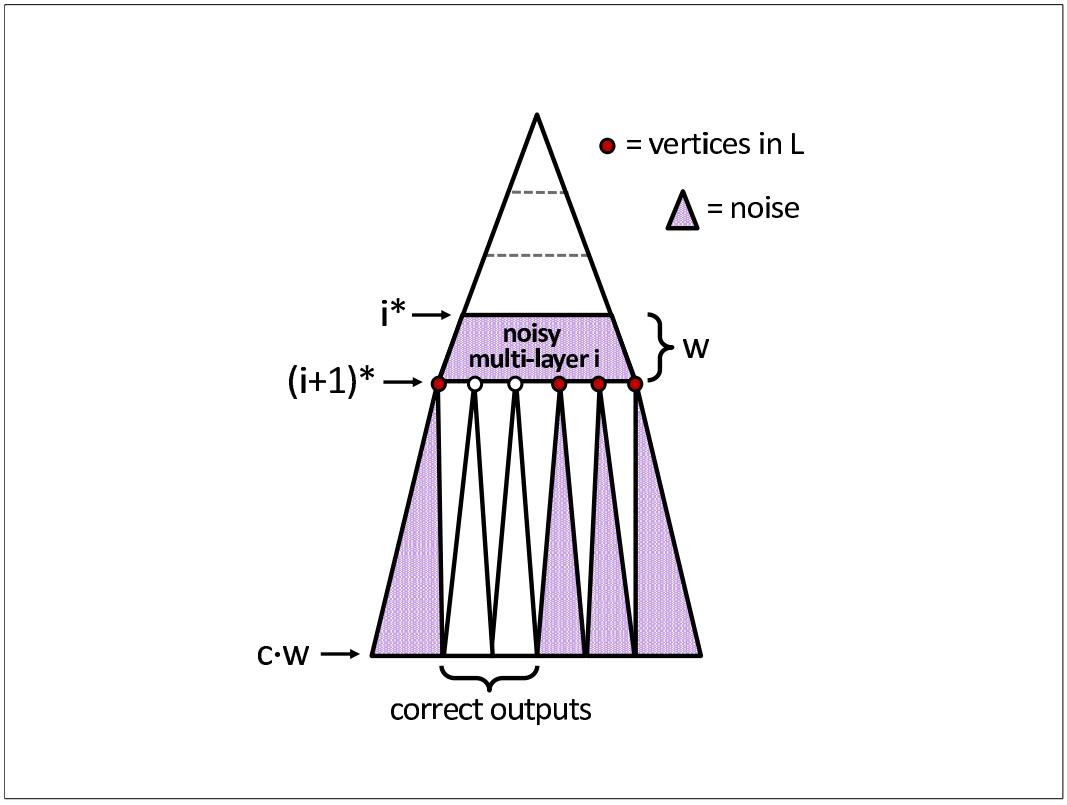

Figure 1: Illustration of Algorithm 1

of Algorithm 1. The bit b is a boolean function of $(x, y) \in \text{supp}(\mu)$, since for the leaf $v \in V$ obtained by following the path of correct children starting from the root, it holds that $x_v \oplus y_v = b$. We denote by

$$f : \text{supp}(\mu) \to \{0, 1\}$$

the function that assigns to every input $(x, y) \in \text{supp}(\mu)$ the corresponding bit b.

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 μ_i is uniformly distributed over $\text{supp}(\mu_i)$ and that $\mu = \frac{1}{c} \sum_{i \in [c]} \mu_i$.

For $i \in [c]$, we denote by μ_i^0 the uniform distribution over $\text{supp}(\mu_i) \cap f^{-1}(0)$. We denote by μ_i^1 the uniform distribution over $\text{supp}(\mu_i) \cap f^{-1}(1)$. Observe that $\mu_i = \frac{1}{2}(\mu_i^0 + \mu_i^1)$, and for every set $S \subseteq \{0, 1\}^V \times \{0, 1\}^V$, it holds that $\mu_i^0(S) = 2\mu_i(S \cap f^{-1}(0))$ and $\mu_i^1(S) = 2\mu_i(S \cap f^{-1}(1))$.

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 noisy multi-layer, and then find a typical leaf with respect to this multi-layer, and thus compute the bursting noise function with communication complexity polynomial in k.*

The protocol with low information cost. Consider the following protocol π' for the bursting noise function. 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 v they output $x_v \oplus y_v$. By the Chernoff bound, the probability that the players reach a leaf that is not typical with respect to the noisy multi-layer is at most $2^{-\Omega(w)}$. Therefore, the error probability of π' is exponentially small in k.

It can be shown that if the protocol π' does not reach a vertex in L (a non-typical vertex with respect to the noisy multi-layer), then it reveals a small amount of information. Intuitively, this follows since in this case, the expected number of vertices reached by the protocol, on which the players' inputs disagree, is $O(k)$ (the disagreement is only on vertices in the noisy multi-layer). However, with exponentially small probability in k, the protocol π' does reach a vertex in L. In this case, the information revealed by the protocol may be double exponential in k (as $c = 2^{4^k}$), making the information cost of π' too large.

For this reason, we consider a variant of π', called π. Informally speaking, the protocol π operates like π' but aborts if too much information about the inputs is revealed. Specifically, a player decides to abort if the bits that she receives differ from the corresponding bits in her input too many times. In Section 9, we formally define π and show that its information cost is $O(k)$.

3 The Relative Discrepancy Bound

In this section we present the *relative discrepancy bound*, a general method for proving communication complexity lower bounds. For our lower bound proof, we will only use Definition 1 and Proposition 3.

Definition 1 (**Relative Discrepancy**). *Let $\epsilon \in (0, 1/2)$ and $\delta \in (0, 1)$. Let μ be a distribution over $\{0,1\}^n \times \{0,1\}^n$ and let $f : \mathrm{supp}(\mu) \to \{0,1\}$ be a function. We say that (f, μ) has the* (ϵ, δ) relative discrepancy property *if there exists a distribution ρ over $\{0,1\}^n \times \{0,1\}^n$ such that for every rectangle $R = A \times B \subseteq \{0,1\}^n \times \{0,1\}^n$ with $\rho(R) \geq \delta$, the following two properties hold:*

$$\mu\left(R \cap f^{-1}(0)\right) \geq \left(\tfrac{1}{2} - \epsilon\right) \cdot \rho(R),$$
$$\mu\left(R \cap f^{-1}(1)\right) \geq \left(\tfrac{1}{2} - \epsilon\right) \cdot \rho(R).$$

Definition 2 (**Adaptive Relative Discrepancy**). *Let $\epsilon \in (0, 1/2)$ and $\delta \in (0, 1)$. Let μ be a distribution over $\{0,1\}^n \times \{0,1\}^n$ and let $f : \mathrm{supp}(\mu) \to \{0,1\}$ be a function. We say that (f, μ) has the* (ϵ, δ) adaptive relative discrepancy property *if for every rectangle partition $\mathcal{R} = \{R^1, \ldots, R^m\}$ of $\{0,1\}^n \times \{0,1\}^n$, there exists a distribution $\rho_{\mathcal{R}}$ over $\{0,1\}^n \times \{0,1\}^n$ such that for every rectangle R^t with $\rho_{\mathcal{R}}(R^t) \geq \delta$, the following two properties hold:*

$$\mu\left(R^t \cap f^{-1}(0)\right) \geq \left(\tfrac{1}{2} - \epsilon\right) \cdot \rho_{\mathcal{R}}(R^t),$$
$$\mu\left(R^t \cap f^{-1}(1)\right) \geq \left(\tfrac{1}{2} - \epsilon\right) \cdot \rho_{\mathcal{R}}(R^t).$$

Observe that the (ϵ, δ) relative discrepancy property implies the (ϵ, δ) adaptive relative discrepancy property. Thus, the following propositions are stated for the adaptive case, but apply for the non-adaptive case as well.

Our first proposition shows how the adaptive relative discrepancy property can be used to obtain distributional communication complexity lower bounds.

Proposition 3. *Let $\epsilon \in (0, 1/2)$ and $\delta \in (0, 1)$. Let μ be a distribution over $\{0,1\}^n \times \{0,1\}^n$ and let $f : \mathrm{supp}(\mu) \to \{0,1\}$ be a function. Assume that (f, μ) has the (ϵ, δ) adaptive relative discrepancy property. Then, every randomized protocol for f with communication complexity at most s, errs with probability at least $\frac{1}{2} - \epsilon - \delta \cdot 2^s$ when the inputs are distributed according to μ (that is, the protocol has advantage at most $\epsilon + \delta \cdot 2^s$). Equivalently, if the protocol has advantage larger than ϵ', then it has communication complexity at least $\log\left(\frac{\epsilon' - \epsilon}{\delta}\right)$.*

Proof. We show that the claim holds for every deterministic protocol, and therefore it also holds for randomized protocols. Let π be a deterministic communication protocol for f with communication complexity at most s. Let $m = 2^s$ and let $\mathcal{R} = \{R^1, \ldots, R^m\}$ be the rectangle partition induced by the protocol π, such that each rectangle R^t is associated with an output $\omega^t \in \{0,1\}$ (the output of the protocol π on this rectangle). Let $\rho = \rho_{\mathcal{R}}$ be the distribution corresponding to $\mathcal{R}$ promised by the (ϵ, δ) adaptive relative discrepancy property.

Let $\mathcal{E}$ be the set of inputs $(x, y) \in \text{supp}(\mu)$ that the protocol π errs on. Our goal is to lower bound $\mu(\mathcal{E})$, that is, the probability that the protocol errs when the inputs are distributed according to μ. Let $L = \{t \in [m] \ : \ \rho(R^t) \geq \delta\}$. It holds that $\sum_{t \in L} \rho(R^t) \geq 1 - \delta m$. By the (ϵ, δ) adaptive relative discrepancy property,

$$\begin{aligned}\mu(\mathcal{E}) &= \sum_{t \in [m]} \mu(R^t \cap f^{-1}(1 - \omega^t)) \geq \sum_{t \in L} \mu(R^t \cap f^{-1}(1 - \omega^t)) \\ &\geq \left(\tfrac{1}{2} - \epsilon\right) \cdot \sum_{t \in L} \rho(R^t) \geq \left(\tfrac{1}{2} - \epsilon\right) \cdot (1 - \delta m) \geq \tfrac{1}{2} - \epsilon - \delta \cdot 2^s.\end{aligned}$$

□

Given the (ϵ, δ) adaptive relative discrepancy property, Proposition 3 can only be used to bound the communication complexity of protocols with advantage greater than ϵ. When ϵ is close to $1/2$, the proposition cannot be used to prove lower bounds for protocol with error, say, $1/3$. The following corollary shows that in the randomized case, by first applying error reduction, the property can be used to prove lower bounds for protocols with error $1/3$, even when ϵ is close to $1/2$.

Corollary 4. *Let $\gamma \in (0, 1/2)$ and $\delta \in (0, 1)$. Let $f : \mathcal{D} \to \{0, 1\}$ be a function, where $\mathcal{D} \subseteq \{0, 1\}^n \times \{0, 1\}^n$. Assume that there exists a distribution μ over $\mathcal{D}$ such that (f, μ) has the $(\frac{1}{2} - \gamma, \delta)$ adaptive relative discrepancy property. Then, every randomized protocol for f with error probability at most $1/3$ (on every input) has communication complexity at least $\Omega\left(\frac{\log(1/2\delta)}{\log(1/\gamma)}\right) - 1$.*

Proof. Consider a protocol for f with error probability at most $1/3$ and communication complexity s. By repeating this protocol $O(\log(1/\gamma))$ times we obtain a protocol with error at most $\gamma/2$ and communication complexity at most $s' = O(s \cdot \log(1/\gamma))$. By Proposition 3, $\gamma - \delta \cdot 2^{s'} \leq \gamma/2$, thus $s' \geq \log(\gamma/2\delta)$. Hence

$$s \geq \Omega\left(\frac{\log(\gamma/2\delta)}{\log(1/\gamma)}\right) = \Omega\left(\frac{\log(1/2\delta)}{\log(1/\gamma)}\right) - 1.$$

□

Connection to the Discrepancy Method

We show that a special case of the relative discrepancy property implies an upper bound on the (regular) discrepancy (defined below), and vice versa.

Definition 3 (Discrepancy). *Let μ be a distribution over $\{0, 1\}^n \times \{0, 1\}^n$ and let $f : \text{supp}(\mu) \to \{0, 1\}$ be a function. The* discrepancy *of f according to μ is*

$$\mathsf{Disc}_\mu(f) = \max_R \mathsf{Disc}_\mu(R, f)$$

where the maximum is taken over all rectangles $R \subseteq \{0,1\}^n \times \{0,1\}^n$ *and*

$$\mathsf{Disc}_\mu(R, f) = \left|\mu\left(R \cap f^{-1}(0)\right) - \mu\left(R \cap f^{-1}(1)\right)\right|.$$

Let $\epsilon \in (0, 1/2)$ and $\delta \in (0, 1)$. If the (ϵ, δ) relative discrepancy property holds when the distribution ρ is equal to μ, then for every rectangle $R \subseteq \{0,1\}^n \times \{0,1\}^n$ with $\mu(R) \geq \delta$ it holds that

$$\left(\tfrac{1}{2} - \epsilon\right) \cdot \mu(R) \leq \mu\left(R \cap f^{-1}(0)\right) \leq \left(\tfrac{1}{2} + \epsilon\right) \cdot \mu(R),$$

where we used the fact that $\mu\left(R \cap f^{-1}(1)\right) = \mu(R) - \mu\left(R \cap f^{-1}(0)\right)$. Therefore,

$$\mathsf{Disc}_\mu(R, f) = \left|2\mu\left(R \cap f^{-1}(0)\right) - \mu(R)\right| \leq 2\epsilon \cdot \mu(R).$$

For a rectangle $R \subseteq \{0,1\}^n \times \{0,1\}^n$ with $\mu(R) < \delta$ it holds that $\mathsf{Disc}_\mu(R, f) < \delta$. Taking the maximum over all rectangles, we get that

$$\mathsf{Disc}_\mu(f) \leq \max\{\delta, 2\epsilon\}.$$

On the other direction, if we have an upper bound of $\epsilon' > 0$ on the discrepancy of f according to μ, then for every rectangle $R \subseteq \{0,1\}^n \times \{0,1\}^n$,

$$\mu\left(R \cap f^{-1}(0)\right) \geq \tfrac{1}{2}\left(\mu(R) - \epsilon'\right),$$

and

$$\mu\left(R \cap f^{-1}(1)\right) \geq \tfrac{1}{2}\left(\mu(R) - \epsilon'\right).$$

Let $\epsilon \in (0, 1/2)$ and $\delta \in (0, 1)$ such that $\epsilon\delta \geq \epsilon'/2$. Then, for every rectangle $R \subseteq \{0,1\}^n \times \{0,1\}^n$ with $\mu(R) \geq \delta$,

$$\mu\left(R \cap f^{-1}(0)\right) \geq \frac{\mu(R)}{2} - \epsilon\delta \geq \left(\tfrac{1}{2} - \epsilon\right) \cdot \mu(R),$$

and similarly,

$$\mu\left(R \cap f^{-1}(1)\right) \geq \left(\tfrac{1}{2} - \epsilon\right) \cdot \mu(R).$$

Taking ρ to be equal to μ, we get that the (ϵ, δ) relative discrepancy property holds.

4 Overview of the Lower Bound Proof

In this section, we overview the proof of the lower bound for the communication complexity of the bursting noise function. The lower bound is proved using the relative discrepancy bound. Let $f : \mathrm{supp}(\mu) \to \{0,1\}$ be the bursting noise function, with parameter k, where μ is the distribution defined by Algorithm 1. Let $m = 2^{2^k}$, $\epsilon = 2^{-2k}$ and $\delta = \epsilon/m$. We show that (f, μ) has the (ϵ, δ) relative discrepancy property. Theorem 1 follows by Proposition 3.

The distribution ρ

For $i \in [c]$, let ρ_i be the uniform distribution over all inputs (x, y) such that $x_{<i} = y_{<i}$, where $x_{<i}$ and $y_{<i}$ are the projections of x, y, respectively, on the vertices in the first $i-1$ multi-layers. We define the distribution ρ as $\frac{1}{c}\sum_{i\in[c]} \rho_i$. Let $R \subseteq \{0,1\}^V \times \{0,1\}^V$ be a rectangle such that $\rho(R) \geq \delta$. We will show the first part of the relative discrepancy property:

$$\mu\left(R \cap f^{-1}(0)\right) \geq \left(\tfrac{1}{2} - \epsilon\right) \cdot \rho(R).$$

The second part of the relative discrepancy property (for $f^{-1}(1)$) is proved in the same way.

Good Rectangles

For $i \in [c]$ and an assignment z to the vertices in the first $i-1$ multi-layers, we denote by $R^z = A^z \times B^z$, the rectangle of all pairs of inputs $(x, y) \in R$, such that the projections of both x, y on the vertices in the first $i-1$ multi-layers are equal to z. Let X^z be a random variable uniformly distributed over A^z, and let Y^z be a random variable uniformly distributed over B^z. We denote by X_i^z, Y_i^z the projections of X^z, Y^z, respectively, on the vertices in multi-layer i. 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. We say that the pair (i, z) is *good* if the following two properties hold:

1. $\mathbf{I}\left(X^z\right), \mathbf{I}\left(Y^z\right) \leq 2\log(m)$.
2. $\mathbf{I}\left(X_i^z\right), \mathbf{I}\left(Y_i^z\right) \leq \frac{m^2}{c}$.

Let $\mathcal{G}$ be the set of all good pairs (i, z).

How the Proof Works

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 start by showing that with high probability, when the inputs are distributed uniformly and independently after the noisy multi-layer, very little information is known on the noisy multi-layer. Then, we show that even when we consider only inputs from $\text{supp}(\mu)$, that agree on all non-noisy vertices after the noisy multi-layer, still very little information is known on the noisy multi-layer.

Formally, recall that μ_i^0 is the uniform distribution over $\text{supp}(\mu_i) \cap f^{-1}(0)$, and therefore,

$$\mu\left(R \cap f^{-1}(0)\right) \geq \frac{1}{2c} \sum_{(i,z)\in\mathcal{G}} \mu_i^0(R^z).$$

Together with Equations (1) and (2) that appear below, the proof is completed.

First, we show that

$$\frac{1}{2c}\left(1 - 2^{-2k}\right) \sum_{(i,z)\in\mathcal{G}} \rho_i(R^z) \geq \left(\tfrac{1}{2} - \epsilon\right) \cdot \rho(R). \tag{1}$$

That is, the probability $\rho(R)$ is concentrated mainly on the rectangles R^z with good (i,z) pairs. To prove Equation (1) we show that with high probability, where every pair (i,z) is sampled with the probability $\frac{\rho_i(R^z)}{c\cdot\rho(R)}$, very little information is known about X_i^z and Y_i^z. Therefore, the sum

$$\sum_{(i,z)\in\mathcal{G}} \frac{\rho_i(R^z)}{c\cdot\rho(R)},$$

is close to 1.

Next, we show that for every pair $(i,z)\in\mathcal{G}$,

$$\mu_i^0(R^z) \geq \left(1-2^{-2k}\right)\rho_i(R^z). \tag{2}$$

Here lies the main difficulty in proving the lower bound. Using the definitions of μ_i^0 and ρ_i, we show that Equation (2) is equivalent to

$$\Pr\left[(X^z,Y^z)\in \mathrm{supp}(\mu_i^0)\right] \geq \left(1-2^{-2k}\right)\rho_i(\mathrm{supp}(\mu_i^0)). \tag{3}$$

That is, the probability for a random pair of inputs $(x,y)\in R^z$ to be in $\mathrm{supp}(\mu_i^0)$ is not much smaller than the probability for a uniformly distributed pair of inputs, that have the same projection on the vertices in the first $i-1$ multi-layers, to be in $\mathrm{supp}(\mu_i^0)$. In what follows, we outline the proof of Equation (3).

Applying the Graph Correlation Lemma

Fix a good (i,z) pair and assume that the noisy multi-layer is i. Let E be the set of all pairs of possible assignments to the vertices in multi-layer i. Observe that an input pair x_i, y_i of assignments to the vertices in multi-layer i determine for every vertex after multi-layer i if it is noisy or not. A pair $(x,y)\in R^z$ is in $\mathrm{supp}(\mu_i^0)$ if and only if x,y agree on all the vertices after multi-layer i that are set to be non-noisy for inputs x_i, y_i. Therefore, the left hand side of Equation (3) is equal to

$$\sum_{(u,w)\in E} \Pr\left[X_i^z=u\right]\cdot\Pr\left[Y_i^z=w\right]$$
$$\cdot\Pr\left[X_{>i}^z \text{ and } Y_{>i}^z \text{ agree on all non-noisy vertices} \mid X_i^z=u, Y_i^z=w\right],$$

where $X_{>i}^z$ and $Y_{>i}^z$ are the projections of X^z, Y^z, respectively, on the vertices after multi-layer i.

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

$$\geq \left(1-2^{-4k}\right)p_i \sum_{(u,w)\in E\setminus\mathcal{D}} \Pr\left[X_i^z=u\right]\cdot\Pr\left[Y_i^z=w\right], \tag{4}$$

where $\mathcal{D}\subset E$ is a small set, compared to the size of E, and p_i is the probability for a uniformly distributed pair of inputs (x,y), that have the same projection on the vertices in

the first $i-1$ multi-layers, to agree on all the vertices after multi-layer i that are set to be non-noisy for inputs x_i, y_i. It holds that $p_i = \rho_i(\text{supp}(\mu_i^0))$. 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^z, Y^z, by the expression in Equation (4) that depends only on the projections of these variables to the vertices in multi-layer i.

We still need to bound from below the expression

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

This sum would be equal to 1 if it was over all pairs of assignments in E, including the pairs in the set $\mathcal{D}$. Since $\mathbf{I}\left(X_i^z\right), \mathbf{I}\left(Y_i^z\right)$ are small, the distributions of X_i^z and Y_i^z are extremely close to uniform, and hence,

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

which is negligible.

5 Definitions and Preliminaries

5.1 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.

5.2 Information Cost

Definition 4 (**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.

5.3 Relative Entropy

Definition 5 (**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*

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

Proposition 5. *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 6 (**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

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

5.4 Information

Definition 6 (**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 $\mathbf{I}(X) = \mathbf{I}(P_X)$.

5.5 Shearer-Like Inequality for Information

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

Lemma 7 (**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 8 (**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,

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

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

$$\sum_{i \in I} \mathbf{I}(X_{T_i}) \leq \sum_{i \in I} \sum_{j \in T_i} \Big(\log(|\Omega_j|) - \mathbf{H}[X_j | X_{<j}] \Big). \tag{5}$$

For every $j \in [M]$, the term $\log(|\Omega_j|) - \mathbf{H}[X_j | X_{<j}]$ appears on the right-hand side of Equation (5) at most $\frac{|I|}{K}$ times. Therefore,

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

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

6 The Graph Correlation Lemma

Lemma 9 (**Graph Correlation Lemma**). [2] *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,

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

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

$$|\Sigma|^T \cdot \sum_{z \in \Sigma^{T(u,w)}} \left(\Pr_{X^u}[X^u_{T(u,w)} = z] - |\Sigma|^{-T} \right) \cdot \left(\Pr_{Y^w}[Y^w_{T(u,w)} = z] - |\Sigma|^{-T} \right). \tag{6}$$

[2] 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 (6) 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 8,

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

in contradiction to the assumption of the lemma. □

7 Communication Lower Bound

In this section we prove Theorem 1. That is, we show that any randomized communication protocol for the bursting noise function with parameter k, that has communication complexity at most 2^k, has error at least $\epsilon \geq \frac{1}{2} - 2^{-k}$ (when the inputs are selected according to the distribution μ). The lower bound is proved using the relative discrepancy bound (Proposition 3).

7.1 Notation

Fix a rectangle $R = A \times B$, for $A, B \subseteq \{0,1\}^V$. Let $m = 2^{2^k}$. Let X be a random variable taking values in $\{0,1\}^V$, that is uniformly distributed over A. Let Y be a random variable taking values in $\{0,1\}^V$, that is uniformly distributed over B.

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. Define $V_{>i} \subseteq V$ to be the set of vertices in multi-layers $i+1$ 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, $Z_{\geq i}$, $Z_{>i}$ to be the projections of Z to $V_{<i}$, V_i, $V_{\geq i}$, $V_{>i}$ (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 $|\Psi^z| = |\{0,1\}^{V_{\geq i}}|$. Define $A^z = A \cap \Psi^z$ and $B^z = B \cap \Psi^z$. Define $R^z = A^z \times B^z$. Let X^z be a random variable taking values in Ψ^z, that is uniformly distributed over A^z. Let Y^z be a random variable taking values in Ψ^z, that is uniformly distributed over B^z.

7.2 The Distribution ρ

For $i \in [c]$, define ρ_i to be the uniform distribution over the set

$$\bigcup_{z \in \{0,1\}^{V_{<i}}} \Psi^z \times \Psi^z.$$

That is, ρ_i is the uniform distribution over all inputs (x, y) such that $x_{<i} = y_{<i}$. Define the distribution ρ over $\{0,1\}^V \times \{0,1\}^V$ as $\frac{1}{c}\sum_{i \in [c]} \rho_i$.

For $i \in [c]$ and $z \in \{0,1\}^{V_{<i}}$, define $\rho_{i,z}$ to be the uniform distribution over the set $\Psi^z \times \Psi^z$. That is, $\rho_{i,z}$ is the uniform distribution over all inputs (x, y) such that $x_{<i} = y_{<i} = z$. In particular,

$$\rho_{i,z}(R) = \rho_{i,z}(R^z) = \frac{|R^z|}{|\Psi^z \times \Psi^z|}. \tag{7}$$

Recall that we fixed a rectangle R. Define $\hat{\rho}$ to be the distribution over the set of pairs (i, z), where $i \in [c]$ and $z \in \{0,1\}^{V_{<i}}$, by

$$\hat{\rho}(i, z) = \frac{\rho_i(R^z)}{c \cdot \rho(R)} = \frac{\rho_{i,z}(R)}{c \cdot |\{0,1\}^{V_{<i}}| \cdot \rho(R)}. \tag{8}$$

7.3 Bounding the Information on the Noisy Multi-Layer

The following lemma shows that, in expectation over (i, z) sampled according to $\hat{\rho}$, the distribution of the projections of inputs in R^z to multi-layer i is close to uniform.

Lemma 11. *It holds that*

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

and similarly,

$$\mathop{\mathbf{E}}_{(i,z) \leftarrow \hat{\rho}} \left[\mathbf{I}\left(Y_i^z\right)\right] \leq \frac{1}{c \cdot \rho(R)}.$$

Proof. First observe that by Equation (8),

$$\rho(R) \mathop{\mathbf{E}}_{(i,z) \leftarrow \hat{\rho}} \left[\mathbf{I}\left(X_i^z\right)\right] = \mathop{\mathbf{E}}_{i \in_R [c]} \mathop{\mathbf{E}}_{z \in_R \{0,1\}^{V_{<i}}} \left[\rho_{i,z}(R) \cdot \mathbf{I}\left(X_i^z\right)\right].$$

Therefore, it suffices to upper-bound the right hand side by $1/c$.

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

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

We have that

$$\begin{aligned}
&\sum_{z \in \{0,1\}^{V_{<i}}} \frac{|A^z|}{|A|} \cdot \mathbf{H}\left(X_i^z\right) = \sum_{z \in \{0,1\}^{V_{<i}}} \frac{|A^z|}{|A|} \cdot \mathbf{H}\left(X_i | X_{<i} = z\right) \\
&= \mathop{\mathbf{E}}_{z \leftarrow X_{<i}} \left[\mathbf{H}\left(X_i | X_{<i} = z\right)\right] = \mathbf{H}\left(X_i | X_{<i}\right).
\end{aligned}$$

By the chain rule for the entropy function,

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

where the last inequality holds as $-x\log(x) < 1$ for $x \in [0,1]$. □

7.4 Good Rectangles

For $i \in [c]$ and $z \in \{0,1\}^{V_{<i}}$, we say that (i,z) is *good* if the following two properties hold:

1. $\mathbf{I}\left(X^z\right), \mathbf{I}\left(Y^z\right) \leq 2\log(m)$.
2. $\mathbf{I}\left(X_i^z\right), \mathbf{I}\left(Y_i^z\right) \leq \frac{m^3}{c}$.

Let $\mathcal{G}$ be the set of all good pairs (i,z).

Lemma 12. *It holds that*

$$\Pr_{(i,z)\leftarrow\hat{\rho}}\left[\ (i,z)\in\mathcal{G}\ \right] \geq 1 - \frac{4}{m^2\cdot\rho(R)}.$$

Proof. We claim that each of the two requirements in the definition of a good pair (i,z) is violated with probability of at most $\frac{2}{m^2\cdot\rho(R)}$:

1. If $\mathbf{I}\left(X^z\right) > 2\log(m)$ or $\mathbf{I}\left(Y^z\right) > 2\log(m)$, then, by Equation (7), $\rho_{i,z}(R) = \frac{|R^z|}{|\Psi^z\times\Psi^z|} \leq 1/m^2$. Since for every $i\in[c]$ there are $|\{0,1\}^{V_{<i}}|$ possibilities for z, by Equation (8), the $\hat{\rho}$-measure of all such pairs is at most $\frac{1}{m^2\cdot\rho(R)}$.
2. By Lemma 11,
$$\mathop{\mathbf{E}}_{(i,z)\leftarrow\hat{\rho}}\left[\mathbf{I}\left(X_i^z\right)\right] \leq \frac{1}{c\cdot\rho(R)},$$
and similarly for Y_i^z. The claim follows by Markov's inequality.

□

7.5 Proof of Theorem 1

In this section we prove Theorem 1. Let $f : \mathrm{supp}(\mu) \to \{0,1\}$ be the bursting noise function, with parameter k, where μ is the distribution defined by Algorithm 1. Let $\epsilon = 2^{-2k}$ and $\delta = \epsilon/m$. Recall that $m = 2^{2^k}$. We will show that (f, μ) has the (ϵ, δ) relative discrepancy property.

For the rest of the lower bound proof, we assume that the rectangle R satisfies $\rho(R) \geq \delta$. We will show the first part of the relative discrepancy property. That is,

$$\mu\left(R \cap f^{-1}(0)\right) \geq \left(\tfrac{1}{2} - \epsilon\right) \cdot \rho(R).$$

The second part of the relative discrepancy property (for $f^{-1}(1)$) is proved in the same way.

Recall that $\mathcal{G}$ is the set of all good pairs (i, z), see Section 7.4. By Lemma 13 (stated and proved in Section 7.6),

$$\mu\left(R \cap f^{-1}(0)\right) = \frac{1}{2c} \sum_{\substack{i \in [c] \\ z \in \{0,1\}^{V_{<i}}}} \mu_i^0(R^z) \geq \frac{1}{2c} \sum_{(i,z) \in \mathcal{G}} \mu_i^0(R^z) \geq \frac{1}{2c}\left(1 - 2^{-2k}\right) \sum_{(i,z) \in \mathcal{G}} \rho_i(R^z).$$

By Equation (8),

$$\frac{1}{c} \sum_{(i,z) \in \mathcal{G}} \rho_i(R^z) = \rho(R) \Pr_{(i,z) \leftarrow \hat{\rho}}\left[\, (i,z) \in \mathcal{G} \,\right].$$

By Lemma 12 and since $\rho(R) \geq \delta$,

$$\mu\left(R \cap f^{-1}(0)\right) \geq \frac{1}{2}\left(1 - 2^{-2k}\right)\left(1 - \frac{4}{m^2 \cdot \rho(R)}\right) \rho(R) \geq \left(\tfrac{1}{2} - \epsilon\right) \rho(R).$$

By Proposition 3, we conclude that every randomized protocol for f with communication complexity at most 2^k, errs with probability at least $\frac{1}{2} - \epsilon - \delta \cdot 2^{2^k} \geq \frac{1}{2} - 2^{-k}$ when the inputs are distributed according to μ.

7.6 Applying the Graph Correlation Lemma

Lemma 13. *Let (i, z) be a good pair. It holds that*

$$\mu_i^0(R^z) \geq \left(1 - 2^{-2k}\right) \rho_i(R^z).$$

Similarly,

$$\mu_i^1(R^z) \geq \left(1 - 2^{-2k}\right) \rho_i(R^z).$$

Proof. We will prove the first part. The second part is proved in the same way. For simplicity of notation, we denote in this proof $R := R^z$, $X := X^z$ and $Y := Y^z$. Note that X and Y are independent random variables over the domain Ψ^z.

We first claim that the first part of the lemma is equivalent to the inequity

$$\Pr\left[(X,Y) \in \mathrm{supp}(\mu_i^0)\right] \geq \left(1 - 2^{-2k}\right) \rho_i\left(\mathrm{supp}(\mu_i^0)\right). \tag{9}$$

This is true as

$$\frac{|\mathrm{supp}(\mu_i^0)|}{|R|} \cdot \mu_i^0(R) = \frac{|\mathrm{supp}(\mu_i^0)|}{|R|} \cdot \frac{|\mathrm{supp}(\mu_i^0) \cap R|}{|\mathrm{supp}(\mu_i^0)|} = \Pr\left[(X,Y) \in \mathrm{supp}(\mu_i^0)\right],$$

while, since $\mathrm{supp}(\mu_i^0) \subseteq \mathrm{supp}(\rho_i)$ and $R \subseteq \mathrm{supp}(\rho_i)$, it holds that

$$\frac{|\mathrm{supp}(\mu_i^0)|}{|R|} \cdot \rho_i(R) = \frac{|\mathrm{supp}(\mu_i^0)|}{|R|} \cdot \frac{|R|}{|\mathrm{supp}(\rho_i)|} = \rho_i\left(\mathrm{supp}(\mu_i^0)\right).$$

The rest of the proof is devoted to proving Equation (9).

Let $G = (U \cup W, E)$ be the complete bipartite graph with sets of vertices U, W and set of edges E, defined as follows: Let $U = W = \{0,1\}^{V_i}$ be the set of all boolean assignments to the vertices in multi-layer i. Let $E = U \times W$.

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. That is, $T(u,w)$ is the set of all typical vertices in layer $(i+1)^*$ with respect to i, u, w. 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, $T(u,w)$ is of the same size T for every u, w. By the definition of the bursting noise function and by the Chernoff bound, 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.

Denote $I := \max\{\mathbf{I}(X), \mathbf{I}(Y), 1\}$, and note that $I \leq 2\log(m) \leq 2^{k+1}$ (by Property (1) of a good pair (i,z)).

Define the bad sets:

$$\mathcal{D}_1 = \left\{u \in U \ : \ \Pr_X[X_i = u] = 0 \ \text{ or } \ \mathbf{I}(X_{>i}|X_i = u) > 2^{4k}\right\},$$

$$\mathcal{D}_2 = \left\{w \in W \ : \ \Pr_Y[Y_i = w] = 0 \ \text{ or } \ \mathbf{I}(Y_{>i}|Y_i = w) > 2^{4k}\right\}.$$

By the chain rule for the entropy function,

$$\begin{aligned} I \geq \mathbf{I}(X) &= \log(|\Psi^z|) - \mathbf{H}(X) \\ &= \log(|U| \cdot |\{0,1\}^{V_{>i}}|) - \mathbf{H}(X_i, X_{>i}) \\ &= \log(|U|) + \log(|\{0,1\}^{V_{>i}}|) - \mathbf{H}(X_i) - \mathbf{H}(X_{>i}|X_i) \\ &= \mathbf{I}(X_i) + \log(|\{0,1\}^{V_{>i}}|) - \mathop{\mathbf{E}}_{u \leftarrow X_i}\left[\mathbf{H}(X_{>i}|X_i = u)\right] \\ &\geq \mathop{\mathbf{E}}_{u \leftarrow X_i}\left[\mathbf{I}(X_{>i}|X_i = u)\right]. \end{aligned}$$

By Markov's inequality,

$$\Pr_X [X_i \in \mathcal{D}_1] \leq \frac{I}{2^{4k}}. \tag{10}$$

By a similar argument,

$$\Pr_Y [Y_i \in \mathcal{D}_2] \leq \frac{I}{2^{4k}}. \tag{11}$$

For $u \notin \mathcal{D}_1$, we define the random variable X^u to be $(X_{>i}|X_i = u)$, that is, X^u has the distribution of $X_{>i}$ conditioned on the event $X_i = u$. For $u \in \mathcal{D}_1$, we define the random variable X^u to be uniformly distributed over $\{0,1\}^{V_{>i}}$. Similarly, for $w \notin \mathcal{D}_2$, we define the random variable Y^w to be $(Y_{>i}|Y_i = w)$, that is, Y^w has the distribution of $Y_{>i}$ conditioned on the event $Y_i = w$. For $w \in \mathcal{D}_2$, we define the random variable Y^w to be uniformly distributed over $\{0,1\}^{V_{>i}}$. Let Σ be the set of all possible boolean assignments to the vertices of a subtree of $\mathcal{T}$ rooted at layer $(i+1)^*$.

By Lemma 9 applied to the graph G, there exists a set $\mathcal{D} \subset E$ such that

$$\frac{|\mathcal{D}|}{|E|} \leq 2^{-4k}, \tag{12}$$

and for every $(u,w) \notin \mathcal{D}$ it holds that

$$\Pr_{X^u, Y^w} \left[X^u_{T(u,w)} = Y^w_{T(u,w)}\right] \geq \left(1 - 2^{-4k}\right) |\Sigma|^{-T}. \tag{13}$$

It holds that

$$\begin{aligned}
&\Pr_{X,Y} \left[(X,Y) \in \mathrm{supp}(\mu_i^0)\right] \\
&= \sum_{(u,w)\in E} \Pr_X [X_i = u] \cdot \Pr_Y [Y_i = w] \cdot \Pr_{X,Y} \left[(X,Y) \in \mathrm{supp}(\mu_i^0) \mid X_i = u, Y_i = w\right] \\
&\geq \sum_{\substack{u \in U \setminus \mathcal{D}_1, \\ w \in W \setminus \mathcal{D}_2, \\ (u,w) \notin \mathcal{D}}} \Pr_X [X_i = u] \cdot \Pr_Y [Y_i = w] \cdot \Pr_{X,Y} \left[(X,Y) \in \mathrm{supp}(\mu_i^0) \mid X_i = u, Y_i = w\right].
\end{aligned}$$

By the definition of the bursting noise function (when the noisy multi-layer is i), for every u, w, the following holds: Conditioned on $X_i = u$ and $Y_i = w$, we have $(X,Y) \in \mathrm{supp}(\mu_i^0)$ if and only if $X_{>i}$ and $Y_{>i}$ agree on the subtrees rooted at vertices in $T(u,w)$ (these are the non-noisy subtrees). Therefore, using Equation (13) and the fact that E contains all

pairs (u, w),

$$\begin{aligned}
&\Pr_{X,Y}\left[(X,Y) \in \mathrm{supp}(\mu_i^0)\right] \\
&\geq \sum_{\substack{u \in U \setminus \mathcal{D}_1, \\ w \in W \setminus \mathcal{D}_2, \\ (u,w) \notin \mathcal{D}}} \Pr_X\left[X_i = u\right] \cdot \Pr_Y\left[Y_i = w\right] \cdot \Pr_{X^u, Y^w}\left[X^u_{T(u,w)} = Y^w_{T(u,w)}\right] \\
&\geq \left(1 - 2^{-4k}\right) |\Sigma|^{-T} \sum_{\substack{u \in U \setminus \mathcal{D}_1, \\ w \in W \setminus \mathcal{D}_2, \\ (u,w) \notin \mathcal{D}}} \Pr_X\left[X_i = u\right] \cdot \Pr_Y\left[Y_i = w\right].
\end{aligned}$$

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

$$\sum_{(u,w) \in U \times W} \Pr_X\left[X_i = u\right] \cdot \Pr_Y\left[Y_i = w\right] = 1.$$

By Equation (10),

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

and by Equation (11),

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

By Lemma 14 (stated and proved below) and Equation (12),

$$\sum_{(u,w) \in \mathcal{D}} \Pr_X\left[X_i = u\right] \cdot \Pr_Y\left[Y_i = w\right] \leq 2^{-3k}.$$

Therefore,

$$\begin{aligned}
\Pr_{X,Y}\left[(X,Y) \in \mathrm{supp}(\mu_i^0)\right] &\geq \left(1 - 2^{-4k}\right) |\Sigma|^{-T} \left(1 - \frac{I}{2^{4k}} - \frac{I}{2^{4k}} - 2^{-3k}\right) \\
&\geq |\Sigma|^{-T} \left(1 - 2^{-2k}\right).
\end{aligned}$$

Finally, note that for every $x, y \in \Psi^z$, such that $x_i = u$ and $y_i = w$, the following holds: $(x, y) \in \mathrm{supp}(\mu_i^0)$ if and only if x and y agree on the subtrees rooted at vertices in $T(u, w)$ (these are the non-noisy subtrees). Therefore,

$$|\Sigma|^{-T} = \Pr_{(x,y) \in_R \Psi^z \times \Psi^z}\left[(x,y) \in \mathrm{supp}(\mu_i^0)\right] = \rho_i\left(\mathrm{supp}(\mu_i^0)\right),$$

and the assertion follows. □

Lemma 14. *Let (i, z) be a good pair. Let $U = W = \{0,1\}^{V_i}$. Let $\mathcal{D} \subseteq U \times W$ be such that*

$\frac{|\mathcal{D}|}{|U|\cdot|W|} \leq 2^{-4k}$. *It holds that*

$$C^z := \sum_{(u,w)\in\mathcal{D}} \Pr\left[X_i^z = u\right] \cdot \Pr\left[Y_i^z = w\right] \leq 2^{-3k}.$$

Proof. For simplicity of notation, we denote in this proof $X := X^z$ and $Y := Y^z$. By Property (1) of a good pair (i,z), we have $\mathbf{I}(X), \mathbf{I}(Y) \leq 2\log(m)$, which means that $|R^z| \geq \frac{1}{m^4}|\Psi^z \times \Psi^z|$. This implies that for every set $L \subseteq \Psi^z$, the probability that X is in L is at most m^4 times the probability that a uniformly distributed variable over Ψ^z obtains a value in L. In particular, for every $u \in U$,

$$\Pr\left[X_i = u\right] \leq \frac{m^4}{|U|}. \tag{14}$$

Similarly, for $w \in W$,

$$\Pr\left[Y_i = w\right] \leq \frac{m^4}{|W|}. \tag{15}$$

Define

$$U' = \left\{u \in U \;:\; \Pr\left[X_i = u\right] \geq \frac{2}{|U|}\right\},$$

$$W' = \left\{w \in W \;:\; \Pr\left[Y_i = w\right] \geq \frac{2}{|W|}\right\}.$$

By Property (2) of a good pair (i,z), we have

$$\mathbf{I}(X_i) \leq \frac{m^3}{c},$$

$$\mathbf{I}(Y_i) \leq \frac{m^3}{c}.$$

Using Lemma 5.12 in [KR13] (stated for convenience at the end of the section, Lemma 15) it holds that

$$\Pr\left[X_i \in U'\right] < 5 \cdot \left(\frac{m^3}{c}\right)^{0.1}, \tag{16}$$

Similarly,

$$\Pr\left[Y_i \in W'\right] < 5 \cdot \left(\frac{m^3}{c}\right)^{0.1}. \tag{17}$$

The expression C^z is a sum over pairs $(u,w) \in \mathcal{D}$. We bound C^z by a sum of three partial sums, and work on each partial sum separately. The first partial sum is over pairs $(u,w) \in U \times W$ with $u \in U'$, the second is over pairs $(u,w) \in U \times W$ with $w \in W'$, the third is over pairs $(u,w) \in \mathcal{D}$ with $u \notin U'$ and $w \notin W'$.

We bound the first partial sum as follows. We use Equation (15) for the first step, and

Equation (16) for the third.

$$\begin{aligned}
&\sum_{\substack{(u,w)\in U\times W\\ u\in U'}} \Pr[X_i = u]\cdot\Pr[Y_i = w]\\
&\leq \frac{m^4}{|W|}\sum_{\substack{(u,w)\in U\times W\\ u\in U'}} \Pr[X_i = u]\\
&= m^4\cdot\sum_{u\in U'}\Pr[X_i = u]\\
&\leq 5m^4\cdot\left(\frac{m^3}{c}\right)^{0.1} \leq c^{-0.05}.
\end{aligned}$$

The second partial sum is bounded in a similar way. We bound the third partial sum using the bound that we have on the size of $\mathcal{D}$,

$$\sum_{\substack{(u,w)\in\mathcal{D}\\ u\notin U',w\notin W'}} \Pr[X_i = u]\cdot\Pr[Y_i = w] \leq |\mathcal{D}|\cdot\frac{2}{|U|}\cdot\frac{2}{|W|} \leq 2^{-4k+2}.$$

We conclude that $C^z \leq 2^{-3k}$. □

Lemma 15 (**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,*

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

8 Bounding Information Cost by Tree Divergence Cost

In this section we give a general method that can be used to upper bound the information cost of any protocol π. Our method uses the notion of *divergence cost of a tree*, a notion that was implicit in [BBCR10] and was formally defined in [BR11].

Let π be a communication protocol between two players. We assume that the first player has the private input x and the second player has the private input y, where (x,y) were chosen according to some joint distribution μ. In this section, we assume without loss of generality that π does not use public randomness (but may use private randomness), as for the purpose of upper bounding the information cost, the public randomness can always be replaced by private randomness. We also assume, without loss of generality, that the players take alternating turns sending bits to each other. That is, in odd rounds, the first player sends a bit to the second player, and in even rounds the second player sends a bit to the first player (if this is not the case, we can add dummy rounds that do not change the information cost).

We denote by $\mathcal{T}_\pi$ the binary tree associated with the communication protocol π. That is, every vertex v of $\mathcal{T}_\pi$ corresponds to a possible transcript of π, and the two edges going

out of v are labeled by 0 and 1, corresponding to the next bit to be transmitted. We think of the first player as owning the vertices in odd layers of $\mathcal{T}_\pi$ (where the root is in layer 1), and of the second player as owning the vertices in even layers of $\mathcal{T}_\pi$. When the protocol π reaches a non-leaf vertex v, the player who owns v sends a bit to the other player.

Every input pair (x, y) for the protocol π induces a distribution $P_v = (p_v, 1-p_v)$ for every non-leaf vertex v of the tree $\mathcal{T}_\pi$, where p_v is the probability that the next bit transmitted by the protocol π on the vertex v and inputs x, y is 0. We think of P_v as a distribution over the two children of the vertex v. Observe that the player who owns v knows P_v. Given the binary tree $\mathcal{T}_\pi$ and the distributions P_v for every non-leaf vertex v of $\mathcal{T}_\pi$, where for each v the player who owns v knows P_v, we can assume without loss of generality that the protocol π operates as follows: 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.

Assume that for every non-leaf vertex v of $\mathcal{T}_\pi$, we have an additional distribution $Q_v = (q_v, 1-q_v)$ that is known to the player who doesn't own v. We think of every P_v as the "correct" distribution over the two children of v. This distribution is known to the player who owns v. We think of Q_v as an estimation of P_v, based on the knowledge of the player who doesn't own v. For the rest of the section, we think of $\mathcal{T}_\pi$ as the tree $\mathcal{T}_\pi$ together with the distributions P_v and Q_v, for every non-leaf vertex v in the tree $\mathcal{T}_\pi$.

To upper bound the information cost of a protocol π it is convenient to use the notion of divergence cost of a tree [BBCR10, BR11].

Definition 7 (**Divergence Cost** [BBCR10, BR11])**.** *Consider 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 non-leaf vertex v in the tree. We think of P_v and Q_v as distributions over the two children of the vertex v. We define the* divergence cost *of the tree $\mathcal{T}$ recursively, as follows. $\mathbf{D}(\mathcal{T}) = 0$ if the tree has depth 0, otherwise,*

$$\mathbf{D}(\mathcal{T}) = \mathbf{D}(P_r \| Q_r) + \mathop{\mathbf{E}}_{v \sim P_r}[\mathbf{D}(\mathcal{T}_v)], \tag{18}$$

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

An equivalent definition of the divergence cost of $\mathcal{T}$ is obtained by following the recursion in Equation (18) and is given by the following equation:

$$\mathbf{D}(\mathcal{T}) = \sum_{v \in V} \tilde{p}_v \cdot \mathbf{D}(P_v \| Q_v), \tag{19}$$

where V is the vertex set of $\mathcal{T}$, and for a vertex $v \in V$, $\tilde{p}_v$ is the probability to reach v by following the distributions P_v, starting from the root. Formally, if v is the root of the tree $\mathcal{T}$, then $\tilde{p}_v = 1$, otherwise,

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

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 of $\mathcal{T}_\pi$, 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 (if π ends before layer d, then Π_d is undefined).

Let the tree $\mathcal{T}'_\pi$ be the same as $\mathcal{T}_\pi$, except that every distribution Q_v, for every non-leaf vertex v in $\mathcal{T}_\pi$, 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 $\mathcal{T}'_\pi$ is at most the divergence cost of $\mathcal{T}_\pi$, in expectation. This is formulated in the following lemma.

Observe that the protocol π induces the distributions P_v (known to the player who owns v) and Q'_v (known to the player who doesn't own v), while the distribution Q_v may be any distribution known to the player who doesn't own v.

Lemma 16. *For every protocol π and distributions Q_v known to the player who doesn't own v, as above, it holds that*

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

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

Proof. By Equation (19),

$$\mathop{\mathbf{E}}_{X,Y,R_1,R_2}[\mathbf{D}(\mathcal{T}_\pi) - \mathbf{D}(\mathcal{T}'_\pi)] = \mathop{\mathbf{E}}_{X,Y,R_1,R_2}\left[\sum_v \tilde{p}_v \left(\mathbf{D}(P_v \| Q_v) - \mathbf{D}(P_v \| Q'_v)\right)\right],$$

where $\tilde{p}_v$ is as in Definition 7. 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,

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

(Recall that V is undefined when the protocol ends before layer d. In that case, for simplicity, we think of P_V, Q_V and Q'_V as all being equal, and hence $\mathbf{D}(P_V\|Q_V)=\mathbf{D}(P_V\|Q'_V)=0$). By the definition of relative entropy,

$$\begin{aligned}
&\mathop{\mathbf{E}}_{A,B,V}\left[\mathbf{D}(P_V\|Q_V)-\mathbf{D}(P_V\|Q'_V)\right]\\
&=\mathop{\mathbf{E}}_{A,B,V}\left[p_V\left(\log\left(\frac{p_V}{q_V}\right)-\log\left(\frac{p_V}{q'_V}\right)\right)+(1-p_V)\left(\log\left(\frac{1-p_V}{1-q_V}\right)-\log\left(\frac{1-p_V}{1-q'_V}\right)\right)\right]\\
&=\mathop{\mathbf{E}}_{A,B,V}\left[p_V\log\left(\frac{q'_V}{q_V}\right)+(1-p_V)\log\left(\frac{1-q'_V}{1-q_V}\right)\right]. \qquad (20)
\end{aligned}$$

Assume that the first player owns the vertices in layer d. The case that the second player owns the vertices in layer d is analogous. Consider the first summand in Equation (20). 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 $q'_V=\mathbf{E}_A\left[p_V|B,V\right]$. 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 (20) 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,

$$\mathop{\mathbf{E}}_{A,B,V}\left[\mathbf{D}(P_V\|Q_V)-\mathbf{D}(P_V\|Q'_V)\right]=\mathop{\mathbf{E}}_{B,V}\left[\mathbf{D}(Q'_V\|Q_V)\right]\geq 0,$$

since the divergence is non-negative. This is true for every layer d in the tree. Therefore, summing over all layers, we get that

$$\mathop{\mathbf{E}}_{A,B}[\mathbf{D}(\mathcal{T}'_\pi)]\leq\mathop{\mathbf{E}}_{A,B}[\mathbf{D}(\mathcal{T}_\pi)].$$

□

The following lemma relates the information cost of π to the expected divergence cost of $\mathcal{T}_\pi$. It was shown in [BR11] (see Lemma 5.3 therein) that $IC_\mu(\pi) = \mathbf{E}[\mathbf{D}(\mathcal{T}'_\pi)]$. Together with Lemma 16 we get:

Lemma 17. *For every protocol π and distributions Q_v known to the player who doesn't own v, as above, it holds that*

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

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

9 Information Upper Bound

In this section we prove Theorem 2. Let $(x, y) \in \text{supp}(\mu)$ be an input pair to the bursting noise function. Consider the following protocol π' for the bursting noise function. 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 v they output $x_v \oplus y_v$. By the Chernoff bound, the probability that the players reach a leaf that is not typical with respect to the noisy multi-layer is at most $2^{-\Omega(w)}$. That is, the error probability of π' is exponentially small in k.

The information cost of the protocol π' is too large. The reason is that if the protocol reaches a non-typical vertex at the end of the noisy multi-layer (with respect to the noisy multi-layer), an event that occurs with probability exponentially small in k, then the rest of the protocol reveals to each player $\Omega\left((c-i)w\right)$ bits of information about the input of the other player, in expectation (as all the vertices below a non-typical vertex are noisy), and note that $\Omega\left((c-i)w\right)$ is double exponentially large (for almost all i). Thus, in expectation, the information revealed to each player about the input of the other player is double exponential in k.

For that reason, we consider a variant of the protocol π', called π. Informally speaking, the protocol π operates like π' but aborts if too much information about the inputs is revealed. Recall that in every round of the protocol π', the players are at a vertex v of $\mathcal{T}$ and the player who owns v sends a bit b_v indicating one of v's children. In the new protocol π, after receiving that bit, the receiving party sends a bit a_v indicating whether they should abort the protocol, where $a_v = 1$ stands for abort and $a_v = 0$ stands for continue. If a bit $a_v = 1$, indicating an abort, was sent, the protocol terminates and both players output an arbitrary bit, say 0. It remains to specify how the receiving party, without loss of generality the second player, decides whether to abort or continue, that is, how she determines the value of a_v.

To determine whether to abort, the second player considers the last $\ell = 2^{100k}$ vertices $v_1, \ldots, v_\ell$, reached by the protocol and owned by the first player, and the corresponding bits $b_{v_1}, \ldots, b_{v_\ell}$ that were sent by the first player (if less than ℓ bits were sent by the first player so far, then the second player does not abort). For every $j \in [\ell]$, the second player compares b_{v_j} and y_{v_j}. The second player decides to abort and sends $a_v = 1$ if and only if less than 0.8ℓ of these pairs are equal (otherwise the second player sends $a_v = 0$).

The following claim shows that the probability that π aborts is exponentially small in k. If π does not abort, it gives the same output as π'. We conclude that the error probability of π is exponentially small in k.

Claim 18. *Let $(x, y) \in \mathrm{supp}(\mu)$ be an input pair to the bursting noise function. The protocol π aborts with probability at most 2^{-10k} on the input (x, y).*

Proof. Fix $(x, y) \in \mathrm{supp}(\mu_i)$ for some $i \in [c]$. Let E be the event that the protocol π reaches a non-typical vertex after multi-layer i (with respect to multi-layer i). By the Chernoff bound, the event E occurs with probability at most 2^{-100k}, as $w = 2^{100}k$. Let A be the event that the protocol π aborts. Assume that E does not occur. By the Chernoff bound, the probability of aborting after each round is at most $2^{-2^{50k}}$, as $\ell = 2^{100k}$ and since if E does not occur then x_v and y_v can only differ for at most w vertices reached by the protocol π. By the union bound, the probability of abort (conditioned on $\neg E$) is at most $cw \cdot 2^{-2^{50k}} < 2^{-100k}$. Therefore,

$$\Pr[A] \leq \Pr[E] + \Pr[A|\neg E] \leq 2 \cdot 2^{-100k}.$$

□

We next upper bound the information cost of the protocol π. Observe that after π reaches a leaf v of the tree $\mathcal{T}$, two additional bits, x_v and y_v are exchanged, which adds at most 2 to the information cost of π. Therefore, for the rest of the section, we ignore the exchange of these last two bits, and think of π as terminating after reaching a leaf of $\mathcal{T}$.

To upper bound the information cost of the protocol π we will use Lemma 17. We denote by $\mathcal{T}_\pi$ the binary tree associated with the communication protocol π, as in Section 8. That is, every vertex v of $\mathcal{T}_\pi$ corresponds to a possible transcript of π, and the two edges going out of v are labeled by 0 and 1, corresponding to the next bit to be transmitted. The non-leaf vertices of the tree $\mathcal{T}_\pi$ have the following structure: Every non-leaf vertex v in an odd layer of $\mathcal{T}_\pi$ corresponds to a non-leaf vertex of $\mathcal{T}$, the binary tree on which the bursting noise game is played. Since the correspondence is one-to-one, we refer to the vertex in $\mathcal{T}$ corresponding to v also as v. The next bit to be transmitted by π on the vertex v is b_v. For a non-leaf vertex v in an even layer of $\mathcal{T}_\pi$, the next bit to be transmitted by π on the vertex v is a_v.

As explained in Section 8, every input pair $(x, y) \in \mathrm{supp}(\mu)$ to the bursting noise function, induces a distribution $P_v = (p_v, 1 - p_v)$ for every non-leaf vertex v of the tree $\mathcal{T}_\pi$, where p_v is the probability that the next bit transmitted by the protocol π on the vertex v and inputs x, y is 0. Namely, if v is in an odd layer of $\mathcal{T}_\pi$ (and recall that in this case we think of v as both a vertex of $\mathcal{T}_\pi$ and of $\mathcal{T}$), the distribution P_v is the following: In the case that the

first player owns v in $\mathcal{T}$, if $x_v = 0$ then $P_v = (0.9, 0.1)$, and if $x_v = 1$ then $P_v = (0.1, 0.9)$. In the case that the second player owns v, if $y_v = 0$ then $P_v = (0.9, 0.1)$, and if $y_v = 1$ then $P_v = (0.1, 0.9)$. If v is in an even layer of $\mathcal{T}_\pi$ then P_v is $P_v = (0, 1)$ if the player sending a_v decides to abort, and $P_v = (1, 0)$ if she decides to continue (note that given x, y, v, this decision is deterministic).

For every non-leaf vertex v of $\mathcal{T}_\pi$, we define an additional distribution $Q_v = (q_v, 1 - q_v)$ (depending on the input (x, y)). We think of every P_v as the "correct" distribution over the two children of v. This distribution is known to the player who sends the next bit on the vertex v. We think of Q_v as an estimation of P_v, based on the knowledge of the player who doesn't send the next bit. For a vertex v in an odd layer of $\mathcal{T}_\pi$ (and recall that in this case we think of v as both a vertex of $\mathcal{T}_\pi$ and of $\mathcal{T}$), the distribution Q_v is the following: In the case that the first player owns v in $\mathcal{T}$, if $y_v = 0$ then $Q_v = (0.9, 0.1)$, and if $y_v = 1$ then $Q_v = (0.1, 0.9)$. In the case that the second player owns v, if $x_v = 0$ then $Q_v = (0.9, 0.1)$, and if $x_v = 1$ then $Q_v = (0.1, 0.9)$. If v is in an even layer of $\mathcal{T}_\pi$ then $Q_v = (1 - \frac{1}{cw}, \frac{1}{cw})$.

For the rest of the section, we think of $\mathcal{T}_\pi$ as the tree $\mathcal{T}_\pi$ together with the distributions P_v and Q_v, for every vertex v in the tree $\mathcal{T}_\pi$.

Proposition 19. *It holds that*

$$\mathbf{D}(\mathcal{T}_\pi) = O(k).$$

Proof. Fix $(x, y) \in \mathrm{supp}(\mu_i)$ for some $i \in [c]$. By Equation (19),

$$\mathbf{D}(\mathcal{T}_\pi) = \sum_v \tilde{p}_v \cdot \mathbf{D}(P_v \| Q_v),$$

where $\tilde{p}_v$ is the probability that the protocol π reaches the vertex v on input (x, y). We will bound the last sum separately for vertices v in odd layers and for vertices v in even layers.

We first sum over vertices in even layers. For every vertex v in an even layer of $\mathcal{T}_\pi$, if $P_v = (0, 1)$ (protocol aborts) we have $\mathbf{D}(P_v \| Q_v) = \log(cw)$, and if $P_v = (1, 0)$ (protocol continues) we have $\mathbf{D}(P_v \| Q_v) = \log\left(\frac{1}{1-1/cw}\right) = \log\left(1 + \frac{1}{cw-1}\right) < \frac{2}{cw}$. By Claim 18, the probability that π aborts is at most 2^{-10k}. Therefore, the sum in Equation (19) taken over vertices in even layers is at most $cw \cdot \frac{2}{cw} + 2^{-10k} \cdot \log(cw) \leq 3$, as for each of the cw even layers, the probability of reaching a vertex in this layer is at most 1.

We next sum over vertices in odd layers. Recall that each such vertex corresponds to a vertex in $\mathcal{T}$. Let v be a vertex in an odd layer of $\mathcal{T}_\pi$. If v corresponds to a non-noisy vertex in $\mathcal{T}$ we have $\mathbf{D}(P_v \| Q_v) = 0$, and if v corresponds to a noisy vertex in $\mathcal{T}$ we have $\mathbf{D}(P_v \| Q_v) \leq 4$. Recall that i is the noisy multi-layer. Then,

1. The vertices above multi-layer i in $\mathcal{T}$ add nothing to the divergence cost.
2. Multi-layer i of $\mathcal{T}$ adds $O(w)$ to the divergence cost.
3. If $i < c$: Let v be the vertex in layer $i^* + w$ of $\mathcal{T}$ that the players reach during the execution of the protocol π. If v is a typical vertex with respect to multi-layer i, the

vertices below v add nothing to the divergence cost. If v is a non-typical vertex, the protocol aborts after at most 4ℓ rounds in expectation. Since the probability that v is a non-typical vertex with respect to multi-layer i is at most 2^{-1000k} (as $w = 2^{100}k$), the expected divergence cost added by this case is at most $2^{-1000k} \cdot 4\ell \cdot 4 \leq 1$.

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

By Proposition 19 and Lemma 17 we get that $IC_\mu(\pi) \leq O(k)$.

Acknowledgements

We thank Andy Drucker and Mark Braverman for very helpful conversations.

References

[BBCR10] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. In *STOC*, pages 67–76, 2010. 2, 27, 28

[BR11] Mark Braverman and Anup Rao. Information equals amortized communication. In *FOCS*, pages 748–757, 2011. 1, 3, 4, 27, 28, 31

[Bra12] Mark Braverman. Interactive information complexity. In *STOC*, pages 505–524, 2012. 1, 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. 3

[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. 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

[Fan49] Robert M Fano. *The transmission of information*. Massachusetts Institute of Technology, Research Laboratory of Electronics, 1949. 2

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

[GKR14] Anat Ganor, Gillat Kol, and Ran Raz. Exponential separation of information and communication. In *FOCS*, 2014. 1, 3, 4

[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

[Huf52] David A Huffman. A method for the construction of minimum redundancy codes. *proc. IRE*, 40(9):1098–1101, 1952. 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. 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. 3

[KN97] Eyal Kushilevitz and Noam Nisan. Communication complexity. *Cambridge University Press*, 1997. 2

[KR13] Gillat Kol and Ran Raz. Interactive channel capacity. In *STOC*, pages 715–724, 2013. 26, 27

[KRW95] Mauricio Karchmer, Ran Raz, and Avi Wigderson. Super-logarithmic depth lower bounds via the direct sum in communication complexity. *Computational Complexity*, 5(3/4):191–204, 1995. 2

[LS09] Troy Lee and Adi Shraibman. Lower bounds in communication complexity. *Foundations and Trends in Theoretical Computer Science*, 3(4):263–398, 2009. 2

[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. 15

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

[Sha48] C. E. Shannon. A mathematical theory of communication. *The Bell Systems Technical Journal*, 27:July 379–423, October 623–656, 1948. 2

[Sha03] Ronen Shaltiel. Towards proving strong direct product theorems. *Computational Complexity*, 12(1-2):1–22, 2003. 2

ECCC ISSN 1433-8092

http://eccc.hpi-web.de