Electronic Colloquium on Computational Complexity, Report No. 168 (2015)

Interactive Compression for Product Distributions

Gillat Kol*

Abstract

We study the interactive compression problem: Given a two-party communication protocol with small information cost, can it be compressed so that the total number of bits communicated is also small? We consider the case where the parties have inputs that are independent of each other, and give a simulation protocol that communicates $I^2 \cdot \text{polylog}(I)$ bits, where I is the information cost of the original protocol. Our protocol is the first simulation protocol whose communication complexity is bounded by a polynomial in the information cost of the original protocol.

1 Introduction

In seminal works, Shannon, Fano and Huffman consider the *data compression problem*: Alice wants to send a message x to Bob. How many bits does she need to send, so that Bob will be able to retrieve x with high probability? The answer given in [Sha48, Fan49, Huf52] is that she needs to send only $\lceil \mathbb{H}(x) \rceil$ bits, in expectation, where $\mathbb{H}$ denotes Shannon's entropy function. Roughly speaking, this means that every message can be compressed to its information content.

Over the last decades, interactive communication protocols were studied extensively. The *interactive compression problem* [BBCR10] is the analog of the data compression problem in the interactive setting. It asks whether the transcript of any interactive protocol can be compressed to its information content.

The interactive compression problem is formalized in the setting of communication complexity. In the two-party distributional communication complexity model, each player gets an input, where the inputs are sampled from a joint distribution μ that is known to both players. The players' engage in an interactive communication protocol in order to perform some communication task that depends on both inputs. They may use both common and private random strings. The players communicate in rounds, where in each round one of the players sends a message to the other player. The communication complexity of the protocol is the total number of bits communicated by the two players.

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

ISSN 1433-8092

The notion of information cost, introduced by [CSWY01, BYJKS04, BBCR10], measures the amount of information that the players need to reveal about their inputs in order to solve a communication task. This notion was motivated by the fundamental information theoretic question of compressing communication, as well as by fascinating relations to communication complexity, and in particular to the direct sum problem in communication complexity. Roughly speaking, the *(internal) information cost* of a protocol π over the distribution μ, denoted $\mathsf{IC}_\mu(\pi)$, is the number of information bits that the players learn about each other's input, when running the protocol π with inputs sampled from μ.

Using the notions of communication complexity and information cost, the interactive compression problem is stated as follows: Given a communication protocol π, and a distribution μ over the inputs for π, does there exist a protocol π' that simulates π over μ and has communication complexity close to $\mathsf{IC}_\mu(\pi)$ (say, polynomial in $\mathsf{IC}_\mu(\pi)$)? By "simulates" we mean that π' performs the same task as π, except with some small error probability, where the probability is over μ and over the randomness used by the players.

1.1 Previous Works

The interactive compression problem received a lot of attention over the last years, and clever compression protocols were suggested for several cases [HJMR10, BBCR10, BR11, Bra12, BBK+13, BMY15, RR15]. Most relevant to our work is a beautiful result by Barak, Braverman, Chen and Rao, showing that over a product distribution μ, any protocol with information cost I and communication complexity C, can be compressed to a protocol with communication complexity $I \cdot \mathrm{polylog}(C)$ [BBCR10]. (We mention that their result is even stronger and gives an $I' \cdot \mathrm{polylog}(C)$ compression for any protocol over any distribution, where I' is the *external* information cost). In [BMY15], a simulation protocol communicating $O(I^2 \cdot \log\log(C))$ bits is shown for the case where the original protocol does not use private randomness. The general case, compressing any protocol over any distribution, is considered in [Bra12], where a $2^{O(I)}$ compression is given, and in [BBCR10], where a $\tilde{O}(\sqrt{C \cdot I})$ compression is given. A compression to $\tilde{O}(I + r)$ is given by [BR11], where r is the number of rounds of the original protocol.

Another line of works studies the limitations of compression protocols. In [GKR14, GKR15b, RS15, GKR15a], exponential separations between communication complexity and information cost are shown. More precisely, for any sufficiently large I, the above results give examples of communication tasks with communication complexity at least 2^I, for which there exist protocols with information cost $O(I)$. Therefore, over general distributions μ, communication protocols cannot always be compressed to their information cost. We mention that in all the above results, the protocol with low information cost has communication complexity C, that is at least double exponential in I. Thus, the $I \cdot \mathrm{polylog}(C)$ compression by [BBCR10], leaves open the possibility that a similar separation can also be obtained with the underlying distribution being a product distribution.

1.2 Our Result

We consider the interactive compression problem over product distributions $\mu = \mu^A \times \mu^B$ (i.e., Alice's input and Bob's input are chosen independently), and give a simulation protocol that communicates $I^2 \cdot \text{polylog}(I)$ bits, where I is the information cost of the original protocol over μ. Our protocol is the first simulation protocol whose communication complexity is polynomial in the information cost of the original protocol, and does not depend on the communication complexity of the original protocol. We mention that our result implies that an exponential separation between information cost and communication complexity, like the ones given by [GKR14, GKR15b, RS15, GKR15a], cannot be obtained for product distributions. Our result is incomparable to the $I \cdot \text{polylog}(C)$ compression protocol of [BBCR10].

Our main result is given in Theorem 1 below. The theorem uses the following notation: Let ξ be a randomized communication protocol between two parties, Alice and Bob. Let x be an input for Alice, and y be an input for Bob. We denote by $\xi(x,y)$ the random variable representing ξ's transcript (that is, the concatenation of all the messages exchanged between the players during the execution of ξ), when it is run with the input (x,y). We also view $\xi(x,y)$ as a distribution over the set of all binary strings. For a pair of distributions D, D', we denote by $\|D - D'\|$ the statistical distance between D and D'.

Theorem 1. *Let $\varepsilon > 0$. Let π be a randomized protocol that may use private and public coins. Let $\mu = \mu^A \times \mu^B$ be a product distribution over the inputs for π. Then, there exist a public coin protocol τ (that takes the same inputs as π), and a pair of "transcript reconstruction" functions g^A, g^B such that: The function g^A takes as inputs $x \in \text{supp}(\mu^A)$ and a possible transcript of τ, and returns a possible transcript of π. The function g^B takes as inputs $y \in \text{supp}(\mu^B)$ and a possible transcript of τ, and returns a possible transcript of π. In addition, the followings hold:*

1. *The worst case communication complexity of τ is $\mathsf{IC}^2_\mu(\pi) \cdot \text{polylog}\left(\mathsf{IC}_\mu(\pi)\right)/\varepsilon^5$.*
2. *$\forall (x,y) \in \text{supp}(\mu)$: $\Pr\left[g^A(x,\tau(x,y)) \neq g^B(y,\tau(x,y))\right] \leq \varepsilon$, where the probability is over the random coins of the protocol τ.*
3. *$\mathbf{E}_{(x,y)\leftarrow\mu}\left[\left\|g^A(x,\tau(x,y)) - \pi(x,y)\right\|\right] \leq \varepsilon$.*

2 Proof Sketch

Let π be a communication protocol between two players, Alice and Bob. Alice has a private input x and Bob has a private input y, where (x,y) is chosen according to some publicly known joint distribution μ. For the rest of this sketch, $\mu = \mu^A \times \mu^B$ is a product distribution. We assume, without loss of generality, that π does not use public randomness (but may use private randomness), as the public randomness can always be replaced by private randomness

without increasing the information cost. We also assume that the players take turns sending bits to one another. Alice sends bits in odd rounds and Bob in the even rounds.

The *external information cost* of a protocol over a distribution μ is the number of information bits that an external observer, who watches the execution of the protocol, learns about the players' inputs, when the inputs are sampled from μ. We next present a sketch of the public coin protocol τ that simulates π, such that the communication complexity of τ is $I^2 \cdot \text{polylog}(I)$, where I is the external information of π over μ. This proves Theorem 1, as over a product distribution μ, the internal and external information costs of a protocol π are the same (see Fact 3). Intuitively, this is true since x conveys no information about y, thus Alice and an external observer who doesn't know x, y have the same information about y at any point in the execution of the protocol. The same is true for Bob. Our protocol builds over the $I \cdot \text{polylog}(C)$ compression by [BBCR10], and parts of this sketch follow their description.

2.1 Divergence Tree

Consider the (directed) binary tree associated with π, where each vertex v of the binary tree corresponds to a possible transcript of π. The two edges going out of v are labeled by 0 and 1, corresponding to the next bit to be transmitted. We think of Alice as owning the non-leaf vertices in the odd layers, and of Bob as owning the non-leaf vertices in the even layers. The protocol π proceeds as follows: Starting from the root, when π reaches a non-leaf vertex v, the player who owns v sends a bit to the other player. The players follow the edge indicated by the sent bit and reach a new vertex.

Let $b \in \{0,1\}$. For every non-leaf vertex w, we denote by $O_w(b)$ the probability of transmitting the bit b at w, conditioned on reaching that vertex (without taking into consideration the actual values of the inputs x, y). We denote by $O_{wx}(b)$, $O_{wy}(b)$ the probabilities of transmitting the bit b at w, conditioned on a particular fixing of x or y (respectively), and conditioned on the event of reaching w during the run of the protocol. We denote by $O_{wxy}(b)$ the probability of transmitting the bit b at w, conditioned on a particular fixing of both x, y, and conditioned on the event of reaching w. We view $O_{wxy} = (O_{wxy}(0), O_{wxy}(1))$ as the "true" probability distribution at w. Observe that Alice can compute the distributions O_{wxy} and O_{wy} for vertices w that she owns: Since Alice is the one deciding on the bit to be sent at w, it holds that $O_{wxy} = O_{wx}$. Since μ is a product distribution, it also holds that $O_{wy} = O_w$. Similarly, Bob can compute O_{wxy} and O_{wx} for vertices w that he owns.

We define the divergence at w with respect to x, denoted by $\mathbb{D}_{wx}$, as $\mathbb{D}(O_{wx} \| O_w)$, where $\mathbb{D}\left((p, 1-p) \parallel (q, 1-q)\right) = p \log(p/q) + (1-p) \log((1-p)/(1-q))$ is the *divergence* (also known as *relative entropy* or the *Kullback-Leibler distance*) between the distributions $(p, 1-p)$ and $(q, 1-q)$. Observe that for a vertex w owned by Bob, $\mathbb{D}_{wx} = \mathbb{D}(O_{wx} \| O_w) = \mathbb{D}(O_w \| O_w) = 0$. Similarly, the divergence at w with respect to y, denoted by $\mathbb{D}_{wy}$, is define as $\mathbb{D}(O_{wy} \| O_w)$.

2.2 Frontiers

Let v be a vertex and let $\mathcal{C}$ be a subset of descendants of v. The set $\mathcal{C}$ is a *frontier* (or a *cut*) with respect to v, if every path from v to a leaf that is a descendant of v, contains exactly one element of $\mathcal{C}$.

Let $\beta = 1/\text{polylog}(I)$. Given a vertex v, we define $\mathcal{C}_{vx}$ to be the set of descendants w of v such that if we sum up $\mathbb{D}_{w'x}$ for all intermediate vertices w' on the path from v to w we get a total $< \beta$, but adding $\mathbb{D}_{wx}$ makes the total at least β, or w is a leaf. Intuitively, vertices in $\mathcal{C}_{vx}$ correspond to the shortest transcripts for which Alice reveals β additional bits of information about x (in addition to the information revealed at v). The set $\mathcal{C}_{vy}$ is defined similarly. Observe that $\mathcal{C}_{vx}$ and $\mathcal{C}_{vy}$ are frontiers with respect to v. In addition, Alice knows the frontier $\mathcal{C}_{vx}$ and Bob knows the frontier $\mathcal{C}_{vy}$.

Let w be a vertex. We define $D_{vxy}(w)$ to be the probability that π reaches w, when the inputs are x, y, conditioned on it reaching v. Note that $D_{vxy}(w)$ is obtained by multiplying the probabilities $O_{w'xy}(b_{w'})$ for vertices w' along the path from v to w, where $b_{w'} = 0$ if the path from v to w passes through the edge going out of w' that is labeled by 0, and $b_{w'} = 1$ otherwise. In other words, following the probabilities $O_{w'xy} = O_{w'x}$ for vertices owned by Alice, and $O_{w'xy} = O_{w'y}$ for vertices owned by Bob.

We define $D_{vx}(w)$ to be the "best estimate" of $D_{vxy}(w)$ by Alice. That is, $D_{vx}(w)$ is induced by following the probabilities $O_{w'x}$ for vertices owned by Alice, and $O_{w'x} = O_{w'}$ for vertices owned by Bob. The value $D_{vy}(w)$ is defined similarly. We define $D_v(w)$ to be the "best estimate" of $D_{vxy}(w)$ by an observer who doesn't know x, y. As before, $D_v(w)$ is induced by following the probabilities $O_{w'}$ along the path. Note that $D_v(w)$ is known to both players, as well as to an external observer. Also note that by restricting each of the functions $D_{vxy}, D_{vx}, D_{vy}, D_v$ (defined over the set of all descendants of v) to any frontier with respect to v, a probability distribution is obtained. We denote by $\tilde{D}_{vxy}, \tilde{D}_{vx}, \tilde{D}_{vy}, \tilde{D}_v$ the distributions obtained by restricting the functions $D_{vxy}, D_{vx}, D_{vy}, D_v$ (respectively) to the leaves that are descendants of v. When we omit the vertex v from the notation (e.g., $D_{xy}, \tilde{D}$ etc.), we mean that v is the root of the tree.

2.3 The Simulation Protocol τ

The simulation protocol proceeds as follows: Initially v is set to the root of the tree, t below is some large constant, $r = \text{poly}(I)$, and $\eta = 1/\text{poly}(I)$.

1. **Selecting a Leader:** Players sample a bit p from the public randomness. If $p = 1$ (Bob leads), they run Steps $2 - 4$ as written. Otherwise, if $p = 0$ (Alice leads), they run Steps $2 - 4$ with the roles of Alice and Bob, and the roles of x and y, switched.

2. **Correlated Sampling:** Players jointly sample a leaf u according to $\tilde{D}_{vy}$. (Recall that Bob knows the distribution $\tilde{D}_{vy}$, while Alice's best estimate of $\tilde{D}_{vy}$ is $\tilde{D}_v$). This is done by running the correlated sampling protocol of [BR11] with $P = \tilde{D}_{vy}$, $Q = \tilde{D}_v$,

and error probability η, and requires the exchange of $\approx \mathbb{D}(\tilde{D}_{vy} \| \tilde{D}_v) + \log(1/\eta)$ bits (see Lemma 14).

3. **Finding a Separation:** Consider the unique vertex w' in the intersection of $\mathcal{C}_{vx}$ and the path from v to u. We define $c_{vx}(u)$ to be the index (number) of w' on the path from v to u. We define $c_{vy}(u)$ similarly. The goal of this step is for the players to agree on an index k such that $\min\{c_{vx}(u), c_{vy}(u)\} \leq k \leq \max\{c_{vx}(u), c_{vy}(u)\}$. This is done with error $1/r$ by sending $O(\log(r))$ bits, as follows:

   (a) Let $c_1 \leq \ldots \leq c_r \in \mathbb{N}$ be such that c_i is the first index satisfying $\Pr_{x'}[c_{vx'}(u) \leq c_i] \geq i/r$, where the probability is over x' that is sampled according to the current distribution over Alice's inputs (i.e., the original distribution μ^A conditioned on the current transcript of τ). Observe that this distribution is still a product distribution and that the indices c_i are known to both players.

   (b) Alice sends the first index $a \in [r]$ such that $c_{vx}(u) \leq c_a$. Bob sends the first index $b \in [r]$ such that $c_{vy}(u) \leq c_b$.

   (c) Return $k = \min\{c_a, c_b\}$. Let w be the k^{th} vertex on the path from v to u.

4. **Rejection Sampling:** Alice sends a bit a' that equals 0 (reject) if w has an ancestor in $\mathcal{C}_{vx}$ (intuitively, the transcript w reveals $\geq \beta$ additional bits of information about x). Otherwise, a' equals 1 (accept) with probability $\min\{1, D_{vx}(w)/tD_v(w)\}$.
   If $a' = 1$, then the players set $v = w$. If v is a leaf they end the protocol, otherwise they go back to Step 1.

2.4 Protocol Analysis

To get a rough idea of why this protocol works, fix x, y and the transcript of τ so far, and assume that vertex v was reached. We first consider the case where Step 1 selects $p = 1$ (Bob leads).

Consider a particular execution of Step 3, and let k be the returned index. Let E be the event that in this execution of Step 3, $c_a = c_b$ and $c_a \neq c_{vx}(u)$. We claim that $\min\{c_{vx}(u), c_{vy}(u)\} \leq k \leq \max\{c_{vx}(u), c_{vy}(u)\}$, unless E occurs (see Lemma 17):

1. If $c_a > c_b$ then $k = \min\{c_a, c_b\} = c_b$. By the definition of c_a as the first index for which $c_{vx}(u) \leq c_a$, it holds that $k = c_b < c_{vx}(u)$. Therefore, $c_{vy}(u) \leq c_b = k < c_{vx}(u)$, and we are done. The case where $c_b > c_a$ is similar.

2. If $c_a = c_b$ and $c_a = c_{vx}(u)$, then $c_{vx}(u) = c_a = k = c_b \geq c_{vy}(u)$, and we are also done.

Fix v, u. The current distribution over the players' inputs is obtained by conditioning the original distribution μ on the current transcript of the protocol τ. Observe that this is still a product distribution, as it is obtained by conditioning a product distribution on a transcript of a protocol. We will prove that since this distribution is a product distribution,

the event E occurs with probability at most $1/r$ (over the selection of inputs). For intuition, consider the case where the index a is (almost) uniformly distributed in $[r]$ (this always happens when the indices c_i are all distinct). Since Alice's input and Bob's input are still independent, we have that $c_b = c_a$ with probability $1/r$, and thus E occurs with probability at most $1/r$. Another extreme case is when there exists $c \in [r]$, such that it is always the case that $c_a = c$ (this always happens when all the indices c_i coincide). Then, except with probability $1/r$ over the selection of x, we have that $c_{vx}(u) = c = c_a$, and thus E does not occur. The actual proof follows from these two intuitions, see Lemmas 18 and 29.

Fix x, y and the transcript of τ so far, and assume that vertex v was reached. Consider the set $\mathcal{C} = \mathcal{C}_{vxy}$ of vertices $w = w_u$ obtained by Step 3 of the protocol τ for the different leaves u that are descendants of v. We claim that $\mathcal{C}$ is "close" to being a frontier with respect to v: Consider the protocol τ' that operates the same as τ, except that Step 3 is changed so whenever E occurs, the protocol τ' has Alice and Bob exchanging $c_{vx}(u)$ and $c_{vy}(u)$, and returns $k = \min\{c_{vx}(u), c_{vy}(u)\}$. It can be shown that the set $\mathcal{C}' = \mathcal{C}'_{vxy}$ of vertices $w = w_u$ obtained by Step 3 of the protocol τ' for the different leaves u that are descendants of v, is an actual frontier with respect to v (see Lemma 16). We also note that $\mathcal{C}'$ is "always between" $\mathcal{C}_{vx}$ and $\mathcal{C}_{vy}$: As claimed before, if E does not occur, the returned value k satisfies $\min\{c_{vx}(u), c_{vy}(u)\} \leq k \leq \max\{c_{vx}(u), c_{vy}(u)\}$. If E occurs then Step 3 of τ' returns $k = \min\{c_{vx}(u), c_{vy}(u)\}$. That is, a vertex in $\mathcal{C}'$ has either an ancestor in $\mathcal{C}_{vx}$ and a descendant in $\mathcal{C}_{vy}$, or has an ancestor in $\mathcal{C}_{vy}$ and a descendant in $\mathcal{C}_{vx}$.

Assume for simplicity that $\mathcal{C} = \mathcal{C}'$, and, in particular, that $\mathcal{C}$ is a frontier that is "always between" $\mathcal{C}_{vx}$ and $\mathcal{C}_{vy}$ ($\mathcal{C}$ and $\mathcal{C}'$ are "close" anyway, as E is an event that occurs with probability $\leq 1/r$). Also assume that all the vertices in $\mathcal{C}$ are two levels below v, with the first vertex (i.e., v) owned by Alice, and the vertices in the intermediate level owned by Bob. Let $w \in \mathcal{C}$. Recall that $D_{vxy}(w)$ is the true probability of arriving at w conditioned on reaching v, and that $D_v(w)$ is the best estimate of $D_{vxy}(w)$ by an observer who does not know x, y. Fixing w, we write $D_{vxy}(w) = D_1 D_2$ and $D_v(w) = D'_1 D'_2$, where D_i denotes the true probability that step i is taken according to x, y, and D'_i denotes this probability as estimated by an observer who does not know x, y. Observe that $D_{vx}(w) = D_1 D'_2$, $D_{vy}(w) = D'_1 D_2$. Also note that the probability that w is selected by Step 3 is $D_{vy}(w) = D'_1 D_2$, as the correlated sampling of Step 2 outputs a leaf u distributed according to $\tilde{D}_{vy}$.

We first consider the case where the frontier $\mathcal{C}_{vy}$ is "always above" $\mathcal{C}_{vx}$. That is, every vertex in $\mathcal{C}_{vx}$ is a descendant of some vertex in $\mathcal{C}_{vy}$. Intuitively, this means that upon reaching v, Bob always gives β bits of information about y before Alice gives β bits of information about x. Since we assume that $\mathcal{C}_{vy}$ is "always above" $\mathcal{C}_{vx}$, and since $\mathcal{C}$ is "always between" $\mathcal{C}_{vy}$ and $\mathcal{C}_{vx}$, it holds that w has a descendant in $\mathcal{C}_{vx}$. This intuitively means that the transcript w reveals $< \beta$ additional bits of information about x, thus $D_{vx}(w)$ and $D_v(w)$ (which is the estimation of $D_{vx}(w)$ by an observer who doesn't know x) tend to be close. Now assume that the threshold t is set high enough so that $tD_v(w) \geq D_{vx}(w)$ with high

probability. In this case the probability that w is accepted by Step 4 equals

$$\Pr[a' = 1] = \frac{D_{vx}(w)}{tD_v(w)} = \frac{D_1 D_2'}{tD_1' D_2'} = \frac{D_1}{tD_1'}.$$

This implies that the total probability that w is selected by Step 4 to be the new v, is $D_1' D_2$ times D_1/tD_1', which is exactly its correct probability $D_1 D_2$ divided by t. Hence, we get an overhead of t steps, but output the right distribution over w.

Now, it may be the case that $\mathcal{C}_{vy}$ is not always above $\mathcal{C}_{vx}$. Actually, it may even be the case that $\mathcal{C}_{vy}$ is not always above $\mathcal{C}_{vx}$ and not always below it, but rather, some vertices in $\mathcal{C}_{vy}$ have descendants in $\mathcal{C}_{vx}$, and others have ancestors in $\mathcal{C}_{vx}$. This means that after reaching v, Alice may give $\geq \beta$ additional bits of information about her input before Bob does so for his input, or it may be the other way around. Step 1 randomly chooses one of the players in every iteration and assumes that this player is going to be the first to give β additional bits of information about his input at the vertex w that will be selected by Step 3.

Assume that Bob was selected by Step 1 ($p = 1$). If w has an ancestor in $\mathcal{C}_{vy}$ (equivalently, w has a descendant in $\mathcal{C}_{vx}$), then Bob gives β additional bits of information before Alice does, and the assumption of Step 1 is true. In this case, Alice just "corrects" the probability of reaching w according to her own view, by accepting it with probability $D_{vx}(w)/tD_v(w)$. Observe that Bob does not need to correct w's probability, as it was already sampled according to D_{vy}. Otherwise, if w has an ancestor in $\mathcal{C}_{vx}$, then Alice gives β additional bits of information before Bob does. Thus, the assumption of Step 1 is false, and w gets rejected by Alice. This case, where Alice gives information first, will be handled when Step 1 selects $p = 0$ (Alice leads). Since the protocol τ cannot predict a priori (before u is selected by Step 2) which player is going to be the first to give β information, Step 1 randomly selects the leader.

2.5 The Communication Complexity of the Protocol

Let w be the vertex selected by Step 3. Since w is between $\mathcal{C}_{vx}$ and $\mathcal{C}_{vy}$, if we sum up the term $\mathbb{D}_{w'x} + \mathbb{D}_{w'y}$ for all intermediate vertices w' on the path from v to w we get a total $\geq \beta$ (unless w is a leaf). Intuitively, this means that $\geq \beta$ bits of information about x, y are revealed to the external observer when we go from the v to w. Since the external information of π over μ is only I, and since we set $\beta = 1/\text{polylog}(I)$, this means that the vertex v gets updated by Step 4 at most $m \approx I/\beta = I \cdot \text{polylog}(I)$ times, on average. Assume again that Bob was selected by Step 1 ($p = 1$) and that $\mathcal{C}_{vy}$ is always above $\mathcal{C}_{vx}$. As we claimed above, the probability that w is selected by Step 4 to be the new v, is roughly $D_{vxy}(w)/t$. Therefore, summing over all possible vertices w, in each iteration of the protocol τ, the vertex v gets updated by Step 4 with probability roughly $1/t$. This implies that the protocol τ runs for at most $m' \approx tm = I \cdot \text{polylog}(I)$ iterations, on average.

Each iteration consists of the following steps: Step 2 has error η. Since we run this step at most m' times, in order for the total error introduced by this step to be small, we set

$\eta = 1/\text{poly}(I) \ll 1/m'$. Step 2 communicates $\approx \mathbb{D}(\tilde{D}_{vy}\|\tilde{D}_v) + \log(1/\eta)$ bits if $p = 1$, and $\approx \mathbb{D}(\tilde{D}_{vx}\|\tilde{D}_v) + \log(1/\eta)$ bits if $p = 0$. Proposition 9 shows that for every x and y,

$$\mathbb{D}(\tilde{D}_x\|\tilde{D}) \leq \mathop{\mathbf{E}}_{y\leftarrow Y}\left[\mathbb{D}(\tilde{D}_{xy}\|\tilde{D})\right]; \quad \mathbb{D}(\tilde{D}_y\|\tilde{D}) \leq \mathop{\mathbf{E}}_{x\leftarrow X}\left[\mathbb{D}(\tilde{D}_{xy}\|\tilde{D})\right].$$

Loosely stated, the first equation above says that the distribution $\tilde{D}_x$ is "closer" to $\tilde{D}$ than $\tilde{D}_{xy}$. This is true as both $\tilde{D}_x$ and $\tilde{D}_{xy}$ are induced by following the probabilities $O_{w'x}$ for vertices owned by Alice. However, for vertices owned by Bob, $\tilde{D}_x$ follows the probability $O_{w'}$, while $\tilde{D}_{xy}$ follows $O_{w'y}$. The distribution $\tilde{D}$ always follows $O_{w'}$.

Proposition 8 gives an alternative definition of external information, by showing that I, the external information of π over μ, satisfies

$$I = \mathop{\mathbf{E}}_{(x,y)\leftarrow\mu}\left[\mathbb{D}(\tilde{D}_{xy}\|\tilde{D})\right].$$

Therefore, Step 2 communicates $O(I)$ bits, in expectation.

Step 3 has error $1/r$. Since we run this step at most m' times, in order for the total error introduced by this step to be small, we set $r = \text{poly}(I) \gg m'$. The communication required by this step is then $O(\log(r)) = O(\log(I))$ bits, as the players only exchange indices $a, b \in [r]$. Step 4 requires 1 bit of communication. Hence, the total communication per round is $O(I)$, in expectation, and the protocol communicates $O(m'I) = I^2 \cdot \text{polylog}(I)$ bits.

3 Preliminaries

3.1 Notation

Let π be a randomized communication protocol. Let (x, y) be a possible input for π. We denote by $\pi(x, y)$ the random variable representing π's transcript (messages exchanged during π's execution), when it is run with the input (x, y). We sometimes confuse random variables and their distribution. For example, we often view $\pi(x, y)$ as a distribution.

Let $s \in \{0, 1\}^*$ be a binary string. We denote by $|s| \in \mathbb{N} \cup \{0\}$ the length of s. For example, $|\pi(x, y)|$ is the length of the transcript represented by the random variable $\pi(x, y)$. For $i \in [|s|]$, we denote by $s_i \in \{0, 1\}$ the i^{th} bit of s, and by $s_{\leq i} \in \{0, 1\}^i$ the prefix of length i of s. For $i \leq 0$, set $s_{\leq i} = \phi$, where ϕ denotes the empty string (the string of length 0).

For a pair of distributions D, D' over the same domain Ω, we denote by $\|D - D'\|$ the *statistical/total variation* distance between D and D', given by $\|D - D'\| = \max_{\mathcal{S}\subseteq\Omega}\{|D(\mathcal{S}) - D'(\mathcal{S})|\}$. We sometimes consider the statistical distance $\|D-D'\|$, where D is a distribution over a set $\mathcal{S}$ of binary strings, and D' is a distribution over a set $\mathcal{S}'$ of binary strings (e.g., $\|\pi(x, y) - \pi'(x, y)\|$, where π, π' are protocols and (x, y) is an input for both π and π'). For this statistical distance to be defined, we view both D and D' as distributions over the set of all binary strings. For a function $D : \Omega \to \mathbb{R}$ (not necessarily a distribution) and a set $\mathcal{S} \subseteq \Omega$, we define $D(\mathcal{S}) = \sum_{x\in\mathcal{S}} D(x)$.

For a pair of random variables X, Y, we denote by $\mathbb{I}(X;Y)$ the *mutual information* between X and Y. For a pair of distributions D, D' over the same domain, we denote by $\mathbb{D}(D\|D')$ the *relative entropy* between D and D', given by

$$\mathbb{D}(D\|D') = \mathop{\mathbf{E}}_{x\leftarrow D}\left[\log\left(\frac{D(x)}{D'(x)}\right)\right].$$

3.2 Information Cost

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

$$\mathsf{IC}_\mu(\pi) = \mathbb{I}(\pi(X,Y); X \mid Y) + \mathbb{I}(\pi(X,Y); Y \mid X).$$

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

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

Fact 2. *Let π be a protocol and let μ be a distribution over the inputs for π. Then,* $\mathsf{IC}_\mu(\pi) \leq \mathsf{Ext}_\mu(\pi)$.

Fact 3 (Fact 4.16 in [BBCR10])**.** *Let π be a protocol and let $\mu = \mu^A \times \mu^B$ be a* product *distribution over the inputs for π. Then,* $\mathsf{IC}_\mu(\pi) = \mathsf{Ext}_\mu(\pi)$.

3.3 Divergence Tree

Transcript tree. Let π be a communication protocol between two players, Alice and Bob. Alice has a private input x and Bob has a private input y, where (x,y) is chosen according to some publicly known joint distribution μ. In this work, we consider the case where $\mu = \mu^A \times \mu^B$ is a product distribution. We assume, without loss of generality, that π does not use public randomness (but may use private randomness), as the public randomness can always be replaced by private randomness without increasing the (internal or external) information cost. We also assume that the players take turns sending bits to one another. Alice send bits in odd rounds and Bob in the even rounds. We further assume, without loss of generality, that the players communicate the same number of bits in every execution of the protocol.

We denote by $\mathcal{T}$ the (directed) binary tree associated with the communication protocol π. That is, every vertex v of $\mathcal{T}$ corresponds to a possible transcript of π. The two edges going out of v are labeled by 0 and 1, corresponding to the next bit to be transmitted. We assume that $\mathcal{T}$ is a complete binary tree. We denote by $\mathcal{V}$ the set of vertices of $\mathcal{T}$, by v_0 the root of $\mathcal{T}$, and by $\mathcal{L}$ the set of leaves of $\mathcal{T}$. Since a vertex $v \in \mathcal{V}$ corresponds to a possible transcript of π, we often think of v as a string. That is, we view v as the binary string induced by the labels of the edges on the path from the root to v.

We think of Alice as owning the non-leaf vertices in the odd layers of $\mathcal{T}$, and of Bob as owning the non-leaf vertices in the even layers of $\mathcal{T}$. We denote by $\mathcal{V}^A \subseteq \mathcal{V} \setminus \mathcal{L}$ the set of vertices owned by Alice and by $\mathcal{V}^B \subseteq \mathcal{V} \setminus \mathcal{L}$ the set of vertices owned by Bob. The protocol π proceeds as follows: Starting from the root, when the protocol π reaches a non-leaf vertex v, the player who owns v sends a bit to the other player. The players follow the edge indicated by the sent bit and reach a new vertex.

For a vertex $v \in \mathcal{V}$, we denote by $\mathcal{V}(v) \subseteq \mathcal{V}$ the set of vertices in the subtree rooted in v (including v itself), and by $\mathcal{L}(v)$ the set of leaves in the subtree rooted in v (including v itself). Let $v, w \in \mathcal{V}$. We say that w is a *descendant* of v, if w is in the subtree rooted in v (in particular, v is a descendant of itself). We say that w is a *strict descendant* of v, if w is in the subtree rooted in v and $w \neq v$. We say that v is an *ancestor* of w, if w is a descendant of v. We say that v is a *strict ancestor* of w, if w is a strict descendant of v.

Distributions associated with a vertex. Let (X, Y) be a pair of random variables distributed according to μ, representing the players' inputs. Since μ is a product distribution, X and Y are independent. We view $\pi(X, Y)$ as the random variable representing the leaf of $\mathcal{T}$ reached by the protocol π.

Let $v \in \mathcal{V}$. We assume that for every $(x, y) \in \mathrm{supp}(\mu)$ it is possible for the protocol π to reach the vertex v. That is, $\Pr[\pi(X, Y)_{\leq |v|} = v \mid X = x, Y = y] > 0$. This can be assumed without loss of generality: Consider the private coin protocol π^*, in which Alice follows π with probability $1 - \delta$, and, with probability δ, she sends a random bit for each vertex that she owns, for some sufficiently small $\delta > 0$. Bob acts similarly. Observe that π^* has the required property and $\|\pi(X, Y) - \pi^*(X, Y)\| \leq 2\delta$.

Let $v \in \mathcal{V} \setminus \mathcal{L}$ and $(x, y) \in \mathrm{supp}(\mu)$. We define the following distributions over the two edges going out of v: O_{vxy} is the distribution over the edges going out of v according to the protocol π. We think of O_{vxy} as the "real" distribution at v. The distribution O_{vx} is Alice's best estimate O_{vxy}, and the distribution O_{vy} is Bob's best estimate O_{vxy}. The distribution O_v is the best estimate of O_{vxy} by an external observer who doesn't know neither x nor y. Formally, for $b \in \{0, 1\}$,

$$\begin{aligned}
O_{vxy}(b) &= \Pr\left[\pi(X, Y)_{|v|+1} = b \mid \pi(X, Y)_{\leq |v|} = v, X = x, Y = y\right], \\
O_{vx}(b) &= \Pr\left[\pi(X, Y)_{|v|+1} = b \mid \pi(X, Y)_{\leq |v|} = v, X = x\right], \\
O_{vy}(b) &= \Pr\left[\pi(X, Y)_{|v|+1} = b \mid \pi(X, Y)_{\leq |v|} = v, Y = y\right], \\
O_{v}(b) &= \Pr\left[\pi(X, Y)_{|v|+1} = b \mid \pi(X, Y)_{\leq |v|} = v\right].
\end{aligned}$$

We note that

$$\forall v \in \mathcal{V}^B : \; O_{vxy}(b) = O_{vy}(b); \qquad \forall v \in \mathcal{V}^A : \; O_{vxy}(b) = O_{vx}(b). \tag{1}$$

Furthermore, since X and Y are independent even conditioned on any transcript v of the

protocol π, we get

$$\forall v \in \mathcal{V}^B : \ O_{vx}(b) = O_v(b); \qquad \forall v \in \mathcal{V}^A : \ O_{vy}(b) = O_v(b). \tag{2}$$

Probabilities of reaching a vertex. Let $v \in \mathcal{V}$ and $(x, y) \in \text{supp}(\mu)$. We define the functions $D_{vxy}, D_{vx}, D_{vy}, D_v : \mathcal{V}(v) \to [0, 1]$ by:

$$D_{vxy}(v) = D_{vx}(v) = D_{vy}(v) = D_v(v) = 1,$$

and for $w \in \mathcal{V}(v) \setminus \{v\}$,

$$\begin{aligned}
D_{vxy}(w) &= \prod_{i \in \{|v|,\ldots,|w|-1\}} O_{w_{\le i}xy}(w_{i+1}), \\
D_{vx}(w) &= \prod_{i \in \{|v|,\ldots,|w|-1\}} O_{w_{\le i}x}(w_{i+1}), \\
D_{vy}(w) &= \prod_{i \in \{|v|,\ldots,|w|-1\}} O_{w_{\le i}y}(w_{i+1}), \\
D_v(w) &= \prod_{i \in \{|v|,\ldots,|w|-1\}} O_{w_{\le i}}(w_{i+1}).
\end{aligned}$$

Let $\tilde{D}_{vxy}, \tilde{D}_{vx}, \tilde{D}_{vy}, \tilde{D}_v : \mathcal{L}(v) \to [0, 1]$ be the restrictions of the functions $D_{vxy}, D_{vx}, D_{vy}, D_v$ to $\mathcal{L}(v)$ (respectively). Observe that $\tilde{D}_{vxy}, \tilde{D}_{vx}, \tilde{D}_{vy}, \tilde{D}_v$ are distributions over $\mathcal{L}(v)$. Whenever we omit the subscript v, we mean that v is v_0, the root of $\mathcal{T}$. E.g., $D_{xy} = D_{v_0xy}$, $\tilde{D}_{xy} = \tilde{D}_{v_0xy}$.

Proposition 4. *Let $v \in \mathcal{V}$, $w \in \mathcal{V}(v)$, and $(x, y) \in \text{supp}(\mu)$. Then,*

$$\frac{D_{vy}(w) \cdot D_{vx}(w)}{D_v(w)} = D_{vxy}(w).$$

Proof.

$$\begin{aligned}
&\frac{D_{vy}(w) \cdot D_{vx}(w)}{D_v(w)} \\
&= \prod_{i \in \{|v|,\ldots,|w|-1\}} \frac{O_{w_{\le i}y}(w_{i+1}) \cdot O_{w_{\le i}x}(w_{i+1})}{O_{w_{\le i}}(w_{i+1})} \\
&= \prod_{\substack{i \in \{|v|,\ldots,|w|-1\} \\ s.t.\ w_{\le i} \in \mathcal{V}^A}} O_{w_{\le i}x}(w_{i+1}) \cdot \prod_{\substack{i \in \{|v|,\ldots,|w|-1\} \\ s.t.\ w_{\le i} \in \mathcal{V}^B}} O_{w_{\le i}y}(w_{i+1}) && \text{(by Equation (2))} \\
&= \prod_{\substack{i \in \{|v|,\ldots,|w|-1\} \\ s.t.\ w_{\le i} \in \mathcal{V}^A}} O_{w_{\le i}xy}(w_{i+1}) \cdot \prod_{\substack{i \in \{|v|,\ldots,|w|-1\} \\ s.t.\ w_{\le i} \in \mathcal{V}^B}} O_{w_{\le i}xy}(w_{i+1}) && \text{(by Equation (1))} \\
&= D_{vxy}(w).
\end{aligned}$$

□

Let $v \in \mathcal{V}$, $w \in \mathcal{V}(v)$, and $(x, y) \in \text{supp}(\mu)$. If $w = v$, we define $\mathbb{D}_{vwxy} = \mathbb{D}_{vwx} = \mathbb{D}_{vwy} = 0$. If $w \neq v$, we define

$$\mathbb{D}_{vwxy} = \sum_{i \in \{|v|,\ldots,|w|-1\}} \mathbb{D}(O_{w_{\leq i}xy} \| O_{w_{\leq i}}),$$
$$\mathbb{D}_{vwx} = \sum_{i \in \{|v|,\ldots,|w|-1\}} \mathbb{D}(O_{w_{\leq i}x} \| O_{w_{\leq i}}),$$
$$\mathbb{D}_{vwy} = \sum_{i \in \{|v|,\ldots,|w|-1\}} \mathbb{D}(O_{w_{\leq i}y} \| O_{w_{\leq i}}),$$

By Equations (1) and (2) it holds that $\mathbb{D}_{vwxy} = \mathbb{D}_{vwx} + \mathbb{D}_{vwy}$.

Proposition 5 (Corollary 6 in [Gan12])**.** *Let $v \in \mathcal{V}$ and $(x, y) \in \text{supp}(\mu)$. Then,*

$$\mathbb{D}(\tilde{D}_{vxy} \| \tilde{D}_v) = \sum_{w \in \mathcal{V}(v) \setminus \mathcal{L}(v)} D_{vxy}(w) \cdot \mathbb{D}(O_{wxy} \| O_w) = \sum_{u \in \mathcal{L}(v)} \tilde{D}_{vxy}(u) \cdot \mathbb{D}_{vuxy}.$$

$$\mathbb{D}(\tilde{D}_{vx} \| \tilde{D}_v) = \sum_{w \in \mathcal{V}(v) \setminus \mathcal{L}(v)} D_{vx}(w) \cdot \mathbb{D}(O_{wx} \| O_w) = \sum_{u \in \mathcal{L}(v)} \tilde{D}_{vx}(u) \cdot \mathbb{D}_{vux}.$$

$$\mathbb{D}(\tilde{D}_{vy} \| \tilde{D}_v) = \sum_{w \in \mathcal{V}(v) \setminus \mathcal{L}(v)} D_{vy}(w) \cdot \mathbb{D}(O_{wy} \| O_w) = \sum_{u \in \mathcal{L}(v)} \tilde{D}_{vy}(u) \cdot \mathbb{D}_{vuy}.$$

Proof. We will prove the first equality, the others are similar. Assume without loss of generality that the subtree of $\mathcal{T}$ rooted at v is a complete binary tree, where v is in level 1 and the leaves are in level c, for some $c \in \mathbb{N}$. Let $k \in [c]$. Denote by $\mathcal{V}^k(v) \subseteq \mathcal{V}(v)$ the vertices in level k of the subtree rooted at v. Let D^k_{vxy} and D^k_v be the distributions obtained by restricting the functions D_{vxy} and D_v to $\mathcal{V}^k(v)$ (respectively).

Let $1 \leq k \leq c - 1$. By definition,

$$\mathbb{D}\left(D^{k+1}_{vxy} \| D^{k+1}_v\right) = \mathop{\mathbf{E}}_{w' \leftarrow D^{k+1}_{vxy}} \left[\log\left(\frac{D^{k+1}_{vxy}(w')}{D^{k+1}_v(w')}\right)\right]$$

To go over the vertices w' in level $k+1$, we consider each of the vertices w in level k, and look at its children in level $k+1$. Let w_0 be the vertex led to from w by the edge labeled by 0, and let w_1 be the vertex led to by the edge labeled by 1. It holds that $D^{k+1}_{vxy}(w_0) = D^k_{vxy}(w) \cdot O_{wxy}(0)$ and $D^{k+1}_{vxy}(w_1) = D^k_{vxy}(w) \cdot O_{wxy}(1)$. Similarly, $D^{k+1}_v(w_0) = D^k_v(w) \cdot O_w(0)$

and $D_v^{k+1}(w_1) = D_v^k(w) \cdot O_w(1)$. Therefore,

$$\begin{aligned}
&\mathbb{D}\left(D_{vxy}^{k+1} \| D_v^{k+1}\right) \\
&= \mathop{\mathbf{E}}_{w \leftarrow D_{vxy}^k} \left[O_{wxy}(0) \cdot \log\left(\frac{D_{vxy}^k(w) \cdot O_{wxy}(0)}{D_v^k(w) \cdot O_w(0)}\right) + O_{wxy}(1) \cdot \log\left(\frac{D_{vxy}^k(w) \cdot O_{wxy}(1)}{D_v^k(w) \cdot O_w(1)}\right)\right] \\
&= \mathop{\mathbf{E}}_{w \leftarrow D_{vxy}^k} \left[\log\left(\frac{D_{vxy}^k(w)}{D_v^k(w)}\right)\right] + \mathop{\mathbf{E}}_{w \leftarrow D_{vxy}^k} \left[O_{wxy}(0) \cdot \log\left(\frac{O_{wxy}(0)}{O_w(0)}\right) + O_{wxy}(1) \cdot \log\left(\frac{O_{wxy}(1)}{O_w(1)}\right)\right] \\
&= \mathbb{D}(D_{vxy}^k \| D_v^k) + \mathop{\mathbf{E}}_{w \leftarrow D_{vxy}^k} \left[\mathbb{D}(O_{wxy} \| O_w)\right].
\end{aligned}$$

By applying the above equation recursively to every level, we get

$$\begin{aligned}
\mathbb{D}(\tilde{D}_{vxy} \| \tilde{D}_v) &= \mathbb{D}(\mathbb{D}_{vxy}^c \| \mathbb{D}_v^c) = \mathbb{D}\left(D_{vxy}^1 \| D_v^1\right) + \sum_{k=1}^{c-1} \mathop{\mathbf{E}}_{w \leftarrow D_{vxy}^k} \left[\mathbb{D}(O_{wxy} \| O_w)\right] \\
&= 0 + \sum_{k=1}^{c-1} \sum_{w \in \mathcal{V}^k(v)} D_{vxy}^k(w) \cdot \mathbb{D}(O_{wxy} \| O_w) = \sum_{w \in \mathcal{V}(v) \setminus \mathcal{L}(v)} D_{vxy}(w) \cdot \mathbb{D}(O_{wxy} \| O_w).
\end{aligned}$$

□

Proposition 6. *For every $(x, y) \in \mathrm{supp}(\mu)$ and $w \in \mathcal{V}$,*

$$\begin{aligned}
D_{xy}(w) &= \Pr\left[\pi(X,Y)_{\le |w|} = w \mid X = x, Y = y\right], \\
D_x(w) &= \Pr\left[\pi(X,Y)_{\le |w|} = w \mid X = x\right], \\
D_y(w) &= \Pr\left[\pi(X,Y)_{\le |w|} = w \mid Y = y\right], \\
D(w) &= \Pr\left[\pi(X,Y)_{\le |w|} = w\right].
\end{aligned}$$

Proof. We will prove the first equality, the rest are similar. Fix $(x, y) \in \mathrm{supp}(\mu)$. Let $w \in \mathcal{V}$. By the probability chain rule,

$$\begin{aligned}
D_{xy}(w) &= \prod_{i \in \{0,\ldots,|w|-1\}} O_{w_{\le i} xy}(w_{i+1}) \\
&= \prod_{i \in \{0,\ldots,|w|-1\}} \Pr\left[\pi(X,Y)_{i+1} = w_{i+1} \mid \pi(X,Y)_1 = w_1, \ldots, \pi(X,Y)_i = w_i, X = x, Y = y\right] \\
&= \Pr\left[\pi(X,Y)_{\le |w|} = w \mid X = x, Y = y\right].
\end{aligned}$$

□

Proposition 7. *For every $x \in \mathrm{supp}(\mu^A)$, $y \in \mathrm{supp}(\mu^B)$, and $w \in \mathcal{V}$,*

$$\begin{aligned}
D_x(w) &= \Pr_{y \leftarrow Y}\left[D_{xy}(w)\right], \\
D_y(w) &= \Pr_{x \leftarrow X}\left[D_{xy}(w)\right].
\end{aligned}$$

Proof. We will prove the first equality, the second is similar. Fix $x \in \mathrm{supp}(\mu^A)$. Let $w \in \mathcal{V}$.

$$\begin{aligned}
D_x(w) &= \Pr\left[\pi(X,Y)_{\le|w|} = w \mid X = x\right] && \text{(by Proposition 6)}\\
&= \sum_{y\in\mathrm{supp}(\mu^B)} \Pr[Y=y|X=x]\cdot \Pr\left[\pi(X,Y)_{\le|w|} = w \mid X=x, Y=y\right]\\
&= \sum_{y\in\mathrm{supp}(\mu^B)} \Pr[Y=y]\cdot \Pr\left[\pi(X,Y)_{\le|w|} = w \mid X=x, Y=y\right] && (X,Y \text{ are independent})\\
&= \Pr_{y\leftarrow Y}\left[\pi(X,Y)_{\le|w|} = w \mid X=x, Y=y\right]\\
&= \Pr_{y\leftarrow Y}\left[D_{xy}(w)\right]
\end{aligned}$$

□

The following claim gives a useful alternative definition of external information.

Proposition 8. *It holds that*

$$\mathsf{Ext}_\mu(\pi) = \mathop{\mathbf{E}}_{(x,y)\leftarrow\mu}\left[\mathbb{D}(\tilde{D}_{xy}\|\tilde{D})\right].$$

Proof. By Proposition 6,

$$\tilde{D}_{xy} = (\pi(X,Y) \mid X=x, Y=y); \qquad \tilde{D} = \pi(X,Y).$$

Recall that for any two random variables A, B,

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

Therefore,

$$\begin{aligned}
\mathsf{Ext}_\mu(\pi) &= \mathbb{I}(X,Y;\pi(X,Y))\\
&= \mathop{\mathbf{E}}_{(x,y)\leftarrow\mu}\left[\mathbb{D}\left((\pi(X,Y)|X=x,Y=y) \parallel \pi(X,Y)\right)\right] = \mathop{\mathbf{E}}_{(x,y)\leftarrow\mu}\left[\mathbb{D}(\tilde{D}_{xy}\|\tilde{D})\right].
\end{aligned}$$

□

Proposition 9. *For every $x \in \mathrm{supp}(\mu^A)$ and $y \in \mathrm{supp}(\mu^B)$,*

$$\begin{aligned}
\mathbb{D}(\tilde{D}_x\|\tilde{D}) &\le \mathop{\mathbf{E}}_{y\leftarrow Y}\left[\mathbb{D}(\tilde{D}_{xy}\|\tilde{D})\right],\\
\mathbb{D}(\tilde{D}_y\|\tilde{D}) &\le \mathop{\mathbf{E}}_{x\leftarrow X}\left[\mathbb{D}(\tilde{D}_{xy}\|\tilde{D})\right].
\end{aligned}$$

Proof. We will prove the first inequality, the second is similar. Fix $x \in \mathrm{supp}(\mu^A)$. It holds

that

$$
\begin{aligned}
&\mathbb{D}(\tilde{D}_x \| \tilde{D}) \\
&= \sum_{w \in \mathcal{V} \setminus \mathcal{L}} D_x(w) \cdot \mathbb{D}(O_{wx} \| O_w) && \text{(by Proposition 5)} \\
&= \sum_{w \in \mathcal{V}^A \setminus \mathcal{L}} D_x(w) \cdot \mathbb{D}(O_{wx} \| O_w) + \sum_{w \in \mathcal{V}^B \setminus \mathcal{L}} D_x(w) \cdot \mathbb{D}(O_w \| O_w) && \text{(by Equation (2))} \\
&\leq \sum_{w \in \mathcal{V}^A \setminus \mathcal{L}} \mathop{\mathbf{E}}_{y \leftarrow Y} [D_{xy}(w)] \cdot \mathbb{D}(O_{wx} \| O_w) + \sum_{w \in \mathcal{V}^B \setminus \mathcal{L}} \mathop{\mathbf{E}}_{y \leftarrow Y} [D_{xy}(w) \cdot \mathbb{D}(O_{wy} \| O_w)] \\
& && \text{(by Proposition 7 and as } \mathbb{D}(O_w \| O_w) = 0, \mathbb{D}(O_{wy} \| O_w) \geq 0) \\
&= \mathop{\mathbf{E}}_{y \leftarrow Y} \left[\sum_{w \in \mathcal{V}^A \setminus \mathcal{L}} D_{xy}(w) \cdot \mathbb{D}(O_{wx} \| O_w) + \sum_{w \in \mathcal{V}^B \setminus \mathcal{L}} D_{xy}(w) \cdot \mathbb{D}(O_{wy} \| O_w) \right] \\
&= \Pr_{y \leftarrow Y} \left[\sum_{w \in \mathcal{V} \setminus \mathcal{L}} D_{xy}(w) \cdot \mathbb{D}(O_{wxy} \| O_w) \right] && \text{(by Equation (1))} \\
&= \Pr_{y \leftarrow Y} \left[\mathbb{D}(\tilde{D}_{xy} \| \tilde{D}) \right] && \text{(by Proposition 5)}
\end{aligned}
$$

□

3.4 Frontiers

Definition 3 (**Frontier**). *Let $v \in \mathcal{V}$ and $\mathcal{C} \subseteq \mathcal{V}(v)$. The set $\mathcal{C}$ is a* frontier *with respect to v, if very path from v to a leaf $u \in \mathcal{L}(v)$ contains exactly one element of $\mathcal{C}$.*

Let $(x, y) \in \text{supp}(\mu)$. Let $v \in \mathcal{V}$ and let $\mathcal{C} \subseteq \mathcal{V}(v)$ be a frontier with respect to v. Observe that when restricting each of the functions $D_{vxy}, D_{vx}, D_{vy}, D_v$ to the frontier $\mathcal{C}$, we get a distribution. In particular, since $\mathcal{L}(v)$ is a frontier with respect to v, $\tilde{D}_{vxy}, \tilde{D}_{vx}, \tilde{D}_{vy}, \tilde{D}_v$ are distributions.

Proposition 10. *Let $x \in \text{supp}(\mu^A)$ and $y \in \text{supp}(\mu^B)$. Let $\mathcal{C}$ be a frontier with respect to the root v_0. Then,*

$$
\begin{aligned}
\mathbb{D}(\tilde{D}_x \| \tilde{D}) &\geq \mathop{\mathbf{E}}_{v \leftarrow D_x|_{\mathcal{C}}} \left[\mathbb{D}(\tilde{D}_{vx} \| \tilde{D}_v) \right] = \sum_{v \in \mathcal{C}} D_x(v) \cdot \mathbb{D}(\tilde{D}_{vx} \| \tilde{D}_v), \\
\mathbb{D}(\tilde{D}_y \| \tilde{D}) &\geq \mathop{\mathbf{E}}_{v \leftarrow D_y|_{\mathcal{C}}} \left[\mathbb{D}(\tilde{D}_{vy} \| \tilde{D}_v) \right] = \sum_{v \in \mathcal{C}} D_y(v) \cdot \mathbb{D}(\tilde{D}_{vy} \| \tilde{D}_v),
\end{aligned}
$$

where $D_x|_{\mathcal{C}}$ and $D_y|_{\mathcal{C}}$ denote the distributions obtained by restricting the functions D_x and D_y (respectively) to the frontier $\mathcal{C}$.

Proof. We will prove the first inequality, the second is similar. Fix $x \in \text{supp}(\mu^A)$. It holds

that

$$\begin{aligned}
&\mathop{\mathbf{E}}_{v\leftarrow D_x|_{\mathcal{C}}}\left[\mathbb{D}(\tilde{D}_{vx}\|\tilde{D}_v)\right] \\
&= \mathop{\mathbf{E}}_{v\leftarrow D_x|_{\mathcal{C}}}\left[\sum_{u\in\mathcal{L}(v)} D_{vx}(u)\cdot\mathbb{D}_{vux}\right] && \text{(by Proposition 5)} \\
&= \sum_{v\in\mathcal{C}}\sum_{u\in\mathcal{L}(v)} D_x(v)\cdot D_{vx}(u)\cdot\mathbb{D}_{vux} = \sum_{v\in\mathcal{C}}\sum_{u\in\mathcal{L}(v)} D_x(u)\cdot\mathbb{D}_{vux} \\
&\leq \sum_{v\in\mathcal{C}}\sum_{u\in\mathcal{L}(v)} D_x(u)\cdot\mathbb{D}_{v_0ux} = \sum_{u\in\mathcal{L}} D_x(u)\cdot\mathbb{D}_{v_0ux} \\
&= \mathbb{D}(\tilde{D}_x\|\tilde{D}). && \text{(by Proposition 5)}
\end{aligned}$$

□

The sets $\mathcal{V}_{vx},\mathcal{V}_{vy}$ and the frontiers $\mathcal{C}_{vx},\mathcal{C}_{vy}$. Assume that the leaves of $\mathcal{T}$ are (all) in level d, and that d is odd. Let $\mathcal{L}^B$ be set of vertices in the level $d-1$, the last level before the leaves. Note that level $d-1$ is owned by Bob. Let $\mathcal{L}^A$ be set of vertices in the level $d-2$, the second to last level before the leaves. Note that level $d-2$ is owned by Alice. Let $\mathcal{L}^+ = \mathcal{L}\cup\mathcal{L}^A\cup\mathcal{L}^B$.

We use a parameter $\beta > 0$ that will be set later. Let $v\in\mathcal{V}\setminus\mathcal{L}^+$, $x\in\mathrm{supp}(\mu^A)$ and $y\in\mathrm{supp}(\mu^B)$. We define the set $\mathcal{V}_{vx}$ as the set of all $w\in\mathcal{V}(v)\setminus\mathcal{L}$ satisfying $\mathbb{D}_{vw_{\leq|w|-1}x} < \beta$. That is, we include w in $\mathcal{V}_{vx}$ if $\mathbb{D}_{vwx} < \beta$, or if $\mathbb{D}_{vwx}\geq\beta$ and w is the first vertex on the path from v to w for which $\mathbb{D}_{vwx}\geq\beta$. Similarly, we define the set $\mathcal{V}_{vy}$ as the set of all $w\in\mathcal{V}(v)\setminus(\mathcal{L}\cup\mathcal{L}^B)$ satisfying $\mathbb{D}_{vw_{\leq|w|-1}y} < \beta$. Note that Alice knows the set $\mathcal{V}_{vx}$ and Bob knows the set $\mathcal{V}_{vy}$.

We define the set $\mathcal{C}_{vx}\subseteq\mathcal{V}_{vx}$ as the set of all $w\in\mathcal{V}_{vx}$ such that w's children are not in $\mathcal{V}_{vx}$ (observe that if one child is not in $\mathcal{V}_{vx}$, then the other child is not in $\mathcal{V}_{vx}$ either). In other words, the set $\mathcal{C}_{vx}$ is the "border" of the set $\mathcal{V}_{vx}$. Similarly, we define the set $\mathcal{C}_{vy}\subseteq\mathcal{V}_{vy}$ as the set of all $w\in\mathcal{V}_{vy}$ such that w's children are not in $\mathcal{V}_{vy}$. Observe that both $\mathcal{C}_{vx}$ and $\mathcal{C}_{vy}$ are frontiers with respect to v. Note that Alice known the frontier $\mathcal{C}_{vx}$ and Bob knows the frontier $\mathcal{C}_{vy}$. In addition, $\mathcal{C}_{vx}\subseteq\mathcal{V}^B$ and $\mathcal{C}_{vy}\subseteq\mathcal{V}^A$, thus

$$\mathcal{C}_{vx}\cap\mathcal{C}_{vy} = \phi. \tag{3}$$

The indices $c_{vx}(u), c_{vy}(u)$. Let $v\in\mathcal{V}\setminus\mathcal{L}^+$ and $u\in\mathcal{L}(v)$. We denote by $P(v,u)$ the set of vertices on the path from v to u. Let $x\in\mathrm{supp}(\mu^A)$ and $y\in\mathrm{supp}(\mu^B)$. Consider the unique vertex w in the intersection of $\mathcal{C}_{vx}$ and $P(v,u)$. We set $c_{vx}(u)$ to be the index (number) of w on the path from v to u. That is, if the vertices on the path from v to u are $w_1,\ldots,w_k$ (where $w_1 = v$ and $w_k = u$), and $w_c\in\mathcal{C}_{vx}$ ($c\in[k]$), then $c_{vx}(u) = c$. Similarly, we consider the unique vertex w' in the intersection of $\mathcal{C}_{vy}$ and $P(v,u)$. We set $c_{vy}(u)$ to be the index (number) of w' on the path from v to u.

The frontiers $\mathcal{C}^{\min}_{vxy}$ and $\mathcal{C}^{\max}_{vxy}$. Let $v \in \mathcal{V} \setminus \mathcal{L}^+$ and $(x,y) \in \text{supp}(\mu)$. We define a set $\mathcal{C}^{\min}_{vxy} \subseteq \mathcal{C}_{vx} \cup \mathcal{C}_{vy}$ as follows: Include $w \in \mathcal{C}_{vx}$ in $\mathcal{C}^{\min}_{vxy}$ if w has a descendant in $\mathcal{C}_{vy}$. In addition, include $w \in \mathcal{C}_{vy}$ in $\mathcal{C}^{\min}_{vxy}$ if w has a descendant in $\mathcal{C}_{vx}$. We define a set $\mathcal{C}^{\max}_{vxy} \subseteq \mathcal{C}_{vx} \cup \mathcal{C}_{vy}$ as follows: Include $w \in \mathcal{C}_{vx}$ in $\mathcal{C}^{\max}_{vxy}$ if w has an ancestor in $\mathcal{C}_{vy}$. In addition, include $w \in \mathcal{C}_{vy}$ in $\mathcal{C}^{\max}_{vxy}$ if w has an ancestor in $\mathcal{C}_{vx}$. Observe that $\mathcal{C}^{\min}_{vxy}$ and $\mathcal{C}^{\max}_{vxy}$ are frontiers.

Let $v \in \mathcal{V}$ and $(x,y) \in \text{supp}(\mu)$. Let $w \in \mathcal{V}(v)$. We write $\mathcal{C}^{\min}_{vxy} < w \leq \mathcal{C}^{\max}_{vxy}$, if w is a strict descendant of a vertex $w' \in \mathcal{C}^{\min}_{vxy}$, and w is an ancestor of a vertex $w'' \in \mathcal{C}^{\max}_{vxy}$ (note that we allow $w = w''$, but do not allow $w = w'$). Observe that if $\mathcal{C}^{\min}_{vxy} < w \leq \mathcal{C}^{\max}_{vxy}$, then $w \in \mathcal{V}_{vx} \cup \mathcal{V}_{vy}$. For a set $\mathcal{C} \subseteq \mathcal{V}(v)$, we write $\mathcal{C}^{\min}_{vxy} < \mathcal{C} \leq \mathcal{C}^{\max}_{vxy}$, if for every $w \in \mathcal{C}$, it holds that $\mathcal{C}^{\min}_{vxy} < w \leq \mathcal{C}^{\max}_{vxy}$.

Proposition 11. *Let $v \in \mathcal{V} \setminus \mathcal{L}^+$ and $(x,y) \in \text{supp}(\mu)$. Let $w \in \mathcal{V}(v)$ such that $\mathcal{C}^{\min}_{vxy} < w \leq \mathcal{C}^{\max}_{vxy}$. Then, either $w \in \mathcal{V}_{vx} \setminus \mathcal{V}_{vy}$ or $w \in \mathcal{V}_{vy} \setminus \mathcal{V}_{vx}$.*

Proof. Since $\mathcal{C}^{\min}_{vxy} < w \leq \mathcal{C}^{\max}_{vxy}$, it holds that $w \in \mathcal{V}_{vx} \cup \mathcal{V}_{vy}$ and that there exists $w' \in \mathcal{C}^{\min}_{vxy}$ such that w is a strict descendant of w'. Since $w' \in \mathcal{C}^{\min}_{vxy}$, either $w' \in \mathcal{C}_{vx}$ or $w' \in \mathcal{C}_{vy}$. First consider the case where $w' \in \mathcal{C}_{vx}$. By the definition of $\mathcal{C}_{vx}$, since w is a strict descendant of w', we get $w \notin \mathcal{V}_{vx}$. Conclude that $w \in \mathcal{V}_{vy} \setminus \mathcal{V}_{vx}$. The case where $w' \in \mathcal{C}_{vy}$ is analyzed similarly and gives $w \in \mathcal{V}_{vx} \setminus \mathcal{V}_{vy}$. □

Proposition 12. *Let $v \in \mathcal{V} \setminus \mathcal{L}^+$ and $(x,y) \in \text{supp}(\mu)$. Let $\mathcal{C}', \mathcal{C}'' \subseteq \mathcal{V}(v)$ be two frontiers with respect to v such that $\mathcal{C}^{\min}_{vxy} < \mathcal{C}', \mathcal{C}'' \leq \mathcal{C}^{\max}_{vxy}$. Let*

$$\mathcal{C} = (\mathcal{C}' \cap \mathcal{V}_{vx}) \cup (\mathcal{C}'' \cap \mathcal{V}_{vy}).$$

Then, $\mathcal{C}$ is a frontier with respect to v. In addition,

$$(\mathcal{C}' \cap \mathcal{V}_{vx}) \cap (\mathcal{C}'' \cap \mathcal{V}_{vy}) = \phi.$$

Proof. Let $w \in \mathcal{C}' \cap \mathcal{V}_{vx}$. Since $\mathcal{C}^{\min}_{vxy} < \mathcal{C}' \leq \mathcal{C}^{\max}_{vxy}$, then also $\mathcal{C}^{\min}_{vxy} < w \leq \mathcal{C}^{\max}_{vxy}$. By Proposition 11, since $w \in \mathcal{V}_{vx}$ then $w \notin \mathcal{V}_{vy}$, and in particular, $w \notin \mathcal{C}'' \cap \mathcal{V}_{vy}$. Conclude that $(\mathcal{C}' \cap \mathcal{V}_{vx}) \cap (\mathcal{C}'' \cap \mathcal{V}_{vy}) = \phi$.

We next prove that $\mathcal{C}$ is a frontier with respect to v. Clearly, $\mathcal{C} \subseteq \mathcal{V}(v)$. Let $u \in \mathcal{L}(v)$. We need to show that $|\mathcal{C} \cap P(v,u)| = 1$. We first show that $\mathcal{C} \cap P(v,u) \neq \phi$. Recall that $\mathcal{C}', \mathcal{C}''$ are frontiers with respect to v, and let $\mathcal{C}' \cap P(v,u) = \{w'\}$ and $\mathcal{C}'' \cap P(v,u) = \{w''\}$. Since $w', w'' \in P(v,u)$, one is a descendant of the other. Assume without loss of generality that w'' is a descendant of w'. If $w'' \in \mathcal{V}_{vy}$, then $w'' \in \mathcal{C} \cap P(v,u)$, and we are done. Consider the case where $w'' \notin \mathcal{V}_{vy}$. Since $w'' \in \mathcal{C}''$, it holds that $\mathcal{C}^{\min}_{vxy} < w'' \leq \mathcal{C}^{\max}_{vxy}$, thus $w'' \in \mathcal{V}_{vx} \cup \mathcal{V}_{vy}$. Since we assume that $w'' \notin \mathcal{V}_{vy}$, it holds that $w'' \in \mathcal{V}_{vx}$. Since w'' is a descendant of w', by the definition of the set $\mathcal{V}_{vx}$, it holds that $w' \in \mathcal{V}_{vx}$. Hence, $w' \in \mathcal{C} \cap P(v,u)$, and we are done.

We now show that $|\mathcal{C} \cap P(v,u)| \leq 1$. Let $\mathcal{S}' = (\mathcal{C}' \cap \mathcal{V}_{vx}) \cap P(v,u)$ and $\mathcal{S}'' = (\mathcal{C}'' \cap \mathcal{V}_{vy}) \cap P(v,u)$. Since, $\mathcal{C} \cap P(v,u) = \mathcal{S}' \cup \mathcal{S}''$, we need to show that $|\mathcal{S}' \cup \mathcal{S}''| \leq 1$. Since each of the sets $\mathcal{C}' \cap \mathcal{V}_{vx}$ and $\mathcal{C}'' \cap \mathcal{V}_{vy}$ is subset of a frontier, $|\mathcal{S}'|, |\mathcal{S}''| \leq 1$. We will

show that either $\mathcal{S}' = \phi$ or $\mathcal{S}'' = \phi$. Assume for contradiction that $\mathcal{S}' = \{w'\}$ and $\mathcal{S}'' = \{w''\}$ for some $w', w'' \in \mathcal{V}(v)$. Since $w' \in \mathcal{C}'$ and $w'' \in \mathcal{C}''$, then $\mathcal{C}^{\min}_{vxy} < w', w'' \leq \mathcal{C}^{\max}_{vxy}$. Since $w', w'' \in P(v, u)$, one is a descendant of the other. Assume without loss of generality that w'' is a descendant of w'. By Proposition 11, since $w' \in \mathcal{V}_{vx}$ it holds that $w' \notin \mathcal{V}_{vy}$. By the definition of $\mathcal{V}_{vy}$, since w'' is a descendant of w', we get that $w'' \notin \mathcal{V}_{vy}$, a contradiction. □

3.5 Smooth Simulation

Like in [BBCR10], it will be convenient for us to assume that the protocol to be simulated has "smooth" messages, in the sense that every bit in the protocol is relatively close to being unbiased, even conditioned on every fixing of the inputs and the prior transcript. We next argue that any protocol can be transformed into a smooth protocol without increasing the external information cost by much.

Definition 4 (**Smooth Protocol**). *Let $\beta > 0$. The protocol π is β-smooth if for every vertex $v \in \mathcal{V} \setminus \mathcal{L}$, every possible input (x, y) for π, and any $b \in \{0, 1\}$, it holds that*

$$O_{vxy}(b) \in [1/2 - \beta, 1/2 + \beta].$$

Lemma 13 (**Smooth Simulation**). *Let $\kappa, \beta > 0$. Let π' be a private coin protocol and μ be a distribution over the inputs for π'. Then, there exists a β-smooth private coin protocol π, and a transcript reconstruction function f that takes as an input a possible transcript of π, and returns a possible transcript of π', such that for every $(x, y) \in \text{supp}(\mu)$,*

$$\|f(\pi(x, y)) - \pi'(x, y)\| \leq \kappa.$$

In addition,

$$\mathsf{Ext}_\mu(\pi) \leq \mathsf{Ext}_\mu(\pi') + \kappa.$$

Proof. Let C be the worst-case communication complexity of π', that is, the maximal number of bits communicated by π' in any execution. Let $k = \left\lceil \log\left(\frac{C \cdot |\text{supp}(\mu)|}{\kappa}\right) / 2\beta^2 \right\rceil$. Consider the protocol π that operates as follows: Every time a player wants to send a bit b in π', he sends a sequence of bits $b_1, \ldots, b_k$ which are each independently chosen to be the correct value with probability $1/2 + \beta$. The players then proceed assuming that the majority of the bits $b_1, \ldots, b_k$ was the real transmission.

Fix $(x, y) \in \text{supp}(\mu)$. Let $i \in [C]$. By Hoeffding bound, the probability that the i^{th} transmission was received incorrectly is at most $\exp(-2\beta^2 k) < \frac{\kappa}{C \cdot |\text{supp}(\mu)|}$. Therefore, with probability at least $1 - \frac{\kappa}{|\text{supp}(\mu)|}$, all the transmissions were received correctly. Let f be the function that gets as an input a transcript of π and returns the string whose first bit is the majority of the first k bits of the given transcript, the string's second bit is the majority of the next k bits of the transcript, and so on. It holds that, $\Pr[f(\pi(x, y)) \neq \pi'(x, y)] \leq \frac{\kappa}{|\text{supp}(\mu)|}$, where the probability is over the randomness used by the players. In particular, $\|f(\pi(x, y)) - \pi'(x, y)\| \leq \kappa$.

As for the external information of π, observe that π may only reveal information about the inputs that was not revealed by π', if for some $i \in [C]$, the majority of the bits sent by π to simulate the i^{th} transition of π' is different than the actual i^{th} transition (as only in this case the players reach a different vertex of $\mathcal{T}$). However, even in this case, since the entropy of the distribution μ is at most $\log(|\text{supp}(\mu)|)$, no more than $\log(|\text{supp}(\mu)|)$ bits of information are revealed by π. Conclude that

$$\mathsf{Ext}_\mu(\pi) \leq \mathsf{Ext}_\mu(\pi') + \mathop{\mathbf{E}}_{(x,y)\leftarrow\mu}[\Pr[f(\pi(x,y)) \neq \pi'(x,y)] \cdot \log(|\text{supp}(\mu)|)] \leq \mathsf{Ext}_\mu(\pi') + \kappa.$$

□

3.6 Correlated Sampling

Our simulation protocol uses the correlated sampling protocol *CorrelatedSampling*, promised by the following lemma proved in [BR11]:

Lemma 14 (Theorems 2.1 and 4.1 in [BR11])**.** *Let $\eta > 0$. Suppose that Alice is given a distribution P and Bob is given a distribution Q, both over the same universe U (the distributions are described by the probabilities assigned to each point). There is a public coin protocol, CorrelatedSampling(P, Q, η), such that at the end of the protocol:*

- *Alice outputs an element a distributed according to P.*
- *Bob outputs b such that $\Pr[b \neq a] < \eta$.*
- *The expected communication complexity of the protocol is at most*

$$10\mathbb{D}(P\|Q) + 2\log(1/\eta) + 10,$$

where the expectation is over the randomness used by the players.

We mention that we only apply Lemma 14 for the special case where Alice knows both P and Q.

4 The Simulation Protocol

Let π be a β-smooth private coin communication protocol, for β to be specified next. Let μ be a distribution over the inputs for π. In this section we present the protocol τ^{ST} that simulates π over μ. We first present a related protocol τ (Section 4.1). The actual simulation protocol, τ^{ST}, is then easily obtained from τ (Section 4.2).

Parameters. Fix a proximity parameter $\varepsilon \in (0, 1)$. Let $I = \mathsf{Ext}_\mu(\pi)$. The protocols τ and τ^{ST} use the following parameters that depend on ε and I: Below $c \in \mathbb{N}$ is some sufficiently large constant.

$$\beta = \frac{\varepsilon}{\log^c(I)} = \frac{\varepsilon}{\text{polylog}(I)}; \quad t = O(1);$$

$$m = \frac{10^5 I}{\varepsilon^2 \beta} = I\text{poly}(\log(I)/\varepsilon); \quad T = \frac{10^{20} t I^2 \log^3(I)}{\varepsilon^5 \beta} = I^2\text{poly}(\log(I)/\varepsilon);$$

$$\eta = (\varepsilon/10T)^{10} = \text{poly}(\varepsilon/I); \quad r = 10^4 t m^2/\varepsilon^2 = \text{poly}(I/\varepsilon);$$

$$\alpha = (0.001\varepsilon/m)^2 = \text{poly}(\varepsilon/I).$$

4.1 The Protocol τ

The protocol τ is formally presented in Algorithms 1-4. Below is an informal description of the input and output of each of the (sub)protocols used by τ. For the rest of the paper we consider the tree $\mathcal{T}$ corresponding to π, and use the sets, functions, and distributions defined with respect to π in Section 3.

The protocol τ (Algorithm 1). At the beginning of the protocol it is assumed that Alice knows the input x and Bob knows the input y. The players' goal is to agree on a leaf of $\mathcal{T}$ distributed according to a distribution that is close (in statistical distance) to $\tilde{D}_{xy}$. The protocol starts at the root of $\mathcal{T}$ and proceeds in a sequence of m iterations. In every iteration a new vertex v is reached. For $i \in [m]$, the vertex v reached at the end of iteration i is a descendant of the vertex v reached at the end of iteration $i-1$. The vertex v reached by iteration m is a leaf with high probability.

Definitions. Consider the protocol τ (Algorithm 1). For $i \in \{0, \ldots, m\}$, let $\hat{\mathcal{V}}_i$ be the set of possible transcripts of the first i iterations of the loop in Line 1 of τ. Let $\hat{\mathcal{V}} = \bigcup_{i \in \{0,\ldots,m\}} \hat{\mathcal{V}}_i$. In particular, any $\hat{v} \in \hat{\mathcal{V}}$ is a (partial) transcript of τ.

Let $(x, y) \in \text{supp}(\mu)$. As will be justified later (see the last paragraph of Section 4.2), we may assume that after any transcript $\hat{v} \in \hat{\mathcal{V}}_i$ of the first i iterations of τ, both Alice and Bob (each using his private input) know the vertex $v \in \mathcal{V}$ reached by τ at the end of this iteration. Let $\hat{v}$ be a possible (partial) transcript of τ, and assume that $\hat{v}'$ is the longest prefix of $\hat{v}$ contained in $\hat{\mathcal{V}}$. We denote by $\text{msg}_{\pi x}(\hat{v}) \in \mathcal{V}$ the vertex v reached by the protocol τ when the transcript of τ is $\hat{v}'$ and Alice's input is x. We denote by $\text{msg}_{\pi y}(\hat{v}) \in \mathcal{V}$ the vertex v reached by the protocol τ when the (partial) transcript of τ is $\hat{v}'$ and Bob's input is y. Since we assume that for every $(x, y) \in \text{supp}(\mu)$ it is the case that $\text{msg}_{\pi x}(\hat{v}) = \text{msg}_{\pi y}(\hat{v})$

Algorithm 1: The Protocol $\tau(x, y)$

1 **for** $i \in [m]$ **do**

2 Players run $Chunk(\hat{v}, x, y)$, where $\hat{v}$ is the transcript of τ so far, to get a vertex v.

3 **return** v.

Algorithm 2: The Protocol $Chunk(\hat{v}, x, y)$

while *no output was given* **do**
  $p \leftarrow$ random bit sampled from the public randomness.
  if $p = 1$ **then**
    Players run $Sample^B(\hat{v}, x, y)$ to get an output w.
  else
    Players run $Sample^A(\hat{v}, x, y)$ to get an output w.
  if $w \neq failure$ **then**
    return w.

Algorithm 3: The Protocol $Sample^B(\hat{v}, x, y)$

$v \leftarrow \text{msg}_{\pi xy}(\hat{v})$.
if $v \in \mathcal{L}^+$ **then**
  Players return a leaf sampled according to $\tilde{D}_{vxy}$, by exchanging one bit for every level in $\mathcal{L}^+$.
Players run $CorrelatedSampling(P = \tilde{D}_{vy}, Q = \tilde{D}_v, \eta)$ to get a leaf u.
Players run $Separate^B(\hat{v}, u, x, y)$ to get a vertex w.
Alice samples and sends a bit a: If $w \notin \mathcal{V}_{vx}$, then $a = 0$. Otherwise, $a = 1$ with probability $\min\{1, D_{vx}(w)/tD_v(w)\}$.
return w if $a = 1$ and failure otherwise.

Algorithm 4: The Protocol $Separate^B(\hat{v}, u, x, y)$

$v \leftarrow \text{msg}_{\pi xy}(\hat{v})$.
Let $c_1 \leq \ldots \leq c_r \in \mathbb{N}$ be s.t. c_i is the first index satisfying $\Pr_{x' \leftarrow \mu^A_{\hat{v}}}[c_{vx'}(u) \leq c_i] \geq i/r$.
Alice sends the first index $a \in [r]$ such that $c_{vx}(u) \leq c_a$.
Alice sends a bit a' which is 1 iff $c_{vx}(u) = c_a$.
Bob sends the first index $b \in [r]$ such that $c_{vy}(u) \leq c_b$.
if $a \neq b$ **then**
  return vertex number $\min\{c_a, c_b\} + 1$ on the path from v to u.
else if $a' = 1$ **then**
  return vertex number c_a on the path from v to u.
else
  Alice sends $c_{vx}(u)$, Bob sends $c_{vy}(u)$.
  return vertex number $\max\{c_{vx}(u), c_{vy}(u)\}$ on the path from v to u.

(both Alice and Bob know the vertex $v \in \mathcal{V}$ reached), we often write $\mathrm{msg}_{\pi xy}(\hat{v})$ instead of $\mathrm{msg}_{\pi x}(\hat{v}), \mathrm{msg}_{\pi y}(\hat{v})$. We note that $\mathrm{msg}_{\pi xy}(\hat{v})$ is well defined: The vertex v reached by the protocol only depends on the inputs x, y and the transcript $\hat{v}$ (specifically, given $x, y, \hat{v}$, the vertex v is independent of the players' private randomness).

As before, (X, Y) is a pair of random variables distributed according to μ, representing the players' inputs. Recall that $\pi(X, Y)$ is the random variable representing the leaf of $\mathcal{T}$ reached by the protocol π. As explained in Section 3.3, we may assume without loss of generality that for every $(x, y) \in \mathrm{supp}(\mu)$ and $\hat{v} \in \hat{\mathcal{V}}$, the transcript of the protocol τ may be $\hat{v}$. That is, we assume that $\Pr[\tau(X, Y)_{\leq|\hat{v}|} = \hat{v} \mid X = x, Y = y] > 0$.

Let $\mu_{\hat{v}}$ be the distribution μ conditioned on the event that the transcript of τ is $\hat{v}$. That is, $\mu_{\hat{v}} = ((X, Y) \mid \tau(X, Y)_{\leq|\hat{v}|} = \hat{v})$. Let $\mu^A_{\hat{v}}$ and $\mu^B_{\hat{v}}$ be the distributions μ^A and μ^B (respectively), conditioned on the event that the transcript of τ is $\hat{v}$. That is, $\mu^A_{\hat{v}} = (X \mid \tau(X, Y)_{\leq|\hat{v}|} = \hat{v})$ and $\mu^B_{\hat{v}} = (Y \mid \tau(X, Y)_{\leq|\hat{v}|} = \hat{v})$. Since $\hat{v}$ is a possible transcript of a protocol, and since $\mu = \mu^A \times \mu^B$ is a product distribution, $\mu_{\hat{v}}$ is also a product distribution and $\mu_{\hat{v}} = \mu^A_{\hat{v}} \times \mu^B_{\hat{v}}$. Observe that given $\hat{v}$, both players can compute $\mu_{\hat{v}}, \mu^A_{\hat{v}}, \mu^B_{\hat{v}}$.

The Protocol *Chunk* (Algorithm 2). At the beginning of the protocol it is assumed that Alice knows the input x, Bob knows the input y, and that both players know the transcript $\hat{v}$ of τ so far. Let $v = \mathrm{msg}_{\pi xy}(\hat{v})$. The players' goal is to agree on a vertex $w \in \mathcal{V}(v)$ satisfying $\mathbb{D}_{vwxy} \geq \beta$ (if such w exists), while the following properties hold: The set $\mathcal{C}_{\hat{v}xy}$ of all the vertices w returned by the protocol is a frontier. In addition, every vertex $w \in \mathcal{C}_{\hat{v}xy}$ is selected by the players with probability close to $D_{vxy}(w)$.

The Protocol $Sample^B$ (Algorithm 3). At the beginning of the protocol it is assumed that Alice knows the input x, Bob knows the input y, and that both players know the transcript of τ up until the beginning of most recent execution of *Chunk*. This transcript is denoted $\hat{v} \in \hat{\mathcal{V}}$. Let $v = \mathrm{msg}_{\pi xy}(\hat{v})$. The players' goal is to agree on a vertex $w \in \mathcal{V}(v)$ such that $w \in \mathcal{V}_{vx} \setminus \mathcal{V}_{vy}$ (in particular, $\mathbb{D}_{vwy} \geq \beta$ and $\mathbb{D}_{vwx} \leq 2\beta$, unless w is a leaf), while the following properties hold: The set of all the vertices w returned by the protocol is a subset of a frontier. This subset is "maximal" in the sense that one cannot add a new vertex $w \in \mathcal{V}_{vx} \setminus \mathcal{V}_{vy}$ to the set, while keeping it a subset of some frontier. In addition, every vertex w in this subset is selected by the players with probability close to $D_{vxy}(w)$. The protocol may fail.

The Protocol $Sample^A$. The protocol $Sample^A$ is obtained from the protocol $Sample^B$ by switching the roles Alice and Bob, switching the roles of x and y, and running $Separate^A$ instead of $Separate^B$.

The Protocol $Separate^B$ (Algorithm 4). At the beginning of the protocol it is assumed that Alice knows the input x, Bob knows the input y, and that both players know the

transcript of τ up until the beginning of most recent execution of *Chunk*. This transcript is denoted $\hat{v} \in \hat{\mathcal{V}}$. It is also assumed that the players agree on a leaf $u \in \mathcal{L}(v)$, where $v = \mathrm{msg}_{\pi xy}(\hat{v})$. The players' goal is to agree on a vertex $w \in P(v,u)$ such that $w \in (\mathcal{V}_{vx} \setminus \mathcal{V}_{vy}) \cup (\mathcal{V}_{vy} \setminus \mathcal{V}_{vx})$.

The Protocol $Separate^A$. The protocol $Separate^A$ is obtained from the protocol $Separate^B$ by switching the roles Alice and Bob, switching the roles of x and y, and using $\mu^B_{\hat{v}}$ instead of $\mu^A_{\hat{v}}$.

4.2 The Protocol τ^{ST}

The protocol τ^{ST} gets the same parameters as τ. It operates the same as τ, expect for the following changes:

1. Add the following line at the beginning of the protocol: "At any point in the execution of this protocol, after the players exchange T bits, τ^{ST} terminates and returns failure".
2. Lines 11 and 12 of the protocols $Separate^B$ and $Separate^A$ are replaced by the single line "τ^{ST} returns failure". That is, if the "else" part in Line 10 of either $Separate^B$ (Algorithm 3) or $Separate^A$ is reached, then τ^{ST} terminates immediately and returns failure.

We mention that the protocol *CorrelatedSampling* is run by $Sample^B$ and $Sample^A$ with error parameter η, in order to sample a leaf $u \in \mathcal{L}(v)$. By Lemma 14, such an execution may result in the players getting two different leaves u^A and u^B with probability η. We ignore this possibility and assume that Alice and Bob always sample the same leaf u. Since *CorrelatedSampling* is run at most T times by τ^{ST}, the probability that Alice and Bob sample a different leaf u in one of the executions of *CorrelatedSampling*, is at most $T\eta$. We select η to be significantly smaller than $1/T$, thus $T\eta$ is negligible.

5 Protocol Analysis

In this section we prove Theorem 1. We restate Theorem 1 with slight change of notation, in order to simplify the notation in the proof (the original protocol is called here π', and the simulation protocol is τ^{ST}).

Theorem (Theorem 1 restated). *Let $\varepsilon > 0$. Let π' be a randomized protocol that may use private and public coins. Let $\mu = \mu^A \times \mu^B$ be a product distribution over the inputs for π'. Then, there exist a public coin protocol τ^{ST} (that takes the same inputs as π'), and a pair of "transcript reconstruction" functions g^A, g^B such that: The function g^A takes as inputs $x \in \mathrm{supp}(\mu^A)$ and a possible transcript of τ^{ST}, and returns a possible transcript of π'. The function g^B takes as inputs $y \in \mathrm{supp}(\mu^B)$ and a possible transcript of τ^{ST}, and returns a possible transcript of π'. In addition, the followings hold:*

1. *The worst case communication complexity of* τ *is* $\mathsf{IC}_\mu^2(\pi') \cdot \text{polylog}\left(\mathsf{IC}_\mu(\pi')\right)/\varepsilon^5$.

2. $\forall (x,y) \in \text{supp}(\mu):\ \Pr\left[g^A(x,\tau^{ST}(x,y)) \neq g^B(y,\tau^{ST}(x,y))\right] \leq \varepsilon$, *where the probability is over the random coins of the protocol* τ^{ST}.

3. $\mathbf{E}_{(x,y)\leftarrow\mu}\left[\left\|g^A(x,\tau^{ST}(x,y)) - \pi'(x,y)\right\|\right] \leq \varepsilon$.

Let π' be the given randomized communication protocol, and let μ be a distribution over the inputs for π'. As explained before, we may assume that π' is a private coin protocol, as the public randomness can always be replaced by private randomness without increasing the information cost. Let $I' = \mathsf{Ext}_\mu(\pi')$, and assume that I' is sufficiently large. Let π be the β'-smooth protocol promised by Lemma 13 for π', when applied with $\kappa = \varepsilon/10$ and $\beta' = \frac{\varepsilon}{\log^c(I'+\varepsilon/10)}$ (the same constant c as in the definition of the parameter β). Let $I = \mathsf{Ext}_\mu(\pi)$. By Lemma 13, $I \leq I' + \varepsilon/10$. Thus, $\beta' \leq \beta = \frac{\varepsilon}{\log^c(I)}$, and, in particular, π is also β-smooth. The simulation protocol, τ^{ST}, is used to simulate the protocol π. This section is devoted to proving the following main lemma:

Lemma 15. *It holds that*

$$\mathop{\mathbf{E}}_{(x,y)\leftarrow\mu}\left[\left\|\text{msg}_{\pi x}(\tau^{ST}(x,y)) - \pi(x,y)\right\|\right] \leq \varepsilon/2.$$

The proof of Lemma 15 is deferred to the end of the section. We next show how Theorem 1 follows from Lemma 15.

Proof of Theorem 1. The first item in Theorem 1 is satisfied as the worst case communication complexity of π^{ST} is $T \leq I^2\text{polylog}(I)/\varepsilon^5 \leq (I')^2\text{polylog}(I')/\varepsilon^5$, and as, by Fact 3, $I' = \mathsf{Ext}_\mu(\pi') = \mathsf{IC}_\mu(\pi')$.

Let f^A be the function that gets an input $x \in \text{supp}(\mu^A)$ and a possible transcript $\hat{v}$ for τ^{ST}, and returns the transcript $\text{msg}_{\pi x}(\hat{v}) \in \mathcal{V}$ for π. Let f^B be the function that gets an input $y \in \text{supp}(\mu^B)$ and a possible transcript $\hat{v}$ for τ^{ST}, and returns the transcript $\text{msg}_{\pi y}(\hat{v}) \in \mathcal{V}$ for π. By Lemma 13, there exists a function f such that $\|f(\pi(x,y)) - \pi'(x,y)\| \leq \varepsilon/10$. Let $g^A = f \circ f^A$ and $g^B = f \circ f^B$, where '$\circ$' denotes function composition.

The second item in Theorem 1 is satisfied, because, as explained in the last paragraph of Section 4.2, Alice and Bob reach a different vertex of $\mathcal{T}$ with probability at most $T\eta \ll \varepsilon$.

The third item in Theorem 1 is satisfied as by Lemma 15,

$$\begin{aligned}
&\mathop{\mathbf{E}}_{(x,y)\leftarrow\mu}\left[\left\|g^A(x,\tau^{ST}(x,y)) - \pi'(x,y)\right\|\right] \\
&\leq \mathop{\mathbf{E}}_{(x,y)\leftarrow\mu}\left[\left\|f(f^A(x,\tau^{ST}(x,y))) - f(\pi(x,y))\right\| + \|f(\pi(x,y)) - \pi'(x,y)\|\right] \\
&\leq \mathop{\mathbf{E}}_{(x,y)\leftarrow\mu}\left[\left\|\text{msg}_{\pi x}(\tau^{ST}(x,y)) - \pi(x,y)\right\|\right] + \mathop{\mathbf{E}}_{(x,y)\leftarrow\mu}\left[\|f(\pi(x,y)) - \pi'(x,y)\|\right] \\
&\leq \varepsilon/2 + \varepsilon/10 \leq \varepsilon.
\end{aligned}$$

□

5.1 The Protocols $Separate^B$ and $Separate^A$

The frontiers $\mathcal{C}^B_{\hat{v}xy}$ and $\mathcal{C}^A_{\hat{v}xy}$. Let $\hat{v} \in \hat{\mathcal{V}}$ and $(x,y) \in \text{supp}(\mu)$. Let $v = \text{msg}_{\pi xy}(\hat{v})$ and assume that $v \in \mathcal{V} \setminus \mathcal{L}^+$. Let $\mathcal{C}^B_{\hat{v}xy}$ be the set of all possible vertices $w \in \mathcal{V}$ that may be returned when running $Separate^B(\hat{v}, u, x, y)$, for some $u \in \mathcal{L}(v)$. Let $\mathcal{C}^A_{\hat{v}xy}$ be the set of all possible vertices $w \in \mathcal{V}$ that may be returned when running $Separate^A(\hat{v}, u, x, y)$, for some $u \in \mathcal{L}(v)$.

Lemma 16. *Let $\hat{v} \in \hat{\mathcal{V}}$ and $(x,y) \in \text{supp}(\mu)$. Let $v = \text{msg}_{\pi xy}(\hat{v})$ and assume that $v \in \mathcal{V} \setminus \mathcal{L}^+$. Then, $\mathcal{C}^B_{\hat{v}xy}$ and $\mathcal{C}^A_{\hat{v}xy}$ are frontiers with respect to v.*

Proof. We show that $\mathcal{C}^B_{\hat{v}xy}$ is a frontier with respect to v. A similar argument can be applied to $\mathcal{C}^A_{\hat{v}xy}$. Let $u \in \mathcal{L}(v)$. Observe that $Separate^B(\hat{v}, u, x, y)$ always returns a vertex in $P(v,u)$. Thus, in particular, $\mathcal{C}^B_{\hat{v}xy} \subseteq \mathcal{V}(v)$ and $|\mathcal{C}^B_{\hat{v}xy} \cap P(v,u)| \geq 1$. It remains to show that $|\mathcal{C}^B_{\hat{v}xy} \cap P(v,u)| \leq 1$.

Let $c(u)$ be the index (number) of the vertex that $Separate^B(\hat{v}, u, x, y)$ returns on the path from v to u. Let $u' \in \mathcal{L}(v)$. Let $c(u')$ be the index (number) of the vertex that $Separate^B(\hat{v}, u', x, y)$ returns on the path from v to u'. Let $k \in \mathbb{N}$ be such that the path from v to u and the path from v to u' agree on the first k vertices and disagree on the $(k+1)^{th}$ vertex (if $u = u'$ then $k = |P(v,u)|$). We next show that if $c(u) \leq k$, then $c(u) = c(u')$. Since the roles of u and u' are symmetric, we also get that if $c(u') \leq k$, then $c(u) = c(u')$. This in turn implies that $|\mathcal{C}^B_{\hat{v}xy} \cap P(v,u)| \leq 1$, as follows: Let w be the vertex returned by the execution of $Separate^B(\hat{v}, u, x, y)$, and let w' be the vertex returned by the execution of $Separate^B(\hat{v}, u', x, y)$. If $c(u') > k$ then $\{w'\} \cap P(v,u) = \phi$. Otherwise $c(u') \leq k$, thus $c(u) = c(u')$, implying $w = w'$.

Let $c_a(u), c_b(u)$ be the c_a, c_b indices used when running $Separate^B(\hat{v}, u, x, y)$. Similarly, let $c_a(u'), c_b(u')$ be the c_a, c_b indices used when running $Separate^B(\hat{v}, u', x, y)$. Observe that if $c_{vx}(u) \leq k$ or $c_{vx}(u') \leq k$ (see the definition of c_{vx} in Section 3.4), then $c_{vx}(u') = c_{vx}(u)$, as $\mathcal{C}_{vx}$ is a frontier. Similarly, if $c_{vy}(u) \leq k$ or $c_{vy}(u') \leq k$, then $c_{vy}(u') = c_{vy}(u)$. In particular, if $c_a(u) \leq k$ or $c_a(u') \leq k$, then $c_a(u') = c_a(u)$. Similarly, if $c_b(u) \leq k$ or $c_b(u') \leq k$, then $c_b(u') = c_b(u)$.

Recall that we assume that $c(u) \leq k$, and want to show that $c(u) = c(u')$. We consider the following cases:

Case 1: $c_a(u) > c_b(u)$ (thus $c(u) = c_b(u) + 1$). Since we assume that $c_b(u) < c(u) \leq k$, then $c_b(u') = c_b(u) \leq k$.

We next show that $c_a(u') > c_b(u')$. Assume for contradiction that $c_a(u') \leq c_b(u')$. Then, since $c_a(u') \leq c_b(u') \leq k$, it holds that $c_a(u') = c_a(u)$. But then, $c_a(u) = c_a(u') \leq c_b(u') = c_b(u)$, a contradiction to the assumption of this case that $c_a(u) > c_b(u)$.

Since $c_a(u') > c_b(u')$, it holds that $c(u') = c_b(u') + 1 = c_b(u) + 1 = c(u)$.

Case 2: $c_b(u) > c_a(u)$ (thus $c(u) = c_a(u) + 1$). Proof is similar to Case 1.

Case 3: $c_{vx}(u) = c_a(u) = c_b(u)$ **(thus** $c(u) = c_a(u)$**).** Since $c_b(u) = c_a(u) = c(u) \leq k$, then $c_a(u') = c_a(u) = c_b(u) = c_b(u')$. Since $c_{vx}(u) = c_a(u) \leq k$, it holds that $c_{vx}(u) = c_{vx}(u')$. Conclude that $c_{vx}(u') = c_{vx}(u) = c_a(u) = c_a(u')$. Because $c_{vx}(u') = c_a(u') = c_b(u')$, we get $c(u') = c_a(u') = c_a(u) = c(u)$.

Case 4: $c_a(u) = c_b(u)$ **and** $c_{vx}(u) < c_a(u)$ **(thus** $c(u) = \max\{c_{vx}(u), c_{vy}(u)\}$**).** Since $c_{vx}(u), c_{vy}(u) \leq c(u) \leq k$, then $c_{vy}(u') = c_{vy}(u)$ and $c_{vx}(u') = c_{vx}(u) < c_a(u)$.

We next show that $c_{vx}(u') < c_a(u')$. If $c_a(u') \leq k$, then $c_a(u) = c_a(u')$ and $c_{vx}(u') < c_a(u) = c_a(u')$. Otherwise, $c_a(u') > k$, and since $c_{vx}(u') = c_{vx}(u) \leq k$, we get that $c_{vx}(u') < c_a(u')$.

We next show that $c_a(u') = c_b(u')$. If $c_a(u') \leq k$ then $c_a(u') = c_a(u) = c_b(u) \leq k$, thus also $c_b(u) = c_b(u')$, and we get $c_a(u') = c_b(u')$. Similarly, if $c_b(u') \leq k$, we get $c_a(u') = c_b(u')$. Consider the case where $c_a(u'), c_b(u') > k$. Since $c_{vx}(u), c_{vy}(u) \leq k < c_a(u'), c_b(u')$, and as the a, b indices used by $Separate^B(\hat{v}, u', x, y)$ are the first indices for which $c_{vx}(u') \leq c_a(u')$ and $c_{vy}(u') \leq c_b(u')$ (respectively), it holds that $c_a(u') = c_b(u')$.

Because $c_a(u') = c_b(u')$ and $c_{vx}(u') < c_a(u')$, we get that $c(u') = \max\{c_{vx}(u'), c_{vy}(u')\} = \max\{c_{vx}(u), c_{vy}(u)\} = c(u)$.

□

Lemma 17. *Let* $\hat{v} \in \hat{\mathcal{V}}$ *and* $(x, y) \in \text{supp}(\mu)$*. Let* $v = \text{msg}_{\pi xy}(\hat{v})$ *and assume that* $v \in \mathcal{V} \backslash \mathcal{L}^+$*. Then,*

$$\mathcal{C}^{\min}_{vxy} < \mathcal{C}^B_{\hat{v}xy}, \mathcal{C}^A_{\hat{v}xy} \leq \mathcal{C}^{\max}_{vxy}.$$

Proof. We show that $\mathcal{C}^{\min}_{vxy} < \mathcal{C}^B_{\hat{v}xy} \leq \mathcal{C}^{\max}_{vxy}$. A similar argument can be applied to $\mathcal{C}^A_{\hat{v}xy}$. Let $u \in \mathcal{L}(v)$. Let w be the vertex returned by $Separate^B(\hat{v}, u, x, y)$. We need to show that $\mathcal{C}^{\min}_{vxy} < w \leq \mathcal{C}^{\max}_{vxy}$.

Let w_y be the $c_{vy}(u)$ vertex on the path from v to u, and let w_x be the $c_{vx}(u)$ vertex on the path from v to u. By definition, $w_y \in \mathcal{C}_{vy}$ and $w_x \in \mathcal{C}_{vx}$. Observe that if w_x is a descendant of w_y, then $w_y \in \mathcal{C}^{\min}_{vxy}$ (as w_y has a descendant w_x in $\mathcal{C}_{vx}$), and $w_x \in \mathcal{C}^{\max}_{vxy}$ (as w_x has an ancestor w_y in $\mathcal{C}_{vy}$).

Recall that $Separate^B(\hat{v}, u, x, y)$ always returns a vertex in $P(v, u)$. Let $c(u)$ be the index (number) of w on the path from v to u. Let $c_a(u), c_b(u)$ be the c_a, c_b indices used when running $Separate^B(\hat{v}, u, x, y)$. We consider the following cases:

Case 1: $c_a(u) > c_b(u)$ **(thus** $c(u) = c_b(u) + 1$**).** Observe that $c_{vx}(u) \geq c_b(u) + 1$, as if $c_{vx}(u) \leq c_b(u)$, then by the definitions of c_a, c_b, we have $c_a(u) \leq c_b(u)$, a contradiction. Since $c_a(u) > c_b(u)$, we get that w_x is a descendant w_y. Thus, $w_y \in \mathcal{C}^{\min}_{vxy}$ and $w_x \in \mathcal{C}^{\max}_{vxy}$. It holds that $\mathcal{C}^{\min}_{vxy} < w \leq \mathcal{C}^{\max}_{vxy}$, as w_y is a strict ancestor of w (because $c(u) > c_b(u) \geq c_{vy}(u)$) and w_x is a descendant of w (because $c(u) = c_b(u) + 1 \leq c_{vx}(u)$).

Case 2: $c_b(u) > c_a(u)$ **(thus** $c(u) = c_a(u) + 1$**).** Proof is similar to Case 1.

Case 3: $c_{vx}(u) = c_a(u) = c_b(u)$ **(thus** $c(u) = c_a(u)$**).** By Equation (3), $\mathcal{C}_{vx} \cap \mathcal{C}_{vy} = \phi$. Therefore, $c_{vx}(u) \neq c_{vy}(u)$. Since $c_{vx}(u) = c_a(u)$, and $c_{vy}(u) \neq c_{vx}(u) = c_a(u) = c_b(u)$, and by the definitions of c_a, c_b, it must be the case that $c_{vx}(u) > c_{vy}(u)$. Conclude that w_x is a strict descendant of w_y. Thus, $w_y \in \mathcal{C}^{\min}_{vxy}$ and $w_x \in \mathcal{C}^{\max}_{vxy}$. Since $c(u) = c_a(u) = c_{vx}(u)$, we get $w = w_x$. It holds that $\mathcal{C}^{\min}_{vxy} < w \leq \mathcal{C}^{\max}_{vxy}$, as w_y is a strict ancestor of $w_x = w$, and as $w = w_x$ (thus, in particular, w_x is a descendant of w).

Case 4: $c_a(u) = c_b(u)$ **and** $c_{vx}(u) < c_a(u)$ **(thus** $c(u) = \max\{c_{vx}(u), c_{vy}(u)\}$**).** Assume without loss of generality that $c_{vx}(u) > c_{vy}(u)$ (otherwise, since $c_{vy}(u) \neq c_{vx}(u)$, we have $c_{vy}(u) > c_{vx}(u)$, and the argument is symmetric). In this case, w_x is a strict descendant of w_y. Thus, $w_y \in \mathcal{C}^{\min}_{vxy}$ and $w_x \in \mathcal{C}^{\max}_{vxy}$. Since $c(u) = \max\{c_{vx}(u), c_{vy}(u)\} = c_{vx}(u)$, we get $w = w_x$. It holds that $\mathcal{C}^{\min}_{vxy} < w \leq \mathcal{C}^{\max}_{vxy}$, as w_y is a strict ancestor of $w_x = w$, and as $w = w_x$ (thus, in particular, w_x is a descendant of w). □

Let $\hat{v} \in \hat{\mathcal{V}}$, $(x, y) \in \text{supp}(\mu)$, $v = \text{msg}_{\pi xy}(\hat{v})$, and $u \in \mathcal{L}(v)$. Let $F^B(\hat{v}, u, x, y) \in \{0, 1\}$ be the value 1 if and only if when running the protocol $Separate^B$ (Algorithm 3) with the parameters $\hat{v}, u, x, y$, the "else" part in Line 10 is reached (the "bad event"). Note that since the protocol $Separate^B$ is deterministic, $F^B(\hat{v}, u, x, y)$ only depends on $\hat{v}, u, x, y$. The value $F^A(\hat{v}, u, x, y) \in \{0, 1\}$ is defined similarly.

Lemma 18. *Let* $\hat{v} \in \hat{\mathcal{V}}$, $x \in \text{supp}(\mu^A)$, $y \in \text{supp}(\mu^B)$, *and* $u \in \mathcal{L}(v)$. *Then,*

$$\Pr_{x \leftarrow \mu^A_{\hat{v}}}\left[F^B(\hat{v}, u, x, y) = 1\right] < 1/r,$$
$$\Pr_{y \leftarrow \mu^B_{\hat{v}}}\left[F^A(\hat{v}, u, x, y) = 1\right] < 1/r.$$

Proof. We will prove the first inequality, the second is similar. Fix $\hat{v} \in \hat{\mathcal{V}}$, $y \in \text{supp}(\mu^B)$, and $u \in \mathcal{L}(v)$. Let $v = \text{msg}_{\pi y}(\hat{v})$. For $x \in \text{supp}(\mu^A)$, let c_i, a_x, a'_x, b be values c_i, a, a', b computed by the execution of $Separate^B(\hat{v}, u, x, y)$. Let $c_0 = 0$. Let $i \in [r]$. Recall that c_i is the first index satisfying $\Pr_{x' \leftarrow \mu^A_{\hat{v}}}[c_{vx'}(u) \leq c_i] \geq i/r$. Therefore, $\Pr_{x' \leftarrow \mu^A_{\hat{v}}}[c_{vx'}(u) < c_i] < i/r$. Since $\Pr_{x' \leftarrow \mu^A_{\hat{v}}}[c_{vx'}(u) \leq c_{i-1}] \geq (i-1)/r$, it holds that

$$\Pr_{x' \leftarrow \mu^A_{\hat{v}}}[c_{i-1} < c_{vx'}(u) < c_i] < 1/r.$$

For $x \in \text{supp}(\mu^A)$, it is the case that $F^B(\hat{v}, u, x, y) = 1$ if and only if $a_x = b$ and $a'_x = 0$. We get

$$\begin{aligned}\Pr_{x \leftarrow \mu^A_{\hat{v}}}\left[F^B(\hat{v}, u, x, y) = 1\right] &= \Pr_{x \leftarrow \mu^A_{\hat{v}}}[(a_x = b) \wedge (a'_x = 0)] \\ &= \Pr_{x \leftarrow \mu^A_{\hat{v}}}[(a_x = b) \wedge (c_{vx}(u) < c_{a_x})] \leq \Pr_{x \leftarrow \mu^A_{\hat{v}}}[c_{b-1} < c_{vx}(u) < c_b] < 1/r.\end{aligned}$$

□

5.2 The Protocols $Sample^B$ and $Sample^A$

Let $\hat{v} \in \hat{\mathcal{V}}$ and $(x,y) \in \mathrm{supp}(\mu)$. Let $v = \mathrm{msg}_{\pi xy}(\hat{v})$. We define the set $\mathcal{C}_{\hat{v}xy}$: If $v \in \mathcal{L}^+$, define $\mathcal{C}_{\hat{v}xy} = \mathcal{L}(v)$. If $v \in \mathcal{V} \setminus \mathcal{L}^+$, define

$$\mathcal{C}_{\hat{v}xy} = \left(\mathcal{C}^B_{\hat{v}xy} \cap \mathcal{V}_{vx}\right) \cup \left(\mathcal{C}^A_{\hat{v}xy} \cap \mathcal{V}_{vy}\right). \tag{4}$$

For the rest of this subsection we assume that $v \in \mathcal{V} \setminus \mathcal{L}^+$. By Lemmas 16,17 and Proposition 12, $\mathcal{C}_{\hat{v}xy}$ is a frontier with respect to v and

$$\left(\mathcal{C}^B_{\hat{v}xy} \cap \mathcal{V}_{vx}\right) \cap \left(\mathcal{C}^A_{\hat{v}xy} \cap \mathcal{V}_{vy}\right) = \phi. \tag{5}$$

By Lemma 17, $\mathcal{C}^{\min}_{vxy} < \mathcal{C}_{\hat{v}xy} \leq \mathcal{C}^{\max}_{vxy}$.

Let $D_{\hat{v}xy}$ be the distribution obtained by restricting the function D_{vxy} to the frontier $\mathcal{C}_{\hat{v}xy}$. Denote

$$\gamma_{\hat{v}xy} = D_{\hat{v}xy}\left(\mathcal{C}^B_{\hat{v}xy} \cap \mathcal{V}_{vx}\right).$$

By Equation (5),

$$1 - \gamma_{\hat{v}xy} = D_{\hat{v}xy}\left(\mathcal{C}^A_{\hat{v}xy} \cap \mathcal{V}_{vy}\right).$$

Assume $\gamma_{\hat{v}xy} > 0$. Let $D^B_{\hat{v}xy}$ be the distribution obtained by conditioning the distribution $D_{\hat{v}xy}$ on the set $\mathcal{C}^B_{\hat{v}xy} \cap \mathcal{V}_{vx}$. That is, for $w \in \mathcal{C}^B_{\hat{v}xy} \cap \mathcal{V}_{vx}$ we have

$$D^B_{\hat{v}xy}(w) = \frac{D_{\hat{v}xy}(w)}{\gamma_{\hat{v}xy}} = \frac{D_{vxy}(w)}{\gamma_{\hat{v}xy}}.$$

Assume $\gamma_{\hat{v}xy} < 1$. Let $D^A_{\hat{v}xy}$ be the distribution obtained by conditioning the distribution $D_{\hat{v}xy}$ on the set $\mathcal{C}^A_{\hat{v}xy} \cap \mathcal{V}_{vy}$. That is, for $w \in \mathcal{C}^A_{\hat{v}xy} \cap \mathcal{V}_{vy}$ we have

$$D^A_{\hat{v}xy}(w) = \frac{D_{\hat{v}xy}(w)}{1 - \gamma_{\hat{v}xy}} = \frac{D_{vxy}(w)}{1 - \gamma_{\hat{v}xy}}.$$

We use the following version of a claim proved by [BBCR10]. The proof of this variant is very similar to their proof. For completeness, we give the proof in Appendix A.

Lemma 19 (Claim 8.9 in [BBCR10])**.** *Let $\hat{v} \in \hat{\mathcal{V}}$ and $(x,y) \in \mathrm{supp}(\mu)$. Let $v = \mathrm{msg}_{\pi xy}(\hat{v})$ and assume that $v \in \mathcal{V} \setminus \mathcal{L}^+$. Then, for*

$$\alpha = \exp\left(-\frac{(\log(t) - 20\beta)^2}{10^5\beta}\right),$$

the followings hold: If $\gamma_{\hat{v}xy} > 0$, then

$$\Pr_{w \leftarrow D^B_{\hat{v}xy}}\left[\frac{D_{vx}(w)}{D_v(w)} \geq t\right] \leq \frac{\alpha}{\gamma_{\hat{v}xy}}.$$

If $\gamma_{\hat{v}xy} < 1$*, then*

$$\Pr_{w \leftarrow D^A_{\hat{v}xy}}\left[\frac{D_{vy}(w)}{D_v(w)} \geq t\right] \leq \frac{\alpha}{1-\gamma_{\hat{v}xy}}.$$

We set the parameters β and t such that $\beta = \frac{\varepsilon}{\text{polylog}(I)}$, $t = O(1)$, and such that the α value promised by Lemma 19 satisfies $\alpha \leq (0.001\varepsilon/m)^2$.

Let $\hat{v} \in \hat{\mathcal{V}}$ and $(x,y) \in \text{supp}(\mu)$. Let $G^B_{\hat{v}xy}$ be the event that the execution of $Sample^B(\hat{v},x,y)$ does not return failure (the "good event"). Let $G^A_{\hat{v}xy}$ be the event that the execution of $Sample^A(\hat{v},x,y)$ does not return failure (the "good event").

Lemma 20. *Let* $\hat{v} \in \hat{\mathcal{V}}$ *and* $(x,y) \in \text{supp}(\mu)$*. Let* $v = \text{msg}_{\pi xy}(\hat{v})$ *and assume that* $v \in \mathcal{V} \setminus \mathcal{L}^+$*. Then, the execution of* $Sample^B(\hat{v},x,y)$ *returns either a vertex in* $\mathcal{C}^B_{\hat{v}xy} \cap \mathcal{V}_{vx}$ *or failure. The execution of* $Sample^A(\hat{v},x,y)$ *returns either a vertex in* $\mathcal{C}^A_{\hat{v}xy} \cap \mathcal{V}_{vy}$ *or failure.*

Assume $\gamma_{\hat{v}xy} \geq 10\alpha$*. Then, for every* $w \in \mathcal{C}^B_{\hat{v}xy} \cap \mathcal{V}_{vx}$*,*

$$\Pr\left[Sample^B(\hat{v},x,y) = w \mid G^B_{\hat{v}xy}\right] \leq (1 + 2\alpha/\gamma_{\hat{v}xy}) \cdot D^B_{\hat{v}xy}(w),$$

Assume $1 - \gamma_{\hat{v}xy} \geq 10\alpha$*. Then, for every* $w \in \mathcal{C}^A_{\hat{v}xy} \cap \mathcal{V}_{vy}$*,*

$$\Pr\left[Sample^A(\hat{v},x,y) = w \mid G^A_{\hat{v}xy}\right] \leq (1 + 2\alpha/(1-\gamma_{\hat{v}xy})) \cdot D^A_{\hat{v}xy}(w).$$

The probabilities are over the randomness used by the players.

Proof. We prove the claims for $Sample^B$. A similar argument can be applied to prove the claims for $Sample^A$. The execution of $Separate^B(\hat{v},u,x,y)$ by Line 5 of $Sample^B$ returns a vertex $w \in \mathcal{C}^B_{\hat{v}xy}$. If $w \notin \mathcal{V}_{vx}$, the vertex w is rejected by Alice. Thus, the returned value is either a vertex in $\mathcal{C}^B_{\hat{v}xy} \cap \mathcal{V}_{vx}$ or failure.

We use the following version of a claim proved by [BBCR10].

Claim 21 (Proposition B.3 in [BBCR10])**.** *Let* $\mathcal{C}$ *be a set and let* D *be a distribution over* $\mathcal{C}$*. Let* $c : \mathcal{C} \to [0,1]$*. Let* ξ *be the protocol:*

- *Sample an element* w *according to* D*.*
- *Return* w *with probability* $c(w)$*, else return failure.*

Define the distribution D_c *over* $\mathcal{C}$ *by*

$$D_c(w) = \Pr\left[\xi \text{ returns } w \mid \xi \text{ does not return failure}\right].$$

Let $c' : \mathcal{C} \to [0,1]$ *be such that* $c'(w) \geq c(w)$ *for every* $w \in \mathcal{C}$*. Let* $p_{c'} \geq 1$ *be such that* $D_{c'} = p_{c'} \cdot D \cdot c'$ *is a distribution. Then, if* $\Pr_{w \leftarrow D_{c'}}[c'(w) > c(w)] < 1$ *then for every* $w \in \mathcal{C}$*,*

$$D_c(w) \leq \frac{1}{1 - \Pr_{w \leftarrow D_{c'}}[c'(w) > c(w)]} \cdot D_{c'}(w).$$

Proof. There exists $p_c \geq 1$ such that $D_c = p_c \cdot D \cdot c$. Let $b : \mathcal{C} \to [0,1]$ be given by $b = c' - c$, and let $p_b \geq 1$ be such that $D_b = p_b \cdot D \cdot b$ is a distribution. Then, there exists $q \in [0,1]$ such that $D_{c'} = q \cdot D_b + (1-q) \cdot D_c$. In particular, $D_{c'} \geq (1-q) \cdot D_c$, and thus for every $w \in \mathcal{C}$, we have $D_c(w) \leq D_{c'}(w)/(1-q)$. We next show that $q \leq \Pr_{w \leftarrow D_{c'}}[c'(w) > c(w)]$, and the assertion will follow.

The term $q \cdot D_b$ in the expansions of $D_{c'}$ shows that at least q of the weight of the distribution $D_{c'}$ is on elements in $\mathrm{supp}(D_b)$. That is, $q \leq \Pr_{w \leftarrow D_{c'}}[w \in \mathrm{supp}(D_b)]$. Since $\mathrm{supp}(D_b) \subseteq \{w \in \mathcal{C} : \ c'(w) > c(w)\}$, it holds that $q \leq \Pr_{w \leftarrow D_{c'}}[w \in \mathrm{supp}(D_b)] \leq \Pr_{w \leftarrow D_{c'}}[c'(w) > c(w)]$. □

Let D be the distribution obtained by restricting the function D_{vy} to the set $\mathcal{C}^B_{\hat{v}xy} \cap \mathcal{V}_{vx}$ and normalizing. Let $w \in \mathcal{C}^B_{\hat{v}xy} \cap \mathcal{V}_{vx}$. Let

$$c(w) = \min\left\{1, \frac{D_{vx}(w)}{tD_v(w)}\right\}; \qquad c'(w) = \frac{D_{vx}(w)}{tD_v(w)}.$$

Let $D_c, D_{c'}$ be the two distributions defined by Claim 21 when it is applied to D, c, c'. Observe that

$$D_c(w) = \Pr\left[Sample^B(\hat{v}, x, y) = w \mid G^B_{\hat{v}xy}\right].$$

In addition, by Proposition 4, for $w \in \mathcal{C}^B_{\hat{v}xy} \cap \mathcal{V}_{vx}$,

$$D_{c'}(w) = \frac{t}{\gamma_{\hat{v}xy}} \cdot D_{vy}(w) \cdot \frac{D_{vx}(w)}{tD_v(w)} = \frac{D_{vxy}(w)}{\gamma_{\hat{v}xy}} = D^B_{\hat{v}xy}(w).$$

By Claim 21, for every $w \in \mathcal{C}^B_{\hat{v}xy} \cap \mathcal{V}_{vx}$,

$$D_c(w) \leq \frac{1}{1 - \Pr_{w \leftarrow D_{c'}}[c'(w) > c(w)]} \cdot D_{c'}(w).$$

By Lemma 19,

$$\Pr_{w \leftarrow D^B_{\hat{v}xy}}[c'(w) > c(w)] \leq \frac{\alpha}{\gamma_{\hat{v}xy}}.$$

Since $1/(1-z) \leq 1+2z$ for $z \in [0, 1/10]$, and since $\alpha/\gamma_{\hat{v}xy} \leq 1/10$, the assertion follows. □

Lemma 22. *Let $\hat{v} \in \hat{\mathcal{V}}$ and $(x,y) \in \mathrm{supp}(\mu)$. Let $v = \mathrm{msg}_{\pi xy}(\hat{v})$ and assume that $v \in \mathcal{V} \setminus \mathcal{L}^+$. Then,*

$$(\gamma_{\hat{v}xy} - \alpha)/t \leq \Pr\left[G^B_{\hat{v}xy}\right] \leq \gamma_{\hat{v}xy}/t,$$
$$(1 - \gamma_{\hat{v}xy} - \alpha)/t \leq \Pr\left[G^A_{\hat{v}xy}\right] \leq (1 - \gamma_{\hat{v}xy})/t,$$

where the probabilities are over the randomness used by the players. The lower bound on $\Pr\left[G^B_{\hat{v}xy}\right]$ holds provided that $\gamma_{\hat{v}xy} > 0$, and the lower bound on $\Pr\left[G^A_{\hat{v}xy}\right]$ holds provided that $\gamma_{\hat{v}xy} < 1$.

Proof. We prove the bounds on $\Pr[G^B_{\hat{v}xy}]$. A similar argument can be applied to show the

bounds on $\Pr[G^A_{\hat{v}xy}]$. Let $v = \mathrm{msg}_{\pi xy}(\hat{v})$. Let $w \in \mathcal{C}^B_{\hat{v}xy} \cap \mathcal{V}_{vx}$. Denote

$$D'_{\hat{v}xy}(w) = \Pr\left[Sample^B(\hat{v}, x, y) = w\right],$$

where the probability is over the randomness used by the players.

We first prove the upper bound on $\Pr[G^B_{\hat{v}xy}]$. Every vertex $w \in \mathcal{C}^B_{\hat{v}xy} \cap \mathcal{V}_{vx}$ is selected with probability $D_{vy}(w)$ by Line 5 of $Sample^B$, as every leaf $u \in \mathcal{L}(v)$ is selected by Line 4 with probability $\tilde{D}_{vy}(u)$ (by Lemma 14), and as $\mathcal{C}^B_{\hat{v}xy}$ is a frontier (by Lemma 16). By Proposition 4,

$$D'_{\hat{v}xy}(w) = D_{vy}(w) \cdot \min\left\{1, \frac{D_{vx}(w)}{tD_v(w)}\right\} \leq D_{vy}(w) \cdot \frac{D_{vx}(w)}{tD_v(w)} = D_{\hat{v}xy}(w)/t. \tag{6}$$

By Lemma 20, the execution of $Sample^B(\hat{v}, x, y)$ returns either a vertex in $\mathcal{C}^B_{\hat{v}xy} \cap \mathcal{V}_{vx}$ or failure. Therefore,

$$\Pr\left[G^B_{\hat{v}xy}\right] = D'_{\hat{v}xy}(\mathcal{C}^B_{\hat{v}xy} \cap \mathcal{V}_{vx}) \leq D_{\hat{v}xy}(\mathcal{C}^B_{\hat{v}xy} \cap \mathcal{V}_{vx})/t = \gamma_{\hat{v}xy}/t.$$

We next prove the lower bound on $\Pr[G^B_{\hat{v}xy}]$. Denote by $\mathcal{S}_{\hat{v}xy} \subseteq \mathcal{C}^B_{\hat{v}xy} \cap \mathcal{V}_{vx}$ the set of vertices $w \in \mathcal{C}^B_{\hat{v}xy} \cap \mathcal{V}_{vx}$ for which $D'_{\hat{v}xy}(w) = D_{\hat{v}xy}(w)/t$. Observe that due to Equation (6), $w \in \mathcal{C}^B_{\hat{v}xy} \cap \mathcal{V}_{vx}$ is in $\mathcal{S}_{\hat{v}xy}$ if and only if $\frac{D_{vx}(w)}{D_v(w)} \leq t$. By Lemma 19, $D^B_{\hat{v}xy}\left((\mathcal{C}^B_{\hat{v}xy} \cap \mathcal{V}_{vx}) \setminus \mathcal{S}_{\hat{v}xy}\right) \leq \frac{\alpha}{\gamma_{\hat{v}xy}}$, thus, $D^B_{\hat{v}xy}\left(\mathcal{S}_{\hat{v}xy}\right) \geq 1 - \frac{\alpha}{\gamma_{\hat{v}xy}}$. This implies $D_{\hat{v}xy}(\mathcal{S}_{\hat{v}xy}) \geq \gamma_{\hat{v}xy} - \alpha$. Conclude that,

$$\Pr\left[G^B_{\hat{v}xy}\right] = D'_{\hat{v}xy}(\mathcal{C}^B_{\hat{v}xy} \cap \mathcal{V}_{vx}) \geq D'_{\hat{v}xy}(\mathcal{S}_{\hat{v}xy}) = D_{\hat{v}xy}(\mathcal{S}_{\hat{v}xy})/t \geq \left(\gamma_{\hat{v}xy} - \alpha\right)/t.$$

□

5.3 The Protocol *Chunk*

Let $N_{\hat{v}xy}$ be a random variable that counts the number of times either $Sample^A$ or $Sample^B$ are executed by $Chunk(\hat{v}, x, y)$. Note that $N_{\hat{v}xy}$ depends on the randomness used by the players.

Lemma 23. *Let $\hat{v} \in \hat{\mathcal{V}}$ and $(x, y) \in \mathrm{supp}(\mu)$. Then,*

$$\Pr\left[N_{\hat{v}xy} = 1\right] \geq \frac{1 - 2\alpha}{2t}; \qquad \mathbf{E}\left[N_{\hat{v}xy}\right] < 3t,$$

where the probability and expectation are over the randomness used by the players.

Proof. Let $\mathrm{msg}_{\pi xy}(\hat{v}) = v$. If $v \in \mathcal{L}^+$, then $\Pr\left[N_{\hat{v}xy} = 1\right] = 1$. Assume that $v \in \mathcal{V} \setminus \mathcal{L}^+$. Let P be the random variable representing the bit p selected in Line 2 during the first iteration of the loop in Line 1 of *Chunk*. Then, by Lemma 22: If $\gamma_{\hat{v}xy} = 1$, then

$$\Pr[N_{\hat{v}xy} = 1] \geq \Pr[P = 1] \cdot \Pr\left[G^B_{\hat{v}xy}\right] \geq \frac{1}{2} \cdot \left(\gamma_{\hat{v}xy} - \alpha\right)/t = \frac{1 - \alpha}{2t}.$$

If $\gamma_{\hat{v}xy} = 0$, then

$$\Pr[N_{\hat{v}xy} = 1] \geq \Pr[P = 0] \cdot \Pr\left[G^A_{\hat{v}xy}\right] \geq \frac{1}{2} \cdot (1 - \gamma_{\hat{v}xy} - \alpha)/t = \frac{1-\alpha}{2t}.$$

If $\gamma_{\hat{v}xy} \in (0,1)$, then

$$\begin{aligned}\Pr[N_{\hat{v}xy} = 1] &= \Pr[P = 0] \cdot \Pr\left[G^A_{\hat{v}xy}\right] + \Pr[P = 1] \cdot \Pr\left[G^B_{\hat{v}xy}\right] \\ &\geq \frac{1}{2} \cdot (1 - \gamma_{\hat{v}xy} - \alpha)/t + \frac{1}{2} \cdot (\gamma_{\hat{v}xy} - \alpha)/t = \frac{1-2\alpha}{2t}.\end{aligned}$$

Therefore, in any case, $\Pr[N_{\hat{v}xy} = 1] \geq \frac{1-2\alpha}{2t}$.

Observe that for every $k \in \mathbb{N}$, it holds that $\Pr[N_{\hat{v}xy} = k \mid N_{\hat{v}xy} > k-1] = \Pr[N_{\hat{v}xy} = 1]$. Therefore, $\mathbf{E}[N_{\hat{v}xy}] = 1/\Pr[N_{\hat{v}xy} = 1] \leq 3t$. □

Let $\hat{v} \in \hat{\mathcal{V}}$ and $(x,y) \in \mathrm{supp}(\mu)$. Let $v = \mathrm{msg}_{\pi xy}(\hat{v})$. An execution of $Chunk(\hat{v}, x, y)$ returns a vertex in $\mathcal{C}_{\hat{v}xy}$: If $v \in \mathcal{L}^+$, then Line 3 of the algorithms $Sample^A$ and $Sample^B$ returns a vertex in $\mathcal{L}(v) = \mathcal{C}_{\hat{v}xy}$. Otherwise, if $v \in \mathcal{V} \setminus \mathcal{L}^+$, then by Lemma 20, both $Sample^A(\hat{v}, x, y)$ and $Sample^B(\hat{v}, x, y)$ return a vertex in $\mathcal{C}_{\hat{v}xy}$. Let $S_{\hat{v}xy}$ be the distribution over $\mathcal{C}_{\hat{v}xy}$ induced by the return value of $Chunk(\hat{v}, x, y)$ (the "simulation's distribution"). Recall that $D_{\hat{v}xy}$ is the function D_{vxy} restricted to the frontier $\mathcal{C}_{\hat{v}xy}$.

Lemma 24. *Let $\hat{v} \in \hat{\mathcal{V}}$ and $(x,y) \in \mathrm{supp}(\mu)$. It holds that*

$$\|S_{\hat{v}xy} - D_{\hat{v}xy}\| \leq 10\sqrt{\alpha}.$$

Proof. Let $\mathrm{msg}_{\pi xy}(\hat{v}) = v$. If $v \in \mathcal{L}^+$, then $S_{\hat{v}xy} = D_{\hat{v}xy}$. Assume that $v \in \mathcal{V} \setminus \mathcal{L}^+$. We show that there exists a set $\mathcal{V}' = \mathcal{V}'_{\hat{v}xy} \subseteq \mathcal{V}(v)$ such that $S_{\hat{v}xy}(\mathcal{V}') \leq 2\sqrt{\alpha}$, and for $w \in \mathcal{C}_{\hat{v}xy} \setminus \mathcal{V}'$ it holds that $S_{\hat{v}xy}(w) \leq (1 + 8\sqrt{\alpha}) \cdot D_{\hat{v}xy}(w)$. The claim then follows as follows: Let $\mathcal{S} \subseteq \mathcal{C}_{\hat{v}xy}$ be the set of all vertices $w \in \mathcal{C}_{\hat{v}xy}$ such that $S_{\hat{v}xy}(w) > D_{\hat{v}xy}(w)$. It holds that

$$\begin{aligned}&S_{\hat{v}xy}(\mathcal{S}) - D_{\hat{v}xy}(\mathcal{S}) \leq (S_{\hat{v}xy}(\mathcal{S} \setminus \mathcal{V}') - D_{\hat{v}xy}(\mathcal{S} \setminus \mathcal{V}')) + S_{\hat{v}xy}(\mathcal{V}') \\ &\leq \left((1 + 8\sqrt{\alpha}) \cdot D_{\hat{v}xy}(\mathcal{S} \setminus \mathcal{V}') - D_{\hat{v}xy}(\mathcal{S} \setminus \mathcal{V}')\right) + 2\sqrt{\alpha} \\ &= 8\sqrt{\alpha} \cdot D_{\hat{v}xy}(\mathcal{S} \setminus \mathcal{V}') + 2\sqrt{\alpha} \leq 10\sqrt{\alpha}.\end{aligned}$$

Since $S_{\hat{v}xy}$ and $D_{\hat{v}xy}$ are distributions,

$$D_{\hat{v}xy}(\mathcal{C}_{\hat{v}xy} \setminus \mathcal{S}) - S_{\hat{v}xy}(\mathcal{C}_{\hat{v}xy} \setminus \mathcal{S}) = S_{\hat{v}xy}(\mathcal{S}) - D_{\hat{v}xy}(\mathcal{S}) \leq 10\sqrt{\alpha}.$$

Let $\gamma = \gamma_{\hat{v}xy}$ and $N = N_{\hat{v}xy}$. Let P be a random variable representing the bit p selected in Line 2, during the first iteration of the loop in Line 1 of $Chunk$. By Lemmas 22 and 23,

$$\Pr[P = 1 | N = 1] \leq \frac{\Pr[P = 1 \wedge N = 1]}{\Pr[N = 1]} \leq \frac{\frac{1}{2} \cdot \frac{\gamma}{t}}{\frac{1-2\sqrt{\alpha}}{2t}} = \frac{\gamma}{1 - 2\sqrt{\alpha}}, \tag{7}$$

$$\Pr[P=0|N=1] \leq \frac{\Pr[P=0 \wedge N=1]}{\Pr[N=1]} \leq \frac{\frac{1}{2}\cdot\frac{1-\gamma}{t}}{\frac{1-2\sqrt{\alpha}}{2t}} = \frac{1-\gamma}{1-2\sqrt{\alpha}}. \tag{8}$$

We define the set $\mathcal{V}'$ as follows: If $\gamma \leq \sqrt{\alpha}$ then $\mathcal{V}' = \mathcal{C}^B_{\hat{v}xy} \cap \mathcal{V}_{vx}$. If $\gamma \geq 1-\sqrt{\alpha}$ then $\mathcal{V}' = \mathcal{C}^A_{\hat{v}xy} \cap \mathcal{V}_{vy}$. Else, if $\sqrt{\alpha} < \gamma < 1-\sqrt{\alpha}$, we set $\mathcal{V}' = \phi$. Consider the case that $\gamma \leq \sqrt{\alpha}$. Let W be a random variable distributed according to $S_{\hat{v}xy}$. By Lemma 20 and Equation (5), conditioned on $N=1$, W is in $\mathcal{C}^B_{\hat{v}xy} \cap \mathcal{V}_{vx}$ only if $P=1$ ($Sample^B$ is called), thus

$$\Pr[W \in \mathcal{V}'|N=1] \leq \Pr[P=1|N=1] \leq \frac{\gamma}{1-2\sqrt{\alpha}} \leq \frac{\sqrt{\alpha}}{1-2\sqrt{\alpha}} \leq 2\sqrt{\alpha}.$$

Similarly, if $\gamma \geq 1-\sqrt{\alpha}$, then

$$\Pr[W \in \mathcal{V}'|N=1] \leq \Pr[P=0|N=1] \leq \frac{1-\gamma}{1-2\sqrt{\alpha}} \leq \frac{\sqrt{\alpha}}{1-2\sqrt{\alpha}} \leq 2\sqrt{\alpha}.$$

Observe that W is independent of N, since for any $n, n' \in \mathbb{N}$, it holds that $W|N=n$ has the same distribution as $W|N=n'$. Therefore, in any case,

$$S_{\hat{v}xy}(\mathcal{V}') = \Pr[W \in \mathcal{V}'] = \Pr[W \in \mathcal{V}'|N=1] \leq 2\sqrt{\alpha}.$$

Assume that $(\mathcal{C}^B_{\hat{v}xy} \cap \mathcal{V}_{vx}) \setminus \mathcal{V}' \neq \phi$. Let $w \in (\mathcal{C}^B_{\hat{v}xy} \cap \mathcal{V}_{vx}) \setminus \mathcal{V}'$. It holds that $\gamma > \sqrt{\alpha}$, as otherwise $(\mathcal{C}^B_{\hat{v}xy} \cap \mathcal{V}_{vx}) \setminus \mathcal{V}' = \phi$. By Equations (5) and (7), Lemma 20, and since $\gamma > \sqrt{\alpha} > 10\alpha$,

$$\begin{aligned}
&\Pr[W=w] = \Pr[W=w|N=1] \\
&= \Pr[P=1|N=1]\cdot\Pr[W=w|P=1,N=1] + \Pr[P=0|N=1]\cdot\Pr[W=w|P=0,N=1] \\
&= \Pr[P=1|N=1]\cdot\Pr[W=w|P=1,N=1] + \Pr[P=0|N=1]\cdot 0 \\
&\leq \frac{\gamma}{1-2\sqrt{\alpha}}\cdot(1+2\alpha/\gamma)\cdot\frac{D_{vxy}(w)}{\gamma} \leq \frac{1+2\alpha/\sqrt{\alpha}}{1-2\sqrt{\alpha}}\cdot D_{vxy}(w) \leq (1+8\sqrt{\alpha})\cdot D_{vxy}(w).
\end{aligned}$$

For $w \in (\mathcal{C}^A_{\hat{v}xy} \cap \mathcal{V}_{vy}) \setminus \mathcal{V}'$, a similar argument shows that $\Pr[W=w] \leq (1+8\sqrt{\alpha})\cdot D_{vxy}(w)$. By the definition of $\mathcal{C}_{\hat{v}xy}$ (see Equation (4)), for every $w \in \mathcal{C}_{\hat{v}xy} \setminus \mathcal{V}'$ it holds that $S_{\hat{v}xy}(w) \leq (1+8\sqrt{\alpha})\cdot D_{\hat{v}xy}(w)$, as required. □

Let $Chunk^S$ be the protocol obtained from $Chunk$ by replacing Lines 11 and 12 of the protocols $Separate^A$ and $Separate^B$ by "$Chunk^S$ returns failure". Roughly speaking, $Chunk^S$ is the variant of $Chunk$ run by the protocol τ^{ST}.

Lemma 25. *Let $\hat{v} \in \hat{\mathcal{V}}$ and $(x,y) \in \mathrm{supp}(\mu)$. Let $v = \mathrm{msg}_{\pi xy}(\hat{v})$. Then,*

$$\mathbf{E}\left[|Chunk^S(\hat{v},x,y)|\right] \leq 30t\left(\mathbb{D}\left(\tilde{D}_{vx}\|\tilde{D}_v\right) + \mathbb{D}\left(\tilde{D}_{vy}\|\tilde{D}_v\right) + \log(1/\eta) + \log(r)\right),$$

where the expectation is over the randomness used by the players.

Proof. By Lemma 23, an execution of $Chunk^S(\hat{v},x,y)$ runs $Sample^A(\hat{v},x,y)$ and $Sample^B(\hat{v},x,y)$ at most $3t$ times in expectation. An execution of $Sample^B(\hat{v},x,y)$ has

expected communication complexity of at most $10\mathbb{D}\left(\tilde{D}_{vy}\|\tilde{D}_v\right) + 2\log(1/\eta) + 3\log(r) + 10$: It runs $CorrelatedSampling(D_{vy}, D_v, \eta)$, which, by Lemma 14, has expected communication complexity of at most $10\mathbb{D}\left(\tilde{D}_{vy}\|\tilde{D}_v\right) + 2\log(1/\eta) + 10$. It then runs the variant of $Separate^B(\hat{v}, x, y)$ obtained by replacing Lines 11 and 12 of the protocol $Separate^B$ by "return failure". This variant has expected communication of at most $3\log(r)$, as the players only communicate two indices $a, b \in [r]$ and a bit $a' \in \{0, 1\}$. Similarly, an execution of $Sample^A(\hat{v}, x, y)$ has expected communication complexity of at most $10\mathbb{D}\left(\tilde{D}_{vx}\|\tilde{D}_v\right) + 2\log(1/\eta) + 3\log(r) + 10$. □

5.4 The Protocol τ^{ST}

Lemma 26. *It holds that*

$$\mathop{\mathbf{E}}_{(x,y)\leftarrow\mu}\left[\left\|\mathrm{msg}_{\pi xy}(\tau(x,y)) - \pi(x,y)\right\|\right] \leq 0.1\varepsilon.$$

Proof. Define the bad set

$$\mathcal{I}' = \left\{(x,y) \in \mathrm{supp}(\mu) :\ \mathbb{D}(\tilde{D}_{xy}\|\tilde{D}) > 100I/\varepsilon\right\}.$$

By Proposition 8 and by Markov's inequality,

$$\mu(\mathcal{I}') < 0.01\varepsilon. \tag{9}$$

Fix $(x,y) \in \mathrm{supp}(\mu) \setminus \mathcal{I}'$. Let $i \in [m]$. Let $\hat{v}' \in \hat{\mathcal{V}}_{i-1}$ be a possible transcript of the first $i-1$ executions of $Chunk$ by τ. Let $v \in \mathcal{V}$ be such that there exists a transcript $\hat{v} \in \hat{\mathcal{V}}_i$ satisfying: (i) $v = \mathrm{msg}_{\pi xy}(\hat{v})$; (ii) $\hat{v}'$ is a prefix of $\hat{v}$. Assume that the transcript of the first $i-1$ executions of $Chunk$ by τ is $\hat{v}'$. Let $\hat{\mathcal{W}}$ be the (multi-)set of all possible transcripts for the i^{th} execution of $Chunk$ by τ, after which v is reached. That is, for $\hat{w} \in \hat{\mathcal{W}}$, it holds that $\mathrm{msg}_{\pi xy}(\hat{v}' \circ \hat{w}) = v$, where '$\circ$' denotes string concatenation. Let $k \in \mathbb{N}$ be a large enough constant to be specified next. We denote the elements of the (multi-)set $\hat{\mathcal{W}}$ by $\mathrm{trans}_1(\hat{v}', v), \ldots, \mathrm{trans}_k(\hat{v}', v)$. It will be convenient for us to assume that for every $j \in [k]$, once $\hat{v}'$ is reached, the probability that the transcript of the i^{th} execution of $Chunk$ is $\mathrm{trans}_j(\hat{v}', v)$, is $1/k$. For this assumption to hold without loss of generality, we allow every transcript to appear many times in the (multi)-set $\hat{\mathcal{W}}$ (that is, we allow $\mathrm{trans}_j(\hat{v}', v) = \mathrm{trans}_{j'}(\hat{v}', v)$ for $j \neq j' \in [k]$), and choose the parameter k to be large enough.

Let $r_1, \ldots, r_m \in [k]$. We define the sequence of frontiers, $\mathcal{B}_{r_1}, \mathcal{B}_{r_1 r_2}, \ldots, \mathcal{B}_{r_1 \ldots r_m} \subseteq \mathcal{V}$, with respect to v_0, inductively. The frontier are such that every vertex in $\mathcal{B}_{r_1 \ldots r_i}$ has an ancestor in $\mathcal{B}_{r_1 \ldots r_{i-1}}$ (that is, $\mathcal{B}_{r_1 \ldots r_i}$ is "below" $\mathcal{B}_{r_1 \ldots r_{i-1}}$). Very roughly, $\mathcal{B}_{r_1 \ldots r_i}$ is the frontier reached by τ after i executions of $Chunk$, when the randomness used is $r_1 \ldots r_i$. Formally, let $\mathcal{B} = \{v_0\}$. For $v \in \mathcal{B}$ and any $r_1 \in [k]$, define $\mathrm{trans}_{r_1}(v) = \phi \in \hat{\mathcal{V}}_0$, where ϕ is the empty transcript. Define $\mathcal{B}_{r_1} = \mathcal{C}_{\mathrm{trans}_{r_1}(v)xy} = \mathcal{C}_{\phi xy}$. Let $i \geq 2$. Assume that the frontiers $\mathcal{B}_{r_1}, \ldots, \mathcal{B}_{r_1,\ldots,r_{i-1}}$ were defined. Let $v \in \mathcal{B}_{r_1,\ldots,r_{i-1}}$ and let v' be the ancestor of v in $\mathcal{B}_{r_1,\ldots,r_{i-2}}$. We first define the

transcript $\text{trans}_{r_1\ldots r_i}(v) \in \hat{\mathcal{V}}_{i-1}$ induced by v and $r_1 \ldots r_i$, by

$$\text{trans}_{r_1\ldots r_i}(v) = \text{trans}_{r_1\ldots r_{i-1}}(v') \circ \text{trans}_{r_i}(\text{trans}_{r_1\ldots r_{i-1}}(v'), v).$$

Observe that $\text{msg}_{\pi xy}(\text{trans}_{r_1\ldots r_i}(v)) = v$. We then define

$$\mathcal{B}_{r_1\ldots r_i} = \bigcup_{v\in\mathcal{B}_{r_1\ldots r_{i-1}}} \mathcal{C}_{\text{trans}_{r_1\ldots r_i}(v)xy}.$$

Since $\text{msg}_{\pi xy}(\text{trans}_{r_1\ldots r_i}(v)) = v$, we get that $\mathcal{C}_{\text{trans}_{r_1\ldots r_i}(v)xy}$ is a frontier with respect to v. Therefore, every vertex in $\mathcal{B}_{r_1\ldots r_i}$ has an ancestor in $\mathcal{B}_{r_1\ldots r_{i-1}}$. In addition, by induction, since $\mathcal{C}_{\text{trans}_{r_1\ldots r_i}(v)xy}$ is a frontier with respect to v, and assuming that $\mathcal{B}_{r_1\ldots r_{i-1}}$ is a frontier, we get that $\mathcal{B}_{r_1\ldots r_i}$ is a frontier too.

Recall that $S_{\hat{v}xy}$ is the distribution over $\mathcal{C}_{\hat{v}xy}$ induced by the return value of $Chunk(\hat{v}, x, y)$ (see Section 5.3), and that $D_{\hat{v}xy}$ is the distribution obtained by restricting the function D_{vxy} to the frontier $\mathcal{C}_{\hat{v}xy}$. Let $i \in [m]$, and $r_1, \ldots, r_i \in [k]$. Let $D_{r_1\ldots r_i}$ be the distribution obtained by restricting the function D_{xy} to the frontier $\mathcal{B}_{r_1\ldots r_i}$. In other words, for $i = 1$, $D_{r_1} = D_{\phi xy}$. For $i \in \{2, \ldots, m\}$ and $v \in \mathcal{B}_{r_1\ldots r_i}$, let v' be the ancestor of v in $\mathcal{B}_{r_1,\ldots,r_{i-1}}$, and set

$$D_{r_1\ldots r_i}(v) = D_{r_1\ldots r_{i-1}}(v') \cdot D_{\text{trans}_{r_1\ldots r_i}(v')xy}(v).$$

Let $S_{r_1\ldots r_i}$ be the distribution over $\mathcal{B}_{r_1\ldots r_i}$ define by: For $i = 1$, define $S_{r_1} = S_{\phi xy}$. For $i \in \{2, \ldots, m\}$ and $v \in \mathcal{B}_{r_1\ldots r_i}$, let v' be the ancestor of v in $\mathcal{B}_{r_1,\ldots,r_{i-1}}$, and set

$$S_{r_1\ldots r_i}(v) = S_{r_1\ldots r_{i-1}}(v') \cdot S_{\text{trans}_{r_1\ldots r_i}(v')xy}(v).$$

Claim 27. *For every $i \in [m]$ and $r_1, \ldots, r_i \in [k]$,*

$$\|S_{r_1\ldots r_i} - D_{r_1\ldots r_i}\| \leq 10i\sqrt{\alpha}.$$

Proof. The proof is by induction.
Induction base: For $i = 1$, by Lemma 24, $\|S_{r_1} - D_{r_1}\| = \|S_{\phi xy} - D_{\phi xy}\| \leq 10\sqrt{\alpha}$.
Induction step: For $i \geq 2$ and $v \in \mathcal{B}_{r_1\ldots r_i}$, let v' be the ancestor of v in $\mathcal{B}_{r_1,\ldots,r_{i-1}}$, and consider the hybrid $H_{r_1,\ldots,r_i}$ defined by

$$H_{r_1\ldots r_i}(v) = D_{r_1\ldots r_{i-1}}(v') \cdot S_{\text{trans}_{r_1\ldots r_i}(v')xy}(v).$$

Observe that $H_{r_1\ldots r_i}$ is a distribution over $\mathcal{B}_{r_1\ldots r_i}$ as well. By the induction hypothesis and Lemma 24,

$$\begin{aligned}
&\|S_{r_1\ldots r_i} - D_{r_1\ldots r_i}\| \\
&\leq \|S_{r_1\ldots r_i} - H_{r_1\ldots r_i}\| + \|H_{r_1\ldots r_i} - D_{r_1\ldots r_i}\| \\
&= \left\|S_{r_1\ldots r_{i-1}} - D_{r_1\ldots r_{i-1}}\right\| + \sum_{v'\in\mathcal{B}_{r_1,\ldots,r_{i-1}}} D_{r_1\ldots r_{i-1}}(v') \cdot \left\|S_{\text{trans}_{r_1\ldots r_i}(v')xy} - D_{\text{trans}_{r_1\ldots r_i}(v')xy}\right\| \\
&\leq 10(i-1)\sqrt{\alpha} + 10\sqrt{\alpha} = 10i\sqrt{\alpha}.
\end{aligned}$$

□

Claim 28. *For every $r_1, \ldots, r_m \in [k]$, it holds that*

$$\left\| D_{r_1 \ldots r_m} - \tilde{D}_{xy} \right\| \leq 0.01\varepsilon.$$

Proof. Define the bad set

$$\mathcal{L}' = \left\{ u \in \mathcal{L} : \ \mathbb{D}_{v_0 uxy} > 10^4 I/\varepsilon^2 \right\}.$$

By Proposition 5 and Markov's inequality, since $(x,y) \in \mathrm{supp}(\mu) \setminus \mathcal{I}'$, then $\tilde{D}_{xy}(\mathcal{L}') < 0.01\varepsilon$. For $u \in \mathcal{L} \setminus \mathcal{L}'$, it holds that $D_{r_1 \ldots r_m}(u) = \tilde{D}_{xy}(u)$, as $m > \frac{10^4 I}{\varepsilon^2 \beta} + 1$ and due to the following: Let $i \in \{0, \ldots, m-1\}$. Assume that the transcript of τ after i executions of *Chunk* is $\hat{v} \in \hat{\mathcal{V}}_i$, and that after an additional iteration the transcript of τ is $\hat{w} \in \hat{\mathcal{V}}_{i+1}$. Let $v = \mathrm{msg}_{\pi xy}(\hat{v})$ and $w = \mathrm{msg}_{\pi xy}(\hat{w})$. By the beginnings of Sections 5.3 and 5.2, we have that $w \in \mathcal{C}_{\hat{v}xy}$ and that $\mathcal{C}^{\min}_{vxy} < \mathcal{C}_{\hat{v}xy}$. Therefore, $\mathcal{C}^{\min}_{vxy} < w$. Recall from Section 3.4 that $\mathcal{C}^{\min}_{vxy} \subseteq \mathcal{C}_{vx} \cup \mathcal{C}_{vy}$. By the definitions of $\mathcal{C}_{vx}$ and $\mathcal{C}_{vy}$, if $w \in \mathcal{L} \setminus \mathcal{L}^+$, it holds that $\mathbb{D}_{vwxy} = \mathbb{D}_{vwx} + \mathbb{D}_{vwy} \geq \beta$.

The assertion follows as follows: We view $\tilde{D}_{xy}$ and $D_{r_1 \ldots r_m}$ as distributions over $\mathcal{V}$. Let $\mathcal{S} \subseteq \mathcal{V}$ be the set of all vertices v such that $\tilde{D}_{xy}(v) > D_{r_1 \ldots r_m}(v)$. Observe that $\mathcal{S} \subseteq \mathcal{L}$, as for $v \in \mathcal{V} \setminus \mathcal{L}$ we have $\tilde{D}_{xy}(v) = 0$. It holds that $\tilde{D}_{xy}(\mathcal{S}) - D_{r_1 \ldots r_m}(\mathcal{S}) \leq \tilde{D}_{xy}(\mathcal{S}) \leq \tilde{D}_{xy}(\mathcal{L}') \leq 0.01\varepsilon$. Since $\tilde{D}_{xy}$ and $D_{r_1 \ldots r_m}$ are distributions, $D_{r_1 \ldots r_m}(\mathcal{V} \setminus \mathcal{S}) - \tilde{D}_{xy}(\mathcal{V} \setminus \mathcal{S}) = \tilde{D}_{xy}(\mathcal{S}) - D_{r_1 \ldots r_m}(\mathcal{S}) \leq 0.01\varepsilon$. □

Using the last two claims,

$$\begin{aligned}
&\left\| \mathrm{msg}_{\pi xy}(\tau(x,y)) - \pi(x,y) \right\| = \left\| \left(\mathop{\mathbf{E}}_{r_1,\ldots,r_m \in_R [k]} \left[S_{r_1 \ldots r_m} \right] \right) - \tilde{D}_{xy} \right\| \qquad (10) \\
&\leq \mathop{\mathbf{E}}_{r_1,\ldots,r_m \in_R [k]} \left[\left\| S_{r_1 \ldots r_m} - \tilde{D}_{xy} \right\| \right] \\
&\leq \mathop{\mathbf{E}}_{r_1,\ldots,r_m \in_R [k]} \left[\| S_{r_1 \ldots r_m} - D_{r_1 \ldots r_m} \| \right] + \mathop{\mathbf{E}}_{r_1,\ldots,r_m \in_R [k]} \left[\left\| D_{r_1 \ldots r_m} - \tilde{D}_{xy} \right\| \right] \\
&\leq 0.01\varepsilon + 10m\sqrt{\alpha} \leq 0.02\varepsilon.
\end{aligned}$$

By Equations (9) and (10),

$$\begin{aligned}
&\mathop{\mathbf{E}}_{(x,y) \leftarrow \mu} \left[\left\| \mathrm{msg}_{\pi xy}(\tau(x,y)) - \pi(x,y) \right\| \right] \leq \sum_{(x,y) \in \mathrm{supp}(\mu)} \mu(x,y) \cdot \left\| \mathrm{msg}_{\pi xy}(\tau(x,y)) - \pi(x,y) \right\| \\
&\leq \mu(\mathcal{I}') + \sum_{(x,y) \in \mathrm{supp}(\mu) \setminus \mathcal{I}'} \mu(x,y) \cdot \left\| \mathrm{msg}_{\pi xy}(\tau(x,y)) - \pi(x,y) \right\| \leq 0.1\varepsilon.
\end{aligned}$$

□

We define an additional variant, τ^S, of the protocol τ. The protocol τ^S gets the same parameters as τ. It operates the same as τ, except for the second change described in Section 4.2. That is, Lines 11 and 12 of the protocols $Separate^A$ and $Separate^B$ are replaced by "τ^S returns failure".

Lemma 29. *It holds that*

$$\mathop{\mathbf{E}}_{(x,y)\leftarrow\mu}\left[\left\|\tau^S(x,y)-\tau(x,y)\right\|\right]<0.1\varepsilon.$$

Proof. Let $i\in[m]$ and $j\in\mathbb{N}$. When we refer to the "(i,j) execution of $Separate^B$" we mean the execution of $Separate^B$ by the j^{th} execution of $Sample^B$ by the i^{th} execution of $Chunk$ by τ, assuming that this execution is reached. Let $\hat{V}_i$ and U_{ij} be random variables that are functions of the inputs X,Y and the randomness used by the players. The random variable $\hat{V}_i$ represents the $\hat{v}$ parameter passed to the (i,j) execution of $Separate^B$. That is, $\hat{V}_i$ is the transcript of τ after the $(i-1)^{th}$ execution of the protocol $Chunk$. Let $V_i=\mathrm{msg}_{\pi Y}(\hat{V}_i)$.

The random variable U_{ij} is defined as follows: If the (i,j) execution of $Separate^B$ is reached, then U_{ij} is the u parameter passed to the (i,j) execution of $Separate^B$. That is, U_{ij} is the leaf sampled by the protocol $CorrelatedSampling$, executed by the j^{th} execution of $Sample^B$ by the i^{th} execution of $Chunk$. If the (i,j) execution of $Separate^B$ is not reached, then we set U_{ij} to the output of the execution of $CorrelatedSampling(P=\tilde{D}_{V_iY},Q=\tilde{D}_{V_i},\eta)$. Since $Sample^B$ runs $CorrelatedSampling$ with parameters that only depend on Y and $\hat{V}_i$, we get that U_{ij} is independent of X given Y and $\hat{V}_i$.

Let $\hat{v}\in\hat{\mathcal{V}}_i$. Recall that $\mu_{\hat{v}}$ was defined in Section 4 by $\mu_{\hat{v}}=((X,Y)\mid\tau(X,Y)_{\leq|\hat{v}|}=\hat{v})$, and that $\mu_{\hat{v}}$ is a product distribution $\mu_{\hat{v}}=\mu^A_{\hat{v}}\times\mu^B_{\hat{v}}$. Observe that $\hat{v}$ determines i, thus $\mu_{\hat{v}}=((X,Y)|\hat{V}_i=\hat{v})$.

Recall that $F^B(\hat{v},u,x,y)\in\{0,1\}$ is the value 1 if and only if when running the protocol $Separate^B$ (Algorithm 3) with the parameters $\hat{v},u,x,y$, the "else" part in Line 10 is reached (F^B is defined just before Lemma 18). It holds that

$$\begin{aligned}
&\Pr_{\hat{V}_i,U_{ij},X,Y}\left[F^B(\hat{V}_i,U_{ij},X,Y)=1\right]\\
&=\mathop{\mathbf{E}}_{\hat{v}\leftarrow\hat{V}_i}\ \mathop{\mathbf{E}}_{(x,y)\leftarrow((X,Y)|\hat{V}_i=\hat{v})}\ \Pr_{u\leftarrow(U_{ij}|X=x,Y=y,\hat{V}_i=\hat{v})}\left[F^B(\hat{v},u,x,y)=1\right]\\
&=\mathop{\mathbf{E}}_{\hat{v}\leftarrow\hat{V}_i}\ \mathop{\mathbf{E}}_{(x,y)\leftarrow((X,Y)|\hat{V}_i=\hat{v})}\ \Pr_{u\leftarrow(U_{ij}|Y=y,\hat{V}_i=\hat{v})}\left[F^B(\hat{v},u,x,y)=1\right]\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad(U_{ij}\text{ and }X\text{ are independent given }Y,\hat{V}_i)\\
&=\mathop{\mathbf{E}}_{\hat{v}\leftarrow\hat{V}_i}\ \mathop{\mathbf{E}}_{(x,y)\leftarrow\mu_{\hat{v}}}\ \Pr_{u\leftarrow(U_{ij}|Y=y,\hat{V}_i=\hat{v})}\left[F^B(\hat{v},u,x,y)=1\right]\qquad(\mu_{\hat{v}}=((X,Y)|\hat{V}_i=\hat{v}))\\
&=\mathop{\mathbf{E}}_{\hat{v}\leftarrow\hat{V}_i}\ \mathop{\mathbf{E}}_{y\leftarrow\mu^B_{\hat{v}}}\ \mathop{\mathbf{E}}_{u\leftarrow(U_{ij}|Y=y,\hat{V}_i=\hat{v})}\ \mathop{\mathbf{E}}_{x\leftarrow\mu^A_{\hat{v}}}\left[F^B(\hat{v},u,x,y)=1\right]\qquad(\mu_{\hat{v}}=\mu^A_{\hat{v}}\times\mu^B_{\hat{v}})\\
&<1/r.\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad(\text{by Lemma 18})
\end{aligned}$$

Consider the protocol $\tau^{S'}$ that operates the same as τ, expect that Lines 11 and 12 of the protocol $Separate^B$ (Algorithm 3) are replaced by "$\tau^{S'}$ returns failure" (the protocol $Separate^A$ is not changed). By Lemma 23 and Markov's inequality, τ runs $Separate^B$ at

most $J = \lceil 90tm/\varepsilon \rceil$ times, except with probability of at most $\varepsilon/30$. Therefore,

$$\begin{aligned}&\mathop{\mathbf{E}}_{(x,y)\leftarrow\mu}\left[\left\|\tau(x,y)-\tau^{S'}(x,y)\right\|\right]\\&\leq \varepsilon/30+\sum_{i\in[m],j\in[J]}\Pr_{\hat{V}_i,U_{ij},X,Y}\left[F^B(\hat{V}_i,U_{ij},X,Y)=1\right]\leq \varepsilon/30+mJ\cdot(1/r)\leq\varepsilon/20.\end{aligned}$$

A similar argument can be used to show that $\mathbf{E}_{(x,y)\leftarrow\mu}\left[\left\|\tau^{S'}(x,y)-\tau^{S}(x,y)\right\|\right]<\varepsilon/20$. The assertion follows as

$$\begin{aligned}&\mathop{\mathbf{E}}_{(x,y)\leftarrow\mu}\left[\left\|\tau(x,y)-\tau^{S}(x,y)\right\|\right]\\&\qquad\leq \mathop{\mathbf{E}}_{(x,y)\leftarrow\mu}\left[\left\|\tau(x,y)-\tau^{S'}(x,y)\right\|\right]+\mathop{\mathbf{E}}_{(x,y)\leftarrow\mu}\left[\left\|\tau^{S'}(x,y)-\tau^{S}(x,y)\right\|\right]\leq\varepsilon/10.\end{aligned}$$

□

Lemma 30. *It holds that*

$$\mathop{\mathbf{E}}_{(x,y)\leftarrow\mu}\left[\left\|\tau^{ST}(x,y)-\tau^{S}(x,y)\right\|\right]\leq 0.3\varepsilon.$$

Proof. Let

$$\mathcal{X}'=\left\{x\in\mathrm{supp}(\mu^A):\ \mathbb{D}(\tilde{D}_x\|\tilde{D})>I\log(I)/\varepsilon\right\},$$

$$\mathcal{Y}'=\left\{y\in\mathrm{supp}(\mu^B):\ \mathbb{D}(\tilde{D}_y\|\tilde{D})>I\log(I)/\varepsilon\right\}.$$

By Propositions 8 and 9,

$$I=\mathop{\mathbf{E}}_{x\leftarrow\mu^A}\mathop{\mathbf{E}}_{y\leftarrow\mu^B}\left[\mathbb{D}(\tilde{D}_{xy}\|\tilde{D})\right]\geq\mathop{\mathbf{E}}_{x\leftarrow\mu^A}\left[\mathbb{D}(\tilde{D}_{x}\|\tilde{D})\right],$$

$$I=\mathop{\mathbf{E}}_{y\leftarrow\mu^B}\mathop{\mathbf{E}}_{x\leftarrow\mu^A}\left[\mathbb{D}(\tilde{D}_{xy}\|\tilde{D})\right]\geq\mathop{\mathbf{E}}_{y\leftarrow\mu^B}\left[\mathbb{D}(\tilde{D}_{y}\|\tilde{D})\right].$$

Therefore, by Markov's inequality,

$$\mu^A(\mathcal{X}'),\mu^B(\mathcal{Y}')\leq\frac{I}{I\log(I)/\varepsilon}=\frac{\varepsilon}{\log(I)}.\tag{11}$$

Let $(x,y)\in\mathrm{supp}(\mu)$. We consider the following sets of vertices

$$\mathcal{V}'_x=\left\{v\in\mathcal{V}:\ \mathbb{D}(\tilde{D}_{vx}\|\tilde{D}_v)>\mathbb{D}(\tilde{D}_x\|\tilde{D})\cdot(\log(I)/\varepsilon)\right\}.$$

$$\mathcal{V}'_y=\left\{v\in\mathcal{V}:\ \mathbb{D}(\tilde{D}_{vy}\|\tilde{D}_v)>\mathbb{D}(\tilde{D}_y\|\tilde{D})\cdot(\log(I)/\varepsilon)\right\}.$$

Let $\mathcal{L}'_x\subseteq\mathcal{L}$ be the set of leaves u such that $P(v_0,u)\cap\mathcal{V}'_x\neq\phi$. Let $\mathcal{L}'_y\subseteq\mathcal{L}$ be the set of leaves u such that $P(v_0,u)\cap\mathcal{V}'_y\neq\phi$.

Claim 31. *It holds that*

$$\mathop{\mathbf{E}}_{(x,y)\leftarrow\mu}\left[\Pr\left[\mathrm{msg}_{\pi xy}(\tau^S(x,y))\in\mathcal{L}'_x\cup\mathcal{L}'_y\right]\right]\leq 0.25\varepsilon.$$

Proof. Fix $(x,y)\in\mathrm{supp}(\mu)$. Consider the set $\mathcal{S}'_x\subseteq\mathcal{V}'_x$ defined as follows: Start from an empty set. For every $u\in\mathcal{L}$, add to $\mathcal{S}'_x$ the first vertex on the path from v_0 to u that is in the set $\mathcal{V}'_x$. Define $\mathcal{S}'_y$ similarly. By definition, for every $u\in\mathcal{L}$, the path from v_0 to u intersects each of the sets $\mathcal{S}'_x$ and $\mathcal{S}'_y$ at most once. Therefore, each of the sets $\mathcal{S}'_x$ and $\mathcal{S}'_y$ is a subset of some frontier. By Proposition 10,

$$\mathbb{D}(\tilde{D}_x\|\tilde{D})\geq\sum_{v\in\mathcal{S}'_x}D_x(v)\cdot\mathbb{D}(\tilde{D}_{vx}\|\tilde{D}_v).$$

$$\mathbb{D}(\tilde{D}_y\|\tilde{D})\geq\sum_{v\in\mathcal{S}'_y}D_y(v)\cdot\mathbb{D}(\tilde{D}_{vy}\|\tilde{D}_v).$$

Since $\mathcal{S}'_x\subseteq\mathcal{V}'_x$, and by the definition of $\mathcal{V}'_x$,

$$D_x(\mathcal{S}'_x),D_y(\mathcal{S}'_y)<\varepsilon/\log(I).$$

This implies

$$\tilde{D}_x(\mathcal{L}'_x)=D_x(\mathcal{S}'_x)<\varepsilon/\log(I);\quad\tilde{D}_y(\mathcal{L}'_y)=D_y(\mathcal{S}'_y)<\varepsilon/\log(I).$$

By Proposition 7,

$$\begin{aligned}\mathop{\mathbf{E}}_{(x,y)\leftarrow\mu}\left[\tilde{D}_{xy}(\mathcal{L}'_x\cup\mathcal{L}'_y)\right]&\leq\mathop{\mathbf{E}}_{x\leftarrow X}\mathop{\mathbf{E}}_{y\leftarrow Y}\left[\tilde{D}_{xy}(\mathcal{L}'_x)\right]+\mathop{\mathbf{E}}_{y\leftarrow Y}\mathop{\mathbf{E}}_{x\leftarrow X}\left[\tilde{D}_{xy}(\mathcal{L}'_y)\right]\\&=\mathop{\mathbf{E}}_{x\leftarrow X}\left[\tilde{D}_x(\mathcal{L}'_x)\right]+\mathop{\mathbf{E}}_{y\leftarrow Y}\left[\tilde{D}_y(\mathcal{L}'_y)\right]\leq 2\varepsilon/\log(I).\end{aligned}$$

By Lemmas 26 and 29,

$$\begin{aligned}&\mathop{\mathbf{E}}_{(x,y)\leftarrow\mu}\left[\Pr\left[\mathrm{msg}_{\pi xy}(\tau^S(x,y))\in\mathcal{L}'_x\cup\mathcal{L}'_y\right]\right]\\&\leq\mathop{\mathbf{E}}_{(x,y)\leftarrow\mu}\left[\Pr\left[\pi(x,y)\in\mathcal{L}'_x\cup\mathcal{L}'_y\right]+\left\|\mathrm{msg}_{\pi xy}(\tau^S(x,y))-\pi(x,y)\right\|\right]\\&\leq\mathop{\mathbf{E}}_{(x,y)\leftarrow\mu}\left[\tilde{D}_{xy}(\mathcal{L}'_x\cup\mathcal{L}'_y)+\left\|\mathrm{msg}_{\pi xy}(\tau^S(x,y))-\mathrm{msg}_{\pi xy}(\tau(x,y))\right\|+\left\|\mathrm{msg}_{\pi xy}(\tau(x,y))-\pi(x,y)\right\|\right]\\&\leq 2\varepsilon/\log(I)+0.1\varepsilon+0.1\varepsilon\leq 0.25\varepsilon.\end{aligned}$$

□

Fix $x\in\mathrm{supp}(\mu^A)\setminus\mathcal{X}'$ and $y\in\mathrm{supp}(\mu^B)\setminus\mathcal{Y}'$. If $u\in\mathcal{L}\setminus(\mathcal{L}'_x\cup\mathcal{L}'_y)$ then for every $v\in P(v_0,u)$, it holds that $\mathbb{D}(\tilde{D}_{vx}\|\tilde{D}_v),\mathbb{D}(\tilde{D}_{vy}\|\tilde{D}_v)\leq I\log^2(I)/\varepsilon^2$. Observe that if an execution of $\tau^S(x,y)$ returned u, then it run $Chunk(\hat{v},x,y)$ with at most m transcripts

$\hat{v} \in \hat{\mathcal{V}}$, such that $v = \mathrm{msg}_{\pi xy}(\hat{v}) \in P(v_0, u)$. Therefore, using Lemma 25,

$$\begin{aligned}
&\mathbf{E}\left[|\tau^S(x,y)| \;\middle|\; \mathrm{msg}_{\pi xy}(\tau^S(x,y)) \in \mathcal{L} \setminus (\mathcal{L}'_x \cup \mathcal{L}'_y)\right] \\
&\quad \le 30mt\left(2I\log^2(I)/\varepsilon^2 + \log(1/\eta) + \log(r)\right) \le \frac{10^{20} t I^2 \log^2(I)}{\varepsilon^4 \beta},
\end{aligned}$$

where the expectation is over the randomness used by the players. By Markov's inequality,

$$\begin{aligned}
&\Pr\left[|\tau^S(x,y)| \ge T \;\middle|\; \mathrm{msg}_{\pi xy}(\tau^S(x,y)) \in \mathcal{L} \setminus (\mathcal{L}'_x \cup \mathcal{L}'_y)\right] \\
&\qquad \le \left(\frac{10^{20} t I^2 \log^2(I)}{\varepsilon^4 \beta}\right) / T = \frac{\varepsilon}{\log(I)},
\end{aligned} \tag{12}$$

where the expectation is over the randomness used by the players.

We get

$$\begin{aligned}
&\mathop{\mathbf{E}}_{(x,y)\leftarrow\mu}\left[\left\|\tau^{ST}(x,y) - \tau^S(x,y)\right\|\right] \\
&\le \mathop{\mathbf{E}}_{(x,y)\leftarrow\mu}\left[\Pr\left[|\tau^S(x,y)| \ge T\right]\right] \\
&\le \mu^A(\mathcal{X}') + \mu^B(\mathcal{Y}') + \sum_{\substack{x\in\mathrm{supp}(\mu^A)\setminus\mathcal{X}' \\ y\in\mathrm{supp}(\mu^B)\setminus\mathcal{Y}'}} \mu(x,y)\cdot\Pr\left[|\tau^S(x,y)| \ge T\right] \\
&\le \frac{2\varepsilon}{\log(I)} + \sum_{\substack{x\in\mathrm{supp}(\mu^A)\setminus\mathcal{X}' \\ y\in\mathrm{supp}(\mu^B)\setminus\mathcal{Y}'}} \mu(x,y)\cdot\left(\frac{\varepsilon}{\log(I)} + \Pr\left[\mathrm{msg}_{\pi xy}(\tau^S(x,y)) \in \mathcal{L}'_x \cup \mathcal{L}'_y\right]\right) \\
&\qquad\qquad \text{(by Equations (11) and (12))} \\
&\le \frac{3\varepsilon}{\log(I)} + \mathop{\mathbf{E}}_{(x,y)\leftarrow\mu}\left[\Pr\left[\mathrm{msg}_{\pi xy}(\tau^S(x,y)) \in \mathcal{L}'_x \cup \mathcal{L}'_y\right]\right] \\
&\le \frac{3\varepsilon}{\log(I)} + 0.25\varepsilon \qquad\qquad \text{(by Claim 31)} \\
&\le 0.3\varepsilon.
\end{aligned}$$

□

5.4.1 Proof of Lemma 15

Proof of Lemma 15. By Lemmas 30, 29 and 26,

$$\begin{aligned}
&\mathop{\mathbf{E}}_{(x,y)\leftarrow\mu}\left[\left\|\mathrm{msg}_{\pi xy}(\tau^{ST}(x,y)) - \pi(x,y)\right\|\right] \\
&\le \mathop{\mathbf{E}}_{(x,y)\leftarrow\mu}\left[\left\|\tau^{ST}(x,y) - \tau^S(x,y)\right\|\right] + \mathop{\mathbf{E}}_{(x,y)\leftarrow\mu}\left[\left\|\tau^S(x,y) - \tau(x,y)\right\|\right] \\
&\qquad + \mathop{\mathbf{E}}_{(x,y)\leftarrow\mu}\left[\left\|\mathrm{msg}_{\pi xy}(\tau(x,y)) - \pi(x,y)\right\|\right] \le \varepsilon/2.
\end{aligned}$$

□

Acknowledgements

We thank Avi Wigderson for numerous helpful discussions.

References

[BBCR10] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. In *STOC*, pages 67–76, 2010. 1, 2, 3, 4, 10, 19, 29, 30, 43

[BBK+13] Joshua Brody, Harry Buhrman, Michal Koucký, Bruno Loff, Florian Speelman, and Nikolay K. Vereshchagin. Towards a reverse newman's theorem in interactive information complexity. In *CCC*, pages 24–33, 2013. 2

[BMY15] Balthazar Bauer, Shay Moran, and Amir Yehudayoff. Internal compression of protocols to entropy. In *APPROX/RANDOM*, pages 481–496, 2015. 2

[BR11] Mark Braverman and Anup Rao. Information equals amortized communication. In *FOCS*, pages 748–757, 2011. 2, 5, 20

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

[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. 1

[Gan12] Anat Ganor. *Efficient Communication (Master's Thesis)*. Weizmann Institute of Science, 2012. 13

[GKR14] Anat Ganor, Gillat Kol, and Ran Raz. Exponential separation of information and communication. In *FOCS*, pages 176–185, 2014. 2, 3

[GKR15a] Anat Ganor, Gillat Kol, and Ran Raz. Exponential separation of communication and external information. *Electronic Colloquium on Computational Complexity (ECCC)*, 22:88, 2015. 2, 3

[GKR15b] Anat Ganor, Gillat Kol, and Ran Raz. Exponential separation of information and communication for boolean functions. In *STOC*, pages 557–566, 2015. 2, 3

[HJMR10] Prahladh Harsha, Rahul Jain, David A. McAllester, and Jaikumar Radhakrishnan. The communication complexity of correlation. *IEEE Transactions on Information Theory*, 56(1):438–449, 2010. 2

[Huf52] David A Huffman. A method for the construction of minimum redundancy codes. *proc. IRE*, 40(9):1098–1101, 1952. 1

[RR15] Sivaramakrishnan Natarajan Ramamoorthy and Anup Rao. How to compress asymmetric communication. In *CCC*, pages 102–123, 2015. 2

[RS15] Anup Rao and Makrand Sinha. Simplified separation of information and communication. *Electronic Colloquium on Computational Complexity (ECCC)*, 22:57, 2015. 2, 3

[Sha48] C. E. Shannon. A mathematical theory of communication. *The Bell Systems Technical Journal*, 27:July 379–423, October 623–656, 1948. 1

A Proof of Lemma 19

This section is devoted to proving Lemma 19. We prove the first equation in the lemma. That is, we prove that

$$\Pr_{w \leftarrow D^B_{\hat{v}xy}}\left[\frac{D_{vx}(w)}{D_v(w)} \geq t\right] \leq \frac{\alpha}{\gamma_{\hat{v}xy}}.$$

The second equation can be shown in a similar way. The proof follows the lines of the proof of Claim 8.9 in [BBCR10] and uses the following generalization of Azuma's inequality proven in [BBCR10].

Lemma 32 (Theorem A.1 in [BBCR10])**.** *Let $T_1, \ldots, T_k$ be real valued random variables such that for every $i \in [k]$, we have $\mathbf{E}[T_i | T_{i-1}, \ldots, T_1] \leq 0$. Set $A_i = (\sup(T_i) - \inf(T_i) \mid T_{i-1}, \ldots, T_1)^2$. Then, if $\sum_{i=1}^k A_i \leq c$, for every $\alpha > 0$,*

$$\Pr\left[\sum_{i=1}^k T_i \geq \alpha\right] \leq \exp(-2\alpha^2/c).$$

Let $\mathcal{C}'$ be the following set: For every $w \in \mathcal{C}_{vx}$, if $w \in \mathcal{V} \setminus \mathcal{C}^{\min}_{vxy}$, add w to $\mathcal{C}'$. If $w \in \mathcal{C}^{\min}_{vxy}$, add w's children to $\mathcal{C}'$. Since $\mathcal{C}_{vx}$ is a frontier with respect to v, so is $\mathcal{C}'$. In addition, $\mathcal{C}^{\min}_{vxy} < \mathcal{C}' \leq \mathcal{C}^{\max}_{vxy}$. Consider the set $\mathcal{C} = (\mathcal{C}^B_{\hat{v}xy} \cap \mathcal{V}_{vx}) \cup (\mathcal{C}' \cap \mathcal{V}_{vy})$. By Lemmas 16 and 17, and by Proposition 12, $\mathcal{C}$ is a frontier with respect to v. Also note that every vertex in $\mathcal{C}$ is either in $\mathcal{V}_{vx}$ or its parent is in $\mathcal{V}_{vx}$. Let $\bar{D}$ be the distribution obtained by restricting the

function D_{vxy} to the frontier $\mathcal{C}$. We will prove

$$\Pr_{w\leftarrow\bar{D}}\left[\frac{D_{vx}(w)}{D_v(w)}\geq t\right]\leq\alpha.$$

The assertion then follows due to the following: Let $\mathcal{S}=\left\{w\in\mathcal{C}\mid\frac{D_{vx}(w)}{D_v(w)}\geq t\right\}$. Then

$$\Pr_{w\leftarrow D^B_{\hat{v}xy}}\left[\frac{D_{vx}(w)}{D_v(w)}\geq t\right]=D^B_{\hat{v}xy}\left(\mathcal{S}\cap(\mathcal{C}^B_{\hat{v}xy}\cap\mathcal{V}_{vx})\right)=\frac{\bar{D}\left(\mathcal{S}\cap(\mathcal{C}^B_{\hat{v}xy}\cap\mathcal{V}_{vx})\right)}{\gamma_{\hat{v}xy}}\leq\frac{\bar{D}(\mathcal{S})}{\gamma_{\hat{v}xy}}\leq\frac{\alpha}{\gamma_{\hat{v}xy}}.$$

Let W be a random variable distributed according to $\bar{D}$. Recall that we assume that the leaves of $\mathcal{T}$ are (all) in level d. Let $Z_{|v|+1},\ldots,Z_d$ be real valued random variables, such that for $i\in\{|v|+1,\ldots,d\}$, if $|W|\geq i$

$$Z_i=\log\left(\frac{O_{W_{\leq i-1}x}(W_i)}{O_{W_{\leq i-1}}(W_i)}\right).$$

If $|W|<i$, set $Z_i=0$. Let $Z_{|v|}=0$. Let $w\in\mathcal{V}\setminus\mathcal{L}$ and $i\in\{1,\ldots,d\}$. If $|w|\geq i$, denote

$$\mathbb{D}_{x,i-1}(w)=\mathbb{D}\left(O_{w_{\leq i-1}x}\|O_{w_{\leq i-1}}\right).$$

If $|w|<i$, denote $\mathbb{D}_{x,i-1}(w)=0$.

Let $i\in\{|v|+1,\ldots,d\}$. Let $w\in\mathcal{C}$ be such that $|w|\geq i$. We claim that

$$\mathbf{E}[Z_i\mid W_{\leq i-1}=w_{\leq i-1}]=\mathbb{D}_{x,i-1}(w),\tag{13}$$

due to the followings: By definition,

$$\mathbf{E}[Z_i\mid W_{\leq i-1}=w_{\leq i-1}]=\sum_{b\in\{0,1\}}\Pr[W_i=b\mid W_{\leq i-1}=w_{\leq i-1}]\cdot\log\left(\frac{O_{w_{\leq i-1}x}(b)}{O_{w_{\leq i-1}}(b)}\right).$$

By the definition of $\bar{D}$ and since $\mathcal{C}$ is a frontier, $\Pr[W_{\leq i-1}=w_{\leq i-1}]=D_{vxy}(w_{\leq i-1})$. Thus, for $b\in\{0,1\}$, it holds that $\Pr[W_i=b\mid W_{\leq i-1}=w_{\leq i-1}]=O_{w_{\leq i-1}xy}(b)$. If $w_{\leq i-1}\in\mathcal{V}^A$, then, by Equation (1), $\Pr[W_i=b\mid W_{\leq i-1}=w_{\leq i-1}]=O_{w_{\leq i-1}xy}(b)=O_{w_{\leq i-1}x}(b)$, and Equation (13) holds. If $w_{\leq i-1}\in\mathcal{V}^B$, then, by Equation (2), $O_{w_{\leq i}x}=O_{w_{\leq i}}$ and both sides of Equation (13) equal 0.

We also have that

$$\sum_{i=|v|+1}^{d}Z_i=\sum_{i=|v|+1}^{|W|}\log\left(\frac{O_{W_{\leq i-1}x}(W_i)}{O_{W_{\leq i-1}}(W_i)}\right)=\log\left(\frac{\prod_{i=|v|+1}^{|W|}O_{W_{\leq i-1}x}(W_i)}{\prod_{i=|v|+1}^{|W|}O_{W_{\leq i-1}}(W_i)}\right)=\log\left(\frac{D_{vx}(W)}{D_v(W)}\right).\tag{14}$$

Next, for $i\in\{|v|+1,\ldots,d\}$, we define

$$T_i=Z_i-\mathbf{E}[Z_i|Z_{i-1},\ldots,Z_{|v|}].$$

Fix $i \in \{|v|+1, \ldots, d\}$. Note that $\mathbf{E}[T_i | T_{i-1}, \ldots, T_1] = 0$. Fix $w \in \mathcal{C}$. By Equation (13), if $|w| \geq i-1$,

$$\begin{aligned}(T_i | W_{\leq i-1} = w_{\leq i-1}) &= (Z_i | W_{\leq i-1} = w_{\leq i-1}) - \mathbf{E}[Z_i | Z_{i-1}, \ldots, Z_1, W_{\leq i-1} = w_{\leq i-1}] \\ &= (Z_i | W_{\leq i-1} = w_{\leq i-1}) - \mathbf{E}[Z_i | W_{\leq i-1} = w_{\leq i-1}] \\ &= (Z_i | W_{\leq i-1} = w_{\leq i-1}) - \mathbb{D}_{x,i-1}(w).\end{aligned}$$

Since $\mathbb{D}_{x,i}(w) \geq 0$,

$$\begin{aligned}\sup(T_i | W_{\leq i-1} = w_{\leq i-1}) &\leq \max_{b \in \{0,1\}} \left\{ \log\left(\frac{O_{w_{\leq i-1}x}(b)}{O_{w_{\leq i-1}}(b)} \right) \right\}, \\ \inf(T_i | W_{\leq i-1} = w_{\leq i-1}) &\geq \min_{b \in \{0,1\}} \left\{ \log\left(\frac{O_{w_{\leq i-1}x}(b)}{O_{w_{\leq i-1}}(b)} \right) - \mathbb{D}_{x,i-1}(w) \right\}.\end{aligned}$$

Let $b \in \{0,1\}$. Recall from the beginning of Section 5 that π is β-smooth. By Definition 4,

$$O_{w_{\leq i-1}xy}(b) \in [1/2 - \beta, 1/2 + \beta].$$

Therefore,

$$\begin{aligned}O_{w_{\leq i-1}x}(b) &= \mathop{\mathbf{E}}_{y \leftarrow \mu^B}[O_{w_{\leq i-1}xy}(b)] \in [1/2 - \beta, 1/2 + \beta], \\ O_{w_{\leq i-1}}(b) &= \mathop{\mathbf{E}}_{x \leftarrow \mu^A}[O_{w_{\leq i-1}x}(b)] \in [1/2 - \beta, 1/2 + \beta].\end{aligned}$$

We claim that

$$\mathbb{D}_{x,i-1}(w) \leq 5\beta. \tag{15}$$

This follows from the fact that $O_{w_{\leq i-1}}(0), O_{w_{\leq i-1}x}(0) \in [1/2 - \beta, 1/2 + \beta]$, as the largest the divergence between two distributions that lie in this range can be is at most $\log\left(\frac{1/2+\beta}{1/2-\beta}\right) \leq \log(1 + 5\beta) \leq 5\beta$, where the last equality is since $\log(1+z) \leq z$ for $z \geq 0$.

Using Pinsker's inequality and since $\log(1+z) \leq z$ for $z \geq 0$,

$$\begin{aligned}&\sup(T_i | W_{\leq i-1} = w_{\leq i-1}) \leq \max_{b \in \{0,1\}} \left\{ \log\left(\frac{O_{w_{\leq i-1}}(b) + \left\| O_{w_{\leq i-1}x} - O_{w_{\leq i-1}} \right\|}{O_{w_{\leq i-1}}(b)} \right) \right\} \\ &\leq \max_{b \in \{0,1\}} \left\{ \log\left(1 + \frac{4\sqrt{\mathbb{D}_{x,i-1}(w)}}{O_{w_{\leq i-1}}(b)} \right) \right\} \leq \log\left(1 + 10\sqrt{\mathbb{D}_{x,i-1}(w)} \right) \leq 10\sqrt{\mathbb{D}_{x,i-1}(w)}.\end{aligned}$$

In addition, by Pinsker's inequality, Equation (15), and since $\log(1-z) \geq -2z$ for every

$z \in (0, 0.5)$,

$$\begin{aligned}
\inf(T_i | W_{\leq i-1} = w_{\leq i-1}) &\geq \max_{b\in\{0,1\}} \left\{ \log\left(\frac{O_{w_{\leq i-1}x}(b)}{O_{w_{\leq i-1}x}(b) + \left\| O_{w_{\leq i-1}x} - O_{w_{\leq i-1}} \right\|} \right) - \mathbb{D}_{x,i-1}(w) \right\} \\
&\geq \max_{b\in\{0,1\}} \left\{ \log\left(\frac{O_{w_{\leq i-1}x}(b)}{O_{w_{\leq i-1}x}(b) + 4\sqrt{\mathbb{D}_{x,i-1}(w)}} \right) - \mathbb{D}_{x,i-1}(w) \right\} \\
&\geq \log\left(1 - 10\sqrt{\mathbb{D}_{x,i-1}(w)}\right) - \mathbb{D}_{x,i-1}(w) \geq -30\sqrt{\mathbb{D}_{x,i-1}(w)}.
\end{aligned}$$

Recall that the set $\mathcal{V}_{vx}$ was defined in Section 3.4 as the set of all $w' \in \mathcal{V}(v) \setminus \mathcal{L}$ satisfying $\mathbb{D}_{vw'_{\leq |w'|-1}x} < \beta$. Since for every $w \in \mathcal{C}$ it holds that either w or its parent is in $\mathcal{V}_{vx}$, since π is β-smooth, and by Equation (15), $\sum_{i=|v|+1}^{d} \mathbb{D}_{x,i-1}(w) = \mathbb{D}_{vwx} \leq 20\beta$. Therefore, for $w \in \mathcal{C}$,

$$\sum_{i=|v|+1}^{d} \left(\sup(T_i) - \inf(T_i) \mid W_{\leq i-1} = w_{\leq i-1}\right)^2 \leq \sum_{i=|v|+1}^{d} 1600 \cdot \mathbb{D}_{x,i-1}(w) \leq 10^5\beta. \tag{16}$$

For $w \in \mathcal{C}$, by Equations (13) and (14),

$$\begin{aligned}
\left(\sum_{i=|v|+1}^{d} T_i \middle| W = w \right) &= \left(\sum_{i=|v|+1}^{d} Z_i \middle| W = w \right) - \sum_{i=|v|+1}^{d} \mathbf{E}\left[Z_i | W_{\leq i-1} = w_{\leq i-1} \right] \qquad (17) \\
&= \log\left(\frac{D_{vx}(w)}{D_v(w)} \right) - \sum_{i=|v|+1}^{d} \mathbb{D}_{x,i-1}(w) \geq \log\left(\frac{D_{vx}(w)}{D_v(w)} \right) - 20\beta.
\end{aligned}$$

The assertion follows as

$$\begin{aligned}
&\Pr_{w\leftarrow\bar{D}}\left[\frac{D_{vx}(w)}{D_v(w)} \geq t \right] = \Pr_{w\leftarrow\bar{D}}\left[\log\left(\frac{D_{vx}(w)}{D_v(w)} \right) \geq \log(t) \right] \\
&\leq \Pr\left[\sum_{i=|v|+1}^{d} T_i \geq \log(t) - 20\beta \right] && \text{(by Equation (17))} \\
&\leq \exp\left(-\frac{(\log(t) - 20\beta)^2}{10^5\beta} \right). && \text{(by Lemma 32 and Equation (16))}
\end{aligned}$$

ECCC ISSN 1433-8092
http://eccc.hpi-web.de