A strong direct product theorem in terms of the smooth rectangle bound

Rahul Jain
Centre for Quantum Technologies
and Department of Computer Science
National U. Singapore
E-mail: rahul@comp.nus.edu.sg.

Penghui Yao
Centre for Quantum Technologies
National U. Singapore
E-mail: pyao@nus.edu.sg.

Abstract

A strong direct product theorem states that, in order to solve k instances of a problem, if we provide less than k times the resource required to compute one instance, then the probability of overall success is exponentially small in k. In this paper, we consider the model of two-way public-coin communication complexity and show a strong direct product theorem for all relations in terms of the *smooth rectangle bound*, introduced by Jain and Klauck [16] as a generic lower bound method in this model. Our result therefore implies a strong direct product theorem for all relations for which an (asymptotically) optimal lower bound can be provided using the smooth rectangle bound. In fact we are not aware of any relation for which it is known that the smooth rectangle bound does not provide an optimal lower bound. This lower bound subsumes many of the other known lower bound methods, for example the *rectangle bound* (a.k.a the *corruption bound*) [31], the *smooth discrepancy bound* (a.k.a the γ_2 *bound* [28] which in turn subsumes the *discrepancy bound*), the *subdistribution bound* [17] and the *conditional min-entropy bound* [14].

As a consequence, our result reproves some of the known strong direct product results, for example for the Inner Product [26] function and the Set-Disjointness [25, 14] function. Recently smooth rectangle bound has been used to provide new tight lower bounds for several functions, for example for the Gap-Hamming Distance [8, 33] partial function, the Greater-Than function [35] and the Tribes [12] function. These results, along with our result, imply strong direct product for these functions. Smooth rectangle bound has also been used to provide near optimal lower bounds for several important functions and relations used to show exponential separations between classical and quantum communication complexity for example by Raz [30], Gavinsky [11] and Klartag and Regev [32]. These results combined with our result imply near optimal strong direct product results for these functions and relations.

We show our result using information theoretic arguments. A key tool we use is a sampling protocol due to Braverman [5], in fact a modification of it used by Kerenidis, Laplante, Lerays, Roland and Xiao [23].

1 Introduction

Given a model of computation, suppose solving one instance of a given problem f with probability of success $p < 1$ requires c units of some resource. A natural question that may be asked is: how much resource is needed to solve f^k, k instances of the same problem, simultaneously. A naive way is by running the optimal protocol for f, k times in parallel, which requires $c \cdot k$ units of resource, however the probability of overall success is p^k (exponentially small in k). A *strong direct product conjecture* for f states that this is essentially optimal, that is if only $o(k \cdot c)$ units of resource are provided for any protocol solving f^k, then the probability of overall success is at most $p^{\Omega(k)}$.

Proving or disproving strong direct product conjectures in various models of computation has been a central task in theoretical computer science, notable examples of such results being Yao's *XOR lemma* [38] and Raz's [29] theorem for two-prover games. Readers may refer to [25, 14, 18] for a good discussion of known results in different models of computation. In the present work, we consider the model of two-party two-way public-coin communication complexity [37] and consider the direct product question in this model. In this model, there are two parties who wish to compute a joint function (more generally a relation) of their input, by doing local computation, sharing public coins and exchanging messages. The resource counted is the number of bits communicated between them. The textbook by Kushilevitz and Nisan [26] is an excellent reference for communication complexity. Much effort has been made towards investigating direct product questions in this model and strong direct product theorems have been shown for many different functions, for example Set-Disjointness [25, 14], Inner Product [27], Pointer Chasing [18] etc. To the best of our knowledge, it is not known if the strong direct product conjecture fails to hold for any function or relation in this model. Therefore, whether the strong direct product conjecture holds for all relations in this model, remains one of the major open problems in communication complexity. In the model of constant-round public-coin communication complexity, recently a strong direct product result has been shown to hold for all relations by Jain, Perezlényi and Yao [18]. The work [18] built on a previous result due to Jain [14] showing a strong direct product result for all relations in the model of one-way public-coin communication complexity (where a single message is sent from Alice to Bob, who then determines the answer).

The weaker *direct sum* conjecture, which states that solving k independent instances of a problem with constant success probability requires k times the resource needed to compute one instance with constant success probability, has also been extensively investigated in different models of communication complexity and has met a better success. Direct sum theorems have been shown to hold for all relations in the public-coin one-way model [21], the entanglement-assisted quantum one-way model [20], the public-coin simultaneous message passing model [21], the private-coin simultaneous message passing model [15], the constant-round public-coin two-way model [6] and the model of two-way distributional communication complexity under product distributions [3]. Again please refer to [25, 14, 18] for a good discussion.

Another major focus in communication complexity has been to investigate generic lower bound methods, that apply to all functions (and possibly to all relations). In the model we are concerned with, various generic lower bound methods are known, for example the *partition bound* [16], the *information complexity* [9], the *smooth rectangle bound* [16] (which in turn subsumes the *rectangle bound* a.k.a the *corruption bound*) [36, 1, 31, 24, 4], the *smooth discrepancy bound* a.k.a the γ_2 *bound* [28] (which in turn subsumes the *discrepancy bound*), the *subdistribution bound* [17] and the *conditional min-entropy bound* [14]. Proving strong direct product results in terms of these lower bound methods is a reasonable approach to attacking the general question. Indeed, many lower bounds have been shown to satisfy strong direct product theorems, example the discrepancy bound [27], the subdistribution bound under product distributions [17], the smooth discrepancy bound [34] and the conditional min-entropy bound [14].

Our result

In present work, we show a strong direct product theorem in terms of the smooth rectangle bound, introduced by Jain and Klauck [16], which generalizes the rectangle bound (a.k.a. the corruption bound) [36, 1, 31, 24, 4]. Roughly speaking, the rectangle bound for relation $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ under a distribution μ, with respect to an element $z \in \mathcal{Z}$, and error ε, tries to capture the size (under μ) of a largest rectangle for which z is a right answer for $1-\varepsilon$ fraction of inputs inside the rectangle. It is not hard to argue that the rectangle bound forms a lower bound on the *distributional* communication complexity of f under μ. The smooth rectangle bound for f further captures the maximum, over all relations g that are close to f under μ, of the rectangle bound of g under μ. The distributional error setting can eventually be related to the worst case error setting via the well known Yao's principle [36].

Jain and Klauck showed that the smooth rectangle bound is stronger than every lower bound method we mentioned above except the partition bound and the information complexity. Jain and Klauck showed that the partition bound subsumes the smooth rectangle bound and in a recent work Kerenidis, Laplante, Lerays, Roland and Xiao [23] showed that the information complexity subsumes the smooth rectangle bound (building on the work of Braverman and Weinstein [7] who showed that the information complexity subsumes the discrepancy bound). New lower bounds for specific functions have been discovered using the smooth rectangle bound, for example Chakrabarti and Regev's [8] optimal lower bound for the Gap-Hamming Distance partial function. Klauck [25] used the smooth rectangle bound to show a strong direct product result for the Set-Disjointness function, via exhibiting a lower bound on a related function. On the other hand, as far as we know, no function (or relation) is known for which its smooth rectangle bound is (asymptotically) strictly smaller than its two-way public-coin communication complexity. Hence establishing whether or not the smooth rectangle bound is a tight lower bound for all functions and relations in this model is an important open question. Our result is as follows.

Theorem 1.1. [1][2] *Let $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ be finite sets, $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. Let μ be a distribution on $\mathcal{X} \times \mathcal{Y}$. Let $z \in \mathcal{Z}$ and $\beta \stackrel{\text{def}}{=} \Pr_{(x,y)\leftarrow\mu}[f(x,y) = \{z\}]$. Let $\varepsilon', \delta > 0$. There exists a small enough $\varepsilon > 0$ such that the following holds. For all integers $t \geq 1$,*

$$\mathrm{R}^{\mathrm{pub}}_{1-(1-\varepsilon)^{\lfloor \varepsilon^2 t/32 \rfloor}}\left(f^t\right) \geq \frac{\varepsilon^2}{32} \cdot t \cdot \left(11\varepsilon \cdot \widetilde{\mathrm{srec}}^{z,\mu}_{(1+\varepsilon')\delta/\beta,\delta}\left(f\right) - 2\right).$$

Above $\mathrm{R}^{\mathrm{pub}}(\cdot)$ represents the two-way public-coin communication communication complexity and $\widetilde{\mathrm{srec}}\left(\cdot\right)$ represents the smooth rectangle bound (please refer to Section 2 for precise definitions). Our result implies a strong direct product theorem for all relations for which an (asymptotically) optimal lower bound can be provided using the smooth rectangle bound.

Jain and Klauck [16] provided an alternate definition (Definition .13) of the smooth rectangle bound for (partial) functions in terms of linear programs and show a relationship between the two definitions. We show a tighter relationship between the two definitions in Lemma .14. Combining Lemma .14 with Theorem 1.1 we get the following result.

Theorem 1.2. *Let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ be a (partial) function. For every $\epsilon \in (0,1)$, there exists small enough $\eta \in (0, 1/3)$ such that the following holds. For all integers $t \geq 1$,*

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

[1] For a relation f with $\mathrm{R}^{\mathrm{pub}}_{\varepsilon}(f) = \mathcal{O}(1)$, a strong direct product result can be shown via direct arguments [14].

[2] With a slight abuse of notation, we write $f(x,y) \stackrel{\text{def}}{=} \{z \in \mathcal{Z} |\ (x,y,z) \in f\}$, and $f^{-1}(z) \stackrel{\text{def}}{=} \{(x,y) : (x,y,z) \in f\}$.

As a consequence, our results reprove some of the known strong direct product results, for example for Inner Product [26] and Set-Disjointness [25, 14]. Recently smooth rectangle bound has been used to provide new tight lower bounds for several functions, for example for the Gap-Hamming Distance [8, 33] partial function and the Greater-Than function [35]. These results, along with our result, imply strong direct product for these functions. Smooth rectangle bound has also been used to provide near optimal lower bounds for several important functions and relations used to show exponential separations between classical and quantum communication complexity for example by Raz [30], Gavinsky [11] and Klartag and Regev [32]. These results combined with our result imply near optimal strong direct product results for these functions and relations.

In a recent work, Harsha and Jain [12] have shown that the smooth-rectangle bound provides an optimal lower bound of $\Omega(n)$ for the Tribes function. For this function all other weaker lower bound methods mentioned before like the rectangle bound, the sub-distribution bound, the smooth discrepancy bound, the conditional min-entropy bound etc. fail to provide an optimal lower bound since they are all $O(\sqrt{n})$. Earlier Jayram, Kumar and Sivakumar [22] had shown a lower bound of $\Omega(n)$ using information complexity. The result of [12] along with Theorem 1.2 implies a strong direct product result for the Tribes function. This adds to the growing list of functions for which a strong direct product result can be shown via Theorem 1.2.

In [23], Kerenidis et. al. introduced the *relaxed partition bound* (a weaker version of the partition bound [16]) and showed it to be stronger than the smooth rectangle bound. It is easily seen (by comparing the corresponding linear-programs) that the smooth rectangle bound and the relaxed partition bound are in-fact equivalent for boolean functions (and more generally when the size of output set is a constant). Thus our result also implies a strong direct product theorem in terms of the relaxed partition bound for boolean functions (and more generally when the size of output set is a constant).

Our techniques

The broad argument of the proof of our result is as follows. We show our result in the distributional error setting and translate it to the worst case error setting using the well known Yao's principle [36]. Let f be a relation, μ be a distribution on $\mathcal{X} \times \mathcal{Y}$, and c be the smooth rectangle bound of f under the distribution μ with output $z \in \mathcal{Z}$. Consider a protocol Π which computes f^k with inputs drawn from distribution μ^k and communication $o(c \cdot k)$ bits. Let $\mathcal{C}$ be a subset of the coordinates $\{1, 2, \ldots, k\}$. If the probability that Π computes all the instances in $\mathcal{C}$ correctly is as small as desired, then we are done. Otherwise, we exhibit a new coordinate $j \notin \mathcal{C}$, such that the probability, conditioned on success in $\mathcal{C}$, of the protocol Π answering correctly in the j-th coordinate is bounded away from 1. Since μ could be a non-product distribution we introduce a new random variable R_j, such that conditioned on it and X_jY_j (input in the jth coordinate), Alice and Bob's inputs in the other coordinates become independent. Use of such a variable to handle non product distributions has been used in many previous works, for example [2, 13, 3, 14, 18].

Let the random variables $X_j^1 Y_j^1 R_j^1 M^1$ represent the inputs in the jth coordinate, the new variable R_j and the message transcript of Π, conditioned on the success on $\mathcal{C}$. The first useful property that we observe is that the joint distribution of $X_j^1 Y_j^1 R_j^1 M^1$ can be written as,

$$\Pr\left[X_j^1 Y_j^1 R_j^1 M^1 = xym\right] = \frac{1}{q}\mu(x,y)u_x(r_j,m)u_y(r_j,m),$$

where u_x, u_y are functions and q is a positive real number. The marginal distribution of $X_j^1 Y_j^1$ is no longer μ though. However (using arguments as in [14, 18]), one can show that the distribution of $X_j^1 Y_j^1$ is close, in ℓ_1 distance, to μ and $\mathrm{I}\left(X_j^1; R_j^1 M^1 \middle| Y_j^1\right) + \mathrm{I}\left(Y_j^1; R_j^1 M^1 \middle| X_j^1\right) \leq o(c)$, where $\mathrm{I}(;)$ represents the mutual information (please refer to Section 2 for precise definitions) .

Now, assume for contradiction that the success in the jth coordinate in Π is large, like 0.99, conditioned on success in $\mathcal{C}$. Using the conditions obtained in the previous paragraph, we argue that there exists a zero-communication public-coin protocol Π', between Alice and Bob, with inputs drawn from μ. In Π' Alice and Bob are allowed to abort the protocol or output an element in $\mathcal{Z}$. We show that the probability of non-abort for this protocol is large, like 2^{-c}, and conditioned on non-abort, the probability that Alice and Bob output a correct answer for their inputs is also large, like 0.99. This allows us to exhibit (by fixing the public coins of Π' appropriately), a large rectangle (with weight under μ like 2^{-c}) such that z is a correct answer for a large fraction (like 0.99) of the inputs inside the rectangle. This shows that the rectangle bound of f, under μ with output z, is smaller than c. With careful analysis we are also able to show that the smooth rectangle bound of f under μ, with output z, is smaller than c, reaching a contradiction to the definition of c.

The sampling protocol that we use to obtain the public-coin zero communication protocol, is the same as that in Kerenidis et al. [23], which in turn is a modification of a protocol due to Braverman [5][3] (a variation of which also appears in [7]). However our analysis of the protocol's correctness deviates significantly in parts from the earlier works [23, 5, 7] due to the fact that for us the marginal distribution of X^1Y^1 need not be the same as that of μ, in fact for some inputs (x, y), the probability under the two distributions can be significantly different.

There is another important original contribution of our work, not present in the previous works [23, 5, 7]. We observe a crucial property of the protocol Π' which turns out to be very important in our arguments. The property is that the bad inputs (x, y) for which the distribution of Π''s sample for $R_j^1 M^1$, conditioned on non-abort, deviates a lot from the desired $R_j^1 M^1 | \, (X^1 Y^1 = xy)$, their probability is nicely reduced (as compared to $\Pr\big[X^1 Y^1 = xy\big]$) in the final distribution of Π', conditioned on non-abort. This helps us to argue that the distribution of inputs and outputs in Π', conditioned on non-abort, is close in ℓ_1 distance to $X_j^1 Y_j^1 R_j^1 M^1$, implying good success in Π', conditioned on non-abort.

Organization. In Section 2, we present some necessary background, definitions and preliminaries. In Section 3, we prove our main result Theorem 1.1. We defer some proofs to Appendix.

2 Preliminary

Information theory

We use capital letters e.g. X, Y, Z or letters in bold e.g. $\mathbf{a}, \mathbf{b}, \boldsymbol{\alpha}, \boldsymbol{\beta}$ to represent random variables and use calligraphic letters e.g. $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ to represent sets. For integer $n \geq 1$, let $[n]$ represent the set $\{1, 2, \ldots, n\}$. Let $\mathcal{X}, \mathcal{Y}$ be finite sets and k be a natural number. Let $\mathcal{X}^k$ be the set $\mathcal{X} \times \cdots \times \mathcal{X}$, the cross product of $\mathcal{X}$, k times. Let μ be a (probability) distribution on $\mathcal{X}$. Let $\mu(x)$ represent the probability of $x \in \mathcal{X}$ according to μ. For any subset $S \subseteq \mathcal{X}$, define $\mu(S) \stackrel{\text{def}}{=} \sum_{x \in S} \mu(x)$. Let X be a random variable distributed according to μ, which we denote by $X \sim \mu$. We use the same symbol to represent a random variable and its distribution whenever it is clear from the context. The expectation value of function f on $\mathcal{X}$ is denoted as $\mathbb{E}_{x \leftarrow X}[f(x)] \stackrel{\text{def}}{=} \sum_{x \in \mathcal{X}} \Pr[X = x] \cdot f(x)$. The entropy of X is defined as $\mathrm{H}(X) \stackrel{\text{def}}{=} -\sum_x \mu(x) \cdot \log \mu(x)$ ($\log, \ln$ represent logarithm to the base $2, e$ repectively). For two distributions μ, λ on $\mathcal{X}$, the distribution $\mu \otimes \lambda$ is defined as $(\mu \otimes \lambda)(x_1, x_2) \stackrel{\text{def}}{=} \mu(x_1) \cdot \lambda(x_2)$. Let $\mu^k \stackrel{\text{def}}{=} \mu \otimes \cdots \otimes \mu$, k times. The ℓ_1 distance between μ and λ is defined to be half of the ℓ_1 norm of $\mu - \lambda$; that is, $\|\lambda - \mu\|_1 \stackrel{\text{def}}{=} \frac{1}{2} \sum_x |\lambda(x) - \mu(x)| = \max_{S \subseteq \mathcal{X}} |\lambda(S) - \mu(S)|$. We say that λ is ε-close to μ if $\|\lambda - \mu\|_1 \leq \varepsilon$. The relative entropy

[3] A protocol, achieving similar task, however working only for product distributions on inputs was first shown by Jain, Radhakrishnan and Sen [20].

between distributions X and Y on $\mathcal{X}$ is defined as $\mathrm{S}(X\|Y) \stackrel{\text{def}}{=} \mathbb{E}_{x\leftarrow X}\left[\log \frac{\Pr[X=x]}{\Pr[Y=x]}\right]$. The relative min-entropy between them is defined as $\mathrm{S}_\infty(X\|Y) \stackrel{\text{def}}{=} \max_{x\in\mathcal{X}}\left\{\log \frac{\Pr[X=x]}{\Pr[Y=x]}\right\}$. It is easy to see that $\mathrm{S}(X\|Y) \le \mathrm{S}_\infty(X\|Y)$. Let X, Y, Z be jointly distributed random variables. Let Y_x denote the distribution of Y conditioned on $X = x$. The conditional entropy of Y conditioned on X is defined as $\mathrm{H}(Y|X) \stackrel{\text{def}}{=} \mathbb{E}_{x\leftarrow X}[\mathrm{H}(Y_x)] = \mathrm{H}(XY) - \mathrm{H}(X)$. The mutual information between X and Y is defined as: $\mathrm{I}(X;Y) \stackrel{\text{def}}{=} \mathrm{H}(X) + \mathrm{H}(Y) - \mathrm{H}(XY) = \mathbb{E}_{y\leftarrow Y}[\mathrm{S}(X_y\|X)] = \mathbb{E}_{x\leftarrow X}[\mathrm{S}(Y_x\|Y)]$. The conditional mutual information between X and Y, conditioned on Z, is defined as: $\mathrm{I}(X;Y\,|\,Z) \stackrel{\text{def}}{=} \mathbb{E}_{z\leftarrow Z}[\mathrm{I}(X;Y\,|\,Z=z)] = \mathrm{H}(X|Z) + \mathrm{H}(Y|Z) - \mathrm{H}(XY|Z)$. The following *chain rule* for mutual information is easily seen : $\mathrm{I}(X;YZ) = \mathrm{I}(X;Z) + \mathrm{I}(X;Y\,|\,Z)$.

We will need the following basic facts. A very good text for reference on information theory is [10].

Fact 2.1. Relative entropy is jointly convex in its arguments. That is, for distributions $\mu, \mu^1, \lambda, \lambda^1 \in \mathcal{X}$ and $p \in [0,1]$: $\mathrm{S}\big(p\mu + (1-p)\mu^1\big\|\lambda + (1-p)\lambda^1\big) \le p\cdot\mathrm{S}(\mu\|\lambda) + (1-p)\cdot\mathrm{S}\big(\mu^1\big\|\lambda^1\big)$.

Fact 2.2. Relative entropy satisfies the following chain rule. Let XY and X^1Y^1 be random variables on $\mathcal{X}\times\mathcal{Y}$. It holds that: $\mathrm{S}\big(X^1Y^1\big\|XY\big) = \mathrm{S}\big(X^1\big\|X\big) + \mathbb{E}_{x\leftarrow X^1}\big[\mathrm{S}\big(Y^1_x\big\|Y_x\big)\big]$. In particular, $\mathrm{S}\big(X^1Y^1\big\|X\otimes Y\big) = \mathrm{S}\big(X^1\big\|X\big) + \mathbb{E}_{x\leftarrow X^1}\big[\mathrm{S}\big(Y^1_x\big\|Y\big)\big] \ge \mathrm{S}\big(X^1\big\|X\big) + \mathrm{S}\big(Y^1\big\|Y\big)$. The last inequality follows from Fact 2.1.

Fact 2.3. Let XY and X^1Y^1 be random variables on $\mathcal{X}\times\mathcal{Y}$. It holds that

$$\mathrm{S}\big(X^1Y^1\big\|X\otimes Y\big) \ge \mathrm{S}\big(X^1Y^1\big\|X^1\otimes Y^1\big) = \mathrm{I}\big(X^1;Y^1\big).$$

The following fact follows from Fact 2.2 and Fact 2.3.

Fact 2.4. Given random variables XY and $X'Y'$ on $\mathcal{X}\times\mathcal{Y}$, it holds that

$$\mathop{\mathbb{E}}_{x\leftarrow X'}[\mathrm{S}(Y'_x\|Y)] \ge \mathop{\mathbb{E}}_{x\leftarrow X'}[\mathrm{S}(Y'_x\|Y')] = \mathrm{I}(X';Y').$$

Fact 2.5. For distributions λ and μ: $\quad 0 \le \|\lambda-\mu\|_1 \le \sqrt{\mathrm{S}(\lambda\|\mu)}$.

Fact 2.6. (**Classical substate theorem** [19]) Let X, X' be two distributions on $\mathcal{X}$. For any $\delta \in (0,1)$, it holds that

$$\Pr_{x\leftarrow X'}\left[\frac{\Pr[X'=x]}{\Pr[X=x]} \le 2^{\left(\mathrm{S}(X'\|X)+1\right)/\delta}\right] \ge 1-\delta.$$

We will need the following lemma. Its proof is deferred to Appendix.

Lemma 2.7. *Given random variables A, A' and $\varepsilon > 0$, if $\|A - A'\|_1 \le \varepsilon$, then for any $r \in (0,1)$,*

$$\Pr_{a\leftarrow A}\left[\left|1 - \tfrac{\Pr[A'=a]}{\Pr[A=a]}\right| \le \tfrac{\varepsilon}{r}\right] \ge 1-2r; \text{ and}$$
$$\Pr_{a\leftarrow A'}\left[\left|1 - \tfrac{\Pr[A'=a]}{\Pr[A=a]}\right| \le \tfrac{\varepsilon}{r}\right] \ge 1-2r-\varepsilon.$$

Communication complexity

Let $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ be finite sets, $f \subseteq \mathcal{X}\times\mathcal{Y}\times\mathcal{Z}$ be a relation and $\varepsilon > 0$. In a two-way public-coin communication protocol, Alice is given $x \in \mathcal{X}$, and Bob is given $y \in \mathcal{Y}$. They are supposed to output $z \in \mathcal{Z}$ such that $(x,y,z) \in f$ via exchanging messages and doing local computations. They may share public coins before the inputs are revealed to them. We assume that the last

$\lceil \log |\mathcal{Z}| \rceil$ bits of the transcript is the output of the protocol. Let $\mathrm{R}^{\mathrm{pub}}_{\varepsilon}(f)$ represent the two-way public-coin randomized communication complexity of f with the worst case error ε, that is the communication of the best two-way public-coin protocol for f with error for each input (x, y) being at most ε. Let μ be a distribution on $\mathcal{X} \times \mathcal{Y}$. Let $\mathrm{D}^{\mu}_{\varepsilon}(f)$ represent the two-way distributional communication complexity of f under distribution μ with distributional error ε, that is the communication of the best two-way deterministic protocol for f, with average error over the distribution of the inputs drawn from μ, at most ε. Following is Yao's min-max principle which connects the worst case error and the distributional error settings, see. e.g., [26, Theorem 3.20, page 36].

Fact 2.8. [37] $\mathrm{R}^{\mathrm{pub}}_{\varepsilon}(f) = \max_{\mu} \mathrm{D}^{\mu}_{\varepsilon}(f)$.

The following fact can be easily verified by induction on the number of message exchanges in a private-coin protocol (please refer for example to [5] for an explicit proof). It is also implicit in the *cut and paste* property of private-coins protocol used in Bar-Yossef, Jayram, Kumar and Sivakumar [2].

Lemma 2.9. *For any private-coin two-way communication protocol, with input $XY \sim \mu$ and transcript $M \in \mathcal{M}$, the joint distribution can be written as*

$$\Pr[XYM = xym] = \mu(x, y) u_x(m) u_y(m),$$

where $u_x : \mathcal{M} \to [0, 1]$ and $u_y : \mathcal{M} \to [0, 1]$, for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$.

Smooth rectangle bound

Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation and $\varepsilon, \delta \geq 0$. With a slight abuse of notation, we write $f(x, y) \stackrel{\text{def}}{=} \{z \in \mathcal{Z} |\ (x, y, z) \in f\}$, and $f^{-1}(z) \stackrel{\text{def}}{=} \{(x, y) : (x, y, z) \in f\}$.

Definition 2.10. (**Smooth-rectangle bound** [16]) The (ε, δ)-smooth rectangle bound of f, denoted by $\widetilde{\mathrm{srec}}_{\varepsilon,\delta}(f)$, is defined as follows:

$$\begin{aligned}
&\widetilde{\mathrm{srec}}_{\varepsilon,\delta}(f) \stackrel{\text{def}}{=} \max\{\widetilde{\mathrm{srec}}^{\lambda}_{\varepsilon,\delta}(f) \mid \lambda \text{ a distribution over } \mathcal{X} \times \mathcal{Y}\};\\
&\widetilde{\mathrm{srec}}^{\lambda}_{\varepsilon,\delta}(f) \stackrel{\text{def}}{=} \max\{\widetilde{\mathrm{srec}}^{z,\lambda}_{\varepsilon,\delta}(f) \mid z \in \mathcal{Z}\};\\
&\widetilde{\mathrm{srec}}^{z,\lambda}_{\varepsilon,\delta}(f) \stackrel{\text{def}}{=} \max\{\widetilde{\mathrm{rec}}^{z,\lambda}_{\varepsilon}(g) \mid g \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z};\\
&\quad \Pr_{(x,y)\leftarrow\lambda}[f(x, y) \neq g(x, y)] \leq \delta\};\\
&\widetilde{\mathrm{rec}}^{z,\lambda}_{\varepsilon}(g) \stackrel{\text{def}}{=} \min\{\mathrm{S}_{\infty}(\lambda_R \| \lambda) \mid R \text{ is a rectangle in } \mathcal{X} \times \mathcal{Y},\\
&\lambda(g^{-1}(z) \cap R) \geq (1 - \varepsilon)\lambda(R)\}.
\end{aligned}$$

When $\delta = 0$, the smooth rectangle bound equals the rectangle bound (a.k.a. the corruption bound) [36, 1, 31, 24, 4]. Definition 2.10 is a generalization of the one in [16], where it is only defined for boolean functions. The smooth rectangle bound is a lower bound on the two-way public-coin communication complexity. The proof of the following lemma appears in Appendix.

Lemma 2.11. *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. Let $\lambda \in \mathcal{X} \times \mathcal{Y}$ be a distribution and let $z \in \mathcal{Z}$. Let $\beta \stackrel{\text{def}}{=} \Pr_{(x,y)\leftarrow\lambda}[f(x, y) = \{z\}]$. Let $\varepsilon, \varepsilon', \delta > 0$ be such that $\frac{\delta+\varepsilon}{\beta-2\varepsilon} < (1 + \varepsilon')\frac{\delta}{\beta}$. Then,*

$$\mathrm{R}_{\varepsilon}(f) \geq \mathrm{D}^{\lambda}_{\varepsilon}(f) \geq \widetilde{\mathrm{srec}}^{z,\lambda}_{(1+\varepsilon')\delta/\beta,\delta}(f) - \log \frac{4}{\varepsilon}.$$

3 Proof

The following lemma builds a connection between the zero-communication protocols and the smooth rectangle bound.

Lemma 3.1. *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, $X'Y' \in \mathcal{X} \times \mathcal{Y}$ be a distribution and $z \in \mathcal{Z}$. Let $\beta \stackrel{\text{def}}{=} \Pr_{(x,y)\leftarrow X'Y'}[f(x,y) = \{z\}]$. Let $c \geq 1$. Let $\varepsilon, \varepsilon', \delta > 0$ be such that $(\delta+2\varepsilon)/(\beta-3\varepsilon) < (1+\varepsilon')\delta/\beta$. Let Π be a zero-communication public-coin protocol with input $X'Y'$, public coin R, Alice's output $A \in \mathcal{Z} \cup \{\perp\}$, and Bob's output $B \in \mathcal{Z} \cup \{\perp\}$. Let $X^1Y^1A^1B^1R^1 \stackrel{\text{def}}{=} (X'Y'ABR|\ A = B \neq \perp)$. Let*

1. $\Pr[A = B \neq \perp] \geq 2^{-c}$; 2. $\left\|X^1Y^1 - X'Y'\right\| \leq \varepsilon$.
3. $\Pr\left[(X^1, Y^1, A^1) \in f\right] \geq 1 - \varepsilon$.

Then $\widetilde{\text{rec}}^{z,X'Y'}_{(1+\varepsilon')\delta/\beta,\delta}(f) < \frac{c}{\varepsilon}$.

Proof. Let $g \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, satisfy $\Pr_{(x,y)\leftarrow X'Y'}[f(x,y) \neq g(x,y)] \leq \delta$. It suffices to show that $\widetilde{\text{rec}}^{z,X'Y'}_{(1+\varepsilon')\delta/\beta}(g) \leq \frac{c}{\varepsilon}$.
Since $\Pr[A = B \neq \perp] \geq 2^{-c}$,

$$\begin{aligned} c &\geq \text{S}_\infty\left(X^1Y^1R^1A^1B^1\middle\|X'Y'RAB\right) \\ &\geq \text{S}\left(X^1Y^1R^1A^1B^1\middle\|X'Y'RAB\right) \geq \mathop{\mathbb{E}}_{r\leftarrow R^1, a\leftarrow A^1}\left[\text{S}\left((X^1Y^1)_{r,a}\middle\|X'Y'\right)\right] \quad \text{(from Fact 2.2).} \quad (1) \end{aligned}$$

Since $\left\|X^1Y^1 - X'Y'\right\| \leq \varepsilon$,

$$\Pr_{xyr\leftarrow X^1Y^1R^1}[f(x,y) = \{z\}] \geq \Pr_{xy\leftarrow X'Y'}[f(x,y) = \{z\}] - \varepsilon \geq \beta - \varepsilon. \quad (2)$$

Since $\Pr\left[(X^1, Y^1, A^1) \in f\right] \geq 1 - \varepsilon$, hence $\Pr\left[A^1 = B^1 = z\right] \geq \beta - 2\varepsilon$. Since

$$\Pr_{(x,y)\leftarrow X'Y'}[f(x,y) \neq g(x,y)] \leq \delta,$$

by item 2 of this lemma, we have

$$\Pr_{xyra\leftarrow X^1Y^1R^1A^1}[(x,y,a) \in g] \geq \Pr_{xyra\leftarrow X^1Y^1R^1A^1}[(x,y,a) \in f] - \delta - \varepsilon \geq 1 - 2\varepsilon - \delta. \quad (3)$$

By standard application of Markov's inequality on equations (1), (2), (3), we get an r_0, such that

$$\begin{aligned} \text{S}\left((X^1Y^1)_{r_0,z}\middle\|X'Y'\right) &\leq \frac{c}{\varepsilon}, \\ \Pr_{xy\leftarrow(X^1Y^1)_{r_0,z}}[z \notin g(x,y)] &\leq (\delta + 2\varepsilon)/(\beta - 3\varepsilon) \leq (1+\varepsilon')\delta/\beta. \end{aligned}$$

Here, $(X^1Y^1)_{r_0,z} = (X^1Y^1|(R^1 = r_0, A^1 = z)$. Note that the distribution of $(X^1Y^1)_{r_0,z}$ is the distribution of $X'Y'$ restricted to some rectangle and then rescaled to make a distribution. Hence

$$\text{S}\left((X^1Y^1)_{r_0,z}\middle\|X'Y'\right) = \text{S}_\infty\left((X^1Y^1)_{r_0,z}\middle\|X'Y'\right).$$

Thus $\widetilde{\text{rec}}^{z,X'Y'}_{(1+\varepsilon')\delta/\beta}(g) < \frac{c}{\varepsilon}$. □

The following is our main lemma. A key tool that we use here is a sampling protocol that appears in [23] (protocol Π' as shown in Figure 1), which is a variant of a sampling protocol that appears in [7], which in turn is a variant of a sampling protocol that appears in [5]. Naturally

similar arguments and calculations, as in this lemma, are made in previous works [5, 7, 23], however with a key difference. In their setting $\sum_m u_x(m)u_y(m) = 1$ for all (x, y). However in our setting this number could be much smaller than one for different (x, y). Hence our arguments and calculations deviate from previous works at several places significantly. Another important original contribution of our work is Claim 3.7 which is used in the proof of the main lemma. We highlight its importance later just before its proof.

Lemma 3.2. ***(Main Lemma)*** *Let $c \geq 1$. Let p be a distribution over $\mathcal{X} \times \mathcal{Y}$ and $z \in \mathcal{Z}$. Let $\beta \stackrel{\text{def}}{=} \Pr_{(x,y)\leftarrow p}[f(x, y) = \{z\}]$. Let $0 < \varepsilon < 1/3$ and $\delta, \varepsilon' > 0$ be such that $\frac{\delta+22\varepsilon}{\beta-33\varepsilon} < (1+\varepsilon')\frac{\delta}{\beta}$. Let XYM be random variables jointly distributed over the set $\mathcal{X}\times\mathcal{Y}\times\mathcal{M}$ such that the last $\lceil \log |\mathcal{Z}| \rceil$ bits of M represents an element in $\mathcal{Z}$. Let $u_x : \mathcal{M} \to [0, 1]$, $u_y : \mathcal{M} \to [0, 1]$ be functions for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$. If it holds that,*

1. *For all $(x, y, m) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{M}$, $\Pr[XYM = xym] = \frac{1}{q}p(x, y)u_x(m)u_y(m)$, where $q \stackrel{\text{def}}{=} \sum_{xym} p(x, y)u_x(m)u_y(m)$;*
2. *$\mathrm{S}(XY\|p) \leq \varepsilon^2/4$;*
3. *$\mathrm{I}(X; M|Y) + \mathrm{I}(Y; M|X) \leq c$;*
4. *$\mathrm{err}_f(XYM) \leq \varepsilon$, where $\mathrm{err}_f(XYM) \stackrel{\text{def}}{=} \Pr_{xym\leftarrow XYM}[(x, y, \tilde{m}) \notin f]$, and $\tilde{m}$ represents the last $\lceil \log |\mathcal{Z}| \rceil$ bits of m;*

then $\widetilde{\mathrm{srec}}^{z,p}_{(1+\varepsilon')\delta/\beta,\delta}(f) < \frac{2c}{11\varepsilon^3}$. $\square$

Proof. Note by direct calculations,

$$\Pr[XY = xy] = \frac{1}{q}p(x, y)\alpha_{xy}, \quad \text{where } \alpha_{xy} \stackrel{\text{def}}{=} \sum_m u_x(m)u_y(m);; \tag{4}$$

$$\Pr[X = x] = \frac{1}{q}p(x)\alpha_x, \quad \text{where } \alpha_x \stackrel{\text{def}}{=} \sum_y p(y|x)\alpha_{xy}; \tag{5}$$

$$\Pr[Y = y] = \frac{1}{q}p(y)\alpha_y, \quad \text{where } \alpha_y \stackrel{\text{def}}{=} \sum_x p(x|y)\alpha_{xy}; \tag{6}$$

$$\Pr[X_y = x] = \frac{p(x|y)\alpha_{xy}}{\alpha_x}, \quad \Pr[Y_x = y] = \frac{p(y|x)\alpha_{xy}}{\alpha_y}; \tag{7}$$

$$\Pr[M_{xy} = m] = u_x(m)u_y(m)/\alpha_{xy}; \tag{8}$$

$$\Pr[M_x = m] = \frac{u_x(m)v_x(m)}{\alpha_x}, \quad \text{where } v_x(m) \stackrel{\text{def}}{=} \textstyle\sum_y p(y|x)u_y(m); \tag{9}$$

$$\Pr[M_y = m] = \frac{u_y(m)v_y(m)}{\alpha_y}, \quad \text{where } v_y(m) \stackrel{\text{def}}{=} \textstyle\sum_x p(x|y)u_x(m). \tag{10}$$

Define

$$G_1 \stackrel{\text{def}}{=} \{(x, y) : \left|1 - \frac{\alpha_{xy}}{q}\right| \leq \frac{1}{2} \text{ and } \left|1 - \frac{\alpha_x}{q}\right| \leq \frac{1}{2} \text{ and} \left|1 - \frac{\alpha_{xy}}{q}\right| \leq \frac{1}{2}\}; \tag{11}$$

$$G_2 \stackrel{\text{def}}{=} \{(x, y) : \mathrm{S}(M_{xy}\|M_x) + \mathrm{S}(M_{xy}\|M_y) \leq c/\varepsilon\}; \tag{12}$$

$$G \stackrel{\text{def}}{=} \{(x, y) : \Pr_{m\leftarrow M_{xy}}\left[\frac{u_y(m)}{v_x(m)} \leq 2^{\Delta} \text{ and } \frac{u_x(m)}{v_y(m)} \leq 2^{\Delta}\right] \geq 1 - 2\varepsilon\}. \tag{13}$$

We begin by showing that $G_1 \cap G_2$ is a large set and also $G_1 \cap G_2 \subseteq G$.

Claim 3.3. 1. $\Pr_{(x,y)\leftarrow p}[(x, y) \in G_1] > 1 - 6\varepsilon$,

2. $\Pr_{(x,y)\leftarrow p}[(x, y) \in G_2] \geq 1 - 3\varepsilon/2$,

Alice's input is x. Bob's input is y. Common input is $c, \varepsilon, q, \mathcal{M}$.

1. Alice and Bob both set $\Delta \stackrel{\text{def}}{=} \frac{c/\varepsilon+1}{\varepsilon} + 2, T \stackrel{\text{def}}{=} \frac{2}{q}|\mathcal{M}|2^{\Delta} \ln \frac{1}{\varepsilon}$ and $k \stackrel{\text{def}}{=} \log(\frac{3}{\varepsilon}(\ln \frac{1}{\varepsilon}))$.
2. For $i = 1, \cdots, T$:
   (a) Alice and Bob, using public coins, jointly sample $\mathbf{m_i} \leftarrow \mathcal{M}, \boldsymbol{\alpha_i}, \boldsymbol{\beta_i} \leftarrow [0, 2^{\Delta}]$, uniformly.
   (b) Alice accepts $\mathbf{m_i}$ if $\boldsymbol{\alpha_i} \leq u_x(\mathbf{m_i})$, and $\boldsymbol{\beta_i} \leq 2^{\Delta} v_x(\mathbf{m_i})$.
   (c) Bob accepts $\mathbf{m_i}$ if $\boldsymbol{\alpha_i} \leq 2^{\Delta} v_y(\mathbf{m_i})$, and $\boldsymbol{\beta_i} \leq u_y(\mathbf{m_i})$.
3. Let $\mathcal{A} \stackrel{\text{def}}{=} \{i \in [T] : \text{Alice accepts } \mathbf{m_i}\}$ and $\mathcal{B} \stackrel{\text{def}}{=} \{i \in [T] : \text{Bob accepts } \mathbf{m_i}\}$.
4. Alice and Bob, using public coins, choose a uniformly random function $\mathbf{h} : \mathcal{M} \to \{0,1\}^k$ and a uniformly random string $\mathbf{r} \in \{0,1\}^k$.
   (a) Alice outputs $\perp$ if either $\mathcal{A}$ is empty or $\mathbf{h}(\mathbf{m_i}) \neq \mathbf{r}$ (where i is the smallest element in non-empty $\mathcal{A}$). Otherwise, she outputs the element in $\mathcal{Z}$, represented by the last $\lceil \log |\mathcal{Z}| \rceil$ bits of $\mathbf{m_i}$.
   (b) Bob finds the smallest $j \in \mathcal{B}$ such that $\mathbf{h}(\mathbf{m_j}) = \mathbf{r}$. If no such j exists, he outputs $\perp$. Otherwise, he outputs the element in $\mathcal{Z}$, represented by the last $\lceil \log |\mathcal{Z}| \rceil$ bits of $\mathbf{m_j}$.

Figure 1: Protocol Π'

3. $\Pr_{(x,y)\leftarrow p}[(x,y) \in G_1 \cap G_2] \geq 1 - 15\varepsilon/2$,
4. $G_1 \cap G_2 \subseteq G$.

Proof. Note item 1. and item 2. imply item 3. Now we show 1. Note that (using item 2. of Lemma 3.2 and Fact 2.5) $\|XY - p\|_1 \leq \varepsilon/2$. From Lemma 2.7 and (4), we have

$$\Pr_{(x,y)\leftarrow p}\left[\left|1 - \frac{\alpha_{xy}}{q}\right| \leq 1/2\right] \geq 1 - 2\varepsilon.$$

By the monotonicity of ℓ_1-norm, we have $\|X - p_{\mathcal{X}}\|_1 \leq \frac{\varepsilon}{2}$ and $\|X - p_{\mathcal{Y}}\|_1 \leq \frac{\varepsilon}{2}$. Similarly, from (5) and (6) we have

$$\Pr_{(x,y)\leftarrow p}\left[\left|1 - \frac{\alpha_x}{q}\right| \leq 1/2\right] \geq 1 - 2\varepsilon, \quad \text{and} \quad \Pr_{(x,y)\leftarrow p}\left[\left|1 - \frac{\alpha_y}{q}\right| \leq 1/2\right] \geq 1 - 2\varepsilon.$$

By the union bound, item 1. follows.

Next we show 2. From item 3. of Lemma 3.2,

$$\mathop{\mathbb{E}}_{(x,y)\leftarrow XY}[\mathrm{S}(M_{xy}\|M_x) + \mathrm{S}(M_{xy}\|M_y)] = \mathrm{I}(X; M|Y) + \mathrm{I}(Y; M|X) \leq c.$$

Markov's inequality implies $\Pr_{(x,y)\leftarrow XY}[(x,y) \in G_2] \geq 1 - \varepsilon$. Then item 2. follows from the fact that XY and p are $\varepsilon/2$-close.

Finally we show 4. For any $(x,y) \in G_1 \cap G_2$,

$$\begin{aligned}
&\mathrm{S}(M_{xy} \| M_x) \le c/\varepsilon \\
&\Rightarrow \Pr_{m \leftarrow M_{xy}}\left[\frac{\Pr[M_{xy}=m]}{\Pr[M_x=m]} \le 2^{\frac{c/\varepsilon+1}{\varepsilon}}\right] \ge 1-\varepsilon \quad \text{(from Fact 2.6)} \\
&\Rightarrow \Pr_{m \leftarrow M_{xy}}\left[\frac{u_y(m)\alpha_x}{v_x(m)\alpha_{xy}} \le 2^{\frac{c/\varepsilon+1}{\varepsilon}}\right] \ge 1-\varepsilon \quad \text{(from (8) and (9))} \\
&\Rightarrow \Pr_{m \leftarrow M_{xy}}\left[\frac{u_y(m)}{v_x(m)} \le 2^{\Delta}\right] \ge 1-\varepsilon. \\
&\qquad\qquad ((x,y) \in G_1 \text{ and the choice of } \Delta)
\end{aligned}$$

Similarly, $\Pr_{m \leftarrow M_{xy}}\left[\frac{u_x(m)}{v_y(m)} \le 2^{\Delta}\right] \ge 1-\varepsilon$. By the union bound,

$$\Pr_{m \leftarrow M_{xy}}\left[\frac{u_y(m)}{v_x(m)} \le 2^{\Delta} \text{ and } \frac{u_x(m)}{v_y(m)} \le 2^{\Delta}\right] \ge 1-2\varepsilon,$$

which implies $(x,y) \in G$. Hence $G_1 \cap G_2 \subseteq G$. □

Following few claims establish the desired properties of protocol Π' (Figure 1).

Definition 3.4. Define the following events.

- E occurs if the smallest $i \in \mathcal{A}$ satisfies $\mathbf{h}(\mathbf{m_i}) = \mathbf{r}$ and $i \in \mathcal{B}$. Note that E implies $\mathcal{A} \neq \emptyset$.
- B_c (subevent of E) occurs if E occurs and there exist $j \in \mathcal{B}$ such that $\mathbf{h}(\mathbf{m_j}) = \mathbf{r}$ and $\mathbf{m_i} \neq \mathbf{m_j}$, where i is the smallest element in $\mathcal{A}$.
- $H \stackrel{\text{def}}{=} E - B_c$.

Below we use conditioning on (x,y) as shorthand for "Alice's input is x and Bob's input is y".

Claim 3.5. For any $(x,y) \in G_1 \cap G_2$, we have

1. for all $i \in [T]$,

$$\frac{1}{2} \cdot \frac{q}{|\mathcal{M}|2^{\Delta}} \le \Pr_{\mathbf{r}_{\Pi'}}[\text{Alice accepts } \mathbf{m_i} |\ (x,y)] \le \frac{3}{2} \cdot \frac{q}{|\mathcal{M}|2^{\Delta}},$$

and

$$\frac{1}{2} \cdot \frac{q}{|\mathcal{M}|2^{\Delta}} \le \Pr_{\mathbf{r}_{\Pi'}}[\text{Bob accepts } \mathbf{m_i} |\ (x,y)] \le \frac{3}{2} \cdot \frac{q}{|\mathcal{M}|2^{\Delta}},$$

where $\mathbf{r}_{\Pi'}$ is the internal randomness of protocol Π';

2. $\Pr_{\mathbf{r}_{\Pi'}}[B_c |\ (x,y), E] \le \varepsilon$;
3. $\Pr_{\mathbf{r}_{\Pi'}}[H |\ (x,y)] \ge (1-4\varepsilon) \cdot 2^{-k-\Delta-2}$.

Proof. 1. We do the argument for Alice. Similar argument follows for Bob. Note that $u_x(m), v_x(m) \in [0,1]$. Then for all $(x,y) \in \mathcal{X} \times \mathcal{Y}$,

$$\Pr_{\mathbf{r}_{\Pi'}}[\text{Alice accepts } \mathbf{m_i} |\ (x,y)] = \frac{1}{|\mathcal{M}|} \sum_m \frac{u_x(m)v_x(m)}{2^{\Delta}} = \frac{\alpha_x}{|\mathcal{M}|2^{\Delta}}.$$

Item 1 follows by the fact that $(x,y) \in G_1$.

2. Define E_i (subevent of E) when i is the smallest element of $\mathcal{A}$. For all $(x,y) \in G_1 \cap G_2$, we have :

$$
\begin{aligned}
&\Pr_{\mathbf{r}_{\Pi'}}[B_c|\ (x,y), E_i] \\
&= \Pr_{\mathbf{r}_{\Pi'}}[\exists j\ :\ j \in \mathcal{B} \text{ and } \mathbf{h}(\mathbf{m_j}) = \mathbf{r} \text{ and } \mathbf{m_j} \neq \mathbf{m_i}|\ (x,y), E_i] \\
&\leq \sum_{j \in [T], j \neq i} \Pr_{\mathbf{r}_{\Pi'}}[j \in \mathcal{B} \text{ and } \mathbf{h}(\mathbf{m_j}) = \mathbf{r} \text{ and } \mathbf{m_j} \neq \mathbf{m_i}|\ (x,y), E_i] \quad \text{(from the union bound)} \\
&\leq \sum_{j \in [T], j \neq i} \Pr_{\mathbf{r}_{\Pi'}}[j \in \mathcal{B}|\ (x,y), E_i] \cdot \Pr_{\mathbf{r}_{\Pi'}}[\mathbf{h}(\mathbf{m_j}) = \mathbf{r}|\ (x,y), E_i, j \in \mathcal{B}, \mathbf{m_j} \neq \mathbf{m_i}] \\
&\leq T \cdot \frac{3q}{|\mathcal{M}|2^{\Delta+1}} \cdot \frac{1}{2^k} \quad \text{(two-wise independence of } \mathbf{h} \text{ and item 1. of this Claim)} \\
&\leq \varepsilon. \quad \text{(from choice of parameters)}
\end{aligned}
$$

Since above holds for every i, it implies $\Pr_{\mathbf{r}_{\Pi'}}[B_c|\ (x,y), E] \leq \varepsilon$.

3. Consider,

$$
\begin{aligned}
&\Pr_{\mathbf{r}_{\Pi'}}[E|\ (x,y)] = \Pr_{\mathbf{r}_{\Pi'}}[\mathcal{A} \neq \varnothing|\ (x,y)] \cdot \Pr_{\mathbf{r}_{\Pi'}}[E|\ \mathcal{A} \neq \varnothing, (x,y)] \\
&\geq \left(1 - \left(1 - \frac{1}{2} \cdot \frac{q}{|\mathcal{M}|2^{\Delta}}\right)^T\right) \cdot \Pr_{\mathbf{r}_{\Pi'}}[E|\ \mathcal{A} \neq \varnothing, (x,y)] \quad \text{(using item 1. of this Claim)} \\
&\geq (1-\varepsilon) \cdot \Pr_{\mathbf{r}_{\Pi'}}[E|\ \mathcal{A} \neq \varnothing, (x,y)] \ \text{(from choice of parameters)} \\
&= (1-\varepsilon) \cdot \Pr_{\mathbf{r}_{\Pi'}}[\mathbf{h}(\mathbf{m_i}) = \mathbf{r}|\ \mathcal{A} \neq \varnothing, (x,y)] \cdot \Pr_{\mathbf{r}_{\Pi'}}[i \in \mathcal{B}|\ i \in \mathcal{A},\ \mathbf{h}(\mathbf{m_i}) = \mathbf{r}, (x,y)] \\
&\text{(from here on we condition on } i \text{ being the first element of } \mathcal{A}) \\
&= (1-\varepsilon) \cdot 2^{-k} \cdot \Pr_{\mathbf{r}_{\Pi'}}[i \in \mathcal{B}|\ i \in \mathcal{A}, (x,y)] \\
&= (1-\varepsilon) \cdot 2^{-k} \cdot \frac{\Pr_{\mathbf{r}_{\Pi'}}[i \in \mathcal{B} \text{ and } i \in \mathcal{A}|\ (x,y)]}{\Pr_{\mathbf{r}_{\Pi'}}[i \in \mathcal{A}|\ (x,y)]} \\
&\geq \frac{2}{3q}(1-\varepsilon) \cdot 2^{-k} \cdot |\mathcal{M}|2^{\Delta} \cdot \Pr_{\mathbf{r}_{\Pi'}}[i \in \mathcal{B} \text{ and } i \in \mathcal{A}|\ (x,y)] \quad \text{(using item 1. of this claim)} \\
&= \frac{2}{3q}(1-\varepsilon) \cdot 2^{-k} \cdot |\mathcal{M}|2^{\Delta} \cdot \sum_{m \in \mathcal{M}} \frac{1}{|\mathcal{M}|2^{2\Delta}} \min\left\{u_x(m), 2^{\Delta}v_y(m)\right\} \cdot \min\left\{u_y(m), 2^{\Delta}v_x(m)\right\} \\
&\qquad\qquad \text{(from construction of protocol } \Pi') \\
&\geq \frac{2}{3q}(1-\varepsilon) \cdot 2^{-k} \cdot |\mathcal{M}|2^{\Delta} \cdot \sum_{m \in G_{xy}} \frac{u_x(m)u_y(m)}{|\mathcal{M}|2^{2\Delta}} \\
&\qquad\qquad (G_{xy} \stackrel{\text{def}}{=} \{m : u_x(m) \leq 2^{\Delta}v_y(m) \text{ and } u_y(m) \leq 2^{\Delta}v_x(m)\}) \\
&= \frac{2}{3q}(1-\varepsilon) \cdot 2^{-k} \cdot |\mathcal{M}|2^{\Delta} \cdot \frac{\alpha_{xy}}{|\mathcal{M}|2^{2\Delta}} \sum_{m \in G_{xy}} \frac{u_x(m)u_y(m)}{\alpha_{xy}} \\
&\geq \frac{1}{3}(1-\varepsilon) \cdot 2^{-k-\Delta} \cdot \Pr_{m \leftarrow M_{xy}}[m \in G_{xy}] \quad \text{(since } (x,y) \in G_1 \text{ and (8))} \\
&\geq \frac{1}{3}(1-\varepsilon) \cdot 2^{-k-\Delta} \cdot (1-2\varepsilon) \quad \text{(since } (x,y) \in G, \text{ using item 4. of Claim 3.3)} \\
&\geq (1-3\varepsilon) \cdot 2^{-k-\Delta-2}.
\end{aligned}
$$

Finally, using item 2. of this Claim.

$$\Pr_{\mathbf{r}_{\Pi'}}[H|\ (x,y)] = \Pr_{\mathbf{r}_{\Pi'}}[E|\ (x,y)]\,(1 - \Pr_{\mathbf{r}_{\Pi'}}[B_c|\ (x,y), E]) \geq (1-4\varepsilon)\cdot 2^{-k-\Delta-2}.$$

□

Claim 3.6. $\Pr_{p,\mathbf{r}_{\Pi'}}[H] \geq (1 - \frac{23}{2}\varepsilon)\cdot 2^{-k-\Delta-2}$.

Proof.

$$\Pr_{p,\mathbf{r}_{\Pi'}}[H] \geq \sum_{(x,y)\in G_1\cap G_2} p(x,y) \Pr_{\mathbf{r}_{\Pi'}}[H|\ (x,y)] \geq (1-4\varepsilon)\cdot 2^{-k-\Delta-2} \sum_{(x,y)\in G_1\cap G_2} p(x,y)$$
$$\geq (1 - \frac{23}{2}\varepsilon)\cdot 2^{-k-\Delta-2}.$$

The second inequality is by Claim 3.5, item 3, and the last inequality is by Claim 3.3 item 3. □

The following claim is an important original contribution of this work (not present in the previous works [23, 5, 7].) The claim helps us establish a crucial property of Π'. The property is that the bad inputs (x,y) for which the distribution of Π''s sample for M, conditioned on non-abort, deviates a lot from the desired, their probability is nicely reduced in the final distribution of Π', conditioned on non-abort. This helps us to argue that the joint distribution of inputs and the transcript in Π', conditioned on non-abort, is still close in ℓ_1 distance to XYM.

Claim 3.7. Let AB and $A'B'$ be random variables over $\mathcal{A}_1 \times \mathcal{B}_1$ and $h : \mathcal{A}_1 \to [0,+\infty)$ be a function. Suppose for any $a \in \mathcal{A}_1$, there exist functions $f_a, g_a : \mathcal{B}_1 \to [0,+\infty)$, such that

1. $\sum_{a,b} h(a)f_a(b) = 1$, and $\Pr[AB = ab] = h(a)f_a(b)$;
2. $f_a(b) \geq g_a(b)$, for all $(a,b) \in \mathcal{A}_1 \times \mathcal{B}_1$;
3. $\Pr[A'B' = ab] = h(a)g_a(b)/C$, where $C = \sum_{a,b} h(a)g_a(b)$;
4. $\Pr_{a\leftarrow A}[\Pr_{b\leftarrow B_a}[f_a(b) = g_a(b)] \geq 1-\delta_1] \geq 1-\delta_2$, for $\delta_1 \in [0,1), \delta_2 \in [0,1)$.

Then $\|AB - A'B'\|_1 \leq \delta_1 + \delta_2$.

Proof. Set $G \stackrel{\text{def}}{=} \{(a,b) : f_a(b) = g_a(b)\}$. By condition 4, $\Pr_{(a,b)\leftarrow AB}[(a,b)\in G] \geq 1-\delta_1-\delta_2$. Then

$$C = \sum_{a,b} h(a)g_a(b) \geq \sum_{a,b:(a,b)\in G} h(a)f_a(b) = \Pr_{(a,b)\leftarrow AB}[(a,b)\in G] \geq 1-\delta_1-\delta_2. \tag{14}$$

We have

$$\begin{aligned}
&\|AB - A'B'\|_1 = \frac{1}{2}\sum_{a,b} |h(a)f_a(b) - \frac{1}{C}h(a)g_a(b)| \\
\leq\ & \frac{1}{2}\sum_{a,b}\left(|h(a)f_a(b) - h(a)g_a(b)| + |h(a)g_a(b) - \frac{1}{C}h(a)g_a(b)|\right) \\
\leq\ & \frac{1}{2}\left(\sum_{a,b}(h(a)f_a(b) - h(a)g_a(b)) + \frac{1-C}{C}\sum_{a,b} h(a)g_a(b)\right) \quad \text{(using item 2. of this claim)} \\
\leq\ & \frac{1}{2}\left(\sum_{a,b:(a,b)\notin G} h(a)f_a(b) + 1 - C\right) \\
=\ & \frac{1}{2}\left(\Pr_{(a,b)\leftarrow AB}[(a,b)\notin G] + 1 - C\right) \leq \delta_1+\delta_2 \quad \text{(from (14))}
\end{aligned}$$

□

Claim 3.8. Let the input of protocol Π' be drawn according to p. Let $X^1Y^1M^1$ represent the input and the transcript (the part of the public coins drawn from $\mathcal{M}$) conditioned on H. Then we have $\left\|XYM - X^1Y^1M^1\right\|_1 \leq 10\varepsilon$. Note that this implies that $\left\|X^1Y^1A^1B^1 - XY\tilde{M}\tilde{M}\right\|_1 \leq 10\varepsilon$, where $\tilde{M}$ represents the last $\lceil \log |\mathcal{Z}| \rceil$ bits of M and A^1, B^1 represent outputs of Alice and Bob respectively, conditioned on H.

Proof. For any (x, y), define

$$w_{xy}(m) \stackrel{\text{def}}{=} \min\left\{u_x(m), 2^{\Delta}v_y(m)\right\} \cdot \min\left\{u_y(m), 2^{\Delta}v_x(m)\right\}.$$

From step 2 (a),(b),(c), of protocol Π', $\Pr\left[M^1X^1Y^1 = mxy\right] = \frac{1}{C}p(x, y)w_{xy}(m)$, where $C = \sum_{xym} p(x, y)w_{xy}(m)$. Now,

$$\begin{aligned}\Pr_{(x,y)\leftarrow XY}\left[\Pr_{m\leftarrow M_{xy}}\left[w_{xy}(m) = u_x(m)u_y(m)\right] \geq 1 - 2\varepsilon\right] \\ = \Pr_{(x,y)\leftarrow XY}\left[(x, y) \in G\right] \geq 1 - 8\varepsilon.\end{aligned}$$

The last inequality above follows using items 3. and 4. of Claim 3.3 and the fact that XY and p are $\varepsilon/2$-close.

Finally using Claim 3.7 (by substituting $\delta_1 \leftarrow 2\varepsilon, \delta_2 \leftarrow 8\varepsilon, A \leftarrow XY, B \leftarrow M, A' \leftarrow X^1Y^1, B' \leftarrow M^1, h \leftarrow \frac{p}{q}, f_{(x,y)}(m) \leftarrow u_x(m)u_y(m)$ and $g_{(x,y)}(m) \leftarrow w_{xy}(m)$), we get that $\left\|X^1Y^1M^1 - XYM\right\|_1 \leq 10\varepsilon$. □

We are now ready to finish the proof of Lemma 3.2. Consider the protocol Π'. We claim that it satisfies Lemma 3.1 by taking the correspondence between quantities in Lemma 3.1 and Lemma 3.2 as follows : $c \leftarrow (c/\varepsilon^2 + 3/\varepsilon), \varepsilon \leftarrow 11\varepsilon, \beta \leftarrow \beta, \delta \leftarrow \delta, z \leftarrow z, X'Y' \leftarrow p$.

Item 1. of Lemma 3.1 is implied by Claim 3.6 since $(1 - \frac{23}{2}\varepsilon) \cdot 2^{-k-\Delta-2} \geq 2^{-(c/\varepsilon^2+3/\varepsilon)}$, from choice of parameters.

Item 2. of Lemma 3.1 is implied since $\left\|X^1Y^1 - p\right\|_1 \leq \left\|X^1Y^1 - XY\right\|_1 + \|XY - p\|_1 \leq \frac{21}{2}\varepsilon$, using item 2. of Lemma 3.2, Fact 2.5 and Claim 3.8.

Item 3. of Lemma 3.1 is implied since $\mathrm{err}_f\left(X^1Y^1M^1\right) \leq \mathrm{err}_f\left(XYM\right) + \left\|X^1Y^1M^1 - XYM\right\|_1 \leq 11\varepsilon$, using item 4. in Lemma 3.2 and Claim 3.8.

This implies

$$\widetilde{\mathrm{srec}}^{z,p}_{(1+\varepsilon')\delta/\beta,\delta}(f) < \frac{c/\varepsilon^2 + 3/\varepsilon}{11\varepsilon} \leq \frac{2c}{11\varepsilon^3}. \quad \square$$

□

We can now prove our main result.

Theorem 3.9. *Let $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ be finite sets, $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. Let μ be a distribution on $\mathcal{X} \times \mathcal{Y}$. Let $z \in \mathcal{Z}$ and $\beta \stackrel{\text{def}}{=} \Pr_{(x,y)\leftarrow\mu}[f(x, y) = \{z\}]$. Let $0 < \varepsilon < 1/3$ and $\varepsilon', \delta > 0$ be such that $\frac{\delta+22\varepsilon}{\beta-33\varepsilon} < (1 + \varepsilon')\frac{\delta}{\beta}$. For all integers $t \geq 1$,*

$$\mathrm{R}^{\mathrm{pub}}_{1-(1-\varepsilon)^{\lfloor\varepsilon^2 t/32\rfloor}}(f^t) \geq \frac{\varepsilon^2}{32} \cdot t \cdot \left(11\varepsilon \cdot \widetilde{\mathrm{srec}}^{z,\mu}_{(1+\varepsilon')\delta/\beta,\delta}(f) - 2\right).$$

Proof. Set $\delta_1 \stackrel{\text{def}}{=} \varepsilon^2/32$. define $c \stackrel{\text{def}}{=} 11\varepsilon \cdot \widetilde{\mathrm{srec}}^{z,\mu}_{(1+\varepsilon')\delta/\beta,\delta}(f) - 2$ and $XY \sim \mu^t$. By Fact 2.8, it suffices to show $\mathrm{D}^{\mu^t}_{1-(1-\varepsilon)^{\lfloor\varepsilon^2 t/32\rfloor}}(f^t) \geq \delta_1 tc$. Let Π be a deterministic two-way communication protocol, that computes f^t, with total communication $\delta_1 ct$ bits. The following claim implies that the success of Π is at most $(1 - \varepsilon)^{\lfloor\delta_1 t\rfloor}$, and this shows the desired. □

Claim 3.10. For each $i \in [t]$, define a binary random variable $T_i \in \{0,1\}$, which represents the success of Π on the i-th instance. That is, $T_i = 1$ if the protocol computes the i-th instance of f correctly, and $T_i = 0$ otherwise. Let $t' \stackrel{\text{def}}{=} \lfloor \delta_1 t \rfloor$. There exists t' coordinates $\{i_1, \cdots, i_{t'}\}$ such that for each $1 \le r \le t'-1$,

1. either $\Pr\left[T^{(r)} = 1\right] \le (1-\varepsilon)^{t'}$ or
2. $\Pr\left[T_{i_{r+1}} = 1 |\ T^{(r)} = 1\right] \le 1-\varepsilon$, where $T^{(r)} \stackrel{\text{def}}{=} \prod_{j=1}^{r} T_{i_j}$.

Proof. Suppose we have already identified r coordinates, $i_1, \cdots, i_r$ satisfying that $\Pr[T_{i_1}] \le 1-\varepsilon$ and $\Pr\left[T_{i_{j+1}} = 1 |\ T^{(j)} = 1\right] \le 1-\varepsilon$ for $1 \le j \le r-1$. If $\Pr\left[T^{(r)} = 1\right] \le (1-\varepsilon)^{t'}$, then we are done. So from now on we assume $\Pr\left[T^{(r)} = 1\right] > (1-\varepsilon)^{t'} \ge 2^{-\delta_1 t}$. Here we assume $r \ge 1$. Similar arguments also work when $r = 0$, that is for identifying the first coordinate, which we skip for the sake of avoiding repetition.

Let D be a random variable uniformly distributed in $\{0,1\}^t$ and independent of XY. Let $U_i = X_i$ if $D_i = 0$, and $U_i = Y_i$ if $D_i = 1$. For any random variable L, define $L^1 \stackrel{\text{def}}{=} (L | T^{(r)} = 1)$. If $L = L_1 \cdots L_t$, define $L_{-i} \stackrel{\text{def}}{=} L_1 \cdots L_{i-1} L_{i+1} \cdots L_t$. Let $\mathcal{C} \stackrel{\text{def}}{=} \{i_1, \cdots, i_r\}$. Define $R_i \stackrel{\text{def}}{=} D_{-i} U_{-i} X_{\mathcal{C} \cup [i-1]} Y_{\mathcal{C} \cup [i-1]}$.

Now let us apply Lemma 3.2 by substituting $XY \leftarrow X_j^1 Y_j^1, M \leftarrow R_j^1 M^1, p \leftarrow X_j Y_j, z \leftarrow z, \varepsilon \leftarrow \varepsilon, \delta \leftarrow \delta, \beta \leftarrow \beta, \varepsilon' \leftarrow \varepsilon'$ and $c \leftarrow 16\delta_1(c+1)$. Condition 1. in Lemma 3.2 is implied by Claim 3.11. Conditions 2. and 3. are implied by Claim 3.12. Also we have $\widetilde{\mathrm{srec}}^{z,\mu}_{(1+\varepsilon')\delta/\beta,\delta}(f) > \frac{32\delta_1(c+1)}{11\varepsilon^3}$, by our choice of c. Hence condition 4. must be false and hence $\mathrm{err}_f\left(X_j^1 Y_j^1 M^1\right) = \mathrm{err}_f\left(X_j^1 Y_j^1 R_j^1 M^1\right) > \varepsilon$. This shows condition 2. of this Claim. □

Claim 3.11. Let $\mathcal{R}$ denote the space of R_j. There exist functions $u_{x_j}, u_{y_j} : \mathcal{R} \times \mathcal{M} \to [0,1]$ for all $(x_j, y_j) \in \mathcal{X} \times \mathcal{Y}$ and a real number $q > 0$ such that

$$\Pr\left[X_j^1 Y_j^1 R_j^1 M^1 = x_j y_j r_j m\right] = \frac{1}{q} \mu(x_j, y_j) u_{x_j}(r_j, m) u_{y_j}(r_j, m).$$

Proof. Note that $X_j Y_j$ is independent of R_j. Now consider a private-coin two-way protocol Π_1 with input $X_j Y_j$ as follows. Let Alice generate R_j and send to Bob. Alice and Bob then generate $(X_{-j})_{x_j r_j}$ and $(Y_{-j})_{y_j r_j}$, respectively. Then they run the protocol Π. Thus, from Lemma 2.9,

$$\Pr[X_j Y_j R_j M = x y_j r m] = \mu(x_j, y_j) \cdot v_{x_j}(r_j, m) \cdot v_{y_j}(r_j, m),$$

where $v_{x_j}, v_{y_j} : \mathcal{R} \times \mathcal{M} \to [0,1]$, for all $(x_j, y_j) \in \mathcal{X} \times \mathcal{Y}$.

Note that conditioning on $T^{(r)} = 1$ corresponds to choosing a subset, say S, of $\mathcal{R} \times \mathcal{M}$. Let

$$q \stackrel{\text{def}}{=} \sum_{x_j y_j r_j m : (r_j, m) \in S} \mu(x_j, y_j) v_{x_j}(r_j, m) v_{y_j}(r_j, m) \ .$$

Then

$$\Pr\left[X_j^1 Y_j^1 R_j^1 M^1 = x_j y_j r_j m\right] = \frac{1}{q} \mu(x_j, y_j) v_{x_j}(r_j, m) v_{y_j}(r_j, m),$$

for $(r_j, m) \in S$ and $\Pr\left[X_j^1 Y_j^1 R_j^1 M^1 = x_j y_j r_j m\right] = 0$ otherwise.

Now define

$$u_{x_j}(r_j, m) \stackrel{\text{def}}{=} v_{x_j}(r_j, m), \text{ and } u_{y_j}(r_j, m) \stackrel{\text{def}}{=} v_{y_j}(r_j, m),$$

for $(r_j, m) \in S$ and define them to be 0 otherwise. The claim follows. □

Claim 3.12. If $\Pr\left[T^{(r)}=1\right] > 2^{-\delta_1 t}$, then there exists a coordinate $j \notin \mathcal{C}$ such that

$$\mathrm{S}\big(X_j^1 Y_j^1 \big\| X_j Y_j\big) \leq 8\delta_1 = \tfrac{\varepsilon^2}{4}, \tag{15}$$

and

$$\mathrm{I}\big(X_j^1; M^1 R_j^1 \big| Y_j^1\big) + \mathrm{I}\big(Y_j^1; M^1 R_j^1 \big| X_j^1\big) \leq 16\delta_1(c+1). \tag{16}$$

Proof. This follows using Claim III.6 in [18]. We include a proof in Appendix for completeness. □

Conclusion and open problems

We provide a strong direct product result for the two-way public-coin communication complexity in terms of an important and widely used lower bound method, the smooth rectangle bound. Some natural questions that arise are:

1. Is the smooth rectangle bound a tight lower bound for the two-way public-coin communication complexity for all relations? If yes, this would imply a strong direct product result for the two-way public-coin communication complexity for all relations, settling a major open question in this area. To start with we can ask: Is the smooth rectangle bound a polynomially tight lower bound for the two-way public-coin communication complexity for all relations?
2. Or on the other hand, can we exhibit a relation for which the smooth rectangle bound is (asymptotically) strictly smaller than its two-way public-coin communication complexity?
3. Can we show similar direct product results in terms of possibly stronger lower bound methods like the partition bound and the information complexity?
4. It will be interesting to obtain new optimal lower bounds for interesting functions and relations using the smooth rectangle bound, implying strong direct product results for them.

Acknowledgement. We thank Prahladh Harsha for helpful discussions.

References

[1] Laszlo Babai, Peter Frankl, and Janos Simon. Complexity classes in communication complexity theory. In *Proceedings of the 27th Annual Symposium on Foundations of Computer Science*, SFCS '86, pages 337–347, Washington, DC, USA, 1986. IEEE Computer Society.

[2] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. In *Proceedings of the 43rd Symposium on Foundations of Computer Science*, FOCS '02, pages 209–218, Washington, DC, USA, 2002. IEEE Computer Society.

[3] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. In *Proceedings of the 42nd ACM symposium on Theory of computing*, STOC '10, pages 67–76, New York, NY, USA, 2010. ACM.

[4] Paul Beame, Toniann Pitassi, Nathan Segerlind, and Avi Wigderson. A strong direct product theorem for corruption and the multiparty communication complexity of disjointness. *Comput. Complex.*, 15(4):391–432, December 2006.

[5] Mark Braverman. Interactive information complexity. In *Proceedings of the 44th annual ACM symposium on Theory of computing*, STOC '12, pages 505–524, New York, NY, USA, 2012. ACM.

[6] Mark Braverman and Anup Rao. Information equals amortized communication. In *Proceedings of the 52nd Symposium on Foundations of Computer Science*, FOCS '11, pages 748–757, Washington, DC, USA, 2011. IEEE Computer Society.

[7] Mark Braverman and Omri Weinstein. A discrepancy lower bound for information complexity. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, Lecture Notes in Computer Science, Springer Berlin Heidelberg, pages 459–470, volume 7408, 2012, isbn 978-3-642-32511-3.

[8] Amit Chakrabarti and Oded Regev. An optimal lower bound on the communication complexity of gap-hamming-distance. In *Proceedings of the 43rd annual ACM symposium on Theory of computing*, STOC '11, pages 51–60, New York, NY, USA, 2011. ACM.

[9] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *In Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science*, pages 270–278, 2001.

[10] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications. John Wiley & Sons, New York, NY, USA, 1991.

[11] Dmitry Gavinsky. Classical interaction cannot replace a quantum message. In *Proceedings of the 40th annual ACM symposium on Theory of computing*, STOC '08, pages 95–102, New York, NY, USA, 2008. ACM.

[12] Prahladh Harsha and Rahul Jain. *A strong direct product theorem for the tribes function via the smooth-rectangle bound*. Preprint available at arXiv:1302.0275.

[13] Thomas Holenstein. Parallel repetition: simplifications and the no-signaling case. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, STOC '07, pages 411–419, New York, NY, USA, 2007. ACM.

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

[15] Rahul Jain and Hartmut Klauck. New results in the simultaneous message passing model via information theoretic techniques. In *Proceedings of the 2009 24th Annual IEEE Conference on Computational Complexity*, CCC '09, pages 369–378, Washington, DC, USA, 2009. IEEE Computer Society.

[16] Rahul Jain and Hartmut Klauck. The partition bound for classical communication complexity and query complexity. In *Proceedings of the 2010 IEEE 25th Annual Conference on Computational Complexity*, CCC '10, pages 247–258, Washington, DC, USA, 2010. IEEE Computer Society.

[17] Rahul Jain, Hartmut Klauck, and Ashwin Nayak. Direct product theorems for classical communication complexity via subdistribution bounds: extended abstract. In *Proceedings of the 40th annual ACM symposium on Theory of computing*, STOC '08, pages 599–608, New York, NY, USA, 2008. ACM.

[18] Rahul Jain, Attila Pereszlényi, and Penghui Yao. A direct product theorem for the two-party bounded-round public-coin communication complexity. In *Proceedings of the 53rd Annual IEEE Symposium on Foundations of Computer Science*, FOCS '12, pages 167–176, Washington, DC, USA, 2012. IEEE Computer Society.

[19] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. Privacy and interaction in quantum communication complexity and a theorem about the relative entropy of quantum states. In *Proceedings of the 43rd Symposium on Foundations of Computer Science*, FOCS '02, pages 429–438, Washington, DC, USA, 2002. IEEE Computer Society.

[20] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. Prior entanglement, message compression and privacy in quantum communication. In *Proceedings of the 20th Annual IEEE Conference on Computational Complexity*, pages 285–296, Washington, DC, USA, 2005. IEEE Computer Society.

[21] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A direct sum theorem in communication complexity via message compression. In *Proceedings of the 30th international conference on Automata, languages and programming*, ICALP'03, pages 300–315, Berlin, Heidelberg, 2003. Springer-Verlag.

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

[23] Iordanis Kerenidis, Sophie Laplante, Virginie Lerays, Jérémie Roland, and David Xiao. Lower bounds on information complexity via zero-communication protocols and applications. In *Proceedings of the 53rd Annual IEEE Symposium on Foundations of Computer Science*, FOCS '12, pages 500–509, Washington, DC, USA, 2012. IEEE Computer Society.

[24] Hartmut Klauck. Rectangle size bounds and threshold covers in communication complexity. In *Proceedings of the 2003 IEEE 18th Annual Conference on Computational Complexity*, CCC '18, pages 118–134, Washington, DC, USA, 2003. IEEE Computer Society.

[25] Hartmut Klauck. A strong direct product theorem for disjointness. In *Proceedings of the 42nd ACM symposium on Theory of computing*, STOC '10, pages 77–86, New York, NY, USA, 2010. ACM.

[26] Eyal Kushilevitz and Noam Nisan. *Communication Complexity.* Cambridge University Press, 1996.

[27] Troy Lee, Adi Shraibman, and Robert Špalek. A direct product theorem for discrepancy. In *Proceedings of the 2008 IEEE 23rd Annual Conference on Computational Complexity*, CCC '08, pages 71–80, Washington, DC, USA, 2008. IEEE Computer Society.

[28] Nati Linial and Adi Shraibman. Lower bounds in communication complexity based on factorization norms. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, STOC '07, pages 699–708, New York, NY, USA, 2007. ACM.

[29] Ran Raz. A parallel repetition theorem. In *Proceedings of the twenty-seventh annual ACM symposium on Theory of computing*, STOC '95, pages 447–456, New York, NY, USA, 1995. ACM.

[30] 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.

[31] Alexander A. Razborov. On the distributional complexity of disjointness. *Theoretical Computer Science*, 106(2):385–390, 1992.

[32] Oded Regev and Bo'az Klartag. Quantum one-way communication can be exponentially stronger than classical communication. In *Proceedings of the 43rd annual ACM symposium on Theory of computing*, STOC '11, pages 31–40, New York, NY, USA, 2011. ACM.

[33] Alexander A. Sherstov. The communication complexity of gap hamming distance. *Electronic Colloquium on Computational Complexity (ECCC)*, 18:63, 2011.

[34] Alexander A. Sherstov. Strong direct product theorems for quantum communication and query complexity. In *Proceedings of the 43rd annual ACM symposium on Theory of computing*, STOC '11, pages 41–50, New York, NY, USA, 2011. ACM.

[35] Emanuele Viola. The communication complexity of addition. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, SODA'13, 2013.

[36] Andrew C. Yao. Lower bounds by probabilistic arguments. In *Proceedings of the 24th Annual Symposium on Foundations of Computer Science*, SFCS '83, pages 420–428, Washington, DC, USA, 1983. IEEE Computer Society.

[37] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing (preliminary report). In *Proceedings of the eleventh annual ACM symposium on Theory of computing*, STOC '79, pages 209–213, New York, NY, USA, 1979. ACM.

[38] Andrew Chi-Chih Yao. Theory and applications of trapdoor functions. In *Proceedings of the 23RD Symposium on Foundations of Computer Science*, FOCS '82, pages 80–91, Washington, DC, USA, 1982. IEEE Computer Society.

Proof of Lemma 2.7: Let $G = \left\{ a : \left| 1 - \frac{\Pr[A'=a]}{\Pr[A=a]} \right| \leq \frac{\varepsilon}{r} \right\}$, then

$$2\varepsilon \geq \sum_a \Big| \Pr[A=a] - \Pr[A'=a] \Big| \geq \sum_{a \notin G} \Big| \Pr[A=a] - \Pr[A'=a] \Big|$$
$$= \sum_{a \notin G} \Pr[A=a] \left| 1 - \frac{\Pr[A'=a]}{\Pr[A=a]} \right| \geq \Pr_{a \leftarrow A}[a \notin G] \cdot \frac{\varepsilon}{r}.$$

Thus $\Pr_{a \leftarrow A}[a \in G] \geq 1 - 2r$. The second inequality follows immediately. □

Proof of Lemma 2.11: Let $c \stackrel{\text{def}}{=} \widetilde{\text{srec}}^{z,\lambda}_{(1+\varepsilon')\frac{\delta}{\beta},\delta}(f)$. Let g be such that $\widetilde{\text{rec}}^{z,\lambda}_{(1+\varepsilon')\frac{\delta}{\beta}}(g) = c$ and $\Pr_{(x,y)\leftarrow\lambda}[f(x,y) \neq g(x,y)] \leq \delta$. If $\text{D}^{\lambda}_{\varepsilon}(f) \geq c - \log(4/\varepsilon)$ then we are done using Fact 2.8.

So lets assume for contradiction that $\text{D}^{\lambda}_{\varepsilon}(f) < c - \log(4/\varepsilon)$. This implies that there exists a deterministic protocol Π for f with communication $c - \log(4/\varepsilon)$ and distributional error under λ bounded by ε. Since $\Pr_{(x,y)\leftarrow\lambda}[f(x,y) \neq g(x,y)] \leq \delta$, the protocol Π will have distributional error at most $\varepsilon + \delta$ for g. Let M represent the message transcript of Π and let O represent protocol's output. We assume that the last $\lceil \log |Z| \rceil$ bits of M contain O. We have,

1. $\Pr_{m \leftarrow M}[\Pr[M=m] \leq 2^{-c}] \leq \varepsilon/4$, since the total number of message transcripts in Π is at most $2^{c - \log(4/\varepsilon)}$.
2. $\Pr_{m \leftarrow M}[O = z |\ M = m] > \beta - \varepsilon$, since $\Pr_{(x,y)\leftarrow\lambda}[f(x,y) = \{z\}] = \beta$ and distributional error of Π under λ is bounded by ε for f.
3. $\Pr_{m \leftarrow M}\Big[\Pr_{(x,y)\leftarrow(XY)_m}[(x,y,O) \notin g |\ M = m] \geq \frac{\varepsilon+\delta}{\beta - 2\varepsilon}\Big] \leq \beta - 2\varepsilon$, since distributional error of Π under λ is bounded by $\varepsilon + \delta$ for g.

Using all of above we obtain a message transcript m such that $\Pr[M=m] > 2^{-c}$ and $(O = z |\ M = m)$ and

$$\Pr_{(x,y)\leftarrow(XY|M=m)}[(x,y,O) \notin g |\ M = m] \leq \frac{\varepsilon + \delta}{\beta - 2\varepsilon} < (1+\varepsilon')\frac{\delta}{\beta}.$$

This and the fact that the support of $(XY |\ M = m)$ is a rectangle, implies that $\widetilde{\text{rec}}^{z,\lambda}_{(1+\varepsilon')\frac{\delta}{\beta}}(g) < c$, contradicting the definition of c. Hence it must be that $\text{D}^{\lambda}_{\varepsilon}(f) \geq c - \log(4/\varepsilon)$, which using Fact 2.8 shows the desired. □

Proof of Claim 3.12: The following calculations are helpful for achieving (15).

$$\delta_1 k > \text{S}_{\infty}\big(X^1Y^1 \big\| XY\big) \geq \text{S}\big(X^1Y^1 \big\| XY\big) \geq \sum_{i \notin C} \text{S}\big(X^1_i Y^1_i \big\| X_i Y_i\big), \tag{17}$$

where the first inequality follows from the assumption that $\Pr\left[T^{(r)}=1\right] > 2^{-\delta k}$, and the last inequality follows from Fact 2.2. The following calculations are helpful for (16).

$$
\begin{aligned}
&\delta_1 k > \mathrm{S}_\infty\left(X^1Y^1D^1U^1 \middle\| XYDU\right)\\
&\geq \mathrm{S}\left(X^1Y^1D^1U^1 \middle\| XYDU\right)\\
&\geq \mathop{\mathbb{E}}_{\substack{(d,u,x_C,y_C)\\ \leftarrow D^1,U^1,X^1_C,Y^1_C}} \left[\mathrm{S}\left(\left(X^1Y^1\right)_{d,u,x_C,y_C} \middle\| (XY)_{d,u,x_C,y_C}\right)\right] && (18)\\
&= \sum_{i\notin C} \mathop{\mathbb{E}}_{\substack{(d,u,x_{C\cup[i-1]},y_{C\cup[i-1]})\\ \leftarrow D^1,U^1,X^1_{C\cup[i-1]},Y^1_{C\cup[i-1]}}} \left[\mathrm{S}\left(\left(X^1_iY^1_i\right)_{\substack{d,u,x_{C\cup[i-1]},\\ y_{C\cup[i-1]}}} \middle\| (X_iY_i)_{\substack{d,u,x_{C\cup[i-1]},\\ y_{C\cup[i-1]}}}\right)\right] && (19)\\
&= \sum_{i\notin C} \mathop{\mathbb{E}}_{\substack{(d_i,u_i,r_i)\\ \leftarrow D^1_i,U^1_i,R^1_i}} \left[\mathrm{S}\left((X^1_iY^1_i)_{d_i,u_i,r_i} \middle\| (X_iY_i)_{d_i,u_i,r_i}\right)\right] && (20)\\
&= \frac{1}{2}\sum_{i\notin C} \mathop{\mathbb{E}}_{(r_i,x_i)\leftarrow R^1_i,X^1_i} \left[\mathrm{S}\left(\left(Y^1_i\right)_{r_i,x_i} \middle\| (Y_i)_{x_i}\right)\right] + \quad \frac{1}{2}\sum_{i\notin C} \mathop{\mathbb{E}}_{(r_i,y_i)\leftarrow R^1_i,Y^1_i} \left[\mathrm{S}\left(\left(X^1_i\right)_{r_i,y_i} \middle\| (X_i)_{y_i}\right)\right]. && (21)
\end{aligned}
$$

Above, Eq. (18) and Eq. (19) follow from Fact 2.2; Eq. (20) is from the definition of R_i. Eq. (21) follows since D^1_i is independent of R^1_i and with probability half D^1_i is 0, in which case $U^1_i = X^1_i$ and with probability half D^1_i is 1 in which case $U^1_i = Y^1_i$. By Fact 2.4,

$$\sum_{i\notin C}\left(\mathrm{I}\left(X^1_i; R^1_i \middle| Y^1_i\right) + \mathrm{I}\left(Y^1_i; R^1_i \middle| X^1_i\right)\right) \leq 2\delta_1 k. \tag{22}$$

We also need the following calculations, which exhibits that the information carried by messages about sender's input is small.

$$
\begin{aligned}
\delta_1 ck &\geq \left|M^1\right| \geq \mathrm{I}\left(X^1Y^1; M^1 \middle| D^1U^1X^1_CY^1_C\right)\\
&= \sum_{i\notin C} \mathrm{I}\left(X^1_iY^1_i; M^1 \middle| D^1U^1X^1_{C\cup[i-1]}Y^1_{C\cup[i-1]}\right)\\
&= \sum_{i\notin C} \mathrm{I}\left(X^1_iY^1_i; M^1 \middle| D^1_iU^1_iR^1_i\right)\\
&= \frac{1}{2}\sum_{i\notin C}\left(\mathrm{I}\left(X^1_i; M^1 \middle| R^1_iY^1_i\right) + \mathrm{I}\left(Y^1_i; M^1 \middle| R^1_iX^1_i\right)\right). && (23)
\end{aligned}
$$

Above first equality follows from chain rule for mutual information, second equality follows from definition of R^1_i and the third equality follows since with probability half D^1_i is 0, in which case $U^1_i = X^1_i$ and with probability half D^1_i is 1 in which case $U^1_i = Y^1_i$.

Combining Eqs. (17), (21) and (23), and making standard use of Markov's inequality, we can get a coordinate $j \notin C$ such that

$$
\begin{aligned}
&\mathrm{S}\left(X^1_jY^1_j \middle\| X_jY_j\right) \leq 8\delta_1,\\
&\mathrm{I}\left(X^1_j; R^1_j \middle| Y^1_j\right) + \mathrm{I}\left(Y^1_j; R^1_j \middle| X^1_j\right) \leq 16\delta_1,\\
&\mathrm{I}\left(X^1_j; M^1 \middle| R^1_jY^1_j\right) + \mathrm{I}\left(Y^1_j; M^1 \middle| R^1_jX^1_j\right) \leq 16\delta_1 c.
\end{aligned}
$$

The first inequality is exactly the same as Eq. (15). Eq. (16) follows by adding the last two inequalities. □

Alternate definition of smooth rectangle bound

An alternate definition of the smooth rectangle bound was introduced by Jain and Klauck [16], using the following linear program.

Definition .13. For function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$, the ϵ- smooth rectangle bound of f denoted $\mathrm{srec}_\epsilon(f)$ is defined to be $\max\{\mathrm{srec}^z_\epsilon(f) : z \in \mathcal{Z}\}$, where $\mathrm{srec}^z_\epsilon(f)$ is given by the optimal value of the following linear program.

Primal

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

Dual

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

The following lemma lower bounds the natural definition in terms of the linear programming definition of smooth rectangle bound. A similar, but weaker, relationship was shown in [16].

Lemma .14. *Let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ be a function. Let $z \in Z$ and $\varepsilon > 0$. There exists a distribution $\mu \in \mathcal{X} \times \mathcal{Y}$ and $\delta, \beta > 0$ such that*

$$\widetilde{\mathrm{srec}}^{z,\mu}_{(1+\varepsilon^2)\frac{\delta}{\beta},\delta}(f) \geq \log(\mathrm{srec}^z_\varepsilon(f)) + 3\log\varepsilon.$$

Proof. Let $(\lambda'_{x,y}, \phi'_{x,y})$ be an optimal solution to the Dual. For $(x,y) \in f^{-1}(z)$, if $\lambda'_{x,y} > \phi'_{x,y}$ define $\lambda = \lambda'_{x,y} - \phi'_{x,y}$ and $\phi_{x,y} = 0$. Otherwise define $\lambda = 0$ and $\phi_{x,y} = \phi'_{x,y} - \lambda'_{x,y}$. For $(x,y) \in f^{-1} - f^{-1}(z)$ define $\phi_{x,y} = 0$. We note that $(\lambda_{x,y}, \phi_{x,y})$ is an optimal solution to the Dual with potentially higher objective value. Hence $(\lambda_{x,y}, \phi_{x,y})$ is also an optimal solution to the Dual.

Let us define three sets

$$U_1 \stackrel{\text{def}}{=} \{(x,y) |\ f(x,y) = z, \lambda_{x,y} > 0\},$$

$$U_2 \stackrel{\text{def}}{=} \{(x,y) |\ f(x,y) = z, \phi_{x,y} > 0\},$$

$$U_0 \stackrel{\text{def}}{=} \{(x,y) \in f^{-1} |\ f(x,y) \neq z, \lambda_{x,y} > 0\}.$$

Define,

$$\forall (x,y) \in U_1 : \mu'(x,y) \stackrel{\text{def}}{=} \lambda_{x,y},$$

$$\forall (x,y) \in U_2 : \mu'(x,y) \stackrel{\text{def}}{=} \varepsilon \phi_{x,y},$$

$$\forall (x,y) \in U_0 : \mu'(x,y) \stackrel{\text{def}}{=} \varepsilon \lambda_{x,y}.$$

Define $r \stackrel{\text{def}}{=} \sum_{x,y} \mu'(x,y)$ and define probability distribution $\mu \stackrel{\text{def}}{=} \mu'/r$. Let $\text{srec}^z_\epsilon(f) = 2^c$. Define function g such that $g(x,y) = z$ for $(x,y) \in U_1$; $g(x,y) = f(x,y)$ for $(x,y) \in U_0$ and $g(x,y) = z'$ (for some $z' \neq z$) for $(x,y) \in U_2$. Then,

$$2^c = \sum_{(x,y) \in f^{-1}(z)} ((1-\epsilon)\lambda_{x,y} - \phi_{x,y}) - \sum_{(x,y) \notin f^{-1}(z)} \epsilon \cdot \lambda_{x,y} = (1-\epsilon)\mu'(U_1) - \frac{1}{\varepsilon}\mu'(U_2) - \mu'(U_0)$$

This implies $r \geq \mu'(U_1) \geq 2^c$. Consider rectangle W.

$$\begin{aligned}
&\sum_{(x,y) \in f^{-1}(z) \cap W} (\lambda_{x,y} - \phi_{x,y}) - \sum_{(x,y) \in (W - f^{-1}(z))} \lambda_{x,y} \leq 1 \\
&\Rightarrow \sum_{(x,y) \in U_1 \cap W} \mu_{x,y} - \frac{1}{\varepsilon} \sum_{(x,y) \in U_2 \cap W} \mu_{x,y} - \sum_{(x,y) \in U_0 \cap W} \frac{1}{\varepsilon}\mu_{x,y} \leq \frac{1}{r} \\
&\Rightarrow \varepsilon \left(\sum_{(x,y) \in U_1 \cap W} \mu_{x,y} - \frac{1}{r} \right) \leq \sum_{(x,y) \in U_2 \cap W} \mu_{x,y} + \sum_{(x,y) \in U_0 \cap W} \mu_{x,y} \\
&\Rightarrow \varepsilon \left(\sum_{(x,y) \in g^{-1}(z) \cap W} \mu_{x,y} - \frac{1}{r} \right) \leq \sum_{(x,y) \in W - g^{-1}(z)} \mu_{x,y} \\
&\Rightarrow \varepsilon \left(\sum_{(x,y) \in W} \mu_{x,y} - \frac{1}{r} \right) \leq (1+\varepsilon) \cdot \sum_{(x,y) \in W - g^{-1}(z)} \mu_{x,y} \\
&\Rightarrow \varepsilon \left(\sum_{(x,y) \in W} \mu_{x,y} - 2^{-c} \right) \leq (1+\varepsilon) \cdot \sum_{(x,y) \in W - g^{-1}(z)} \mu_{x,y}.
\end{aligned}$$

Now consider a W with $\mu(W) \geq 2^{-c}/\varepsilon^3$. We have $\mu(W - g^{-1}(z)) \geq \frac{(1-\varepsilon^3)\varepsilon}{1+\varepsilon}\mu(W)$. Define $\beta \stackrel{\text{def}}{=} \mu(U_1 \cup U_2)$, $\delta \stackrel{\text{def}}{=} \mu(U_2)$. Now,

$$(1-\varepsilon) r \beta \geq (1-\varepsilon)\mu'(U_1) \geq \frac{1}{\varepsilon}\mu'(U_2) = \frac{1}{\varepsilon} r \delta.$$

Hence we have

$$\mu(W - g^{-1}(z)) \geq \frac{(1-\varepsilon^3)\delta}{(1-\varepsilon^2)\beta}\mu(W) \geq (1+\varepsilon^2)\frac{\delta}{\beta}\mu(W).$$

This implies $\widetilde{\text{rec}}^{z,\mu}_{(1+\varepsilon^2)\delta/\beta}(g) \geq c + 3\log\varepsilon$. This implies that

$$\widetilde{\text{srec}}^{z,\mu}_{(1+\varepsilon^2)\frac{\delta}{\beta},\delta}(f) \geq c + 3\log\varepsilon = \log(\text{srec}^z_\varepsilon(f)) + 3\log\varepsilon.$$

□