# A strong direct product theorem for the tribes function via the smooth-rectangle bound

Prahladh Harsha* Rahul Jain†

**Abstract**

The main result of this paper is an optimal strong direct product result for the two-party public-coin randomized communication complexity of the Tribes function. This is proved by providing an alternate proof of the optimal lower bound of $\Omega(n)$ for the randomized communication complexity of the Tribes function using the so-called smooth-rectangle bound, introduced by Jain and Klauck [JK10]. The optimal $\Omega(n)$ lower bound for Tribes was originally proved by Jayram, Kumar and Sivakumar [JKS03], using a more powerful lower bound technique, namely the information complexity bound. The information complexity bound is known to be at least as strong a lower bound method as the smooth-rectangle bound [KLL+12]. On the other hand, we are not aware of any function or relation for which the smooth-rectangle bound is (asymptotically) smaller than its public-coin randomized communication complexity. The optimal direct product for Tribes is obtained by combining our smooth-rectangle bound for tribes with the strong direct product result of Jain and Yao [JY12] in terms of smooth-rectangle bound.

*School of Technology and Computer Science, Tata Institute of Fundamental Research, Mumbai, INDIA. E-mail: prahladh@tifr.res.in.

†Centre for Quantum Technologies and Department of Computer Science, National University of Singapore. E-mail: rahul@comp.nus.edu.sg.

# 1 Introduction

Study of lower bounds for various natural functions and relations has been a major theme of research in communication complexity from its advent; both for its own intrinsic value and for applications of these bounds towards other areas of theoretical computer science [KN97]. Several lower bound techniques have been developed over the years in communication complexity such as fooling sets, discrepancy method, rectangle bound, information complexity bound, partition bound etc. It is interesting to understand the relative power of these techniques and rank them against each other. Sometimes, we would like to understand what is the weakest technique required to prove a particular lower bound.

An important and extensively used technique in communication complexity is the the so called rectangle bound (a.k.a. the corruption bound). In this technique, one argues that for some output value $z$, and all large rectangles, a constant fraction of inputs in the rectangle have a function value different from $z$. This helps to lower bound the distributional communication complexity of the function, which then translates to a lower bound on the public-coin communication complexity via Yao's minmax principle [Yao83]. This technique has been successfully applied to obtain optimal lower bounds for several problems; Razborov's lower bound proof [Raz92] for the set-disjointness function [KS92] is arguably the most well-known application of this technique.

Another technique that has been extremely useful is the information complexity bound, first introduced by Chakrabarti, Shi, Wirth and Yao [CSWY01]. In this method, one lower bounds the distributional communication complexity by the amount of information the transcript of the protocol reveals about the inputs of Alice and Bob. The tools from information theory then come handy to lower bound the information cost of the protocol. Bar-Yossef, Jayram, Kumar and Sivakumar [BJKS04] successfully used this technique to give an alternate proof of the linear lower bound for the set-disjointness function. This method has also been useful to give an optimal linear lower bound for the Tribes function [JKS03].

Jain and Klauck [JK10], using tools from linear programming and semi-definite programming gave a uniform treatment to several of the existing lower bound techniques and proposed two additional lower bound techniques, the so-called partition bound and the smooth-rectangle bound. These bounds are stronger than almost all other known lower bound techniques including the rectangle bound. The partition bound, as the name suggests, is a linear programming formulation of the number of partitions in a randomized protocol. The smooth-rectangle bound, a weakening of the partition bound, is a robust version of the rectangle bound in the following informal sense: smooth-rectangle bound for a function $f$ under a distribution $\mu$, is the maximum over all functions $g$ , which are close to $f$ under the distribution $\mu$, of the rectangle bound of $g$. In other words, a function $f$ is said to have a large smooth-rectangle bound, if it is close to some other function $g$ (under the distribution $\mu$) which has a large rectangle bound, even though $f$ itself migh not have a large rectangle bound. This suffices to lower bound the communication complexity of $f$. These new lower bound methods have been successfully applied, for example to obtain an optimal lower bound for the Gap-Hamming problem [CR12]. In fact we are not aware (to the best of our knowledge) of any function or relation for which the partition bound or smooth-rectangle bound is (asymptotically) smaller than its public-coin randomized communication complexity. To determine how tight these new lower bounds are, remains an important open question in communication complexity.

Recently, Kerenidis *et al.* [KLL+12] showed that the information complexity is at least as powerful as the relaxed-partition-bound, which is a bound intermediate between the partition bound and the smooth-rectangle bound. The relative strengths of the information complexity and partition bound is

not yet well understood.

Another important theme in communication complexity has been the study of the so called strong direct-product and (the weaker) direct-sum conjectures; again for their own intrinsic value and also for important applications of such results in other areas of theoretical computer science [KRW95]. A strong direct-product conjecture for the public-coin communication complexity of a relation $f$ would state the following. Let $c$ be the public-coin communication complexity of $f$ (with constant error). Suppose $k$ independent instances of $f$ are being solved using communication less than $kc$, then the overall success would be exponentially small in $k$. In fact, the information complexity was introduced initially [CSWY01] as a tool to resolve the direct sum/product question. However, despite the considerable progress made over the last few years [BBCR10, JPY12], the direct product question has not yet been resolved. On the other hand, we are not aware of any function or relation for which this conjecture is false. Settling this conjecture for all relations, again is an important open question in communication complexity.

Recently, Jain and Yao [JY12] proved a direct-product result for all relations in terms of the smooth-rectangle bound (srec). They show that for any relation $f$, if less than $k \cdot \log \mathsf{srec}(f)$ communication (c.f., Definition 2.2) is provided for solving $k$ independent copies of $f$, then the overall success is exponentially small in $k$. This provides a recipe to arrive at strong direct-product results for any relation $f$: by exhibiting that $\log \mathsf{srec}(f)$ provides optimal lower bound for the public-coin communication complexity of $f$. Jain and Yao's result implies (and in some cases reproves) strong direct product result for many interesting functions and relations including that for the set-disjointness function (a strong direct-product result for set-disjointness was first shown by Klauck [Kla10], again via showing that the smooth-rectangle bound of a related function is large).

### 1.1 Our result

In this work we are concerned with the Tribes : $\{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ function, defined as follows.

$$\mathsf{Tribes}(x,y) \stackrel{\text{def}}{=} \bigwedge_{i=1}^{\sqrt{n}} \left( \bigvee_{j=1}^{\sqrt{n}} \left( x_{(i-1)\sqrt{n}+j} \wedge y_{(i-1)\sqrt{n}+j} \right) \right).$$

As mentioned earlier, an optimal linear lower bound for Tribes was shown by Jayram, Kumar and Sivakumar [JKS03] using the information complexity technique. It is to be noted that the rectangle bound proves only a $\Theta(\sqrt{n})$ lower bound and thus fails to provide an optimal lower bound for Tribes. In fact, the primary motivation for Jayram *et al.* [JKS03] to study the Tribes function was the fact that it provided the first example where information complexity techniques were provably stronger than the then known "combinatorial" lower bound techniques. Given the recent surge in combinatorial lower bound techniques, one can ask if information complexity techniques continue to remain provably stronger than the combinatorial techniques (such as smooth-rectangle bound, partition-bound etc) and in particular, if Tribes also separates information complexity bound and smooth-rectangle bound. We consider this question in this work and answer it in the negative.

**Theorem 1.1** (smooth-rectangle bound for Tribes)**.**
*For sufficiently small* $\varepsilon \in (0,1)$, $\mathsf{R}^{\mathrm{pub}}_{\varepsilon}(\mathsf{Tribes}) \geq \log \mathsf{srec}_{\varepsilon}(\mathsf{Tribes}) \geq \Omega(n)$.

Here, $\mathsf{R}^{\mathrm{pub}}_{\varepsilon}(f)$ refers to the $\varepsilon$-error public-coin randomized communication complexity of $f$.

The primary motivation for or work (besides answering the above question) is its consequence to strong direct product. As indicated in the recipe outlined above, combining our smooth-rectangle bound bound for Tribes with the result of Jain and Yao [JY12], we obtain the following.

**Corollary 1.2** (strong direct product for Tribes). $\mathsf{R}^{\mathrm{pub}}_{1-2^{-\Omega(k)}}\left(\mathsf{Tribes}^{(k)}\right) = \Omega(kn)$.

Here, $f^{(k)}$ refers to the $k$-wise direct product of the function $f$. Our result (Theorem 1.1) also exhibits for the first time, an asymptotic separation between the smooth-rectangle bound and the rectangle bound for a total function (previously a quadratic separation was known however for the Gap-Hamming partial function [CR12]).

It is to be noted that the information complexity lower bound for Tribes was generalised to constant depth read-once trees functions [JKR09, LS10]. Given our results, it is interesting to ask if these lower bounds can be obtained using the smooth-rectangle bound instead, which would imply a direct product for these functions. These alternate lower bounds might also help to obtain bounds for super-constant depth read-once formulae.

### 1.2 Our techniques

It will be convenient for us to view the Tribes function as the disjunction of $\sqrt{n}$ set-disjointness functions over $\sqrt{n}$ sized inputs. We refer to the $\sqrt{n}$ sized inputs to each of the disjointness functions as a block. We consider a distribution $\mu$ on the inputs for the Tribes function which has support only on the following type of inputs: in every block, except for one block (say $j$), the inputs to the two parties Alice and Bob are NO instances of the disjointness function (the sets corresponding to the blocks intersect at exactly one location) and in block $j$, there could be $0, 1$ or $2$ intersections which occur at locations $k_j$ and $l_j$. Let's refer to the three types of subsets of inputs based on the number of intersections as $U_0, U_1$, and $U_2$ respectively. Recall that to show that the smooth-rectangle bound of Tribes is large, we need to demonstrate a function $g$, close to Tribes (under $\mu$), whose rectangle bound is large. This function $g$ is constructed as follows: $g$ takes takes value 0 in $U_0 \cup U_2$ and value 1 in $U_1$. Note that Tribes takes value 0 in $U_0$ and value 1 in $U_1 \cup U_2$. I.e., Tribes and $g$ disagree on the inputs in $U_2$. For our choice of distribution $\mu$, this disagreement set $U_2$ will have weight $\mu(U_2) \approx 1/16$ while the weight of the 1-inputs will be approximately $\mu(U_1) \approx 6/16$ (i.e., $U_1$ is 6 times larger than $U_2$).

Observe that for Tribes, there are large rectangles (of size $\approx 2^{-\sqrt{n}}$ under $\mu$) which are monochromatic. We can just fix any one coordinate in each block and force intersection there to create large 1-monochromatic rectangle. Similarly we can choose any one block and force non-intersection in that entire block to create large 0-monochromatic rectangle. Hence the rectangle bound of Tribes is at most $O(\sqrt{n})$. However, note that the 1-monochromatic rectangles described above are not monochromatic in $g$. Indeed, we show that there exists constants $C$ and $D$ such that for every large rectangle $W$ (with $\mu(W) \geq 2^{-\Omega(n)}$), it is $\mu(U_1 \cap W)$ is either dominated by $C \cdot \mu(U_0 \cap W)$ (this is similar to the rectangle bound) or is dominated by $D \cdot \mu(U_2 \cap W)$. This immediately implies the rectangle bound of $g$ is $\Omega(n)$. We will prove the above statement for $D$ strictly smaller than 6. This fact implies that whenever $\mu(U_1 \cap W)$ is not dominated by $C \cdot \mu(U_0 \cap W)$ in $W$, the ratio of $U_2$-inputs to $U_1$-inputs in the rectangle $W$ is considerably more than the similar ratio globally (which is $\approx 1/6$). This fact lets us translate the $\Omega(n)$ rectangle bound for $g$ to a similar smooth-rectangle bound for Tribes.

We consider an exhaustive collection of sub-events such that conditioned on any such sub-event, the non-product distribution $\mu$ becomes a product distribution. Such handling of non-product

distributions, by decomposing them into several product distributions, has been done several times before, for instance in Razborov's proof [Raz92] of the optimal lower bound for the set-disjointness function. Assume such a conditioning exists for the rest of this proof outline.

How does one prove that for all large rectangles $W$, either $\mu(U_1 \cap W) \leq C\mu(U_0 \cap W))$ or $\mu(U_1 \cap W) \leq D\mu(U_2 \cap W)$ for some $D$ strictly smaller than 6. Note that one cannot prove for all rectangles $W$, $\mu(U_1 \cap W) \leq D\mu(U_2 \cap W)$ for some $D$ strictly less than 6, since this is false globally (i.e., $\mu(U_1) \approx 6\mu(U_2)$). Hence, one needs to do a case analysis[1]. And we do this based on the values of $\Pr\left[X_{l_j} = Y_{l_i} = 1\right]$ and $\Pr\left[X_{k_j} = Y_{k_j} = 1\right]$.

Consider the case when $\Pr\left[X_{l_j} = Y_{l_j} = 1\right] \geq \frac{3}{4}\mu(U_1 \cap W)$. Since the rectangle is large, using an entropy argument, we can argue that in most cases, conditioned on the sub-event $(X_{l_j} = Y_{l_j} = 1)$, both $\Pr\left[X_{k_j} = 1\right]$ and $\Pr\left[Y_{k_j} = 1\right]$ are large enough ($\approx 1/2$). Now since the distribution is product it means that conditioned on $(X_{l_j} = Y_{l_j} = 1)$, $\Pr\left[X_{k_j} = Y_{k_j} = 1\right]$ is large enough and hence $\mu(U_2 \cap W)$ is a required fraction of $\mu(U_1 \cap W)$. Similar arguments hold for the case with the roles of $l$ and $k$ reversed.

In the third case, when $\max\{\Pr\left[X_{l_j} = Y_{l_j} = 1\right], \Pr\left[X_{k_j} = Y_{k_j} = 1\right]\} \leq \frac{3}{4}\mu(U_1 \cap W)$, again using the same entropy argument, we can show that $\Pr\left[X_{l_j} = Y_{l_j} = 1, X_{k_j} = Y_{k_j} = 0\right]$ and $\Pr\left[X_{l_j} = Y_{l_j} = 0, X_{k_j} = Y_{k_j} = 1\right]$ are large. Now, since $W$ is a rectangle, using a *cut-and-paste* argument, we can show that $\Pr\left[X_{l_j} = 1, Y_{l_j} = 0, X_{k_j} = 0, Y_{k_j} = 1\right]$ and $\Pr\left[X_{l_j} = 0, Y_{l_j} = 1, X_{k_j} = 1, Y_{k_j} = 0\right]$ are large. This implies that $\mu(U_0 \cap W)$ is a required fraction of $\mu(U_1 \cap W)$. This concludes our proof outline.

We note that our distribution is similar to (and in fact inspired from) the distribution used by Jain and Klauck [JK10] while analyzing the query complexity of the Tribes function. We also note that the distribution used by Jayram, Kumar and Sivakumar [JKS03] in their information complexity lower bound for Tribes is different from our distribution, in particular, their distribution does not put any support on $U_2$, inputs which have intersections of size 2 within block $j$. However, we do add that they also use similar in spirit, albeit different cut-and-paste arguments in their lower bound proof.

## 2 Preliminaries

**Communication Complexity:** We begin by recalling the two-party communication model introduced by Yao [Yao79] (see Kushilevitz and Nisan [KN97] for an excellent introduction to the area). Let $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ be finite non-empty sets, and let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ be a function. A two-party protocol for computing $f$ consists of two parties, Alice and Bob, who get inputs $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ respectively, and exchange messages in order to compute $f(x, y) \in \mathcal{Z}$.

For a distribution $\mu$ on $\mathcal{X} \times \mathcal{Y}$, let the $\varepsilon$-error distributional communication complexity of $f$ under $\mu$ (denoted by $\mathsf{D}^{\mu}_{\varepsilon}(f)$), be the number of bits communicated (for the worst-case input) by the best deterministic protocol for $f$ with average error at most $\varepsilon$ under $\mu$. Let $\mathsf{R}^{\mathrm{pub}}_{\varepsilon}(f)$, the public-coin randomized communication complexity of $f$ with worst case error $\varepsilon$, be the number of bits communicated (for the worst-case input) by the best public-coin randomized protocol, that for each input $(x, y)$ computes $f(x, y)$ correctly with probability at least $1 - \varepsilon$. Randomized and distributional complexity are related by the following celebrated result of Yao [Yao83].

[1] Such a case analysis is not required to prove rectangle bound (c.f., proof of disjointness [Raz92]), but is necessary while proving a smooth-rectangle bound.

**Theorem 2.1** (Yao's minmax principle [Yao83]). $\mathsf{R}^{\mathrm{pub}}_{\varepsilon}(f) = \max_{\mu} \mathsf{D}^{\mu}_{\varepsilon}(f)$.

Given a function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$, the $k$-wise direct product of $f$, denoted by $f^{(k)}$ is the function $f : \mathcal{X}^k \times \mathcal{Y}^k \to \mathcal{Z}^k$ defined as follows: $f^{(k)}((x_1, \ldots, x_k), (y_1, \ldots, y_k)) = (f(x_1, y_1), \ldots, (x_k, y_k))$. The direct product/sum question involves relating $\mathsf{R}^{\mathrm{pub}}(f^{(k)})$ to $\mathsf{R}^{\mathrm{pub}}(f)$. More precisely, the strong direct product result conjectures that $\mathsf{R}^{\mathrm{pub}}_{1-2^{-\Omega(k)}}(f^{(k)}) = \Omega\left(k \cdot \mathsf{R}^{\mathrm{pub}}_{1/3}(f)\right)$.

**The smooth rectangle bound:** The smooth rectangle bound was introduced by Jain and Klauck [JK10], as a generalization of the rectangle bound. Informally, the smooth-rectangle bound for a function $f$ under a distribution $\mu$, is the maximum over all functions $g$ , which are close to $f$ under the distribution $\mu$, of the rectangle bound of $g$. However, it will be more convenient for us to work with the following linear programming formulation of smooth-rectangle bound. Please see [JK10, Lemma 2] and [JY12, Lemma 6] for the relations between the LP formulation and the more "natural" formulation in terms of rectangle bound. A broad connection between the two definitions is that the variable $\varphi$ in the dual of the linear programming definition takes non-zero values precisely at the inputs $(x, y)$ where $f$ and $g$ differ.

**Definition 2.2** (smooth-rectangle bound). *For a total Boolean function $f$, the $\varepsilon$- smooth rectangle bound of $f$ denoted $\mathsf{srec}_{\varepsilon}(f)$ is defined to be $\max\{\mathsf{srec}^{z}_{\varepsilon}(f) : z \in \{0,1\}\}$, where $\mathsf{srec}^{z}_{\varepsilon}(f)$ is given by the optimal value of the following linear program (below $\mathcal{W}$ represents the set of all rectangles in $\mathcal{X} \times \mathcal{Y}$).*

*Primal*

$$\begin{aligned}
&\textit{min:} \quad \sum_{W \in \mathcal{W}} v_W \\
&\forall (x,y) \in f^{-1}(z) : \sum_{W : (x,y) \in W} v_W \geq 1 - \varepsilon, \\
&\forall (x,y) \in f^{-1}(z) : \sum_{W : (x,y) \in W} v_W \leq 1, \\
&\forall (x,y) \notin f^{-1}(z) : \sum_{W : (x,y) \in W} v_W \leq \varepsilon, \\
&\forall W : v_W \geq 0 \ .
\end{aligned}$$

*Dual*

$$\begin{aligned}
&\textit{max:} \quad \sum_{(x,y) \in f^{-1}(z)} ((1-\varepsilon)\lambda_{x,y} - \varphi_{x,y}) - \sum_{(x,y) \notin f^{-1}(z)} \varepsilon \cdot \lambda_{x,y} \\
&\forall W : \sum_{(x,y) \in f^{-1}(z) \cap W} (\lambda_{x,y} - \varphi_{x,y}) - \sum_{(x,y) \in (W \setminus f^{-1}(z))} \lambda_{x,y} \leq 1, \\
&\forall (x,y) : \lambda_{x,y} \geq 0; \varphi_{x,y} \geq 0 \ .
\end{aligned}$$

**Theorem 2.3** ([JK10, Theorem 1]). *For all functions $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ and $\varepsilon \in (0,1)$, we have $\mathsf{R}^{\mathrm{pub}}_{\varepsilon}(f) \geq \log(\mathsf{srec}_{\varepsilon}(f))$.*

Jain and Yao [JY12] proved the following strong direct product theorem in terms of the smooth rectangle bound.

**Theorem 2.4** ([JY12, Theorem 1 and Lemma 6]). *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a Boolean function. For every $\varepsilon \in (0,1)$, there exists small enough $\eta \in (0, 1/3)$ such that the following holds. For all integers $k$,*

$$\mathsf{R}^{\mathrm{pub}}_{1-(1-\eta)^{\lfloor \eta^2 k/32 \rfloor}}(f^{(k)}) \geq \frac{\eta^2}{32} \cdot k \cdot \left(11\eta \cdot \log \mathsf{srec}_{\varepsilon}(f) - 3 \log \frac{1}{\varepsilon} - 2\right).$$

**Information theory:** We need the following basic facts from information theory. Let $\mu$ be a (probability) distribution on a finite set $\mathcal{X}$ and $X$ be a random variable distributed according to $\mu$. Let $\mu(x)$ represent the probability of $x \in \mathcal{X}$ according to $\mu$. The entropy of $X$ is defined as $H(X) \stackrel{\mathrm{def}}{=} \sum_x \mu(x) \cdot \log \frac{1}{\mu(x)}$. Entropy satisfies subadditivity: $H(XY) \leq H(X) + H(Y)$.

## 3 The smooth-rectangle bound for Tribes

In this section, we prove a linear lower bound on the randomized communication of Tribes via the smooth-rectangle bound.

First we introduce some notation. We will prove the result for $n$ of the form $(2r+1)^2$, where $r \geq 2$ is even. Assume the input indices $[n]$ to the Tribes function are partitioned into $\sqrt{n}$ blocks $s_1, \ldots, s_{\sqrt{n}}$, where the $i^{th}$ block $s_i = \{(i-1)\sqrt{n}+1, \ldots, i\sqrt{n}\}$. Thus,

$$\mathsf{Tribes}(x,y) = \bigwedge_{i=1}^{\sqrt{n}} \left( \bigvee_{j \in s_i} (x_j \wedge y_j) \right).$$

A string $x \in \{0,1\}^n$ can be viewed both as an $n$-bit string and as a subset $x \subseteq [n]$. We will use both these interpretations.

Consider the distribution $\mu(x,y)$ on the inputs of the Tribes function defined by the following procedure:

1. Choose $j \in [\sqrt{n}]$ uniformly.
   For each $i \in [\sqrt{n}] \setminus \{j\}$, choose a random partition $s_i = (t_i^A, t_i^B, \{l_i\})$ into 3 disjoint sets such that $|t_i^A| = |t_i^B| = r$ and $l_i \in s_i$.
   For index $j$, choose a random partition $s_j = (\tilde{t}_j^A, \tilde{t}_j^B, \{k_j\}, \{l_j\})$ into 4 disjoint sets such that $|\tilde{t}_j^A| = |\tilde{t}_j^B| = r-1$ and $k_j, l_j \in s_j$. Set $t_j^A = \tilde{t}_j^A \cup \{k_j\}$ and $t_j^B = \tilde{t}_j^B \cup \{k_j\}$.
   Let $t = \left(j, k_j, (t_i^A, t_i^B, l_i)_{i \in [\sqrt{n}]}\right)$.
2. For each $i \neq j \in [\sqrt{n}]$, set the variables in block $s_i$ as follows:
   - Set $x_{l_i} \leftarrow 1$ and $x_{s_i \setminus (t_i^A \cup \{l_i\})} \leftarrow \bar{0}$. Let $x_{t_i^A}$ be a random string of exactly $r/2$ ones.
   - Set $y_{l_i} \leftarrow 1$ and $y_{s_i \setminus (t_i^B \cup \{l_i\})} \leftarrow \bar{0}$. Let $y_{t_i^B}$ be a random string of exactly $r/2$ ones.
3. Set the variables in block $s_j$ as follows:
   - Let $x_{t_j^A \cup \{l_j\}}$ be a random string of exactly $r/2+1$ ones and $x_{s_j \setminus (t_j^A \cup \{l_j\})} \leftarrow \bar{0}$.
   - Let $y_{t_j^B \cup \{l_j\}}$ be a random string of exactly $r/2+1$ ones and $y_{s_j \setminus (t_j^B \cup \{l_j\})} \leftarrow \bar{0}$.

Let $(X,Y)$ be distributed according to $\mu$, where $X$ represents the input to Alice and $Y$ represents the input to Bob. Let $T = \left(J, K_J, (T_i^A, T_i^B, L_i)_{i \in [\sqrt{n}]}\right)$ be the random variable (correlated with $(X,Y)$) representing $t$ distributed as above. Observe that though $(X,Y)$ is not a product distribution, the conditional distribution $((X,Y) \mid T = t)$ is product for each $t$.

Partition the set of inputs (in the support of $\mu$) into 3 sets $U_0, U_1$ and $U_2$ as follows:

$$U_i = \{(x,y) \mid \mu(x,y) > 0 \text{ and sets } x \text{ and } y \text{ have exactly } \sqrt{n} - 1 + i \text{ intersections}\}.$$

Note that $U_0$ are the 0-inputs and $U_1 \cup U_2$ the 1-inputs of the Tribes function while $U_0 \cup U_2$ and $U_1$ are the 0- and 1-inputs respectively of the function $g$ described in Section 1.2.

Let $\beta \stackrel{\text{def}}{=} \frac{r+2}{r+1}$. The following facts can be easily verified from the definition of the distribution $\mu$. For all $t$,

$$\Pr\left[X_{l_j} = 1 \mid T = t\right] = \frac{\beta}{2}; \quad \Pr\left[X_{l_j} = X_{k_j} = 1 \mid T = t\right] = \Pr\left[X_{l_j} = 1, X_{k_j} = 0 \mid T = t\right] = \frac{\beta}{4}.$$

Given this, it can be easily checked that the weights of the sets $U_0, U_1$ and $U_2$ are as follows: $\mu(U_0) = 1 - 7\beta^2/16, \mu(U_1) = 6\beta^2/16$, and $\mu(U_2) = \beta^2/16$.

Our main lemma is the following (we have not optimized the constants).

**Lemma 3.1.** *There exists a constant $\delta \in (0,1)$ such that for sufficiently large $n$, the following holds: for every rectangle $W = A \times B$, we have*

$$0.99\mu(U_1 \cap W) \leq \frac{16}{3(0.99)^2} \cdot \mu(U_2 \cap W) + \frac{16}{(0.99)^2}\mu(U_0 \cap W) + 2^{-\delta n/2+1}.$$

In other words, in any rectangle which contains a significant fraction of inputs from $U_1$ (i.e., at least $2^{-\delta n/2+1}$), the weight of the $U_1$ inputs is dominated by some linear function of the weights of $U_0$ and $U_2$ inputs. Before proving this lemma, let us first see how this lemma implies the smooth-rectangle bound for Tribes, which implies our Main Theorem 1.1

**Theorem 3.2** (smooth-rectangle bound for Tribes)**.** *There exists $\gamma \in (0,1)$ such that for all sufficiently large $n$ and $\varepsilon < 1/1000$, we have:* $\mathsf{srec}^1_\varepsilon(\mathsf{Tribes}) \geq 2^{\gamma \cdot n}$.

*Proof.* We will prove the bound using the dual formulation for smooth-rectangle boundgiven in Definition 2.2. Define the dual variables $\lambda_{x,y}$ and $\varphi_{x,y}$ as follows:

$$\lambda_{x,y} = \begin{cases} 0 & \text{if} \quad (x,y) \in U_2 \\ 0.99\mu(x,y)2^{\delta n/2-1} & \text{if} \quad (x,y) \in U_1 \\ \frac{16}{(0.99)^2}\mu(x,y)2^{\delta n/2-1} & \text{if} \quad (x,y) \in U_0. \end{cases}$$

$$\varphi_{x,y} = \begin{cases} \frac{16}{3(0.99)^2}\mu(x,y)2^{\delta n/2-1} & \text{if} \quad (x,y) \in U_2 \\ 0 & \text{if} \quad (x,y) \in U_1 \cup U_0. \end{cases}$$

From Lemma 3.1 we get

$$\forall \text{ rectangles } W: \sum_{(x,y) \in \mathsf{Tribes}^{-1}(1) \cap W} (\lambda_{x,y} - \varphi_{x,y}) - \sum_{(x,y) \in (W - \mathsf{Tribes}^{-1}(1))} \lambda_{x,y} \leq 1.$$

The objective of the LP can be bounded as follows:

$$\begin{aligned} &\sum_{(x,y) \in \mathsf{Tribes}^{-1}(1)} ((1-\varepsilon)\lambda_{x,y} - \varphi_{x,y}) - \sum_{(x,y) \notin \mathsf{Tribes}^{-1}(1)} \varepsilon \cdot \lambda_{x,y} \\ &\geq \left( (0.999)(0.99)\mu(U_1) - \frac{16}{3(0.99)^2}\mu(U_2) - \frac{16}{1000(0.99)^2}\mu(U_0) \right) 2^{\delta n/2-1} \\ &\geq 0.02 \cdot 2^{\delta n/2-1} \qquad \text{(for sufficiently large n).} \end{aligned}$$

Thus, proved. □

Corollary 1.2 follows by combining the above theorem and Jain-Yao's strong direct product theorem in terms of the smooth-rectangle bound (Theorem 2.4).

### 3.1 Proof of Lemma 3.1

Let $W = A \times B$ be the rectangle. For each $t = \left(j, k_j, (t_i^A, t_i^B, l_i)_{i\in[\sqrt{n}]}\right)$ and $a, b \in \{0,1\}$, define,

$$R(t,a,b) = \Pr\left[X \in A \mid T = t, X_{l_j} = a, X_{k_j} = b\right]; \qquad R(t,a) = \Pr\left[X \in A \mid T = t, X_{l_j} = a\right];$$
$$C(t,a,b) = \Pr\left[Y \in B \mid T = t, Y_{l_j} = a, Y_{k_j} = b\right]; \qquad C(t,a) = \Pr\left[Y \in B \mid T = t, Y_{l_j} = a\right].$$

Define the following random variables (we will set $\delta$ later):

$$\mathsf{BAD}_{\mathsf{A}}(t) = 1 \text{ iff } \min\{R(t,1,1), R(t,1,0)\} < 0.99\left(R(t,1) - 2^{-\delta n}\right),$$

and symmetrically,

$$\mathsf{BAD}_{\mathsf{B}}(t) = 1 \text{ iff } \min\{C(t,1,1), C(t,1,0)\} < 0.99\left(C(t,1) - 2^{-\delta n}\right).$$

For a given $t$, let $t'$ denote a partition identical to $t$ except that the role of the indices $l_j$ and $k_j$ are exchanged (i.e., $k'_j = l_j, l'_j = k_j, (t_j^A)' = \tilde{t}_j^A \cup \{l_j\}$ and $(t_j^B)' = \tilde{t}_j^B \cup \{l_j\}$). To define $\mathsf{BAD}(t)$, we need the following two quantities.

$$\begin{aligned}\rho_l(t) &= \Pr\left[X_{l_j} = Y_{l_j} = 1, X \in A, Y \in B, (X,Y) \in U_1 \mid T = t\right],\\ \rho_k(t) &= \Pr\left[X_{k_j} = Y_{k_j} = 1, X \in A, Y \in B, (X,Y) \in U_1 \mid T = t\right].\end{aligned}$$

Observe that $\mu(U_1 \cap W \mid T = t) = \rho_l(t) + \rho_k(t)$. Hence, it must be the case that exactly one of the following happens: (1) $\rho_l(t) > 3\mu(U_1 \cap W \mid T = t)/4$, (2) $\rho_k(t) > 3\mu(U_1 \cap W \mid T = t)/4$ or (3) $\max\{\rho_l(t), \rho_l(t)\} \le 3\mu(U_1 \cap W \mid T = t)/4$ (equivalently, $\min\{\rho_l(t), \rho_l(t)\} \ge \mu(U_1 \cap W \mid T = t)/4$). We define $\mathsf{BAD}(t)$ based on these cases as follows.

$$\mathsf{BAD}(t) = \begin{cases} \mathsf{BAD}_{\mathsf{A}}(t) \vee \mathsf{BAD}_{\mathsf{B}}(t), & \text{if } \rho_l(t) > 3\mu(U_1 \cap W \mid T = t)/4 \\ \mathsf{BAD}_{\mathsf{A}}(t') \vee \mathsf{BAD}_{\mathsf{B}}(t'), & \text{if } \rho_k(t) > 3\mu(U_1 \cap W \mid T = t)/4 \\ \mathsf{BAD}_{\mathsf{A}}(t) \vee \mathsf{BAD}_{\mathsf{B}}(t) \vee \mathsf{BAD}_{\mathsf{A}}(t') \vee \mathsf{BAD}_{\mathsf{B}}(t'), & \text{otherwise.} \end{cases} \quad (3.1)$$

The following claim shows that the probability that $\mathsf{BAD}_{\mathsf{A}}(T)$ and $\mathsf{BAD}_{\mathsf{B}}(T)$ occurs is small.

**Claim 3.3.** *There exists a small fixed constant $\delta > 0$ such that for sufficiently large $n$, the following holds: for any $(t_i^A, l_i)_{i\in[\sqrt{n}]}$, we have*

$$\Pr\left[\mathsf{BAD}_{\mathsf{A}}(T) = 1 \mid T_i^A = t_i^A, L_i = l_i, \text{ for each } i \in [\sqrt{n}]\right] < \frac{1}{6400}.$$

*(Symmetrically, for any $(t_i^B, l_i)_i$, $\Pr\left[\mathsf{BAD}_{\mathsf{B}}(T) = 1 \mid T_i^B = t_i^B, L_i = l_i, \text{ for each } i \in [\sqrt{n}]\right] < \frac{1}{6400}$.)*

*Proof.* We prove the inequality involving $\mathsf{BAD}_{\mathsf{A}}(T)$. The other inequality is proved similarly. We first consider the easy case when $(t_i^A, l_i)_{i\in[\sqrt{n}]}$ satisfies

$$\Pr\left[X \in A \mid X_{l_i} = 1, T_i^A = t_i^A, L_i = l_i, \text{ for each } i \in [\sqrt{n}]\right] < 2^{-\delta n}.$$

It follows from the definition of the distribution $\mu$, that the above probability is unchanged on further conditioning by $T = t$ for any $t$ consistent with $(t_i^A, l_i)_{i\in[\sqrt{n}]}$. In other words, this probability

is equal to $R(t,1) = \Pr\left[X \in A \mid T = t, X_{l_j} = 1\right]$ for any $t$ consistent with $(t_i^A, l_i)_{i\in[\sqrt{n}]}$. Hence, for any such $t$ we have that $R(t,1) < 2^{-\delta n}$. Thus, in this case $\mathsf{BAD}_\mathsf{A}(t) = 0$ for all such $t$ and we are done.

Now consider the other case when

$$\Pr\left[X \in A \mid X_{l_i} = 1, T_i^A = t_i^A, L_i = l_i, \text{ for each } i \in [\sqrt{n}]\right] \geq 2^{-\delta n}. \tag{3.2}$$

Consider a $t = (j, k_j, (t_i^A, t_i^B, l_i)_{i\in[\sqrt{n}]})$ consistent with $(t_i^A, l_i)_{i\in[\sqrt{n}]}$. We know that the bit $(X_{k_j} \mid T = t, X_{l_j} = 1)$ is a unbiased bit. Now, suppose $\mathsf{BAD}_\mathsf{A}(t) = 1$. Then, for some $a \in \{0,1\}$, we have

$$\Pr\left[X \in A \mid T = t, X_{l_j} = 1, X_{k_j} = a\right] < 0.99\left(\Pr\left[X \in A \mid T = t, X_{l_j} = 1\right]\right).$$

By a simple rewriting of the above inequality, we have

$$\Pr\left[X_{k_j} = a \mid X \in A, T = t, X_{l_j} = 1\right] < 0.99\left(\Pr\left[X_{k_j} = a \mid T = t, X_{l_j} = 1\right]\right) = 0.99/2. \tag{3.3}$$

In other words, the unbiased bit $(X_{k_j} \mid T = t, X_{l_j} = 1)$ when conditioned on the event "$X \in A$" is now more likely to be $1 - a$ than $a$.

Suppose, for contradiction, that

$$\Pr\left[\mathsf{BAD}_\mathsf{A}(T) = 1 \mid T_i^A = t_i^A, L_i = l_i, \text{ for each } i \in [\sqrt{n}]\right] \geq \frac{1}{6400}.$$

Consider the random variable

$$Z \stackrel{\text{def}}{=} (X \mid X_{l_i} = 1, T_i^A = t_i^A, L_i = l_i, \text{ for each } i \in [\sqrt{n}]).$$

Note that the distribution of $Z$ is uniform and each string has probability $\left(\frac{1}{\binom{r}{r/2}}\right)^{\sqrt{n}}$. Consider the event $E \stackrel{\text{def}}{=}$ "$X \in A$", which by (3.2) has probability at least $2^{-\delta n}$. Therefore the probability of each string in the distribution $(Z|E)$ would be at most $2^{\delta n} \cdot \left(\frac{1}{\binom{r}{r/2}}\right)^{\sqrt{n}}$. Therefore, using standard estimates on binomial coefficients,

$$H(Z|E) \geq \sqrt{n} \cdot \log \binom{r}{r/2} - \delta n \geq \sqrt{n} \cdot r(1 - o(1)) - \delta n.$$

Observe that conditioned on $T_i^A = t_i^A, L_i = l_i,$ for each $i \in [\sqrt{n}]$, the index $K_J$ can equally likely be any one of the $r\sqrt{n}$ indices in $\bigcup_i t_i^A$ (each resulting in a different value for $T$). Furthermore, from (3.3), we have that conditioning on $E$ causes $H(X_{K_J}) \leq H(0.99/2)$ if $\mathsf{BAD}_\mathsf{A}(T) = 1$. Using these facts, we can upper bound the entropy of $(Z|E)$ as follows:

$$\begin{aligned} H(Z|E) &\leq \sum_i H(Z_i|E) && \text{[By subadditivity of entropy]} \\ &\leq r\sqrt{n}\left(\frac{H(0.99/2)}{6400} + \left(1 - \frac{1}{6400}\right)\right). \end{aligned}$$

Combining the upper and lower bounds on $H(Z|E)$, we get

$$\delta n \geq (1 - H(0.99/2) - o(1)) \cdot \frac{r\sqrt{n}}{6400}.$$

Thus, if $\delta > 0$ is small enough we get a contradiction. □

The following claim shows that a version of Lemma 3.1 is true when $\mathsf{BAD}(t) = 0$.

**Claim 3.4.** *Let $n$ be large enough. If $\mathsf{BAD}(t) = 0$, then,*

$$\mu(U_1 \cap W \mid T = t) \leq \frac{16}{3(0.99)^2} \cdot \mu(U_2 \cap W \mid T = t) + \frac{16}{(0.99)^2} \cdot \mu(U_0 \cap W \mid T = t) + 2^{-\delta n/2}.$$

The following claim argues that not much probability is lost when $\mathsf{BAD}(T) = 1$. Our proof of this claim differs significantly from the proof of a similar claim in Razborov's result [Raz92] of linear lower bound for set-disjointness. This is because we need to consider several sub-events of $U_1$. Our arguments are more general and in fact can also be used in the context of set-disjointness.

**Claim 3.5.** *Let $n$ be large enough. Then,*

$$\mathop{\mathbb{E}}_{t \leftarrow T}[\mu(U_1 \cap W \mid T = t) \cdot \mathsf{BAD}(t)] \leq \frac{1}{100} \cdot \mathop{\mathbb{E}}_{t \leftarrow T}[\mu(W \cap U_1 \mid T = t)] + 2^{-\delta n + 3}.$$

The proofs of Claims 3.4-3.5 are deferred to the Appendix A. Lemma 3.1 follows by combining Claim 3.4 and Claim 3.5 as follows.

$$\begin{aligned}
0.99\mu(U_1 \cap W) &= 0.99 \mathop{\mathbb{E}}_{t \leftarrow T}[\mu(U_1 \cap W \mid T = t)] \\
&\leq \mathop{\mathbb{E}}_{t \leftarrow T}[\mu(U_1 \cap W \mid T = t) \cdot (1 - \mathsf{BAD}(t))] + 2^{-\delta n + 3} \qquad \text{(from Claim 3.5)} \\
&\leq \mathop{\mathbb{E}}_{t \leftarrow T}\left[\left(\frac{16}{3(0.99)^2} \cdot \mu(U_2 \cap W \mid T = t) + \frac{16}{(0.99)^2} \cdot \mu(U_0 \cap W \mid T = t) + 2^{-\delta n/2}\right) \cdot (1 - \mathsf{BAD}(t))\right] \\
&\qquad \text{(from Claim 3.4)} \\
&\leq \frac{16}{3(0.99)^2} \cdot \mu(U_2 \cap W) + \frac{16}{(0.99)^2} \cdot \mu(U_0 \cap W) + 2^{-\delta n/2 + 1}
\end{aligned}$$

□

## References


[BBCR10] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. *How to compress interactive communication.* In *Proc. 42nd ACM Symp. on Theory of Computing (STOC)*, pages 67–76. 2010. doi:10.1145/1806689.1806701.

[BJKS04] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. *An information statistics approach to data stream and communication complexity.* J. Computer and System Sciences, 68(4):702–732, June 2004. (Preliminary version in *43rd FOCS*, 2002). doi:10.1016/j.jcss.2003.11.006.

[CR12] Amit Chakrabarti and Oded Regev. *An optimal lower bound on the communication complexity of Gap-Hamming-distance.* SIAM J. Computing, 41(5):1299–1317, 2012. (Preliminary version in *43rd STOC*, 2011). arXiv:1009.3460, doi:10.1137/120861072.

[CSWY01] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Chi-Chih Yao. *Informational complexity and the diect sum problem for simultaneous message complexity.* In *Proc. 42nd IEEE Symp. on Foundations of Comp. Science (FOCS)*, pages 270–278. 2001. doi:10.1109/SFCS.2001.959901.

[JK10] Rahul Jain and Hartmut Klauck. *The partition bound for classical communication complexity and query complexity.* In *Proc. 25th IEEE Conference on Computational Complexity*, pages 247–258. 2010. arXiv:0910.4266, doi:10.1109/CCC.2010.31.

[JKR09] T. S. Jayram, Swastik Kopparty, and Prasad Raghavendra. *On the communication complexity of read-once $AC^0$ formulae.* In *Proc. 24th IEEE Conference on Computational Complexity*, pages 329–340. 2009. doi:10.1109/CCC.2009.39.

[JKS03] T. S. Jayram, Ravi Kumar, and D. Sivakumar. *Two applications of information complexity.* In *Proc. 35th ACM Symp. on Theory of Computing (STOC)*, pages 673–682. 2003. doi:10.1145/780542.780640.

[JPY12] Rahul Jain, Attila Pereszlényi, and Penghui Yao. *A direct product theorem for the two-party bounded-round public-coin communication complexity.* In *Proc. 53th IEEE Symp. on Foundations of Comp. Science (FOCS)*, pages 167–176. 2012. arXiv:1201.1666, doi:10.1109/FOCS.2012.42.

[JY12] Rahul Jain and Penghui Yao. *A strong direct product theorem in terms of the smooth rectangle bound*, 2012. arXiv:1209.0263.

[Kla10] Hartmut Klauck. *A strong direct product theorem for disjointness.* In *Proc. 42nd ACM Symp. on Theory of Computing (STOC)*, pages 77–86. 2010. arXiv:0908.2940, doi:10.1145/1806689.1806702.

[KLL+12] Iordanis Kerenidis, Sophie Laplante, Virginie Lerays, Jérémie Roland, and David Xiao. *Lower bounds on information complexity via zero-communication protocols and applications.* In *Proc. 53th IEEE Symp. on Foundations of Comp. Science (FOCS)*, pages 500–509. 2012. arXiv:1204.1505, doi:10.1109/FOCS.2012.68.

[KN97] Eyal Kushilevitz and Noam Nisan. *Communication Complexity.* Cambridge University Press, 1997. doi:10.2277/052102983X.

[KRW95] Mauricio Karchmer, Ran Raz, and Avi Wigderson. *Super-logarithmic depth lower bounds via the direct sum in communication complexity.* Comput. Complexity, 5(3/4):191–204, 1995. (Preliminary version in *6th Structure in Complexity Theory Conference*, 1991). doi:10.1007/BF01206317.

[KS92] Bala Kalyanasundaram and Georg Schnitger. *The probabilistic communication complexity of set intersection.* SIAM J. Discrete Math., 5(4):545–557, 1992. (Preliminary version in *2nd Structure in Complexity Theory Conference*, 1987). doi:10.1137/0405044.

[LS10] Nikos Leonardos and Michael Saks. *Lower bounds on the randomized communication complexity of read-once functions.* cc, 19(2):153–181, 2010. (Preliminary version in *24th IEEE Conference on Computational Complexity*, 2009). eccc:TR09-010, doi:10.1007/s00037-010-0292-2.

[Raz92] Alexander A. Razborov. *On the distributional complexity of disjointness.* Theoretical Comp. Science, 106(2):385–390, 1992. doi:10.1016/0304-3975(92)90260-M.

[Yao79] Andrew Chi-Chih Yao. *Some complexity questions related to distributive computing (preliminary report).* In *Proc. 11th ACM Symp. on Theory of Computing (STOC)*, pages 209–213. 1979. doi:10.1145/800135.804414.

[Yao83] ———. *Lower bounds by probabilistic arguments (extended abstract).* In *Proc. 24th IEEE Symp. on Foundations of Comp. Science (FOCS)*, pages 420–428. 1983. doi:10.1109/SFCS.1983.30.

# A Proofs

*Proof of Claim 3.4.* Recall the definition of $\mathsf{BAD}(t)$ from (3.1). We will consider three cases depending on the relative sizes of $\rho_l(t)$ and $\rho_k(t)$ with respect to $\mu(U_1 \cap W \mid T = t)$.

- $\rho_l(t) > 3\mu(U_1 \cap W \mid T = t)/4$: In this case, we have $\mathsf{BAD}_\mathsf{A}(t) \vee \mathsf{BAD}_\mathsf{B}(t) = 0$. We can now bound $\mu(U_1 \cap W \mid T = t)$ as follows.

$$\begin{aligned}\frac{3}{4} \cdot \mu(U_1 \cap W \mid T = t) < \rho_l(t) &\leq \Pr\left[X_{l_j} = Y_{l_j} = 1, X \in A, Y \in B \mid T = t\right] \\ &= \Pr\left[X_{l_j} = Y_{l_j} = 1 \mid T = t\right] \cdot R(t,1) \cdot C(t,1) \\ &\leq \frac{\beta^2}{4} \cdot \left(\frac{R(t,1,1)}{0.99} + 2^{-\delta n}\right) \cdot \left(\frac{C(t,1,1)}{0.99} + 2^{-\delta n}\right) \\ &\leq \frac{\beta^2}{4(0.99)^2}(R(t,1,1)C(t,1,1)) + 2^{-\delta n} \\ &= \frac{4}{(0.99)^2} \cdot \mu(U_2 \cap W \mid T = t) + 2^{-\delta n}.\end{aligned}$$

- $\rho_k(t) > 3\mu(U_1 \cap W \mid T = t)/4$: Similar arguments as above show

$$\frac{3}{4} \cdot \mu(U_1 \cap W \mid T = t) < \frac{4}{(0.99)^2} \cdot \mu(U_2 \cap W \mid T = t) + 2^{-\delta n}.$$

- $\min\{\rho_l(t), \rho_k(t)\} \geq \mu(U_1 \cap W \mid T = t)/4$: From $\rho_l(t) \geq \mu(U_1 \cap W \mid T = t)/4$, we have

$$\begin{aligned}\frac{1}{4} \cdot \mu(U_1 \cap W \mid T = t) \leq \rho_l(t) &\leq \Pr\left[X_{l_j} = Y_{l_j} = 1, X \in A, Y \in B \mid T = t\right] \\ &= \frac{\beta^2}{4} \cdot R(t,1)C(t,1) \\ &\leq \frac{\beta^2}{4} \cdot \left(\frac{R(t,1,0)}{0.99} + 2^{-\delta n}\right) \cdot \left(\frac{C(t,1,0)}{0.99} + 2^{-\delta n}\right) \\ &\leq \frac{\beta^2}{4(0.99)^2}(R(t,1,0)C(t,1,0)) + 2^{-\delta n}.\end{aligned}$$

  Similarly from $\rho_k(t) \geq \mu(U_1 \cap W \mid T = t)/4$, we have

$$\begin{aligned}\frac{1}{4} \cdot \mu(U_1 \cap W \mid T = t) &\leq \Pr\left[X_{k_j} = Y_{k_j} = 1, X \in A, Y \in B \mid T = t\right] \\ &\leq \frac{\beta^2}{4(0.99)^2}(R(t,0,1)C(t,0,1)) + 2^{-\delta n}.\end{aligned}$$

Multiplying the above two inequalities we have,

$$\begin{aligned}
&\left(\frac{1}{4}\cdot\mu(U_1\cap W\mid T=t)\right)^2\\
&\le\left(\frac{\beta^2}{4(0.99)^2}\cdot(R(t,1,0)C(t,1,0))+2^{-\delta n}\right)\left(\frac{\beta^2}{4(0.99)^2}\cdot(R(t,0,1)C(t,0,1))+2^{-\delta n}\right)\\
&\le\frac{\beta^4}{4^2(0.99)^4}\cdot(R(t,1,0)C(t,1,0)R(t,0,1)C(t,0,1))+2^{-\delta n} &\text{(A.1)}\\
&=\frac{\beta^4}{4^2(0.99)^4}\cdot(R(t,1,0)C(t,0,1)R(t,0,1)C(t,1,0))+2^{-\delta n} &\text{(A.2)}\\
&=\frac{4^2}{(0.99)^4}\cdot\Pr\left[(X_{l_j},X_{k_j},Y_{l_j},Y_{k_j})=(0,1,1,0),X\in A,Y\in B\mid T=t\right]\cdot\\
&\qquad\Pr\left[(X_{l_j},X_{k_j},Y_{l_j},Y_{k_j})=(1,0,0,1),X\in A,Y\in B\mid T=t\right]+2^{-\delta n}\\
&\le\frac{4^2}{(0.99)^4}\cdot(\mu(U_0\cap W\mid T=t))^2+2^{-\delta n}. &\text{(A.3)}
\end{aligned}$$

Observe that (A.2) is obtained from (A.1) by re-ordering the terms, which in communication complexity jargon is more commonly referred to as the cut-and-paste-property. (A.3) implies,

$$\frac{1}{4}\cdot\mu(U_1\cap W\mid T=t)\le\frac{4}{(0.99)^2}\cdot\mu(U_0\cap W\mid T=t)+2^{-\delta n/2}.$$

Combining the three cases yields the claim. □

*Proof of Claim 3.5.* For a partition $t$, define $\mathsf{BAD}_{\mathsf{A}\vee\mathsf{B}}(t)=1$ if either $\mathsf{BAD}_{\mathsf{A}}(t)=1$ or $\mathsf{BAD}_{\mathsf{B}}(t)=1$. We first show that for all partitions $t$,

$$\mu(U_1\cap W\mid T=t)\cdot\mathsf{BAD}(t)\le 4\left(\rho_l(t)\cdot\mathsf{BAD}_{\mathsf{A}\vee\mathsf{B}}(t)+\rho_k(t)\cdot\mathsf{BAD}_{\mathsf{A}\vee\mathsf{B}}(t')\right). \tag{A.4}$$

As before, we consider three cases depending on the relative sizes of $\rho_l(t)$ and $\rho_k(t)$ with respect to $\mu(U_1\cap W\mid T=t)$.

- $\rho_l(t)>3\mu(U_1\cap W\mid T=t)/4$: In this case, we have $\mathsf{BAD}(t)=\mathsf{BAD}_{\mathsf{A}\vee\mathsf{B}}(t)$. Thus, $\mu(U_1\cap W\mid T=t)\cdot\mathsf{BAD}(t)\le\frac{4}{3}\cdot\rho_l(t)\cdot\mathsf{BAD}_{\mathsf{A}\vee\mathsf{B}}(t)$.
- $\rho_k(t)>3\mu(U_1\cap W\mid T=t)/4$: In this case, we have $\mathsf{BAD}(t)=\mathsf{BAD}_{\mathsf{A}\vee\mathsf{B}}(t')$. Thus, $\mu(U_1\cap W\mid T=t)\cdot\mathsf{BAD}(t)\le\frac{4}{3}\cdot\rho_k(t)\cdot\mathsf{BAD}_{\mathsf{A}\vee\mathsf{B}}(t')$.
- $\min\{\rho_l(t),\rho_k(t)\}\ge\mu(U_1\cap W\mid T=t)/4$: In this case, we have $\mathsf{BAD}(t)\le\mathsf{BAD}_{\mathsf{A}\vee\mathsf{B}}(t)+\mathsf{BAD}_{\mathsf{A}\vee\mathsf{B}}(t')$. Hence, we have

$$\begin{aligned}
\mu(U_1\cap W\mid T=t)\cdot\mathsf{BAD}(t)&\le\mu(U_1\cap W\mid T=t)\cdot(\mathsf{BAD}_{\mathsf{A}\vee\mathsf{B}}(t)+\mathsf{BAD}_{\mathsf{A}\vee\mathsf{B}}(t'))\\
&\le 4\left(\rho_l(t)\cdot\mathsf{BAD}_{\mathsf{A}\vee\mathsf{B}}(t)+\rho_k(t)\cdot\mathsf{BAD}_{\mathsf{A}\vee\mathsf{B}}(t')\right).
\end{aligned}$$

The bound in (A.4) follows from combining the three cases.

We now argue that

$$\mathop{\mathbb{E}}_{t\leftarrow T}[\rho_l(t)\cdot\mathsf{BAD}_{\mathsf{A}\vee\mathsf{B}}(t)]\le\frac{1}{800}\cdot\mathop{\mathbb{E}}_{t\leftarrow T}[\mu(W\cap U_1\mid T=t)]+2^{-\delta n}. \tag{A.5}$$

A similar bound holds for $\mathbb{E}_{t\leftarrow T}[\rho_k(t)\cdot \mathsf{BAD}_{\mathsf{A}\vee\mathsf{B}}(t')]$. Combining these two bounds with (A.4) yields the statement of the claim.

We prove (A.5) by first showing that for each partition $t$, we have

$$\rho_l(t)\cdot \mathsf{BAD}_{\mathsf{A}\vee\mathsf{B}}(t) \leq \frac{1}{2}\cdot\left(R(t,1,0)\cdot C(t,1)\cdot \mathsf{BAD}_{\mathsf{B}}(t) + R(t,1)\cdot C(t,1,0)\cdot \mathsf{BAD}_{\mathsf{A}}(t) + 2^{-\delta n}\right). \quad \text{(A.6)}$$

We consider various cases depending on the values of $\mathsf{BAD}_{\mathsf{A}}(t)$ and $\mathsf{BAD}_{\mathsf{B}}(t)$.

- $\mathsf{BAD}_{\mathsf{A}}(t) = \mathsf{BAD}_{\mathsf{B}}(t)$: We first bound $\rho_l(t)$ as follows:

$$\begin{aligned}\rho_l(t) &= \Pr\left[X_{l_j} = Y_{l_j} = 1, X\in A, Y\in B, (X,Y)\in U_1 \mid T = t\right],\\ &\leq \Pr\left[X_{l_j} = Y_{l_j} = 1, X_{k_j} = 0, X\in A, Y\in B, \mid T = t\right]\\ &\quad + \Pr\left[X_{l_j} = Y_{l_j} = 1, , Y_{k_j} = 0, X\in A, Y\in B \mid T = t\right],\\ &= \frac{\beta^2}{8}\left(R(t,1,0)\cdot C(t,1) + R(t,1)\cdot C(t,1,0)\right).\end{aligned}$$

  (A.6) then follows by observing that in this case $\mathsf{BAD}_{\mathsf{A}\vee\mathsf{B}}(t) = \mathsf{BAD}_{\mathsf{A}}(t) = \mathsf{BAD}_{\mathsf{B}}(t)$.

- $\mathsf{BAD}_{\mathsf{A}}(t) = 1, \mathsf{BAD}_{\mathsf{B}}(t) = 0$: Since $\mathsf{BAD}_{\mathsf{B}}(t) = 0$, we have that $C(t,1) \leq C(t,1,0)/0.99 + 2^{-\delta n}$. We now bound $\rho_l(t)$ as follows.

$$\begin{aligned}\rho_l(t) &\leq \Pr\left[X_{l_j} = Y_{l_j} = 1, X\in A, Y\in B, \mid T = t\right],\\ &= \frac{\beta^2}{4}\cdot R(t,1)\cdot C(t,1) \leq \frac{\beta^2}{4(0.99)}\cdot\left(R(t,1)\cdot C(t,1,0) + 2^{-\delta n}\right)\end{aligned}$$

  (A.6) then follows by observing that in this case $\mathsf{BAD}_{\mathsf{A}\vee\mathsf{B}}(t) = \mathsf{BAD}_{\mathsf{A}}(t)$.

- $\mathsf{BAD}_{\mathsf{A}}(t) = 0, \mathsf{BAD}_{\mathsf{B}}(t) = 1$: This case is similar to the above case.

We now bound $\mathbb{E}_{t\leftarrow T}[R(t,1,0)\cdot C(t,1)\cdot \mathsf{BAD}_{\mathsf{B}}(t)]$. We will bound this expectation by setting the random variable $T$ in stages: we first set $t_B = \{t_i^B, l_i : \ i\in[\sqrt{n}]\}$, and then set the variable $k_j\in[n]$ from the distribution $(K_J \mid T^B = t^B)$. We observe that $C(t,1)$ is only a function of $t_B$ and independent of $k_j$; thus, $C(t,1) = c(t_B)$ for some function $c$. Similarly $R(t,1,0)$ is only a function of $t_B$ and is independent of $k_j$; thus, $R(t,1,0) = r(t_B)$ for some function $r$. We have $\mathsf{BAD}_{\mathsf{B}}(t) = b(t_B, k_j)$ for some function $b$. In this notation, Claim 3.3 states that for all $t_B$, $\mathbb{E}_{k_j\leftarrow K_J|T_B = t_B}[b(t_B,k_j)] \leq 1/6400$.

$$\begin{aligned}\mathop{\mathbb{E}}_{t\leftarrow T}[R(t,1,0)\cdot C(t,1)\cdot \mathsf{BAD}_{\mathsf{B}}(t)] &= \mathop{\mathbb{E}}_{t_B\leftarrow T_B}\left[c(t_B)\cdot r(t_B)\cdot \mathop{\mathbb{E}}_{k_j\leftarrow K_J|T_B=t_B}[b(t_B,k_j)]\right]\\ &= \mathop{\mathbb{E}}_{t_B\leftarrow T_B}\left[c(t_B)\cdot r(t_B)\cdot \mathop{\mathbb{E}}_{k_j\leftarrow K_J|T_B=t_B}[b(t_B,k_j)]\right]\\ &\leq \frac{1}{6400}\cdot \mathop{\mathbb{E}}_{t_B\leftarrow T_B}[c(t_B)\cdot r(t_B)]\\ &= \frac{1}{6400}\cdot \mathop{\mathbb{E}}_{t\leftarrow T}[R(t,1,0)\cdot C(t,1)]\\ &\leq \frac{8}{6400}\cdot \mathop{\mathbb{E}}_{t\leftarrow T}[\mu(U_1\cap W|T=t)].\end{aligned}$$

Hence,

$$\mathop{\mathbb{E}}_{t\leftarrow T}[R(t,1,0)\cdot C(t,1)\cdot \mathsf{BAD}_{\mathsf{B}}(t)] \leq \frac{1}{800}\mathop{\mathbb{E}}_{t\leftarrow T}[\mu(U_1\cap W \mid T=t)].$$

A similar bound holds for $\mathbb{E}_{t\leftarrow T}[R(t,1)\cdot C(t,1,0)\cdot \mathsf{BAD}_{\mathsf{A}}(t)]$. Combining these bounds with (A.6) yields (A.5) which completes the proof of the claim.

□