Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 54 (2017)

Lifting randomized query complexity to randomized communication complexity

Anurag Anshu[†] Naresh B. Goud[†] Rahul Jain [*] Srijita Kundu[†]

Priyanka Mukhopadhyay[†]

April 7, 2017

Abstract

We show that for any (partial) query function $f : \{0,1\}^n \to \{0,1\}$, the randomized communication complexity of f composed with Index_m^n (with $m = \mathrm{poly}(n)$) is at least the randomized query complexity of f times $\log n$. Here $\mathrm{Index}_m : [m] \times \{0,1\}^m \to \{0,1\}$ is defined as $\mathrm{Index}_m(x, y) = y_x$ (the xth bit of y).

Our proof follows on the lines of Raz and Mckenzie [RM99] (and its generalization due to [GPW15]), who showed a lifting theorem for deterministic query complexity to deterministic communication complexity. Our proof deviates from theirs in an important fashion that we consider partitions of rectangles into many sub-rectangles, as opposed to a particular sub-rectangle with desirable properties, as considered by Raz and McKenzie. As a consequence of our main result, some known separations between quantum and classical communication complexities follow from analogous separations in the query world.

1 Introduction

Communication complexity and query complexity are two concrete models of computation which are very well studied. In the communication model there are two parties Alice, with input x and Bob, with input y, and they wish to compute a joint function $f(x, y)$ of their inputs. In the query model one party Alice tries to compute a function $f(x)$ by querying bits of a database string x. There is a natural way in which a query protocol can be viewed as a communication protocol between Alice, with no input, and Bob, with input x, in which the only communication allowed is queries to the bits of x and answers to these queries. Given this, we can (informally) view the query model as a "simpler" sub-model of the communication model. Indeed several results in query complexity are easier to argue and obtain than the corresponding results in communication complexity. One interesting technique that is often employed with great success is to first show a result in the query model and then "lift" it to a result in the communication model via some "lifting theorem".

One of the first such lifting theorems was shown by Raz and McKenzie [RM99] (and its generalization by [GPW15]). For a (partial) query function $f : \{0,1\}^n \to \{0,1\}$ and a communication function $g : \{0,1\}^m \times \{0,1\}^k \to \{0,1\}$ let the composed function $f \circ g^n$ be defined as $f \circ g^n((x_1, y_1), \ldots, (x_n, y_n)) = f(g(x_1, y_1), \ldots, g(x_n, y_n))$. Raz and McKenzie [RM99] (and the generalization due to [GPW15]) showed that for every query function $f : \{0,1\}^n \to \{0,1\}$ the

[*] Centre for Quantum Technologies, National University of Singapore and MajuLab, UMI 3654, Singapore.
[†] Center for Quantum Technologies, National University of Singapore, Singapore.

ISSN 1433-8092

deterministic communication complexity of f composed with Index_m (with $m = \mathrm{poly}(n)$) is at least the deterministic query complexity of f times $\log n$. Here $\mathrm{Index}_m : [m] \times \{0,1\}^m \to \{0,1\}$ is defined as $\mathrm{Index}_m(x, y) = y_x$ (the xth bit of y). Subsequently several lifting theorems for different complexity measures have been shown, example lifting *approximate-degree* to *approximate-rank* [She11] and *approximate junta-degree* to *smooth-corruption-bound* [GLM+15] etc.

Our result

In this work we show lifting of (bounded error) randomized query complexity to (bounded error) randomized communication complexity. For a (partial) query function $f : \{0,1\}^n \to \{0,1\}$ let the randomized query complexity with (worst-case) error $\varepsilon > 0$ of f be denoted $\mathrm{R}_\varepsilon(f)$. Similarly for a communication function $g : \{0,1\}^m \times \{0,1\}^k \to \{0,1\}$, let the randomized communication complexity with (worst-case) error $\varepsilon > 0$ of g be denoted $\mathrm{R}_\varepsilon(g)$. We show the following.

Theorem 1. *For all (partial) functions* $f : \{0,1\}^n \to \{0,1\}$,

$$\mathrm{R}_{1/4}(f \circ \mathrm{Index}_m^n) = \Omega(\mathrm{R}_{1/3}(f) \cdot \log n),$$

where $m = \mathrm{poly}(n)$ [1].

On the other hand it is easily seen with a simple simulation of a query protocol using a communication protocol that $\mathrm{R}_{1/3}(f \circ \mathrm{Index}_m^n) = O(\mathrm{R}_{1/3}(f) \cdot \log m)$. This implies $\mathrm{R}_{1/3}(f \circ \mathrm{Index}_m^n) = \Theta(\mathrm{R}_{1/3}(f) \cdot \log n)$ with $m = \mathrm{poly}(n)$.

Our result implies a recent result of [ABBD+16] where they exhibited a power 2.5 separation between classical randomized and quantum communication complexities for a total function. It also implies exponential separation between two-round quantum communication complexity and randomized communication complexity first shown by [Raz99]

Our techniques

Our techniques are largely based on the techniques of Raz and McKenzie [RM99] as presented in [GPW15] with an important modification to deal with distributional error protocols instead of deterministic protocols. Let $\mathcal{T}$ be a deterministic communication protocol tree for $f \circ \mathrm{Index}_m^n$ (with $m = \mathrm{poly}(n)$). We use this to create a randomized query protocol Π (see. Figure 1) for f. Let z be an input for which we are supposed to output $f(z)$. We start with the root of $\mathcal{T}$ and continue to simulate $\mathcal{T}$ (using randomness) till we find a co-ordinate $i \in [n]$ where $\mathcal{T}$ has *worked enough* so that $\mathrm{Index}_m(x_i, y_i)$ is becoming (only slightly) *determined*. Using the properties of Index_m we conclude that $\mathcal{T}$ must have communicated $O(\log n)$ bits by now. We go ahead a query z_i (the ith bit of z) and *synchronize* with z_i, that is go to the appropriate sub-event of the current node in $\mathcal{T}$ consistent with z_i. We then continue to simulate $\mathcal{T}$. This way we do (in expectation) one query in Π for $O(\log n)$ communication bits in $\mathcal{T}$. On reaching a leaf of $\mathcal{T}$ we make the same output as $\mathcal{T}$. This output is correct with high probability since the unqueried bits are sufficiently *un-determined*.

The synchronizing of $\mathcal{T}$ with z_i was done by Raz and McKenzie [RM99] via a *Projection Lemma* by going to a "sub-event" (of small probability) of the current node in $\mathcal{T}$. They could afford to do so since $\mathcal{T}$ was a deterministic protocol and hence it was correct *everywhere*. On the other hand we are forced to work with a "partition" of the node into sub-events where each sub-event is consistent with either z_i being 0 or z_i being 1. This allows us to move according to the "flow" of $\mathcal{T}$ so that we can "capture" correctness of $\mathcal{T}$ wherever it has to offer. This requires us to show a different *Partition Lemma* which works in place of the Projection Lemma of Raz and McKenzie.

[1] We state our result for Boolean functions f, however it holds for general relations.

2 Preliminaries

In this section, we present some notations and basic lemmas needed for the proof of our main result.

Let $f : \{0,1\}^n \to \{0,1\}$ be a (partial) function. Let $\varepsilon > 0$ be an error parameter. Let the randomized query complexity, denoted $\mathrm{R}_\varepsilon(f)$, be the maximum number of queries made by the best randomized query protocol computing f with error at most ε on any input $x \in \{0,1\}^n$. Let θ be a distribution on $\{0,1\}^n$. Let the distributional query complexity, denoted $\mathrm{D}^\theta_\varepsilon(f)$, be the maximum number of queries made by the best deterministic query protocol computing f with average error at most ε under θ. The distributional and randomized query complexities are related by the following Yao's Lemma.

Fact 2 (Yao's Lemma)**.** *Let $\varepsilon > 0$. We have $\mathrm{R}_\varepsilon(f) = \max_\theta \mathrm{D}^\theta_\varepsilon(f)$.*

Similarly, we can define randomized and distributional communication complexities with a similar Yao's Lemma relating them.

Let λ be a *hard* distribution on $\{0,1\}^n$ such that $\mathrm{D}^\lambda_{1/3}(f) = \mathrm{R}_{1/3}(f)$, as guaranteed by Yao's Lemma. Let $m = O(n^{100})$ and let $\mathrm{Bal}_m \subset \{0,1\}^m$ be the set of all strings of length m with equal number of 0's and 1's. Observe that $|\mathrm{Bal}_m| = \binom{m}{m/2} \geq 2^m/\sqrt{m}$. The notation Bal^n_m will refer to the set $\mathrm{Bal}_m \times \mathrm{Bal}_m \times \ldots \mathrm{Bal}_m$. Let $\mathrm{Index}_m : [m] \times \mathrm{Bal}_m \to \{0,1\}$ be defined as $\mathrm{Index}_m(x,y) = y_x$ (the xth bit of y). Consider the following *lifted* distribution for the composed function $f \circ \mathrm{Index}^n_m$: $\mu(x,y) = \lambda(G(x,y))/|G^{-1}(G(x,y))|$, where $G := \mathrm{Index}^n_m$. We observe for this distribution that $\mu(x)$ and $\mu(y)$ are uniform in their support.

Let Alice and Bob's inputs for the composed function be respectively $x = (x_1, \ldots, x_n) \in [m]^n$ and $y = (y_1, \ldots, y_n) \in \mathrm{Bal}^n_m$.

We use the following notation, which coincides largely with the notation used in [GPW15].

- For a node v in a communication protocol tree, let $R^v = X^v \times Y^v$ denote its associated rectangle. If Alice or Bob send the bit b at v, let v_b be the corresponding child of v and $X^{v,b} \subseteq X^v$ and $Y^{v,b} \subseteq Y^v$ be the set of inputs of Alice and Bob respectively on which they do this.
- For an interval $I \in [n]$, let $\bar{I}$ be $[n] \setminus I$.
- For a string $x \in [m]^n$ and an interval $I \subset [n]$, let x_I be the restriction of x to the interval I. We use shorthand x_i for $x_{\{i\}}$. We use similar notation for string $y \in \mathrm{Bal}^n_m$. For a set $A \subset [m]^n$, let $A_I := \{x_I : x \in A\}$ be the restriction of A to the interval I. Moreover, we use A_{x_I} to denote the subset of A in which the bits at coordinates inside I are fixed to those in x_I.
- For a set $U \subseteq [m]$, let $A|_{U,i} := \{x \in A : x_i \in U\}$ (we will omit the substript i from $A|_{U,i}$ when it is clear from context).
- For $A \subseteq [m]^n$, an interval J, let

$$p_A(x_J) := \frac{|\{x_{[n]\setminus J} \in [m]^{n-|J|} : x_J \circ x_{\bar{J}} \in A\}|}{|A|} = \frac{|A_{x_J}|}{|A|},$$

  where $x_J \circ x_{\bar{J}}$ denotes the concatenation of x_J and $x_{\bar{J}}$. We use similar notation for A_I and x_J whenever $J \subseteq I$.
- Fix an $\alpha \in (0, \infty)$. For a probability distribution $p(x,y)$, define the Conditional Renyi Entropy of order α as

$$\mathrm{H}_\alpha(X|Y)_p := \frac{\alpha}{1-\alpha} \log \left(\sum_y p(y) \left(\sum_x p(x|y)^\alpha \right)^{\frac{1}{\alpha}} \right).$$

Fix $\varepsilon = \frac{1}{\log m}$. Let

$$p^{\varepsilon}(A, J) = \left(\sum_{x_J} p_A(x_J)^{1+\varepsilon} \right)^{\frac{1}{\varepsilon}}.$$

- For $A \subseteq [m]^n$, interval I and index $i \in I$,

  $$\mathrm{AvgDeg}_i(A, I) = \sum_{x_{I\setminus\{i\}}} p_A(x_{I\setminus\{i\}}) p^{\varepsilon}(A_{x_{I\setminus\{i\}}}, \{i\}), \quad \mathrm{MinDeg}_i(A, I) = \max_{x_{I\setminus\{i\}}} p^{\varepsilon}(A_{x_{I\setminus\{i\}}}, \{i\}).$$

- $A \subseteq [m]^n$ is called *thick* for $I \subseteq [n]$ if $\mathrm{MinDeg}_i(A, I) \leq m^{-\frac{17}{20}}$ for all $i \in I$.
- For $i \in [n]$, a set $B \subset \mathrm{Bal}_m^n$ and a string $\eta \in \mathrm{Bal}_m$, let $B_{\eta,i} := \{y \in B : y_i = \eta\}$.
- For an interval I, $y \in \mathrm{Bal}_m^{|I|}$ and $z \in \{0,1\}^{|I|}$, define $U(y, z) := \{x \in [m]^{|I|} : \mathrm{Index}_m^{|I|}(x, y) = z\}$.

Following fact is well known.

Fact 3 (Chernoff bound)**.** *Let $X_1, X_2 \ldots X_m$ be random variables such that $0 < X_i < b$. Let $X = \sum_i X_i$ and $\mu = \mathbb{E}(X)$. Then for $\delta > 0$, it holds that*

$$\Pr[|X - \mu| \geq \mu\delta] \leq 2e^{-\frac{\delta^2\mu^2}{mb^2}}.$$

Following is the chain rule for Conditional renyi entropy [Dup15].

Fact 4 (Chain rule for conditional Renyi entropy, [Dup15])**.** *Let XY be jointly correlated random variables with probability distribution $p(x, y)$. Let α, β, γ be such that $\frac{\alpha}{\alpha-1} = \frac{\beta}{\beta-1} + \frac{\gamma}{\gamma-1}$. If $(\alpha - 1)(\beta - 1)(\gamma - 1) < 0$, then*

$$\mathrm{H}_{\alpha}(XY)_p \leq \mathrm{H}_{\beta}(X|Y)_p + \mathrm{H}_{\gamma}(Y)_p.$$

We will need the following claims. First claim is the application of above fact.

Claim 5. *Let XY be jointly correlated random variables with probability distribution $p(x, y)$. It holds that*

$$\mathrm{H}_{1+2\varepsilon}(XY) \leq \mathrm{H}_{1+\varepsilon}(X|Y) + \mathrm{H}_{1+2\varepsilon}(Y) + 6\varepsilon \log^2 m,$$

for the choice of ε.

Proof. Setting $\alpha = 1 + 2\varepsilon, \beta = 1 + \varepsilon, \gamma = \frac{1}{1+2\varepsilon}$ in Fact 4 (the constraints are satisfied), we conclude that

$$\mathrm{H}_{1+2\varepsilon}(XY)_p \leq \mathrm{H}_{1+\varepsilon}(X|Y)_p + \mathrm{H}_{\frac{1}{1+2\varepsilon}}(Y)_p \leq \mathrm{H}_{1+\varepsilon}(X|Y)_p + \mathrm{H}_{1-2\varepsilon}(Y)_p,$$

where the last line follows from the fact that Renyi entropy is monotonically decreasing in Renyi parameter. Now, consider

$$\begin{aligned}
\mathrm{H}_{1-2\varepsilon}(Y)_p - \mathrm{H}_{1+2\varepsilon}(Y)_p &= \frac{1}{2\varepsilon}\log\left(\sum_y p^{1-2\varepsilon}(y)\cdot\sum_y p^{1+2\varepsilon}(y)\right)\\
&= \frac{1}{2\varepsilon}\log\left(\sum_{y,y'} p(y)p(y')e^{2\varepsilon\log\frac{p(y')}{p(y)}}\right)\\
&= \frac{1}{2\varepsilon}\log\left(\sum_{k=0}^{\infty}\frac{(2\varepsilon)^k}{k!}\sum_{y,y'} p(y)p(y')\left(\log\frac{p(y')}{p(y)}\right)^k\right)\\
&\overset{a}{=} \frac{1}{2\varepsilon}\log\left(1+\sum_{k=2}^{\infty}\frac{(2\varepsilon)^k}{k!}\sum_{y,y'} p(y)p(y')\left(\log\frac{p(y')}{p(y)}\right)^k\right)\\
&\le \frac{1}{2\varepsilon}\log\left(1+\sum_{k=2}^{\infty}\frac{(2\varepsilon)^k}{k!}\sum_{y,y'} p(y)p(y')\left(\log\frac{1}{p(y)}\right)^k\right)\\
&= \frac{1}{2\varepsilon}\log\left(1+\sum_{k=2}^{\infty}\frac{(2\varepsilon)^k}{k!}\sum_{y} p(y)\left(\log\frac{1}{p(y)}\right)^k\right)\\
&\le \frac{1}{2\varepsilon}\log\left(1+\sum_{k=2}^{\infty}\frac{(2\varepsilon)^k}{k!}\log^k(m)\right)\\
&\le \frac{1}{2\varepsilon}\log\left(e^{2\varepsilon\log m}-2\varepsilon\log m\right)\overset{b}{\le} 6\varepsilon\log^2 m,
\end{aligned}$$

where (a) follows since $\sum_{y,y'} p(y)p(y')\log\frac{p(y')}{p(y)} = 0$, and (b) follows from the choice of ε. This completes the proof. □

Claim 6. *Let $A \in [m]^n$ be a set and $A' \subseteq A$ be its subset. Let I be an interval and $i \in I$ be an index. Then*

$$\mathrm{AvgDeg}_i(A', I) \le \mathrm{AvgDeg}_i(A, I)\cdot\frac{|A'|}{|A|}.$$

Proof. We expand

$$\begin{aligned}
\mathrm{AvgDeg}_i(A, I) &= \sum_{x_{I\setminus\{i\}}} p_A(x_{I\setminus\{i\}})p^{\varepsilon}(A_{x_{I\setminus\{i\}}},\{i\})\\
&= \sum_{x_{I\setminus\{i\}}}\frac{|A_{x_{I\setminus\{i\}}}|}{|A|}\cdot\max_{x_i}\frac{|A_{x_{I\setminus\{i\}}\circ x_i}|}{|A_{x_{I\setminus\{i\}}}|}\\
&= \sum_{x_{I\setminus\{i\}}}\frac{\max_{x_i}|A_{x_{I\setminus\{i\}}\circ x_i}|}{|A|}.
\end{aligned}$$

Similarly,

$$\mathrm{AvgDeg}_i(A', I) = \sum_{x_{I\setminus\{i\}}}\frac{\max_{x_i}|A'_{x_{I\setminus\{i\}}\circ x_i}|}{|A'|}.$$

But $|A'_{x_{I\setminus\{i\}}\circ x_i}| \le |A_{x_{I\setminus\{i\}}\circ x_i}|$ as $A' \subseteq A$. Thus,

$$\mathrm{AvgDeg}_i(A', I) \le \sum_{x_{I\setminus\{i\}}}\frac{\max_{x_i}|A_{x_{I\setminus\{i\}}\circ x_i}|}{|A'|} = \frac{|A|}{|A'|}\mathrm{AvgDeg}_i(A, I).$$

This completes the proof. □

Claim 7. *Let $A \in [m]^n$ be thick in the interval $I \cup \{i\}$, where $i \notin I$. Then A is also thick in the interval I.*

Proof. Choose any index $j \in I$. For every $x_{I\setminus\{j\}}$, we have that

$$p_{A_{x_{I\setminus\{j\}}}}(x_j) = \sum_{x_i} p_{A_{x_{I\setminus\{j\}}}}(x_i) p_{A_{x_{I\setminus\{j\}\cup\{i\}}}}(x_j).$$

Since A is thick in the interval $I \cup \{i\}$, we have $p_{A_{x_{I\setminus\{j\}\cup\{i\}}}}(x_j) \leq m^{-17/20}$ for all x_j. This implies that $p_{A_{x_{I\setminus\{j\}}}}(x_j) \leq m^{-17/20}$. This proves the claim. □

Claim 8. *Consider a tree with nodes and weighted directed edges such that for every node, the sum of weights of edges going to its children sum to 1. Call a non-leaf node aborted if it has no children. For any node, let the sum of weights of the edges going to aborted children be at most δ. Let the depth of the tree be d. Consider a random walk that starts from the root and goes to the children according to the weights of the edges. Then the overall probability of abort is at most $\delta \cdot d$.*

Proof. We construct a new tree in which nodes at a particular level which do not abort are coarse-grained into a single node and the aborting nodes are coarse grained into another node (which we again call abort node). For this tree, the probability of a node having an aborted child is still at most δ and the overall probability of abort is at least as large as in the original tree, which is

$$\delta + (1-\delta)\cdot\delta + (1-\delta)^2\cdot\delta \ldots + (1-\delta)^{d-1}\delta \leq d\delta.$$

This completes the proof. □

A Thickness Lemma was shown in [GPW15]. We prove our own version below.

Lemma 9 (Thickness Lemma, [GPW15])**.** *If $n \geq 2$ and $A \subseteq [m]^n$ is such that $\mathrm{AvgDeg}_i(A, I) \leq \frac{1}{d}$ for all $i \in I$ for some $I \subseteq [n]$, then there exists $A' \subseteq A$ such that*

(i) $|A'| \geq \left(1 - \frac{1}{n^2}\right)|A|$,

(ii) $\mathrm{MinDeg}_i(A', I) \leq \frac{n^3}{d}$ *for all $i \in I$.*

Proof. For each $i \in I$, let Bad_i be the set of all $x_{I\setminus\{i\}}$ such that $p^{\varepsilon}(A_{x_{I\setminus\{i\}}}, \bar{I} \cup \{i\}) \geq \frac{n^3}{d}$. By Markov's inequality, we have that

$$\begin{aligned} \frac{1}{n^3} \quad &\geq \quad \sum_{x_{I\setminus\{i\}} \in \mathrm{Bad}_i} p_A(x_{I\setminus\{i\}}) = \sum_{x_{I\setminus\{i\}} \in \mathrm{Bad}_i} \frac{|A_{x_{I\setminus\{i\}}}|}{|A|} \\ &= \quad \frac{|\{x \in A : x_{\bar{I}\cup\{i\}} \in \mathrm{Bad}_i\}|}{|A|} \end{aligned}$$

This implies that

$$\frac{|\{x \in A : x_{\bar{I}\cup\{i\}} \in \cup_{i\in I}\mathrm{Bad}_i\}|}{|A|} \leq \frac{1}{n^2}.$$

We define $A' := \{x \in A : x_{\bar{I}\cup\{i\}} \notin \cup_{i\in I}\mathrm{Bad}_i\}$. This completes the construction of A'. □

Now we prove a Partition Lemma which helps to partition a current node when the algorithm Π performs a query.

Lemma 10 (Partition Lemma)**.** *For $I \subseteq [n]$, let $B \subseteq \mathrm{Bal}_m^{|I|}$ be such that $\frac{|B|}{|\mathrm{Bal}_m^{|I|}|} \geq 2^{-n^2}$. If $A \subseteq [m]^{|I|}$ is thick in the interval I, then for all $i \in I$ and for all $z_i \in \{0,1\}$ with probability at least $1 - 2^{-m^{1/4}}$ when $y_i = \eta$ is drawn randomly from the distribution $\frac{|B_{\eta,i}|}{|B|}$ it holds that*

$$\frac{1}{2} + n^{-5} \geq \sum_{x_i \in U(y_i,z_i)} p_{A_{x_{I\setminus\{i\}}}}(x_i) \geq \frac{1}{2} - n^{-5} \quad \forall x_{I\setminus\{i\}} \in [m]^{|I|-1}.$$

We say a $y_i \in \mathrm{Bad}(A,B,i)$ if it does not satisfy above property or $\frac{|B_{y_i}|}{|\mathrm{Bal}_m^{|I|-1}|} < 2^{-n^2}$. Then, for all $y_i \notin \mathrm{Bad}(A,B,i)$, we have that $p_{A|_{U(y_i,z_i)}}(x_{I\setminus\{i\}}) \in [1 \pm 2n^{-5}]p_A(x_{I\setminus\{i\}})$.

Proof. Fix an index $i \in I$ and $z_i \in \{0,1\}$. Fix an δ to be chosen later. For arguments below y is fixed outside I. Let a string y_i satisfy the property $P(x_{I\setminus\{i\}})$ if it holds that

$$\frac{1}{2} + \delta \geq \sum_{x_i \in U(y_i,z_i)} p_{A_{x_{I\setminus\{i\}}}}(x_i) \geq \frac{1}{2} - \delta.$$

For a fixed $x_{I\setminus\{i\}}$, we have $p^{max}(A_{x_{I\setminus\{i\}}}, \{i\}) \leq m^{-17/20}$. We can rewrite

$$\sum_{x_i \in U(y_i,z_i)} p_{A_{x_{I\setminus\{i\}}}}(x_i) = \sum_{x_i} |(y_i)_{x_i} - z_i| p_{A_{x_{I\setminus\{i\}}}}(x_i).$$

If y_i is chosen uniformly from $\{0,1\}^m$, then the expectation value of $\sum_{x_i} |(y_i)_{x_i} - z_i| p_{A_{x_{I\setminus\{i\}}}}(x_i)$ is $\frac{1}{2}$, as expectation value of $|(y_i)_{x_i} - z_i|$ is $\frac{1}{2}$. Thus, using Chernoff bound (Fact 3) and the fact that $0 < |(y_i)_{x_i} - z_i| p_{A_{x_{I\setminus\{i\}}}}(x_i) \leq m^{-17/20}$, we have

$$\Pr[|\sum_{x_i}((y_i)_{x_i} - z_i)p_{A_{x_{I\setminus\{i\}}}}(x_i) - \frac{1}{2}| > \delta] \leq 2e^{-\frac{\delta^2}{4m\cdot m^{-17/10}}} = 2e^{-\frac{m^{7/10}\cdot\delta^2}{4}}.$$

Now for a fixed $y_{I\setminus\{i\}} \in \mathrm{Bal}_m^{|I|-1}$, consider $B_{y_{I\setminus\{i\}}}$ at the i-th block. Above argument implies that the total number of strings in the i-th block of $B_{y_{I\setminus\{i\}}}$ (even after an over counting to include strings not in Bal_m), such that the property $P(x_{I\setminus\{i\}})$ does not hold, is at most

$$2^{m+1} \cdot e^{-\frac{m^{7/10}\cdot\delta^2}{4}}.$$

This implies, for a y drawn from $B_{y_{I\setminus\{i\}}}$, the probability that y_i does not satisfy the property $P(x_{I\setminus\{i\}})$ is at most $e^{-\frac{m^{7/10}\cdot\delta^2}{4}} \frac{2^{m+1}}{|B_{y_{I\setminus\{i\}}}|}$. Thus the probability that y_i does not satisfy any of the properties $P(x_{I\setminus\{i\}})$ for all $x_{I\setminus\{i\}}$ is at most $m^n \cdot e^{-\frac{m^{7/10}\cdot\delta^2}{4}} \frac{2^{m+1}}{|B_{y_{I\setminus\{i\}}}|}$. Hence, the probability that y drawn according to the distribution given in the statement of the lemma satisfies $P(x_{I\setminus\{i\}})$ for all $x_{I\setminus\{i\}}$ is at least

$$\begin{aligned} 1 - m^n \cdot e^{-\frac{m^{7/10}\cdot\delta^2}{4}} \sum_{y_{I\setminus\{i\}}} \frac{2^{m+1}\Pr[y_{I\setminus\{i\}}]}{|B_{y_{I\setminus\{i\}}}|} &\geq 1 - 2^{n\log m + m - m^{7/20}\cdot\delta^2} \sum_{y_{I\setminus\{i\}}} \frac{|B_{y_{I\setminus\{i\}}}|}{|B|\cdot|B_{y_{I\setminus\{i\}}}|} \\ &= 1 - 2^{n\log m + m - m^{7/20}\cdot\delta^2} \sum_{y_{I\setminus\{i\}}} \frac{1}{|B|}. \end{aligned}$$

Since number of strings $y_{I\setminus\{i\}} \leq |\mathrm{Bal}_m|^{|I|-1}$ and $|B| \geq |\mathrm{Bal}_m|^{|I|} \cdot 2^{-n^2}$, we conclude that above probability is at least

$$1 - 2^{n\log m + m - m^{7/20}\cdot\delta^2} \cdot |\mathrm{Bal}_m|^{-1} \cdot 2^{n^2} \geq 1 - \sqrt{m} \cdot 2^{2n^2 - m^{7/20}\cdot\delta^2}.$$

Choosing $\delta = m^{-1/20}$, we conclude the first part of the proof.
For the second part, we observe that

$$\begin{aligned}
p_{A|_{U(y_i,z_i)}}(x_{I\setminus\{i\}}) &= \frac{|(A|_{U(y_i,z_i)})_{x_{I\setminus\{i\}}}|}{|A|_{U(y_i,z_i)}|} \\
&= \frac{\sum_{x_i\in U(y_i,z_i)} |A_{x_{I\setminus\{i\}}\circ x_i}|}{|A|_{U(y_i,z_i)}|} \\
&= \frac{|A_{x_{I\setminus\{i\}}}|\sum_{x_i\in U(y_i,z_i)} p_{A_{x_{I\setminus\{i\}}}}(x_i)}{|A|_{U(y_i,z_i)}|} \\
&= \frac{|A_{x_{I\setminus\{i\}}}|\sum_{x_i\in U(y_i,z_i)} p_{A_{x_{I\setminus\{i\}}}}(x_i)}{\sum_{x_{I\setminus\{i\}}}|A_{x_{I\setminus\{i\}}}|\sum_{x_i\in U(y_i,z_i)} p_{A_{x_{I\setminus\{i\}}}}(x_i)} \\
&\in [1\pm 2n^{-5}]\frac{|A_{x_{I\setminus\{i\}}}|}{\sum_{x_{I\setminus\{i\}}}|A_{x_{I\setminus\{i\}}}|} \in [1\pm 2n^{-5}]p_A(x_{I\setminus\{i\}}),
\end{aligned}$$

where in the last step we have used the first part of the lemma. This proves the second part. □

Following is a Uniformity Lemma, which ensures that as long as B is a large set, the distribution μ on Alice's side, conditioned on Bob's side being in set B, behaves like a uniform distribution. The proof of the lemma is deferred to Appendix A.

Lemma 11 (Uniformity lemma)**.** *Fix an interval $I \subset [n]$, a subset $B \subseteq \mathrm{Bal}_m^n$ and a string $z \in \{0,1\}^n$. Let B have the additional property that $B_{\bar{I}}$ is a fixed string and $\frac{|B|}{|\mathrm{Bal}_m^{|I|}|} \geq 2^{-n^2}$. Define the distribution*

$$\sigma^{I,B,z}(x') := \frac{\sum_{y:G(x',y)=z,y\in B}\mu(x',y)}{\sum_{x,y:G(x,y)=z,y\in B}\mu(x,y)},$$

which is the distribution μ conditioned on z, B. Then following properties hold:

1. *For any $A \subset [m]^n$ such that A is thick in I and $A_{\bar{I}} \subseteq U(z_{\bar{I}}, B_{\bar{I}})$, we have that*

$$\sum_{x\in A}\sigma^{I,B,z}(x) \geq \left(\frac{2}{m}\right)^{|\bar{I}|}\frac{|A|(1-4n^{-5})}{m^{|I|}}\cdot(1-\frac{2n^7\log m}{m^{17/20}})$$

2. *For any $A' \subset A$, where A is thick in I and $A_{\bar{I}} \subseteq U(z_{\bar{I}}, B_{\bar{I}})$, we have*

$$\sum_{x\in A'}\sigma^{I,B,z}(x) \leq \left(\frac{2}{m}\right)^{|\bar{I}|}\frac{1+4n^{-5}}{m^{|I|}}(|A'| + \left(\frac{4n^6\log m}{m^{17/20}}\right)|A|).$$

3 Proof of main result

We restate Theorem 1 and provide its proof below.

Theorem 12. *For all (partial) functions f, it holds that*

$$\mathrm{R}^{\mu}_{1/4}(f\circ \mathrm{Index}^n_m) \geq \Omega(\mathrm{R}_{1/3}(f)\cdot \log n),$$

where $m = \mathrm{poly}(n)$.

Proof. For a given function f, recall the definition of λ (hard distribution for f) and μ (lifted distribution for $f \circ \mathrm{Index}_m^n$) from Section 2. Let $\mathcal{T}$ be a deterministic communication tree for f achieving $\mathrm{D}^{\mu}_{1/4}(f \circ \mathrm{Index}_m^n)$. Let $k := \mathrm{D}^{\mu}_{1/4}(f \circ \mathrm{Index}_m^n)$ be the depth of $\mathcal{T}$. Using our algorithm Π given in Figure 1, we get a randomized query protocol for f which makes an error of at most $\frac{1.1}{4}$ under λ (as implied by Lemma 14) and makes at most $O(k/\log n)$ expected number of queries (as implied by Lemma 18). This can be converted into an algorithm with $O(k/\log n)$ number of queries (in the worst case) and distributional error $\frac{1}{3}$, using standard application of Markov's inequality. This shows that

$$\mathrm{R}_{\frac{1}{3}}(f) = \mathrm{D}^{\lambda}_{\frac{1}{3}}(f) \leq O\left(\frac{k}{\log n}\right).$$

□

We construct a tree $\mathcal{T}_{\mathcal{AB}}$ which represents the evolution of the algorithm Π and is helpful in our analysis. The steps have been depicted in Figures **??** and **??**. All the nodes of the tree are labelled by subsets of $[m]^j, \mathrm{Bal}_m^j$ (where $j \in \{1, 2, \ldots n\}$) and the current interval. The root node is $([m]^n, \mathrm{Bal}_m^n, [n])$ and the rest of the tree is constructed as follows. Consider the step 2(a.i.A) of the algorithm Π when A is going to be partitioned into A_1, A_2. Set the children of (A, B, I) to be $(A_1, B, I), (A_2, B, I)$ and assign the weights of the edges going from parent to children as $\Pr_\mu[A_1 \times B | A \times B], \Pr_\mu[A_2 \times B | A \times B]$. Now,consider the step 2(a.i.C). Let the children be regular nodes obtained by Thickness Lemma. Consider the step 2(b.iii) and let the query at this step be done at i-th index. We shall construct a pair of regular children for every η. Fixing an η, let the partitions of A be A_1, A_2 (corresponding to $U(\eta, 0), U(\eta, 1)$). Set the children to be $(A_1, B_\eta, I \setminus \{i\}), (A_2, B_\eta, I \setminus \{i\})$ and label the edges with η, probability of obtaining η conditioned on $A \times B$, and probability of the child conditioned on fixing η. For the step 2(a.ii) that involves partition of B, the children are constructed in similar way and weights of edges are associated probabilities of transition. This finishes the construction of the tree.

For a node $(A, B, I) \in \mathcal{T}_{\mathcal{AB}}$, let $\mathrm{Par}((A, B, I))$ represent the parent node of (A, B, I) and $\mathrm{Chil}((A, B, I))$ represent the set of children nodes of (A, B, I). With some abuse of notation (as it shall be clear from the context), $\mathrm{Par}_{\mathcal{A}}(A)$ shall represent the set A associated to the parent node, $\mathrm{Par}_{\mathcal{B}}(B)$ shall represent the set B associated to the parent node and $\mathrm{Chil}_{\mathcal{I}}(I)$ represent the set of sets in the children node. Similarly, we shall consider $\mathrm{Par}_{\mathcal{A}}(I)$ and $\mathrm{Chil}_{\mathcal{A}}(I)$. A level in tree shall be represented as $t > 0$, with $t = 1$ representing the root node. In any level t, let $\mathcal{R}(t)$ represent set of all nodes which are at distance t from the root. Let $\mathcal{Q}(t)$ (or query nodes) represent the set of all nodes at level t which were obtained from their parent through the step 2(b). Let $\mathcal{C}(t)$ (or non-query nodes) represent the set of remaining nodes at level t. Let the nodes that did not abort for set A at level t be $\mathcal{N}_{abort}(t)$.

Note that the depth of the tree $\mathcal{T}_{\mathcal{AB}}$ is at most $O(n \log n)$, as (without loss of generality) there are at most $O(n \log n)$ communication steps in $\mathcal{T}$ and at most n query steps and constant number of operations for each of these steps in the algorithm Π.

Error analysis of algorithm Π

We first show the following lemma, which states some conditions that remain invariant under our algorithm.

Lemma 13 (Invariance Lemma)**.** *Throughout the execution of the algorithm* Π*, we show the following invariant:*

1. *A is thick in the current interval I, after leaving the step 2(a) and step 2(b).*
2. *$|B_I| \geq 2^{m|I| - n^2}$ for the current interval I for nodes of $\mathcal{T}_{\mathcal{AB}}$.*

```
1. Initialize $v$ as root of the protocol tree $\mathcal{T}$, initialize the intervals $I = [n], J = \phi$, Alice's part of rectangle $A = [m]^n$ and Bob's part of rectangle $B = \mathrm{Bal}_m^n$.
2. While $v$ is not a leaf do:
   (a) If $\mathrm{AvgDeg}_i(A_I) \leq m^{-19/20}$ for all $i \in I$:
       i. If Alice sends a bit at $v$:
          A. Pick $b \in \{0,1\}$ with probability $\Pr_\mu[(A \cap X^{v,b}) \times B | A \times B]$. If $\frac{|(A \cap X^{v,b})|}{|A|} \leq \frac{1}{n^2}$, for the picked $b$, then Abort.
          B. Set $v \leftarrow v_b$ and $A = A \cap X^{v,b}$.
          C. Apply the Thickness Lemma to get $A'$ satisfying $|A'| > \left(1 - \frac{1}{n^2}\right)|A|$ and $A'$ is thick for $I$. Set $A \leftarrow A'$ with probability $\Pr_\mu[A' \times B | A \times B]$. Set $A \leftarrow A \setminus A'$ with probability $1 - \Pr_\mu[A' \times B | A \times B]$ and Abort.
       ii. If Bob sends a bit at $v$:
          A. Pick $b$ with probability $\Pr_\mu[A \times (B \cap Y^{v,b}) | A \times B]$. If $\frac{|(B \cap Y^{v,b})_I|}{2^{m|I|}} \leq 2^{-n^2}$ for the picked $b$, then Abort.
          B. Set $v \leftarrow v_b$ and $B \leftarrow B \cap Y^{v,b}$.
   (b) If $\mathrm{AvgDeg}_i(A_I) < m^{19/20}$ for some $i \in I$:
       i. Query $z_i$.
       ii. For an $\eta \in \mathrm{Bal}_m$, select $\eta$ with probability $\Pr_\mu[A \times B_\eta | A \times B]$. If the picked $\eta \in \mathrm{Bad}(A, B, i)$, then Abort.
       iii. Set $B \leftarrow B_{\eta,i}$, $A \leftarrow A|_{U(\eta,z_i),i}$.
       iv. Set $I \leftarrow I \backslash \{i\}$.
3. Assume $I = \{1, 2, \ldots |I|\}$ (without loss of generality for the procedure described here). Set $i = 1$.
4. While $i \leq |I|$ do:
   (a) For an $\eta \in \mathrm{Bal}_m$, select $\eta$ with probability $\Pr_\mu[A \times B_\eta | A \times B]$. If the picked $\eta \in \mathrm{Bad}(A, B, i)$, then Abort.
   (b) Set $B \leftarrow B_{\eta,i}$, $i \leftarrow i + 1$.
5. Output as $\mathcal{T}$ does on the leaf $v$.
```

Figure 1: Randomized query algorithm Π for f.

Proof. 1. After the execution of step 2(a.i.B), the size of A reduces by at most $\frac{1}{n^2}$ (otherwise there is abort). Thus, Claim 6 ensures that for all $i \in I$, $\mathrm{AvgDeg}_i(A, I) \leq n^2 \cdot m^{-19/20}$. Further application of thickness lemma produces an A' such that $\mathrm{MinDeg}_i(A', I) \leq n^5 \cdot m^{-19/20} < m^{-17/20}/2$. Thus A' is thick in current interval after leaving step 2(a) due to the application of thickness lemma in step 2(a.C).

   Now we consider step 2(b). A was thick in the interval I just before entering step 2(a). Let the queried index be i. Partition Lemma (which can be applied as η did not abort) ensures that conditioned on z_i, η, the distribution of $A|_{U(\eta,z_i)}$ in the set $I \setminus \{i\}$ is same as the distribution of A in the set $I \setminus \{i\}$ (upto multiplicative factor of $1 \pm 2n^{-5}$). But A is thick in the interval I, which implies that A is thick in the interval $I \setminus \{i\}$ (using Claim 7; the multiplicative factor of $(1 \pm 2n^{-5})$ is insignificant as $\mathrm{MinDeg}_i(A, I)$ was at most $m^{-17/20}/2$, whereas the requirement for being thick is to have $\mathrm{MinDeg}_i(A, I) \leq m^{-17/20}$). Thus, distribution of $A|_{U(\eta,z_i)}$ in the set $I \setminus \{1\}$ is thick in the interval $I \setminus \{i\}$. Thus, the item follows.

2. The item follows since a B which does not satisfy this condition is aborted.

□

Now we are in a position to do the error analysis.

Lemma 14. *The algorithm Π makes an error of at most $1/4 + O(\log n/n)$.*

Proof. We begin with computing the overall probability of abort. We will first compute the probability of abort associated to A subsets. For a node $(A, B, I) \in \mathcal{T}_{\mathcal{AB}}$, consider the quantity $|(A \cap X^{v,b})|/|A|$. It is upper bounded by $\frac{1}{n^2}$ if there is abort. Thus, child of (A, B, I) aborts only if size of $\mathrm{Chil}_{\mathcal{A}}(A)$ is smaller than $\frac{1}{n^2}$ times the size of A. We appeal to Uniformity Lemma 11 (it can be verified that the conditions required for the lemma are satisfied for the considered sets) to conclude that the probability of transition from parent to child is upper bounded by

$$\Pr_{\mu}[(A \cap X^{v,b}) \times B | A \times B] \leq |(A \cap X^{v,b})|/|A| + n^{-67} \leq n^{-2}.$$

Similar argument holds for steps 2.(a.i.D). As noted before, the tree $\mathcal{T}_{\mathcal{AB}}$ has depth at most $O(n \log n)$. Hence, we obtain that the overall probability to abort is at most $O(\log n/n)$, appealing to Claim 8.

Marginalizing over Alice, we now compute the probability of abort associated to B subsets. At the steps 2(b.ii) and 4(a), the sampled η is from $\mathrm{Bad}(A, B, i)$ with probability at most $2^{-n^{35}} + 2^{-n^2}$. At the steps 2(a.ii.A), the abort occurs if $\frac{|B \cap Y^{v,b}|}{2^{m|I|}} \leq 2^{-n^2}$. Without loss of generality, the depth of $\mathcal{T}$ is at most $O(n \log n)$, resulting in at most $2^{O(n \log n)}$ subsets of Bal_m^n. Thus the overall probability of abort of this form is at most $2^{O(n \log n)} \cdot 2^{-n^2} \leq 2^{-n}$.

Thus by union bound, the overall probability of aborting in the algorithm is at most $O(\log n/n)$.

Conditioned on non abort, we proceed as follows. We consider the algorithm when it reaches a leaf L in $\mathcal{T}$ and does not abort. Let I be the set of corresponding un-queried bits (dropping the index L) and let the queried string be $z^*_{n/\{I\}}$. Without loss of generality, let $I = \{1, 2, \ldots |I|\}$. Let the rectangle corresponding to this leaf be $A \times B$. We have the following claim.

Claim 15. *For all $r < |I|$ and the sequence $y_1, y_2, \ldots y_r$, the y_{r+1} drawn from $B_{y_1,y_2,\ldots y_r}$ belongs to $\neg\mathrm{Bad}(A, B, i)$ and $|B_{y_1,y_2,\ldots y_r}| \geq 2^{m(|I|-r)-n^2}$.*

Proof. The property that y_{r+1} drawn from $B_{y_1,y_2,\ldots y_r}$ belongs to $\neg\mathrm{Bad}(A, B, i)$ is guaranteed by Step 4. To lower bound the size, we consider the base case $r = 0$. Then $|B| \geq 2^{m|I|-n^2}$ from the non-abort condition. Moreover, y_1 belongs to $\neg\mathrm{Bad}(A, B, 1)$, which, by definition, implies that $|B_{y_1}| \geq 2^{m|I|-n^2-m} = 2^{m(|I|-1)-n^2}$. Continuing this way, the claim follows. □

We start with the following distribution over the strings $z \in \{0,1\}^n$:

$$\rho^L(z) = \sum_{(x,y)\in A\times B: G(x,y)=z} \frac{1}{|A|\cdot|B|}.$$

Note that $\rho^L(z)$ is only supported on those strings z for which $z_{[n]/I} = z^*_{[n]/I}$. We shall establish the following claim.

Claim 16. *For the strings z such that $z_{[n]/I} = z^*_{[n]/\{I\}}$, it holds that*

$$\rho^L(z) \in \frac{1}{2^{|I|}}\left[1-2n^{-4}, 1+2n^{-4}\right].$$

Proof. We shall keep track of two invariant properties: A is thick in the interval I and $|B| \geq 2^{m|I|-n^2}$. Note that these conditions are true after the execution of step 2, as argued in Invariance Lemma 13. We start with computing

$$\rho^L(z_1) = \frac{1}{|B|}\sum_{y\in B}\sum_{x_{-1}} \frac{|A_{x_{-1}}|}{|A|}\frac{|(A|_{U(y_1,z_1)})_{x_{-1}}|}{|A_{x_{-1}}|} = \frac{1}{|B|}\sum_{y\in B}\sum_{x_{-1}}\frac{|A_{x_{-1}}|}{|A|}\sum_{x_1\in U(y_1,z_1)} p_{A_{x_{-1}}}(x_1)$$

From Partition Lemma (which we can apply due to our invariant), it holds that

$$\sum_{x_1\in U(y_1,z_1)} p_{A_{x_{-1}}}(x_1) \in \left[\frac{1}{2}-n^{-5}, \frac{1}{2}+n^{-5}\right]$$

for every x_{-1}, y. Thus, we conclude that $\rho^L(z_1) \in [\frac{1}{2}-n^{-5}, \frac{1}{2}+n^{-5}]$.

Now we proceed to compute $\rho^L_{z_2|z_1}$. We fix a y_1 and argue in same manner as given in Invariance Lemma 13 (Item 1). Partition Lemma ensures that conditioned on z_1, y_1, the distribution of $A|_{U(y_1,z_1)}$ in the set $I\setminus\{1\}$ is same as the distribution of A in the set $I\setminus\{1\}$ (upto multiplicative factor of $1\pm 2n^{-5}$). But A is thick in the interval I, which implies that A is thick in the interval $I\setminus\{1\}$ (using Claim 7). Thus, distribution of $A|_{U(y_1,z_1)}$ in the set $I\setminus\{1\}$ is thick in the interval $I\setminus\{1\}$. From Claim 15, we have that $|B_{y_1}| \geq 2^{m(|I|-1)-n^2}$. Thus, we have maintained the invariant properties that we started with. Hence, we can apply the Partition Lemma again to obtain that $\rho^L_{z_2|z_1} \in [\frac{1}{2}-n^{-5}, \frac{1}{2}+n^{-5}]$.

Continuing in the same fashion, we see that the invariant properties continue to be maintained. Hence by recursive application of the Partition Lemma, we obtain the claim.

□

A corollary of this claim is the following.

Corollary 17. *Consider the distribution $\tau^L(z)$ over strings z conditioned on reaching the leaf. It holds that $\tau^L(z) \in \lambda(z|z_{n/\{I\}})\cdot[1-4\cdot n^{-4}, 1+4\cdot n^{-4}]\cdot\delta(z^*_{n/\{I\}} = z_{n/\{I\}})$.*

Proof. We have that

$$\tau^L(z) = \frac{\sum_{x,y\in A\times B: G(x,y)=z}\mu(x,y)}{\sum_{x,y\in A\times B}\mu(x,y)} = \frac{\lambda(z)\sum_{x,y\in A\times B:G(x,y)=z}\frac{1}{|A||B|}}{\sum_z \lambda(z)\sum_{x,y\in A\times B:G(x,y)=z}\frac{1}{|A||B|}}.$$

From above claim, we have that

$$\sum_{x,y\in A\times B:G(x,y)=z}\frac{1}{|A||B|} \in \frac{1}{2^{|I|}}[1-2\cdot n^{-4}, 1+2\cdot n^{-4}]$$

as long as $z_{n/\{I\}} = z^*_{n/\{I\}}$. Thus, we find that

$$\tau^L(z) \in \lambda(z|z_{n/\{I\}}) \cdot [1 - 4 \cdot n^{-4}, 1 + 4 \cdot n^{-4}] \cdot \delta(z^*_{n/\{I\}} = z_{n/\{I\}}).$$

This proves the corollary □

Thus we conclude that conditioned on being on a non-aborting leaf L, τ^L differs by at most n^{-4} (in trace distance) with λ conditioned on $z_{n/\{I\}} = z^*_{n/\{I\}}$.

Furthermore, probability of our algorithm reaching a non-abort leaf is as according to the probability in the tree $\mathcal{T}$. This can be seen as follows: by construction of the algorithm, our probabilities of transition into sub-rectangles are as directed by $\mathcal{T}$ during all the non-query steps of the algorithm. During the query steps, the probability of transition of our algorithm to the event corresponding to query outcome z_i is as directed by λ. However, it can be argued identically along the lines of Claim 16 (using Invariance Lemma 13) that this transition probability, up to a multiplicative error of $1 \pm n^{-5}$ for each query step, is as directed by $\mathcal{T}$. Thus, overall error due to this discrepancy is at most n^{-4}.

Since $\mathcal{T}$ made an error of at most 1/4, the query algorithm makes an error of at most $1/4 + 2n^{-4} + O(\log n/n) \leq 1/4 + O(\log n/n)$. This proves the lemma. □

Expected number of queries of Π

We prove the following lemma.

Lemma 18. *The algorithm Π makes at most $\frac{2k}{5 \log n}$ expected number of queries, where the depth of the tree $\mathcal{T}$ is k.*

Proof. We will consider a potential function at the level t as follows, where the quantities are computed with respect to the tree $\mathcal{T}_{\mathcal{AB}}$. For every node $(A, B, I) \in \mathcal{T}_{\mathcal{AB}}$, let the probability of this node, computed by summing over all the conditional probabilities from root to the node be $\Pr(A, B)$. Define

$$P(t) = \sum_{(A,I) \in \mathcal{N}_{abort}(t)} \Pr(A, B) \log\left(\frac{m^{|I|}}{|A_I|}\right). \tag{1}$$

Consider the change in the potential function as the level t progresses.

$$
\begin{aligned}
P(t+1)-P(t) &= \sum_{(A,B,I)\in\mathcal{N}_{abort}(t)} \Pr(A,B)\log\left(\frac{m^{|I|}}{|A_I|}\right) - \sum_{(A,B,I)\in\mathcal{N}_{abort}(t)} \Pr(A,B)\log\left(\frac{m^{|I|}}{|A_I|}\right)\\
&\overset{a}{=} \sum_{(A,B,I)\in\mathcal{C}(t+1)\cap\mathcal{N}_{abort}(t)} \Pr(A,B)\log\left(\frac{m^{|I|}}{|A_I|}\right) - \sum_{(A,B,I)\in\mathrm{Par}(\mathcal{C}(t+1))\cap\mathcal{N}_{abort}(t)} \Pr(A,B)\log\left(\frac{m^{|I|}}{|A_I|}\right)\\
&+ \sum_{(A,B,I)\in\mathcal{Q}(t+1)\cap\mathcal{N}_{abort}(t)} \Pr(A,B)\log\left(\frac{m^{|I|}}{|A_I|}\right) - \sum_{(A,B,I)\in\mathrm{Par}(\mathcal{Q}(t+1))\cap\mathcal{N}_{abort}(t)} \Pr(A,B)\log\left(\frac{m^{|I|}}{|A_I|}\right)\\
&\overset{b}{\leq} \sum_{(A,B,I)\in\mathcal{C}(t+1)\cap\mathcal{N}_{abort}(t)} \Pr(A,B)\log\left(\frac{|\mathrm{Par}_{\mathcal{A}}(A)_{\mathrm{Par}_{\mathcal{A}}(I)}|}{|A_I|}\right)\\
&+ \sum_{(A,B,I)\in\mathcal{Q}(t+1)} \Pr(A,B)\log\left(\frac{|\mathrm{Par}_{\mathcal{A}}(A)_{\mathrm{Par}_{\mathcal{A}}(I)}|m^{|I|}}{|A_I|m^{|\mathrm{Par}_{\mathcal{I}}(I)|}}\right)\\
&\overset{c}{=} \sum_{(A,B,I)\in\mathcal{C}(t+1)\cap\mathcal{N}_{abort}(t)} \Pr(A,B)\log\left(\frac{|\mathrm{Par}_{\mathcal{A}}(A)_I|}{|A_I|}\right)\\
&+ \sum_{(A,B,I)\in\mathcal{Q}(t+1)} \Pr(A,B)\log\left(\frac{|\mathrm{Par}_{\mathcal{A}}(A)_{\mathrm{Par}_{\mathcal{A}}(I)}|}{|A_I|m}\right), \qquad (2)
\end{aligned}
$$

where in (a), we note that parents of non-aborted nodes cannot be aborted nodes. The inequality (b) holds by partitioning the probability of a parent node as a sum of probabilities of its child nodes. An inequality comes instead of an equality since we use the fact that a parent of a non aborted node may have an aborted intermediate child. In (b), we have also used the fact that in query step, every child of a node belongs to $\mathcal{N}_{abort}$ and in any non-query step, the size of I does not change. Equality (c) follows since in a query step, the size of I decreases by 1.

Lets consider the second to last expression above. In a step that involves partitioning of a B set, there is no change in the expression in logarithm. Thus, we consider only the steps involving partitioning of A sets. We begin with steps 2(a.i.A)-2.a(i.C). Consider a node (A^*, B) and its intermediate children $(\mathrm{IntChil}(A^*)_1, B), (\mathrm{IntChil}(A^*)_2, B)$ in a non-query step. Let the collection of A's obtained from $\mathrm{IntChil}(A^*)_1$ and $\mathrm{IntChil}(A^*)_2$ by the Flattening Lemma be $\mathcal{C}_1$ and $\mathcal{C}_2$ respectively. Define $\frac{1}{p} := \frac{|A^*|}{|\mathrm{IntChil}(A^*)_1|}$. For an $A \in \mathcal{C}_1$, define $\frac{1}{q_1} := \frac{|A^*_I|}{|A_I|}$, which is the same for all $A \in \mathcal{C}_1$ as argued in the Flattening Lemma. Similarly define q_2. This allows us to conclude that for $j \in \{1, 2\}$,

$$\sum_{A: A\in\mathcal{C}_j} \Pr(A,B)\log\left(\frac{|A^*_I|}{|A_I|}\right) = \Pr(\mathrm{IntChil}(A^*)_j, B)\log\left(\frac{|A^*_I|}{|A_I|}\right),$$

where the A appearing on the right hand side belongs to $\mathcal{C}_j$. Suppose the node A^* was obtained from a non-query step. Then A^* is thick and uniform. Thus from the statement of the Flattening Lemma, we have that $\frac{1}{q_1} \leq \frac{1}{p}$ and $\frac{1}{q_2} \leq \frac{1}{1-p}$. Thus, we conclude that for $j \in \{1, 2\}$,

$$\sum_{A: A\in\mathcal{C}_j} \Pr(A,B)\log\left(\frac{|A^*_I|}{|A_I|}\right) \leq \Pr(\mathrm{IntChil}(A^*)_j, B)\log\left(\frac{|A^*|}{|\mathrm{IntChil}(A^*)_j|}\right).$$

Now using Uniformity Lemma 11, we have the upper bound $\Pr[A \times B | A^* \times B] \leq \frac{|A|}{|A^*|} + n^{-67} \leq (1+n^{-60})\frac{|A|}{|A^*|}$, due to the fact that $\frac{|A|}{|A^*|} \geq \frac{1}{n^2}$ (else the intermediate child was an aborted child). Thus, the contribution of A^* to second to last expression in Equation 2 is upper bounded by

$$(1+\frac{1}{n^{60}}) \cdot (p\log\frac{1}{p} + (1-p)\log\frac{1}{1-p}) \leq 1.1$$

With this we conclude,

$$\begin{aligned}\sum_{(A,B,I)\in\mathcal{C}(t+1)\cap\mathcal{N}_{abort}(t+1)} \Pr(A)\log\left(\frac{|\mathrm{Par}_{\mathcal{A}}(A)_{\mathrm{Par}_{\mathcal{I}}(I)}|}{|A_I|}\right) &\leq 1.1 \sum_{(A,B,I)\in\mathcal{C}(t+1)} \Pr(A,B) \\ &= 1.1 \sum_{(A^*,B,I)\in\mathcal{C}^*(t)} \Pr(A^*,B),\end{aligned}$$

where by $\mathcal{C}^*(t)$, we represent the set of parent nodes at level t that are partitioned via a non-query step. Same analysis holds for the step 2.(a.i.D).

Now we evaluate the last expression in Equation 2. Let the queried index be i. Since AvgDeg_i has reduced below $m^{19/20}$ and $\mathrm{Par}_{\mathcal{I}}(I) = I \cup \{i\}$, we conclude that $\frac{|\mathrm{Par}_{\mathcal{A}}(A)_{\mathrm{Par}_{\mathcal{I}}(I)}|}{|A_I|} \leq m^{19/20}$. This implies

$$\sum_{(A,B,I)\in\mathcal{Q}(t+1)} \Pr(A,B)\log\left(\frac{|\mathrm{Par}_{\mathcal{A}}(A)_{\mathrm{Par}_{\mathcal{I}}(I)}|}{m|A_I|}\right) \leq \left(-\frac{1}{20}\log m\right) \sum_{(A^*,B,I)\in\mathcal{Q}^*(t)} \Pr(A^*,B),$$

where the set of parent nodes that are partitioned via step 2(b) at level t are represented by $\mathcal{Q}^*_{\mathcal{A}}(t)$.

Finally, we observe that $\sum_{(A^*,B,I)\in\mathcal{C}^*(t)} \Pr(A^*,B)$ is the expected number of communication steps taken by Alice at step t (which is upper bounded by k) and $\sum_{(A^*,B,I)\in\mathcal{Q}^*(t)} \Pr(A^*,B)$ is the expected number of query steps to be taken at step t. Let the last step of the algorithm be last. Using telescopic sum, we have that

$$P(\text{last}) - P(1) = \sum_{t=1}^{\mathsf{last}-1} P(t+1) - P(t).$$

Thus, setting q to be the expected number of queries made by the algorithm, we have that

$$P(\mathsf{last}) - P(1) \leq 2.2k - \frac{q}{20}\cdot\log m.$$

But $P(1) = 0$ and $P(\mathsf{last}) > 0$. Thus setting the value of m, we have that $q \leq \frac{2.2}{5\log n}\cdot k$. This proves the lemma.

□

Acknowledgement

This work is supported by the Singapore Ministry of Education and the National Research Foundation, also through the Tier 3 Grant Random numbers from quantum processes MOE2012-T3-1-009.

References

[ABBD$^+$16] A. Anshu, A. Belovs, S. Ben-David, M. Gs, R. Jain, R. Kothari, T. Lee, and M. Santha. Separations in communication complexity using cheat sheets and information complexity. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 555–564, Oct 2016.

[Dup15] Frdric Dupuis. Chain rules for quantum rnyi entropies. *Journal of Mathematical Physics*, 56(2):022203, 2015.

[GLM+15] Mika Göös, Shachar Lovett, Raghu Meka, Thomas Watson, and David Zuckerman. Rectangles are nonnegative juntas. In *Proceedings of the Forty-seventh Annual ACM Symposium on Theory of Computing*, STOC '15, pages 257–266, New York, NY, USA, 2015. ACM.

[GPW15] Mika Göös, Toniann Pitassi, and Thomas Watson. Deterministic communication vs. partition number. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 1077–1088, Oct 2015.

[Raz99] Ran Raz. Exponential separation of quantum and classical communication complexity. In *Proceedings of the Thirty-first Annual ACM Symposium on Theory of Computing*, STOC '99, pages 358–367, New York, NY, USA, 1999. ACM.

[RM99] Ran Raz and Pierre McKenzie. Separation of the monotone nc hierarchy. *Combinatorica*, 19(3):403–435, 1999.

[She11] Alexander A. Sherstov. The pattern matrix method. *SIAM Journal on Computing*, 40(6):1969–2000, 2011.

A Proof of Uniformity Lemma 11

We shall use the following claim for the proof.

Claim 19. *Let $I \subset [n]$ be an interval and $B \in \mathrm{Bal}_m^{|I|}$ be a set such that $\frac{|B|}{|\mathrm{Bal}_m|^{|I|}} \geq 2^{-k}$ for some $k \geq n^2$. Then there exists a set of indices J with $|J| \geq m|I| - 2n^4 \cdot k$, such that for all $j \in J$, $\frac{|B_{y_j}|}{|B|} \in [\frac{1}{2} \pm \frac{1}{n^2}]$, for $y_j \in \{0,1\}$.*

Proof. Let $Y := Y_1Y_2 \dots Y_{m|I|}$ be the random variable with associated probability distribution $\Pr_Y(y) = \frac{1}{|B|}$ if $y \in B$ and 0 otherwise. We have that

$$H(Y) = \log|B| \geq |I| \log|\mathrm{Bal}_m| - k \geq m|I| - k - n\log m \geq m|I| - 2k.$$

Hence, $\sum_j H(Y_j) \geq H(Y) \geq m|I| - 2k$. Since $H(Y_j) \leq 1$, this can be re-written as $\sum_j (1 - H(Y_j)) \leq 2k$. By Markov's ineqaulity, the number of indices for which $H(Y_j) < 1 - \frac{1}{n^4}$ is at most $n^4 \cdot k$. Let the rest of indices form the set J. By our construction, for every $j \in J$, we have $H(Y_j) \geq 1 - \frac{1}{n^4}$. Thus, $\Pr[Y_j = 0] \in [\frac{1}{2} \pm \frac{1}{n^2}]$. Since $\Pr[Y_j = 0] = \frac{|\{y : y_j = 0, y \in B\}|}{|B|}$, and same argument applies for $Y_j = 1$, the claim follows.

□

Proof of Lemma 11. We start with the observation that

$$\sigma^{I,B,z}(x') = \frac{\sum_{y:G(x',y)=z, y\in B} \mu(x',y)}{\sum_{x,y:G(x,y)=z,y\in B} \mu(x,y)} = \left(\frac{2}{m}\right)^{|\bar{I}|} \delta(x'_{\bar{I}} \in U(z_{\bar{I}}, B_{\bar{I}})) \cdot \frac{|\{y_I : G(x'_I, y_I) = z_I, y_I \in B_I\}|}{\sum_{x_I} |\{y_I : G(x_I, y_I) = z_I, y_I \in B_I\}|},$$

which follows since B is fixed in the interval $\bar{I}$. Without loss of generality, let $I = \{1, 2, \dots |I|\}$. For brevity, we set

$$\sigma^{I,B,z}(x'_I) = \frac{|\{y_I : G(x'_I, y_I) = z_I, y_I \in B_I\}|}{\sum_{x_I} |\{y_I : G(x_I, y_I) = z_I, y_I \in B_I\}|},$$

which is a probability distribution. First we compute the distribution $\sigma^{I,B,z}(x'_1)$. For this, consider

$$\sum_{x'_{-1}} |\{y_I : G(x'_I, y_I) = z_I, y_I \in B_I\}| = \sum_{y_I \in B} |\{x'_{I\setminus\{1\}} : G(x'_1 x'_{I\setminus\{1\}}, y_I) = z_I\}|$$

$$= (\frac{m}{2})^{|I|-1}|\{y_I \in B : \mathrm{Index}_m(x'_1, y_1) = z_1\}|,$$

where second equality holds since for every y_I which is balanced, the number of $x_{I\setminus\{1\}}$ giving a particular $z_{I\setminus\{1\}}$ is $(\frac{m}{2})^{|I|-1}$. Thus, we find that

$$\sum_{x'_{-1}} |\{y_I : G(x'_I, y_I) = z_I, y_I \in B_I\}| = (\frac{m}{2})^{|I|-1}|B_{(z_1,x'_1)}|,$$

where $B_{(z_1,x'_1)}$ is the set $\{y \in B : y_{x'_1} = z_1\}$. This allows us to conclude:

$$\sigma^{I,B,z}(x'_1) = \frac{|B_{(z_1,x'_1)}|}{\sum_{x_1} |B_{(z_1,x_1)}|}.$$

Similarly, we conclude that

$$\sigma^{I,B,z}(x'_2|x_1) = \frac{|(B_{(z_1,x_1)})_{(z_2,x'_2)}|}{\sum_{x_2} |(B_{(z_1,x_1)})_{(z_2,x_2)}|},$$

and so on.

Now we introduce some notations. For every sequence $x_I := x_1x_2\ldots x_{|I|}$, we assign an associated string $D(x_I)$ of length $|I|$ over alphabet $\{\mathrm{VeryGood}, \mathrm{Good}, \mathrm{Bad}, \mathrm{Small}\}$ (which serves as a description of the string x_I) in the following way. We shall drop the brackets whenever x_I is clear from context.

- We let $D_1 = \mathrm{VeryGood}$ if it holds that $\frac{1}{2} - n^{-2} \le \frac{|B_{(z_1,x_1)}|}{|B|} \le \frac{1}{2} + n^{-2}$. If $m^{-|I|+1}/3 \le \frac{|B_{(z_1,x_1)}|}{|B|} \le \frac{1}{2} - n^{-2}$, then we let $D_1 = \mathrm{Good}$. If $\frac{|B_{(z_1,x_1)}|}{|B|} \le m^{-|I|+1}/3$, then we let $D_1 = \mathrm{Small}$. Else we let $D_1 = \mathrm{Bad}$.
- Similarly, let $D_2 = \mathrm{VeryGood}$ if it holds that $\frac{1}{2} - n^{-2} \le \frac{|(B_{(z_1,x_1)})_{(z_2,x_2)}|}{|B_{(z_1,x_1)}|} \le \frac{1}{2} + n^{-2}$. If $m^{-|I|+1}/3 \le \frac{|(B_{(z_1,x_1)})_{(z_2,x_2)}|}{|B_{(z_1,x_1)}|} \le \frac{1}{2} - n^{-2}$, then we let $D_1 = \mathrm{Good}$. If $\frac{|(B_{(z_1,x_1)})_{(z_2,x_2)}|}{|B_{(z_1,x_1)}|} \le m^{-|I|+1}/3$, then we let $D_2 = \mathrm{Small}$. Otherwise let $D_2 = \mathrm{Bad}$.
- Continue this way for whole of the string.

We are now in a position to prove the following claim.

Claim 20. *Fix a $x_I := x_1x_2\ldots x_{|I|}$ and associated description $D := D_1D_2\ldots D_{|I|}$ Following properties hold.*

1. *If there exists an i such that $D_i = \mathrm{Small}$, then $\sigma^{I,B,z}(x_I) \le m^{-|I|}$.*
2. *If for all i, $D_i \neq \mathrm{Small}$, then*
   (a) *If for all i, $D_i = \mathrm{VeryGood}$, then we have $\sigma^{I,B,z}(x_I) \ge \frac{1}{m^{|I|}} \cdot (1 - n^{-1})$.*
   (b) *$\sigma^{I,B,z} \le \frac{1}{m^{|I|}} \cdot (1 + n^{-2})^{|\#i:D_i \neq \mathrm{Bad}|} \cdot 2^{|\#i:D_i = \mathrm{Bad}|}$.*

For any i, a substring $x_1x_2\ldots x_i$ and the associated description $D_1D_2\ldots D_i$ from the alphabet $\{\mathrm{VeryGood}, \mathrm{Good}, \mathrm{Bad}\}$, the number of x'_{i+1} such that $D_{i+1} = \mathrm{VeryGood}$ is at least $m - 2n^6 \log m$.

Proof. Claim 19 ensures that number of x_1 such that $D_1 = \mathrm{VeryGood}$ is at least $m - 2n^6$. Thus,

$$(m + 2n^6)|B|/2 = (m - 2n^6)|B|/2 + 2n^6|B| \ge \sum_{x_1} |B_{(z_1,x_1)}| \ge (m - 2n^6)|B|/2.$$

This gives,

$$\frac{|B_{(z_1,x_1)}|}{m|B|} \cdot (2 - 3n^{-6}) \le \sigma^{I,B,z}(x_1) \le \frac{|B_{(z_1,x_1)}|}{m|B|} \cdot (2 + 3n^{-6})$$

for all x_1.

If x_1 is such that $D_1 = \text{Small}$, then $\frac{|B_{(z_1,x_1)}|}{|B|} \leq m^{-|I|+1}/3$. This automatically ensures that $\sigma^{I,B,z}(x_1 x_{I\setminus\{1\}}) \leq \sigma^{I,B,z}(x_1) \leq m^{-|I|}$ for all $x_{I\setminus\{1\}}$.

Otherwise, we have that $D_1 \neq \text{Small}$, for which $|B_{(z_1,x_1)}| \geq |B| \cdot \frac{m^{-|I|+1}}{3} \geq |\text{Bal}_m|^{|I|} 2^{-n^2 - n\log m}$. In this case, Claim 19 once again ensures that the number of x_2 such that $D_2 = \text{VeryGood}$ is at least $m - 2(n^6 + n^5 \log m)$. Thus, $\sum_{x_2} |(B_{(z_2,x_2)})_{(z_1,x_1)}| \geq (m - 2n^6 - 2n^5 \log m)|B_{(z_1,x_1)}|/2$. This gives,

$$\frac{|(B_{(z_2,x_2)})_{(z_1,x_1)}|}{m|B_{(z_1,x_1)}|} \cdot (2 - 4n^{-6}) \leq \sigma^{I,B,z}(x_2|x_1) \leq \frac{|(B_{(z_2,x_2)})_{(z_1,x_1)}|}{m|B_{(z_1,x_1)}|} \cdot (2 + 4n^{-6}).$$

Moreover, we find the following expression for $\sigma^{I,B,z}(x_1 x_2)$:

$$\frac{1}{m^2}(2-4n^{-6})^2 \cdot \frac{|B_{(z_1,x_1)}|}{|B|} \cdot \frac{|(B_{(z_2,x_2)})_{(z_1,x_1)}|}{|B_{(z_2,x_2)}|} \leq \sigma^{I,B,z}(x_1 x_2) \leq \frac{1}{m^2}(2+4n^{-6})^2 \cdot \frac{|B_{(z_1,x_1)}|}{|B|} \cdot \frac{|(B_{(z_2,x_2)})_{(z_1,x_1)}|}{|B_{(z_2,x_2)}|},$$

for any x_1 for which $D_1 \neq \text{Small}$.

The argument proceeds similarly in an inductive fashion. For any x_I and associated D, if it holds that $D_1 \neq \text{Small}$, $D_2 \neq \text{Small}$, $\ldots D_{|I|} \neq \text{Small}$, then we find

$$\sigma^{I,B,z}(x_1 x_2 \ldots x_{|I|}) \in \frac{1}{m^{|I|}}(2 \pm 4n^{-6})^{|I|} \cdot \frac{|B_{(z_1,x_1)}|}{|B|} \cdot \frac{|(B_{(z_2,x_2)})_{(z_1,x_1)}|}{|B_{(z_1,x_1)}|} \ldots \tag{3}$$

On the other hand, if there exists an i such that $D_i = \text{Small}$, then we automatically have that $\sigma^{I,B,z}(x_I) \leq m^{-|I|}$ (proving Item 1).

In the rest, we prove Item 2 and its subitems.

For 2(a), we observe that $D_i = \text{VeryGood}$ implies that $\frac{|B_{(z_1,x_1)}|}{|B|} \geq \frac{1}{2} - n^{-2}$, $\frac{|(B_{(z_2,x_2)})_{(z_1,x_1)}|}{|B_{(z_1,x_1)}|} \geq \frac{1}{2} - n^{-2}$ and so on. From Equation 3, we get the lower bound. Similar observation proves 2(b).

Last part of the claim follows from Claim 19, as discussed in the inductive steps above.

□

Now, we are in a position to prove the items of the lemma.

1. Consider

$$\begin{aligned}
\sum_{x \in A} \sigma^{I,B,z}(x) &= \sum_{x_I \circ x_{\bar{I}} \in A} \left(\frac{2}{m}\right)^{|\bar{I}|} \delta(x_{\bar{I}} \in U(z_{\bar{I}}, B_{\bar{I}})) \cdot \sigma^{I,B,z}(x_I) \\
&= \left(\frac{2}{m}\right)^{|\bar{I}|} \sum_{x_I \circ x_{\bar{I}} \in A} \sigma^{I,B,z}(x_I) \\
&\geq \left(\frac{2}{m}\right)^{|\bar{I}|} \sum_{x_I \circ x_{\bar{I}} \in A, D(x_I) = \text{VeryGood}^{|I|}} \frac{1 - 4n^{-5}}{m^{|I|}}
\end{aligned}$$

   To evaluate the last expression, we fix the prefix $x_1 x_2 \ldots x_{|I|-1}$ of x_I. Then we need to evaluate

$$|A_{x_1 x_2 \ldots x_{|I|-1}}| \sum_{x_{|I|} : (D(x_1 \ldots x_{|I|}))_{|I|} = \text{VeryGood}} p_{A_{x_1 \ldots x_{|I|-1}}}(x_{|I|}) \frac{1 - 4n^{-5}}{m^{|I|}}.$$

By Claim 20 (Item 4), we have that number of strings $x_{|I|}$ such that $(D(x_1 \dots x_I))_I \neq$ VeryGood is at most $2n^6 \log m$. By definition of Thickness, we have that $p^{\varepsilon}(A_{x_1 x_2 \dots x_{|I|-1}}, \{|I|\}) \leq m^{-17/20}$. Thus,

$$\begin{aligned}
\sum_{x_{|I|}:(D(x_1\dots x_{|I|}))_{|I|}=\text{VeryGood}} p_{A_{x_1\dots x_{|I|-1}}}(x_{|I|}) &\geq 1 - \sum_{x_{|I|}:(D(x_1\dots x_{|I|}))_{|I|}\neq\text{VeryGood}} p_{A_{x_1\dots x_{|I|-1}}}(x_{|I|}) \\
&\geq 1 - \frac{2n^6 \log m}{m^{17\varepsilon/20(1+\varepsilon)}}.
\end{aligned}$$

Thus, we have removed the constraint that $(D(x_1 \dots x_{|I|}))_{|I|} =$ VeryGood at the cost of a multiplicative factor of $(1 - \frac{2n^6 \log m}{m^{17/20}})$. We can continue this way for $x_{|I|-1}$ (constraints on which depend only on $x_1, x_2 \dots x_{|I|-2}$) and so on. This finally gives us

$$\sum_{x\in A} \sigma^{I,B,z}(x) \geq \left(\frac{2}{m}\right)^{|\bar{I}|} |A| \cdot (1 - \frac{2n^6 \log m}{m^{17/20}})^n \frac{1-4n^{-5}}{m^{|I|}}.$$

This proves the item.

2. We now compute $\sum_{x\in A'} \sigma^{I,B,z}(x)$. For this, we shall divide the set of strings x_I into two parts: S_1, S_2. S_1 contains all the strings x_I such that their associated string $D(x_I)$ has at least one alphabet 'Bad' and no alphabet 'Small'. S_2 contains rest of the string. Now we proceed as follows.

$$\begin{aligned}
\sum_{x\in A'} \sigma^{I,B,z}(x) &= \sum_{x\in A':x_I\in S_2} \sigma^{I,B,z}(x) + \sum_{x\in A',x_I\in S_1} \sigma^{I,B,z}(x) \\
&\leq \sum_{x\in A':x_I\in S_2} \sigma^{I,B,z}(x) + \sum_{x\in A,x_I\in S_1} \sigma^{I,B,z}(x) \\
&\leq \left(\frac{2}{m}\right)^{|\bar{I}|} \left(\sum_{x_I\circ x_{\bar{I}}\in A':x_I\in S_2} \frac{1+4n^{-5}}{m^{|I|}} + \sum_{x_I\circ x_{\bar{I}}\in A,x_I\in S_1} \frac{2^{|\#i:(D(x_I))_i=\text{Bad}|}(1+2n^{-5})}{m^{|I|}} \right) \\
&\leq \left(\frac{2}{m}\right)^{|\bar{I}|} \frac{1+4n^{-5}}{m^{|I|}} \left(|A'| + \sum_{x_I\circ x_{\bar{I}}\in A,x_I\in S_1} 2^{|\#i:(D(x_I))_i=\text{Bad}|} \right)
\end{aligned}$$

where the second inequality uses Claim 20. Below, we upper bound the last summation. Consider the set of all strings in S_1 for which the description string has a 'Bad' at the index 1 in I. Let this set be S_1^*. We consider the summation

$$\sum_{x_I\circ x_{\bar{I}}\in A,x_I\in S_1^*} 2^{|\#i:(D(x_I))_i=\text{Bad}|}.$$

Fix a string $x_2 \dots x_{|I|}$. Then summing over all $x_1, x_{\bar{I}}$ such that $x_1 x_2 \dots x_{|I|} \circ x_{\bar{I}} \in A$ and $x_1 \dots x_{|I|} \in S_1^*$, we obtain that

$$\begin{aligned}
&\sum_{x_1,x_{|I|}} 2^{|\#i:(D(x_I))_i=\text{Bad}|} = |A_{x_2\dots x_{|I|}}| \cdot \sum_{x_1} p_{A_{x_2\dots x_{|I|}}}(x_1) 2^{|\#i:(D(x_I))_i=\text{Bad}|} \\
&\leq \frac{2n^6 \log m}{m^{17/20}} 2^{|\#i:(D(x_I))_i=\text{Bad}|} |A_{x_2\dots x_{|I|}}|,
\end{aligned}$$

where the last inequality uses the fact that A is thick in interval I and number of possible x_1 is at most $2n^6 \log m$, from Claim 20. Thus, we have removed the constraint over x_1

that its associated alphabet D_1 is Bad, at the cost of multiplicative factor of $\frac{2n^6 \log m}{m^{17/20}}$. This process can be continued further for every x_i with associated $D_i = \text{Bad}$, leading to the upper bound

$$\sum_{x_I \circ x_{\bar{I}} \in A, x_I \in S_1} 2^{|\# i : (D(x_I))_i = \text{Bad}|} \leq \sum_{x_I \circ x_{\bar{I}} \in A} \left(\frac{4n^6 \log m}{m^{17/20}} \right)^{|\# i : (D(x_I))_i = \text{Bad}|} \leq \left(\frac{4n^6 \log m}{m^{17/20}} \right) |A|,$$

as S_1 contains strings with associated description strings having at least one 'Bad' alphabet. This collectively implies that

$$\sum_{x \in A'} \sigma^{I,B,z}(x) \leq \left(\frac{2}{m} \right)^{|\bar{I}|} \frac{1 + 4n^{-5}}{m^{|I|}} \left(|A'| + \left(\frac{4n^6 \log m}{m^{17/20}} \right) |A| \right).$$

This concludes the item.

Thus, the lemma concludes. □