New Strong Direct Product Results in Communication Complexity

RAHUL JAIN, National University of Singapore

We show two new direct product results in two different models of communication complexity. Our first result is in the one-way public-coin model. Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation and $\varepsilon > 0$ be a constant. Let $\mathsf{R}_{\varepsilon}^{1,\mathsf{pub}}(f)$ represent the communication complexity of f, with worst-case error ε in this model. We show that if for computing f^k (k independent copies of f) in this model, $o(k \cdot \mathsf{R}_{1/3}^{1,\mathsf{pub}}(f))$ communication is used, then the success is exponentially small in k. We show a new tight characterization of communication complexity in this model which strengthens the tight characterization shown in Jain et al. [2008]. We use this new characterization to show our direct product result and this characterization may also be of independent interest.

Our second direct product result is in the model of two-way public-coin communication complexity. We show a direct product result for all relations in this model in terms of a new complexity measure that we define. Our new measure is a generalization to nonproduct distributions, of the two-way product subdistribution bound of Jain et al. [2008]. Our direct product result therefore generalizes to nonproduct distributions, their direct product result in terms of the two-way product subdistribution bound. As an application of our new direct product result, we reproduce (via completely different arguments) strong direct product result for the set-disjointness problem which was previously shown by Klauck [2010]. We show this by proving that our new complexity measure gives a tight lower bound of $\Omega(n)$ for the set-disjointness problem on n-bit inputs (this strengthens the linear lower bound on the rectangle/corruption bound for set-disjointness shown by Razborov [1992]). In addition, we show that many previously known direct product results in this model are uniformly implied and often strengthened by our result.

Categories and Subject Descriptors: F.1.2 [**Computation by Abstract Devices**]: Modes of Computation

General Terms: Theory, Algorithms, Performance

Additional Key Words and Phrases: Communication complexity, strong direct product, information theory

ACM Reference Format:
Rahul Jain. 2015. New strong direct product results in communication complexity. J. ACM 62, 3, Article 20 (June 2015), 27 pages.
DOI: http://dx.doi.org/10.1145/2699432

1. INTRODUCTION

Can computing k simultaneous instances of a problem be done more efficiently than computing them in parallel? This question has been very well studied in many models of computation, first for being a natural, fundamental, and interesting question in itself, and second for the many implications it has for some other important questions. One way to pose this question for bounded error computation models is as follows. Let the resource required for solving a single instance with constant success be c; then if less than $k \cdot c$ resource is provided for solving k instances together, is the overall success exponentially small in k? This is referred to as the direct product question.

This work is supported by the Singapore Ministry of Education Tier 3 Grant and also the Core Grants of the Center for Quantum Technologies, Singapore.

Author's address: R. Jain, S15-04-01, Center for Quantum Technologies and Department of Computer Science, 3 Science Drive 2, National University of Singapore, Singapore 117543; email: rahul.jain.75@gmail.com.

DOI: http://dx.doi.org/10.1145/2699432

We consider this question in two models of communication complexity: the one-way public-coin model and the two-way public-coin model.

The One-Way Public-Coin Model. Let $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ be finite sets. Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a *complete* relation, by which we mean that, for each $(x, y) \in \mathcal{X} \times \mathcal{Y}$, there exists at least one $z \in \mathcal{Z}$ such that $(x, y, z) \in f$. We only consider complete relations in this work. Let $\varepsilon > 0$. Let Alice with input $x \in \mathcal{X}$ and Bob with input $y \in \mathcal{Y}$ wish to compute a $z \in \mathcal{Z}$ such that $(x, y, z) \in f$. In this model, Alice sends a single message to Bob who outputs z and Alice and Bob use pubic coins. Please refer to the excellent text by Kushilevitz and Nisan [1997] for a more formal introduction to the different models of communication complexity. Let $\mathsf{R}_{\varepsilon}^{1,\mathsf{pub}}(f)$ denote the (worst-case) communication of the best protocol $\mathcal{P}$ which achieves this with error at most ε (over the public-coins) for any input (x, y). We answer the direct product question for all relations f in this model in the following manner. Let f^k be the set $\{(x_1, \ldots, x_k, y_1, \ldots, y_k, z_1, \ldots, z_k) : \text{for } i = 1, 2, \ldots, k, (x_i, y_i, z_i) \in f\}$.

THEOREM 1.1. *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation, $\varepsilon > 0$ be a constant and k be a natural number. Then,*

$$\mathsf{R}_{1-2^{-\Omega(\varepsilon^3 k)}}^{1,\mathsf{pub}}(f^k) = \Omega\big(k \cdot \big(\varepsilon^2 \cdot \mathsf{R}_{\varepsilon}^{1,\mathsf{pub}}(f) - O(1)\big)\big).$$

To our knowledge, this is the first time a direct product statement has been made for all relations in any model of communication complexity. We present here some previous results which are now implied and strengthened by our current result.

(1) Jain et al. [2008] introduced the so-called one-way subdistribution bound and showed a direct product result in terms of the one-way subdistribution bound under product distributions. The one-way subdistribution bound forms a lower bound on $\mathsf{R}_{\varepsilon}^{1,\mathsf{pub}}(f)$ and hence our result implies the result of Jain et al. [2008].
(2) Gavinsky [2008] proves direct product for one-way distributional communication complexity of a certain class of relational problems under the uniform distribution. Gavinsky used his result to show communication vs entanglement tradeoff for communication protocols. Since our result holds for all relations, it implies the result of Gavinsky.
(3) de Wolf [2005] proves a strong direct product theorem for the one-way public-coin randomized communication complexity of the Index function. Ben-Aroya et al. [2008] derive a similar direct product theorem for the one-way quantum communication complexity of Index. Since Index captures the notion of VC-dimension, similar results follow for the one-way distributional (classical and quantum) communication complexity of any Boolean function under the worst-case product distribution. These results for classical communication complexity are implied by this result.
(4) Jain et al. [2005] show optimal direct sum for all relations in the one-way public coin communication complexity. A direct sum is a weaker question to direct product question and asks the following. Let the resource required for solving a single instance with constant success be c; then, if less than kc resource is provided for solving k instances together, is the overall success is at most a constant? (In direct product, the overall success is required to be exponentially small in k.) Jain et al. [2005] also show similar optimal direct sum result for one-way entanglement assisted quantum communication complexity of all relations; however quantum communication complexity is beyond the scope of this work.

Remark 1.2. In case $\mathsf{R}^{1,\mathsf{pub}}_{1/3}(f) \geq 1$, our result is trivial. However, in this case, the corresponding direct product statement $\mathsf{R}^{1,\mathsf{pub}}_{1-2^{-\Omega(k)}}(f^k) = \Omega(k)$ can be argued as follows. Since $\mathsf{R}^{1,\mathsf{pub}}_{1/3}(f) \geq 1$, there exists x_0, x_1 and y such that $f(x_0, y) \cap f(x_1, y) = \emptyset$. (We write $f(x, y) = \{z \in Z : (x, y, z) \in f\}$.)

Let $\mathcal{P}$ be a protocol for f^k with communication d bits and overall (worst-case) success probability p. Now consider protocol $\mathcal{P}'$ for f^k, where Alice does not communicate anything and let Bob guess the communication in $\mathcal{P}$ and then act as in $\mathcal{P}$. The worst-case success probability of $\mathcal{P}'$ is at least $2^{-d}p$. Suppose Bob's input for $\mathcal{P}'$ is $\bar{y} = (y, , \ldots, y)$. Since $f(x_0, y) \cap f(x_1, y) = \emptyset$, any answer is correct for at most one input sequence (for Alice) in $\{x_0, x_1\}^k$. Thus, there is an input sequence for Alice of the form $\bar{w} = (w_1, w_2, \ldots, w_k)$, such that the answer returned by Bob is correct for $(\bar{w}, \bar{y})$ with probability at most 2^{-k}. Thus, $2^{-k} \geq 2^{-d}p$. Hence, if $d < k/2$, then $p < 2^{-k/2}$.

The following corollary for the one-way private-coin communication complexity follows immediately from Theorem 1.1 and the well-known relationship

$$\mathsf{R}^{1,\mathsf{pub}}_{\varepsilon}(f) \leq \mathsf{R}^{1,\mathsf{pvt}}_{\varepsilon}(f) \leq \mathsf{R}^{1,\mathsf{pub}}_{\varepsilon}(f) + O(\log|\mathcal{X}| + \log|\mathcal{Y}|),$$

which holds for all relations $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ and constant $\varepsilon > 0$ [Newman 1991].

Corollary 1.3. *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation, $\varepsilon > 0$ be a constant and k be a natural number. Then,*

$$\mathsf{R}^{1,\mathsf{pvt}}_{1-2^{-\Omega(\varepsilon^3 k)}}(f^k) = \Omega\big(k\varepsilon^2 \cdot \big(\mathsf{R}^{1,\mathsf{pvt}}_{\varepsilon}(f) - O(\log|\mathcal{X}| + \log|\mathcal{Y}|)\big)\big).$$

Our Techniques. We follow a natural argument for showing the direct product result. Let us say there are totally k coordinates (instances) and we condition on success on some $l = d \cdot k$ ($d < 1$ is a small constant) coordinates. If the overall success in these l coordinates is already as small as we want, then we are done and stop. Otherwise, we exhibit another coordinate j outside of these l coordinates such that the success in the jth coordinate, even conditioned on the success in the l coordinates, is bounded away from 1. This way, the overall success keeps going down, decreasing by a constant factor per coordinate, and becomes exponentially small in k once $\Omega(k)$ coordinates have been chosen. This broad outline to show direct product results have been used previously, for example, by Raz [1998] in the famous parallel repetition theorem for two-prover games (and in the recent modifications of its proof due to Holenstein [2007]).

We do this argument in the distributional setting where one is concerned with average error over the inputs coming from a specified distribution rather than the worst-case error over all inputs. The distributional setting can then be related to the worst-case setting by the well-known Yao's principle. Let μ be a "hard distribution" on $\mathcal{X} \times \mathcal{Y}$, possibly nonproduct across $\mathcal{X}$ and $\mathcal{Y}$. Let us consider the inputs for f^k drawn from the distribution μ^k (k-fold product of μ). Now consider a one-way protocol $\mathcal{P}$ for f^k with communication $o(kc)$ and condition on a typical message string m from Alice (where c is $\mathsf{rcment}(f)$, please refer to the next section for precise definition). Conditioned on this message m and also on success in l coordinates, we analyze what the distribution of Alice and Bob's inputs on a typical coordinate j (outside the l coordinates) looks like. We argue that this distribution is still "hard enough", that is, Bob will make constant errors on this coordinate whichever way he tries to give an answer. We are able to identify some key properties in such a distribution, concerning its relationship to μ, and argue that any distribution with these properties must be a hard distribution, given that μ is a hard distribution.

We do this last argument by showing a new tight characterization of one-way public-coin communication complexity for all relations. We introduce a new measure of complexity which we call the robust conditional relative min-entropy bound (rcment). We show that this bound is equivalent, up to constants, to $\mathsf{R}_{\varepsilon}^{1,\mathsf{pub}}(f)$ (for any constant $\varepsilon > 0$). This bound forms lower bound on the one-way subdistribution bound of Jain et al. [2008], where they also show that their bound is equivalent, up to constants, to $\mathsf{R}_{\varepsilon}^{1,\mathsf{pub}}(f)$. We make crucial use of a message compression protocol due to Braverman and Rao [2011] while exhibiting this tight characterization.

One key difficulty that is faced in the argument outlined here is that Bob's answer on the jth coordinate can depend on his inputs in other coordinates. Since μ could be a nonproduct distribution, Bob's inputs in other coordinates, because of the correlations between Alice's and Bob's inputs within each instance, potentially provide him information about Alice's inputs, in addition to the information obtained from the message from Alice. This difficulty is overcome by splitting the distribution μ into a convex combination of several product distributions. The particular way in which we split distributions leads us to consider the conditional distributions (conditioned on Bob's inputs) in the definition of rcment. This idea of splitting a nonproduct distribution into convex combination of product distributions has been used in several previous works to handle nonproduct distributions in different settings [Razborov 1992; Raz 1998; Bar-Yossef et al. 2002; Holenstein 2007; Barak et al. 2010].

Another key difficulty faced in the argument outlined here is as follows. Once we condition on success in the l coordinates, and look at the joint distribution X_jY_j (for a typical coordinate j outside the l coordinates), we are able to argue that the joint distribution of X_jY_j has small (less than c) conditional (conditioned on Y_j) relative entropy with μ (please refer to the next section for definitions of information theoretic quantities). However, X_jY_j may no longer be one-way for μ. Nonetheless, we argue that X_jY_j also has small (a very small constant) conditional (conditioned on X_j) relative entropy with μ. This helps us obtain another distribution X^1Y^1 which is close (in total variation distance) to X_jY_j, is one-way for μ and has small (less than c) robust conditional (conditioned on Y) relative min-entropy with μ. This shows that X^1Y^1 (and hence X_jY_j) must have error (since we have exhibited tight characterization for the robust-conditional min-entropy bound earlier).

The Two-Way Public-Coin Model. In this model, Alice on input x and Bob on input y exchange messages using public coins and at the end agree on a common output z. Let $\mathsf{R}_{\varepsilon}^{2,\mathsf{pub}}(f)$ denote the (worst-case) communication of the best protocol $\mathcal{P}$ which achieves this with error at most ε (over the public-coins) for any input (x, y). We show a direct product result in terms of a new complexity measure that we introduce: the ε-error two-way conditional relative entropy bound of f with respect to distribution μ, denoted $\mathsf{crent}_{\varepsilon}^{2,\mu}(f)$. The measure $\mathsf{crent}_{\varepsilon}^{2,\mu}(f)$ forms a lower bound (up to constant factors) on $\mathsf{R}_{\varepsilon}^{2,\mathsf{pub}}(f)$. Although this result is not an optimal direct product result that one may desire, we show how many previously known direct product results in the two-way model follow as a consequence of our result.

(1) Recently Klauck [2010] showed a direct product result for the set disjointness problem. In the set disjointness problem, Alice with input $x \in \{0, 1\}^n$ and Bob with input $y \in \{0, 1\}^n$ are supposed to determine if x and y intersect when viewed as characteristic vectors of subsets of $[n]$. This is arguably one of the most well-studied problems in communication complexity. We show (in Section 5) that our new complexity measure crent gives tight lower bound for the set-disjointness problem. This combined with the direct product in terms of crent, implies strong direct product result for the set disjointness problem for its two-way public-coin communication complexity. We point out here that the arguments used in Klauck [2010] are arguably specifically

geared to handle the set disjointness problem. In contrast, our result is much more general as we further argue in this article.

(2) When μ is a product distribution, $\mathsf{crent}^{2,\mu}_{\varepsilon}(f)$ forms an upper bound (up to constant factors) on the two-way subdistribution bound of Jain et al. [2008]. Hence, our direct product result implies the direct result of Jain et al. [2008] using the two-way product subdistribution bound and in addition generalizes their result to nonproduct distributions. It was pointed out in Jain et al. [2008] that their result provides a unified view of several recent works on the topic, simultaneously generalizing and strengthening them. These works include the strong direct product property for the rectangle/corruption bound for Boolean functions due to Beame et al. [2007].

(3) Shaltiel [2003] gave strong direct product theorem for the discrepancy bound for communication complexity under the uniform distribution. The discrepancy bound under product distributions (in particular, under the uniform distribution) is upper bounded by the rectangle bound which in turn is upper bounded (up to constants) by the crent. Therefore, our result implies and strengthens on Shaltiel's result and in particular implies strong direct product for the Inner Product function ($\mathsf{IP}_n(x, y) = \sum_i x_i \cdot y_i \mod 2$), since, for this function, the discrepancy bound under the uniform distribution is $\Omega(n)$.

Our techniques to show the direct product in the two-way model are quite similar to the techniques in the one-way model. However, in the two-way model, we do not present an upper bound on the public-coin communication complexity in terms of the new measure crent.

Recent Developments in the Direct Product Question. In a recent work, Jain et al. [2012] have extended our direct product result (for one-way protocols) to apply for bounded round public-coin communication protocols. One key difference between our work and theirs is that while we look at the distribution of X_jY_j (for a typical coordinate j outside the l coordinates where success is conditioned), conditioned on a typical message transcript m of the protocol for f^k; they look at the joint distribution of X_jY_jM (where M is the random variable representing the message transcript of the protocol for f^k). This helps them to argue about multiple-round protocols (even in the absence of a tight characterization of a lower bound method for such protocols). They have used and extended many arguments and techniques from our work, implicitly and explicitly.

In a more recent work, Braverman et al. [2013a] have provided a direct product result for bounded round public-coin communication protocols with near optimal dependence on the number of rounds. In another work [Braverman et al. 2013b], the same set of authors have provided a direct product result the two-way unbounded round public-coin protocols, which can be roughly stated as follows. Let c be the communication required for one-copy of f. If communication less than $\sqrt{k} \cdot c$ is provided for f^k, then the success is exponentially small in k. For the two-way unbounded round public-coin protocols, Jain and Yao [2012] have provided a direct product result in terms the smooth rectangle bound and this strengthens our direct product result in terms of crent. The smooth rectangle bound has been introduced by Jain and Klauck [2009] as a lower bound method for two-way public-coin communication complexity and is stronger than many other lower bound methods including crent. Very recently, a strong direct product result has been shown in terms of (internal) *information complexity*, which is stronger lower bound method than smooth rectangle bound, by Braverman and Weinstein [2014].

Other Related Work in Communication Complexity. Parnafes et al. [1997] prove a direct product result when a different algorithm works for each of the different instances and each algorithm is only provided communication at most the communication complexity of a single instance (with constant error). In their result, the bound on the success probability is shown to behave like $2^{-k/c}$ for the communication complexity c of the problem at hand. Lee et al. [2008] have shown strong direct product for the discrepancy bound under arbitrary distributions and Viola and Wigderson [2008] have extended it to the multiparty case. Recently, Sherstov [2011] showed strong direct product for the generalized-discrepancy bound. For deterministic protocols, it is known that k times the square root of the deterministic communication complexity of a function f is needed to compute k instances of f (see, e.g., Kushilevitz and Nisan [1997, Exercise 4.11, page 46]). It is also straightforward to show that the deterministic one-way communication complexity of every function f has the direct sum property. In a sequence of results [Jain et al. 2003; Harsha 2009; Barak et al. 2010; Braverman and Rao 2011], successively tighter direct sum results for all relations in the bounded-round two-way public-coin model have been shown, however optimal direct sum result holding for all relations for the two-way (unbounded round) model is still open. Direct sum result for all relations in the public-coin SMP model has been shown in Jain et al. [2005] both for classical and quantum protocols. In the SMP (simultaneous message passing) model, Alice with input x and Bob with input y, each send a message to a Referee who outputs z. Direct sum results for all relations in the private-coin SMP model has been shown in Jain and Klauck [2009] both for classical and quantum protocols. In a weak direct product theorem, one shows that the success probability of solving k instances of a problem with the resources needed to solve one instance (with constant success) goes down exponentially with k. Klauck [2004] shows such a result for the rectangle/corruption bound for all functions and all distributions in the two-way model and Jain et al. [2008] extend this result for all relations and for all distributions in the same model.

Indeed this is only an incomplete list of the many interesting results concerning direct product and related questions in communication complexity.

Organization. In Section 2, we provide some information theory and communication complexity preliminaries that we need. We refer the reader to Cover and Thomas [1991] and Kushilevitz and Nisan [1997] for good introductions to these topics respectively. In Section 3, we introduce our new bound for one-way communication, show that it tightly characterizes one-way public-coin communication complexity and show our direct product result in the one-way model. In Section 4, we introduce our new bound for two-way communication and show a direct product result in its terms. In Section 5, we present the strong direct product for set disjointness as an application of our two-way direct product result.

2. PRELIMINARIES

Information Theory. Let $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ be finite sets and k be a natural number. Let $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. Let $\mathcal{X}^k$ represent the k-fold Cartesian product of $\mathcal{X}$ with itself. Let $\mathcal{P}(\mathcal{X})$ denote the set of all distributions over $\mathcal{X}$. Let $\mu \in \mathcal{P}(\mathcal{X})$. We let $\mu(x)$ represent the probability of x under μ. The entropy of μ is defined as $S(\mu) = -\sum_{x \in \mathcal{X}} \mu(x) \log \mu(x)$ (all logarithms are to base 2 unless otherwise specified). 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. For distributions $\mu, \mu_1 \in \mathcal{P}(\mathcal{X})$, $\mu \otimes \mu_1$ represents the product distribution $(\mu \otimes \mu_1)(x) = \mu(x) \cdot \mu_1(x)$ and μ^k represents $\mu \otimes \cdots \otimes \mu$, k times. The *total variation* distance between distributions μ and λ is defined as $||\mu - \lambda||_t = \frac{1}{2} \sum_{x \in \mathcal{X}} |\mu(x) - \lambda(x)|$. The relative entropy between λ and μ is defined as $S(\lambda||\mu) = \sum_{x \in \mathcal{X}} \lambda(x) \log \frac{\lambda(x)}{\mu(x)}$. The relative min-entropy between λ and μ is defined as $S_\infty(\lambda||\mu) = \max_{x \in \mathcal{X}} \log \frac{\lambda(x)}{\mu(x)}$. It is easily seen that $S(\lambda||\mu) \leq S_\infty(\lambda||\mu)$. Let $XY \sim \mu$, here we assume that X is distributed over $\mathcal{X}$ and Y is distributed over $\mathcal{Y}$. We use μ_x and Y_x to represent $(Y|\ X = x)$. We use $\mu(x|y)$ to represent $\mu(x, y)/\mu(y)$ (when $\mu(y) \neq 0$). The conditional entropy of Y given X, is defined as $S(Y|X) = \mathbb{E}_{x\leftarrow X} S(Y_x)$. We use the following properties of relative entropy.

FACT 2.1.

(1) *Relative entropy is jointly convex in its arguments, that is for distributions* $\lambda_1, \lambda_2, \mu_1, \mu_2$ *and* $p \in [0, 1]$

$$S(p\lambda_1 + (1 - p)\lambda_2 \ || \ p\mu_1 + (1 - p)\mu_2) \leq p \cdot S(\lambda_1||\mu_1) + (1 - p) \cdot S(\lambda_2||\mu_2).$$

(2) *Let the random variables* XY, X^1Y^1 *be distributed over* $\mathcal{X} \times \mathcal{Y}$. *Relative entropy satisfies the following chain rule*

$$S(XY||X^1Y^1) = S(X||X^1) + \mathbb{E}_{x\leftarrow X} S\big(Y_x||Y^1_x\big).$$

This implies, using joint convexity of relative entropy

$$S(XY||X^1 \otimes Y^1) = S(X||X^1) + \mathbb{E}_{x\leftarrow X} S(Y_x||Y^1) \geq S(X||X^1) + S(Y||Y^1).$$

(3) *For distributions* $\lambda, \mu : S(\lambda||\mu) \geq 0$ *and* $||\lambda - \mu||_t \leq \sqrt{\frac{\ln 2}{2} S(\lambda||\mu)} \leq \sqrt{S(\lambda||\mu)}$. *The latter is referred to as the Pinsker's inequality.*

(4) *Substate theorem [Jain et al. 2002]: Let* λ, μ *be distributions. For every* $\delta > 0$, *there exists a distribution* λ_δ *such that* $S_\infty(\lambda_\delta||\mu) \leq O(\frac{1}{\delta}(S(\lambda||\mu) + 1))$ *and* $||\lambda_\delta - \lambda||_t \leq \delta$.

Let X, Y, Z be random variables. The mutual information between X and Y is defined as

$$I(X : Y) = S(X) + S(Y) - S(XY) = \mathbb{E}_{x\leftarrow X} S(Y_x||Y) = \mathbb{E}_{y\leftarrow Y} S(X_y||X).$$

The conditional mutual information is defined as

$$I(X : Y|\ Z) = \mathbb{E}_{z\leftarrow Z} I(X : Y|\ Z = z) = \mathbb{E}_{z\leftarrow Z}[S(X|Z = z) + S(Y|Z = z) - S(XY|Z = z)].$$

Random variables XYZ form a Markov chain $Z \leftrightarrow X \leftrightarrow Y$ iff $I(Y : Z|\ X = x) = 0$ for each x in the support of X. The following fact is often used in our proofs.

FACT 2.2. *Let* $\lambda, \mu \in \mathcal{P}(\mathcal{X})$ *be distributions. Then* $\sum_{x\in\mathcal{X}:\lambda(x)<\mu(x)} \lambda(x) \log \frac{\lambda(x)}{\mu(x)} \geq -1$.

One-Way Communication Complexity. Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation, where $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ are finite sets. We only consider complete relations that is for each $(x, y) \in \mathcal{X} \times \mathcal{Y}$, there exists at least one $z \in \mathcal{Z}$ such that $(x, y, z) \in f$. In the one-way model of communication, there is a single message, from Alice with input $x \in \mathcal{X}$ to Bob with input $y \in \mathcal{Y}$, at the end of which Bob is supposed to determine an answer z such that $(x, y, z) \in f$. Let $\varepsilon > 0$ and let $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ be a distribution. We let $\mathsf{D}^{1,\mu}_\varepsilon(f)$ represent the distributional one-way communication complexity of f under μ with expected error ϵ, that is, the (worst-case over all inputs) communication of the best deterministic one-way protocol for f, with distributional error (average error over the inputs) at most ε under μ. Let $\mathsf{R}^{1,\mathsf{pub}}_\epsilon(f)$ represent the one-way public-coin communication complexity of f with worst case error ε, that is, the (worst case over all inputs) communication of the best one-way public-coin protocol for f with error for each input (x, y) being at most ε. The following is a consequence of the min-max theorem in game theory [Kushilevitz and Nisan 1997, Theorem 3.20, page 36].

LEMMA 2.3 (YAO PRINCIPLE). $\mathsf{R}^{1,\mathsf{pub}}_\epsilon(f) = \max_\mu \mathsf{D}^{1,\mu}_\epsilon(f)$.

The following result follows quite directly from the arguments in Braverman and Rao [2011]. We provide its proof in Appendix A for completeness.

LEMMA 2.4 [BRAVERMAN AND RAO 2011]. *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation and $\varepsilon > 0, \delta \geq 0$. Let $XY \sim \mu$ be inputs to a one-way private-coin communication protocol $\mathcal{P}$ with distributional error at most ε. Let M represent the message of $\mathcal{P}$. Let θ be the joint distribution of XYM and let*

$$\Pr_{(x,y,i)\leftarrow\theta}\left[\log\frac{\theta(i|x)}{\theta(i|y)} > c\right] \leq \delta. \tag{2.1}$$

There exists a deterministic one-way protocol $\mathcal{P}_1$ for f, such that the communication of $\mathcal{P}_1$ is $c + O(\log(1/\delta))$, and distributional error of $\mathcal{P}_1$, with inputs distributed according to μ, is at most $\varepsilon + 2\delta$.

Two-Way Communication Complexity. Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a complete relation, where $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ are finite sets. In the two-way model of communication, Alice with input $x \in \mathcal{X}$ and Bob with input $y \in \mathcal{Y}$, communicate by exchanging messages with each other over several rounds, at the end of which they are supposed to determine a common answer z (as a function of the message transcript) such that $(x, y, z) \in f$. Let $\varepsilon > 0$ and let $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ be a distribution. We let $\mathsf{D}^{2,\mu}_{\varepsilon}(f)$ represent the two-way distributional communication complexity of f under μ with expected error ϵ, that is, the (worst case over all inputs) communication of the best deterministic two-way protocol for f, with distributional error (average error over the inputs) at most ε under μ. Let $\mathsf{R}^{2,\mathsf{pub}}_{\epsilon}(f)$ represent the two-way public-coin communication complexity of f with worst case error ε, that is, the (worst-case over all inputs) communication of the best two-way public-coin protocol for f with error for each input (x, y) being at most ε. The following is a consequence of the min-max theorem in game theory [Kushilevitz and Nisan 1997, Theorem 3.20, page 36].

LEMMA 2.5 (YAO PRINCIPLE). $\mathsf{R}^{2,\mathsf{pub}}_{\epsilon}(f) = \max_{\mu} \mathsf{D}^{2,\mu}_{\epsilon}(f)$.

3. ONE-WAY COMMUNICATION

Definitions. We make here the necessary definitions for this section. Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation, $\mu, \lambda \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ be distributions and $\varepsilon, \delta > 0$.

The following two definitions were made in Jain et al. [2008].

Definition 3.1 (*One-Way Distributions*). Distribution λ is called *one-way* for distribution μ if, for all (x, y) in the support of λ, we have $\mu(y|x) = \lambda(y|x)$.

Note that in a one-way protocol if the inputs are drawn from μ and we condition on a message transcript from Alice, then the resulting distribution would be one-way for μ.

Definition 3.2 (*Error of a Distribution*). Error of distribution μ with respect to f, denoted $\mathsf{err}_f(\mu)$, is defined as

$$\mathsf{err}_f(\mu) \stackrel{\text{def}}{=} \min\left\{\Pr_{(x,y)\leftarrow\mu}[(x, y, g(y)) \notin f] \mid g : \mathcal{Y} \to \mathcal{Z}\right\}.$$

Let μ be the distribution of inputs for Alice and Bob. Let Bob make an output depending on his input, without any communication from Alice. Then, $\mathsf{err}_f(\mu)$ represents the least error that Bob must make.

The following bound was defined in Jain et al. [2008] where it was referred to as the one-way subdistribution bound. We call it differently here for consistency of nomenclature with the new bound that we define subsequently.

Definition 3.3 (*Relative min-entropy Bound*). The ε-error relative min-entropy bound of f with respect to distribution μ, denoted $\mathsf{ment}^{\mu}_{\varepsilon}(f)$, is defined as

$$\mathsf{ment}^{\mu}_{\varepsilon}(f) \overset{\text{def}}{=} \min\left\{S_{\infty}(\lambda||\mu)|\ \lambda \text{ is one-way for } \mu \text{ and } \mathsf{err}_f(\lambda) \le \varepsilon\right\}.$$

The ε-error relative min-entropy bound of f, denoted $\mathsf{ment}(f)$, is defined as

$$\mathsf{ment}_{\varepsilon}(f) \overset{\text{def}}{=} \max\left\{\mathsf{ment}^{\mu}_{\varepsilon}(f)|\ \mu \text{ is a distribution over } \mathcal{X} \times \mathcal{Y}\right\}.$$

The following two are key new quantities we define in this article. As mentioned previously, while considering the direct product argument, we need to handle distributions which are close (in total variation distance) to distributions which are one-way for μ and have bounded conditional relative entropy with μ. This leads us to consider distributions which are one-way for μ, have potentially very high conditional relative entropy with μ, however for which the relevant ratios are bounded with high probability.

Definition 3.4 (*Robust Conditional Relative min-entropy*). The δ-robust conditional relative min-entropy of λ with respect to μ, denoted $\mathsf{rcment}^{\mu}_{\delta}(\lambda)$, is defined to be the minimum number c such that

$$\Pr_{(x,y)\leftarrow\lambda}\left[\log\frac{\lambda(x|y)}{\mu(x|y)} > c\right] \le \delta.$$

The following can be viewed as a "modified and smoothened" version of the bound in the Definition 3.3. We consider conditional relative min-entropy instead of relative min-entropy and smooth it by allowing a small violation of the conditional relative min-entropy condition.

Definition 3.5 (*Robust Conditional Relative min-entropy Bound*). The ε-error δ-robust conditional relative min-entropy bound of f with respect to distribution μ, denoted $\mathsf{rcment}^{\mu}_{\varepsilon,\delta}(f)$, is defined as

$$\mathsf{rcment}^{\mu}_{\varepsilon,\delta}(f) \overset{\text{def}}{=} \min\left\{\mathsf{rcment}^{\mu}_{\delta}(\lambda)|\ \lambda \text{ is one-way for } \mu \text{ and } \mathsf{err}_f(\lambda) \le \varepsilon\right\}.$$

The ε-error δ-robust conditional relative min-entropy bound of f, denoted $\mathsf{rcment}_{\epsilon,\delta}(f)$, is defined as

$$\mathsf{rcment}_{\varepsilon,\delta}(f) \overset{\text{def}}{=} \max\left\{\mathsf{rcment}^{\mu}_{\varepsilon,\delta}(f)|\ \mu \text{ is a distribution over } \mathcal{X} \times \mathcal{Y}\right\}.$$

We often use this definition in the following way. Let λ be a distribution which is one-way for μ and with $\mathsf{rcment}^{\mu}_{\delta}(\lambda) < \mathsf{rcment}^{\mu}_{\varepsilon,\delta}(f)$. Then, $\mathsf{err}_f(\lambda) > \varepsilon$.

Lemma 3.6. *For every λ, μ and $\delta > 0$ we have $\mathsf{rcment}^{\mu}_{\delta}(\lambda) \le (S_{\infty}(\lambda||\mu) + 1)/\delta$, hence $\mathsf{rcment}^{\mu}_{\varepsilon,\delta}(f) = O(\mathsf{ment}^{\mu}_{\varepsilon}(f))$ and $\mathsf{rcment}_{\varepsilon,\delta}(f) = O(\mathsf{ment}_{\varepsilon}(f))$.*

Proof. Let $XY \sim \lambda$ and $X'Y' \sim \mu$. Then, using Part 2 of Fact 2.1 we have

$$S_{\infty}(\lambda||\mu) \ge S(\lambda||\mu) \ge \mathbb{E}_{y\leftarrow Y} S(X_y||X'_y) = \mathbb{E}_{(x,y)\leftarrow\lambda}\log\frac{\lambda(x|y)}{\mu(x|y)}.$$

Therefore, using Fact 2.2 and Markov's inequality, we get $\Pr_{(x,y)\leftarrow\lambda}[\log\frac{\lambda(x|y)}{\mu(x|y)} > (S_{\infty}(\lambda||\mu)+1)/\delta] < \delta$. Hence, $\mathsf{rcment}^{\mu}_{\delta}(\lambda) \le (S_{\infty}(\lambda||\mu)+1)/\delta$. The other relationships follow from definitions. □

New Characterization. In this section, we show a new characterization of one-way public-coin communication complexity in terms of rcment. The following lemma which lower bounds distributional communication complexity using ment appears in Jain et al. [2008].

LEMMA 3.7. *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation and $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ be a distribution and $\varepsilon, k > 0$. Then,*

$$\mathsf{D}^{1,\mu}_{\epsilon(1-2^{-k})}(f) \quad \geq \quad \mathsf{ment}^{\mu}_{\varepsilon}(f) - k.$$

We show the following key lemma which upper bounds distributional communication complexity using rcment. Its proof is inspired by a similar result in Jain et al. [2008] which upper bounds distributional communication complexity using ment.

LEMMA 3.8. *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation and $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ be a distribution and $\varepsilon, \delta > 0$. Then,*

$$\mathsf{D}^{1,\mu}_{\varepsilon+4\delta}(f) \leq \mathsf{rcment}_{\varepsilon,\delta}(f) + O\left(\log\frac{1}{\delta}\right).$$

PROOF. We start with the following key claim where we produce a desired split of μ into several distributions which are one-way for μ. This will enable us to obtain a one-way protocol with small communication as we show later.

CLAIM 3.9. *There exists a natural number k and a Markov chain $M \leftrightarrow X \leftrightarrow Y$, where M is distributed over $[k]$ and $XY \sim \mu$, such that*

(1) *for each $i \in [k]$: $\mathsf{err}_f(P_i) \leq \varepsilon$, where $P_i = (XY \mid M = i)$ and*
(2) $\Pr_{(x,y,i)\leftarrow\theta}[\log\frac{\theta(i|x)}{\theta(i|y)} > \mathsf{rcment}_{\varepsilon,\delta}(f) + \log\frac{1}{\delta}] \leq 2\delta$*, where θ is the distribution of XYM.*

PROOF. Let $c = \mathsf{rcment}_{\varepsilon,\delta}(f)$. Let us perform a procedure as follows. Start with $i = 1$.

(1) Let us say we have collected distributions $P_1, \ldots, P_{i-1}$, each one-way for μ, and positive numbers $p_1, \ldots, p_{i-1}$ such that $\mu \geq \sum_{j=1}^{i-1} p_j P_j$. If $\mu = \sum_{j=1}^{i-1} p_j P_j$, then set $k = i - 1$ and stop.
(2) Otherwise, let us express $\mu = \sum_{j=1}^{i-1} p_j P_j + q_i Q_i$, where Q_i is a distribution, one-way for μ. Since $\mathsf{rcment}^{Q_i}_{\varepsilon,\delta}(f) \leq c$, we know that there is a distribution R, one-way for Q_i (hence, also one-way for μ), such that $\mathsf{rcment}^{Q_i}_{\delta}(R) \leq c$ and $\mathsf{err}_f(R) \leq \varepsilon$. Let $r = \max\{q \mid Q_i \geq qR\}$. Let $P_i = R$, $p_i = q_i * r$, $i = i + 1$ and go back to step (1).

It can be observed that for each new i, there is a new $x \in \mathcal{X}$ such that $Q_i(x) = 0$. Hence, this process converges after at most $|\mathcal{X}|$ iterations. At the end, we have $\mu = \sum_{i=1}^{k} p_i P_i$.

Let us define random variable $M \in [k]$ such that $\Pr[M = i] = p_i$. Let us define XY, distributed over $\mathcal{X} \times \mathcal{Y}$, and correlated with M such that $(XY \mid M = i) \sim P_i$. It is easily checked that $XY \sim \mu$. Also since each P_i is one-way for μ, XYM form a Markov chain $M \leftrightarrow X \leftrightarrow Y$. This shows Part (1) of Claim 3.9, and it remains to show Part (2). Let θ be the distribution of XYM. Let us define

(1) $B = \{(x, y, i) \mid \log\frac{P_i(x|y)}{\mu(x|y)} > c + \log\frac{1}{\delta}\}$,
(2) $B_1 = \{(x, y, i) \mid \log\frac{P_i(x|y)}{Q_i(x|y)} > c\}$,
(3) $B_2 = \{(x, y, i) \mid \frac{\mu(y)}{q_i Q_i(y)} > \frac{1}{\delta}\}$.

Since $q_i Q(x, y) \leq \mu(x, y)$,

$$\frac{P_i(x|y)}{\mu(x|y)} = \frac{P_i(x|y)}{Q_i(x|y)} \cdot \frac{Q_i(x|y)}{\mu(x|y)} = \frac{P_i(x|y)}{Q_i(x|y)} \cdot \frac{Q_i(x,y)\mu(y)}{Q_i(y)\mu(x,y)} \leq \frac{P_i(x|y)}{Q_i(x|y)} \cdot \frac{\mu(y)}{q_i Q_i(y)}.$$

Therefore, $B \subseteq B_1 \cup B_2$. Since for each i, $\mathsf{rcment}^{Q_i}_{\delta}(P_i) \leq c$, we have

$$\Pr_{(x,y,i)\leftarrow\theta}[(x, y, i) \in B_1] \leq \delta.$$

For a given y, let i_y be the smallest i such that $\frac{\mu(y)}{q_i Q_i(y)} > \frac{1}{\delta}$. Once $q_i Q_i$ is a δ factor smaller than μ at a sample point, it remains so. Therefore,

$$\Pr_{(x,y,i)\leftarrow\theta}[(x,y,i) \in B_2] = \sum_y q_{i_y} Q_{i_y}(y) < \sum_y \delta\mu(y) = \delta.$$

Hence, $\Pr_{(x,y,i)\leftarrow\theta}[(x,y,i) \in B] < 2\delta$. Finally we note that

$$\frac{P_i(x|y)}{\mu(x|y)} = \frac{\theta(x|(y,i))}{\theta(x|y)} = \frac{\theta(i|x)}{\theta(i|y)}. \quad \square$$

Now consider the following one-way private-coin protocol $\mathcal{P}_1$ for f with inputs drawn from distribution μ. In $\mathcal{P}_1$ Alice on input x generates i from the distribution $(M|\ X = x)$ and sends i to Bob. Note that from Part (1). of Claim 3.9, conditioned on Alice's message being i, the joint distribution of the inputs of Alice and Bob is P_i. Bob on input y and receiving message i, gives the best possible output assuming distribution of Alice's inputs being $X|(M = i, Y = y)$. Since for all i, $\mathsf{err}_f(P_i) \leq \varepsilon$, the distributional error of $\mathcal{P}_1$ is at most ε. Now, using Lemma 2.4, we get a deterministic protocol $\mathcal{P}_2$ for f, with distributional error at most $\varepsilon + 4\delta$ and communication at most $d = \mathsf{rcment}_{\varepsilon,\delta}(f) + O(\log\frac{1}{\delta})$.

We can now conclude our characterization.

THEOREM 3.10. *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation and $\varepsilon > 0$. Then,*

$$\mathsf{ment}_{2\varepsilon}(f) - 1 \leq \mathsf{R}^{1,\mathsf{pub}}_{\varepsilon}(f) \leq \mathsf{rcment}_{\varepsilon/5,\varepsilon/5}(f) + O\left(\log\frac{1}{\varepsilon}\right).$$

Hence, for constant ε,

$$\mathsf{R}^{1,\mathsf{pub}}_{\varepsilon}(f) = \Theta(\mathsf{ment}_{\varepsilon}(f)) = \Theta(\mathsf{rcment}_{\varepsilon,\varepsilon}(f)).$$

PROOF. The first inequality follows from Lemma 3.7 (set $k = 1$) and maximizing both sides over all distributions μ and using Lemma 2.3 (Yao principle). The second inequality follows from Lemma 3.8 (set $\delta \leftarrow \varepsilon$) and maximizing both sides over all distributions μ and using Lemma 2.3. The other relations now follow from Lemma 3.6 and from the fact that the error in public-coin randomized one-way communication complexity can be decreased by a constant factor by increasing the communication by a constant factor. □

Strong Direct Product. In this section, we show our strong direct product theorem for one-way public-coin communication complexity. We start with the following key theorem.

THEOREM 3.11 (DIRECT PRODUCT IN TERMS OF ment AND rcment). *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation and $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ be a distribution. Let $0 < 4\sqrt{80\delta} < \varepsilon < 0.5$ be constants, k be a natural number and $\mathsf{rcment}^{\mu}_{\varepsilon,\varepsilon}(f) \geq 4/\delta$. Then,*

$$\mathsf{ment}^{\mu^k}_{1-(1-\varepsilon/2)^{\lfloor \delta k \rfloor}}(f^k) \geq \delta \cdot k \cdot \mathsf{rcment}^{\mu}_{\varepsilon,\varepsilon}(f).$$

Proof. Let $c = \mathsf{rcment}^{\mu}_{\varepsilon,\varepsilon}(f)$. Let $\lambda \in \mathcal{P}(\mathcal{X}^k \times \mathcal{Y}^k)$ be a distribution which is one-way for μ^k and with $S_\infty(\lambda||\mu^k) < \delta c k$. We show that $\mathsf{err}_{f^k}(\lambda) > 1 - (1-\varepsilon/2)^{\lfloor \delta k \rfloor}$. This shows the desired.

Let B be a set. For $\sigma \in B^k$, let σ_i be the ith component of σ; for a set $C \subseteq [k]$, let σ_C denote $< \sigma_i : i \in C >$, and for $i \in [k]$, let σ_{-i} denote $\sigma_{[k]-\{i\}}$. For joint random variables MN, we let M_n to represent $(M|\ N = n)$ and also $(MN|\ N = n)$ (which of these we mean will be clear from the context).

Let $XY \sim \lambda$. Let us fix $g : \mathcal{Y}^k \to \mathcal{Z}^k$. For a coordinate i, let the binary random variable $T_i \in \{0, 1\}$, correlated with XY, denote success in the ith coordinate. That is, $T_i = 1$ iff $(X_i, Y_i, g(Y)_i) \in f$. Using Claim 3.12, we get a sequence of coordinates $i_1, \ldots, i_{k'}$ such that,

$$\Pr[T_1 \times T_2 \times \cdots \times T_k = 1] \leq \Pr[T_{i_1} \times T_{i_2} \times \cdots \times T_{i_{k'}} = 1] < (1-\varepsilon/2)^{k'}.$$

Claim 3.12. *Let $k' = \lfloor \delta k \rfloor$. There exist k' distinct coordinates $i_1, \ldots, i_{k'}$ such that $\Pr[T_{i_1} = 1] \leq 1 - \varepsilon/2$ and for each $r < k'$,*

(1) *either* $\Pr[T_{i_1} \times T_{i_2} \times \cdots \times T_{i_r} = 1] < (1-\varepsilon/2)^{k'}$,
(2) *or* $\Pr[T_{i_{r+1}} = 1|\ (T_{i_1} \times T_{i_2} \times \cdots \times T_{i_r} = 1)] < 1 - \varepsilon/2$.

Proof. Let us say we have identified $r < k'$ coordinates $i_1, \ldots i_r$. Let $C = \{i_1, i_2, \cdots, i_r\}$. Let $T = T_{i_1} \times T_{i_2} \times \cdots \times T_{i_r}$. If $\Pr[T = 1] < (1-\varepsilon/2)^{k'}$, then we will be done. So assume that $\Pr[T = 1] \geq (1-\varepsilon/2)^{k'} > 2^{-\delta k}$.

Let $X'Y' \sim \mu$. Let $X^1Y^1 = XY|(T = 1)$. Note that $S_\infty(X^1Y^1||XY) \leq \log \frac{1}{\Pr[T=1]} \leq \delta k$. Let D be uniformly distributed in $\{0, 1\}^k$ and independent of X^1Y^1 (and $X'Y'$). Let $U_i = X^1_i$ if $D_i = 0$ and $U_i = Y^1_i$ if $D_i = 1$. Let $U = U_1 \cdots U_k$. Here for any random variable $\tilde{X}\tilde{Y}$, we let $\tilde{X}\tilde{Y}_{d,u}$ represent the random variable obtained by following conditioning on $\tilde{X}\tilde{Y}$: for all i, $\tilde{X}_i = u_i$ if $d_i = 0$, otherwise $\tilde{Y}_i = u_i$ if $d_i = 1$. The following calculations help us identify a coordinate i outside C in which the marginal distribution, even conditioned on success on C, is close to μ in terms of the measure we consider, namely the robust conditional min-entropy bound. This helps us argue that the success in that coordinate, even conditioned on success on C, is bounded away from 1. Consider,

$$\begin{aligned}
&\delta k + \delta c k \\
&\quad > S_\infty(X^1Y^1||XY) + S_\infty(XY||(X'Y')^{\otimes k}) \\
&\quad \geq S_\infty(X^1Y^1||(X'Y')^{\otimes k}) \geq S(X^1Y^1||(X'Y')^{\otimes k}) \\
&\quad \geq \mathbb{E}_{(d,u,x_C,y_C)\leftarrow(DUX^1_CY^1_C)} S((X^1Y^1)_{d,u,x_C,y_C}||((X'Y')^{\otimes k})_{d,u,x_C,y_C}) \quad \text{(from Part (2) of Fact 2.1)} \\
&\quad \geq \mathbb{E}_{(d,u,x_C,y_C)\leftarrow(DUX^1_CY^1_C)} S(X^1_{d,u,x_C,y_C}||X'_{d_1,u_1,x_C,y_C} \otimes \ldots \otimes X'_{d_k,u_k,x_C,y_C}) \quad \text{(from Part (2) of Fact 2.1)} \\
&\quad \geq \mathbb{E}_{(d,u,x_C,y_C)\leftarrow(DUX^1_CY^1_C)} \sum_{i \notin C} S\big((X^1_{d,u,x_C,y_C})_i||X'_{d_i,u_i}\big) \quad \text{(from Part (2) of Fact 2.1)} \\
&\quad = \sum_{i \notin C} \mathbb{E}_{(d,u,x_C,y_C)\leftarrow(DUX^1_CY^1_C)} S\big((X^1_{d,u,x_C,y_C})_i||X'_{d_i,u_i}\big). \qquad (3.1)
\end{aligned}$$

Also

$$\begin{aligned}
\delta k &> S_\infty(X^1Y^1||XY) \geq S(X^1Y^1||XY) && (3.2)\\
&\geq \mathbb{E}_{(d,u,x_C,y_C)\leftarrow(DUX^1_CY^1_C)} S\big((X^1Y^1)_{d,u,x_C,y_C}||(XY)_{d,u,x_C,y_C}\big) \quad \text{(from Part (2) of Fact 2.1)}\\
&\geq \mathbb{E}_{(d,u,x_C,y_C)\leftarrow(DUX^1_CY^1_C)} S\big(Y^1_{d,u,x_C,y_C} \;||\; Y_{d_1,u_1,x_C,y_C} \otimes \cdots \otimes Y_{d_k,u_k,x_C,y_C}\big)\\
&\qquad \text{(from Part (2) of Fact 2.1 and using the fact that } XY \text{ is one-way for } \mu^{\otimes k})\\
&\geq \mathbb{E}_{(d,u,x_C,y_C)\leftarrow(DUX^1_CY^1_C)} \sum_{i\notin C} S\big(\big(Y^1_{d,u,x_C,y_C}\big)_i||Y_{d_i,u_i}\big) \quad \text{(from Part (2) of Fact 2.1)}\\
&= \sum_{i\notin C} \mathbb{E}_{(d,u,x_C,y_C)\leftarrow(DUX^1_CY^1_C)} S\big(\big(Y^1_{d,u,x_C,y_C}\big)_i||Y'_{d_i,u_i}\big). && (3.3)
\end{aligned}$$

From Eq. (3.1) and Eq. (3.3) and using Markov's inequality we get a coordinate j outside of C such that

(1) $\mathbb{E}_{(d,u,x_C,y_C)\leftarrow(DUX^1_CY^1_C)} S((X^1_{d,u,x_C,y_C})_j||X'_{d_j,u_j}) \leq \frac{2\delta(c+1)}{(1-\delta)} \leq 4\delta c$, and
(2) $\mathbb{E}_{(d,u,x_C,y_C)\leftarrow(DUX^1_CY^1_C)} S((Y^1_{d,u,x_C,y_C})_j||Y'_{d_j,u_j}) \leq \frac{2\delta}{(1-\delta)} \leq 4\delta$.

Therefore,

$$\begin{aligned}
4\delta c &\geq \mathbb{E}_{(d,u,x_C,y_C)\leftarrow(DUX^1_CY^1_C)} S\big(\big(X^1_{d,u,x_C,y_C}\big)_j||X'_{d_j,u_j}\big)\\
&= \mathbb{E}_{(d_{-j},u_{-j},x_C,y_C)\leftarrow(D_{-j}U_{-j}X^1_CY^1_C)} \mathbb{E}_{(d_j,u_j)\leftarrow(D_jU_j)|\ (D_{-j}U_{-j}X^1_CY^1_C)=(d_{-j},u_{-j},x_C,y_C)}\\
&\qquad S\big(\big(X^1_{d,u,x_C,y_C}\big)_j||X'_{d_j,u_j}\big).
\end{aligned}$$

And,

$$\begin{aligned}
4\delta &\geq \mathbb{E}_{(d,u,x_C,y_C)\leftarrow(DUX^1_CY^1_C)} S\big(\big(Y^1_{d,u,x_C,y_C}\big)_j||Y'_{d_j,u_j}\big)\\
&= \mathbb{E}_{(d_{-j},u_{-j},x_C,y_C)\leftarrow(D_{-j}U_{-j}X^1_CY^1_C)} \mathbb{E}_{(d_j,u_j)\leftarrow(D_jU_j)|\ (D_{-j}U_{-j}X^1_CY^1_C)=(d_{-j},u_{-j},x_C,y_C)} S\big(\big(Y^1_{d,u,x_C,y_C}\big)_j||Y'_{d_j,u_j}\big).
\end{aligned}$$

Now using Markov's inequality, there exists set G_1 with $\Pr[D_{-j}U_{-j}X^1_CY^1_C \in G_1] \geq 1 - 0.2$, such that for all $(d_{-j}, u_{-j}, x_C, y_C) \in G_1$,

(3) $\mathbb{E}_{(d_j,u_j)\leftarrow(D_jU_j)|\ (D_{-j}U_{-j}X^1_CY^1_C)=(d_{-j},u_{-j},x_C,y_C)} S((X^1_{d,u,x_C,y_C})_j||X'_{d_j,u_j}) \leq 40\delta c$, and
(4) $\mathbb{E}_{(d_j,u_j)\leftarrow(D_jU_j)|\ (D_{-j}U_{-j}X^1_CY^1_C)=(d_{-j},u_{-j},x_C,y_C)} S((Y^1_{d,u,x_C,y_C})_j||Y'_{d_j,u_j}) \leq 40\delta$.

Fix $(d_{-j}, u_{-j}, x_C, y_C) \in G_1$. Conditioning on $D_j = 1$ (which happens with probability 1/2) in inequality (3), we get,

$$\mathbb{E}_{y_j\leftarrow Y^1_j|(D_{-j}U_{-j}X^1_CY^1_C)=(d_{-j},u_{-j},x_C,y_C)} S\big(\big(X^1_{d_{-j},u_{-j},y_j,x_C,y_C}\big)_j||X'_{y_j}\big) \leq 80\delta c. \tag{3.4}$$

Conditioning on $D_j = 0$ (which happens with probability 1/2) in inequality (4), we get

$$\mathbb{E}_{x_j\leftarrow X^1_j|(D_{-j}U_{-j}X^1_CY^1_C)=(d_{-j},u_{-j},x_C,y_C)} S\big(\big(Y^1_{d_{-j},u_{-j},x_j,x_C,y_C}\big)_j||Y'_{x_j}\big) \leq 80\delta.$$

Using Part (3) of Fact 2.1 and concavity of square root, we get

$$\mathbb{E}_{x_j\leftarrow X^1_j|(D_{-j}U_{-j}X^1_CY^1_C)=(d_{-j},u_{-j},x_C,y_C)} ||\big(Y^1_{d_{-j},u_{-j},x_j,x_C,y_C}\big)_j - Y'_{x_j}||_t \leq \sqrt{80\delta}. \tag{3.5}$$

Let X^2Y^2 be such that $X^2 \sim (X^1_{d_{-j},u_{-j},x_C,y_C})_j$ and $(Y^2|\ X^2 = x_j) \sim Y'_{x_j}$. From Eq. (3.5), we get

$$||X^2Y^2 - \big((X^1Y^1)_{d_{-j},u_{-j},x_C,y_C}\big)_j||_t \leq \sqrt{80\delta}. \tag{3.6}$$

Note that the distribution $((X^1Y^1)_{d_{-j},u_{-j},x_C,y_C})_j$ is not necessarily one-way for μ. However, by construction the distribution of X^2Y^2 is one-way for μ. The following claim helps us argue that $\mathsf{rcment}^{\mu}_{\varepsilon}(X^2Y^2) < c$. Its proof is deferred.

Claim 3.13.

$$\Pr_{(x,y)\leftarrow X^2Y^2}\left[\log\frac{X^2Y^2(x|y)}{\mu(x|y)} > c\right] \le 200\delta + \sqrt{80\delta} < \varepsilon.$$

This implies, $\mathsf{err}_f(X^2Y^2) \ge \varepsilon$ and therefore using Eq. (3.6) (and the assumption regarding ε, δ),

$$\mathsf{err}_f\big(((X^1Y^1)_{d_{-j},u_{-j},x_C,y_C})_j\big) \ge \varepsilon - \sqrt{80\delta} \ge \frac{3\varepsilon}{4}.$$

Note that conditioned on $(Y^1_{d_{-j},u_{-j},x_C,y_C})_j$, the distribution $(X^1Y^1)_{d_{-j},u_{-j},x_C,y_C}$ is product across the $\mathcal{X}^k$ and $\mathcal{Y}^k$ parts (this is because of the conditioning on D). Due to this, Bob's inputs in coordinates other than j do not matter as far as the correctness of Bob's answer in the jth coordinate is concerned. Therefore,

$$\Pr[T_j = 1|\ (1, d_{-j}, u_{-j}, x_C, y_C) = (TD_{-j}U_{-j}X_CY_C)] \le 1 - \mathsf{err}_f\big(((X^1Y^1)_{d_{-j},u_{-j},x_C,y_C})_j\big).$$

Therefore, overall

$$\Pr[T_j = 1|\ (T = 1)] \le 0.8\left(1 - \frac{3\varepsilon}{4}\right) + 0.2 < 1 - \varepsilon/2. \quad \square$$

We return to the proof of Claim 3.13.

Proof of Claim 3.13. The broad argument used in the proof is as follows. Let $X^3Y^3 = ((X^1Y^1)_{d_{-j},u_{-j},x_C,y_C})_j$. From Eq. (3.4) and Markov's inequality, one can infer that the desired statement is (roughly) true when X^2Y^2 is replaced by X^3Y^3. Since the distributions X^2Y^2 and X^3Y^3 are close in total variation distance (from Eq. (3.5)) we can conclude the desired statement. The details follow.

From Eq. (3.4), we have,

$$80\delta c \ge \mathbb{E}_{y\leftarrow Y^3} S\big(X^3_y || X'_y\big) = \mathbb{E}_{(x,y)\leftarrow X^3Y^3}\log\frac{X^3_y(x)}{\mu(x|y)}.$$

Using Fact 2.2 and Markov's inequality on Claim 3.13, we get

$$161\delta \ge \frac{80\delta c + 1}{c/2 + 1/\delta} \ge \Pr_{(x,y)\leftarrow X^3Y^3}\left[\log\frac{X^3_y(x)}{\mu(x|y)} > \frac{c}{2} + \frac{1}{\delta}\right]. \tag{3.7}$$

Now assume for contradiction that

$$200\delta + \sqrt{80\delta} < \Pr_{(x,y)\leftarrow X^2Y^2}\left[\log\frac{X^2_y(x)}{\mu(x|y)} > c\right].$$

Using Eq. (3.6), we get

$$200\delta < \Pr_{(x,y)\leftarrow X^3Y^3}\left[\log\frac{X^2_y(x)}{\mu(x|y)} > c\right]. \tag{3.8}$$

Using Eq. (3.7), Eq. (3.8), and since $c \geq 4/\delta$, we get

$$39\delta < \Pr_{(x,y)\leftarrow X^3Y^3}\left[\log\frac{X^2_y(x)}{X^3_y(x)} > \frac{c}{2} - \frac{1}{\delta} \geq \frac{1}{\delta}\right].$$

Therefore, there exists a y such that

$$39\delta < \Pr_{x\leftarrow X^3_y}\left[\log\frac{X^2_y(x)}{X^3_y(x)} > \frac{1}{\delta}\right].$$

But this is not possible since both X^2_y and X^3_y are probability distributions. □

We now ready to state and prove the main result of this section.

THEOREM 3.14 (STRONG DIRECT PRODUCT FOR ONE-WAY PUBLIC-COIN COMMUNICATION COMPLEXITY). *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. Let ε, δ be constants such that $0 < 4\sqrt{80\delta} < \varepsilon/5 < 0.5$ and k be a natural number. Let $\delta' = (1-\varepsilon/10)^{\lfloor\delta k\rfloor} + 2^{-k}$. Then,*

$$\mathsf{R}^{1,\mathsf{pub}}_{1-\delta'}(f^k) = \Omega\big(k\cdot(\delta\cdot\mathsf{R}^{1,\mathsf{pub}}_{\varepsilon}(f) - O(1))\big).$$

In other words,

$$\mathsf{R}^{1,\mathsf{pub}}_{1-2^{-\Omega(\varepsilon^3 k)}}(f^k) = \Omega\big(k\cdot\big(\varepsilon^2\cdot\mathsf{R}^{1,\mathsf{pub}}_{\varepsilon}(f) - O(1)\big)\big).$$

PROOF. Let μ_1 be a distribution such that $\mathsf{D}^{1,\mu_1}_{\varepsilon}(f) = \mathsf{R}^{1,\mathsf{pub}}_{\varepsilon}(f)$. Let μ be a distribution such that $\mathsf{rcment}^{\mu}_{\varepsilon/5,\varepsilon/5}(f) = \mathsf{rcment}_{\varepsilon/5,\varepsilon/5}(f)$. Then,

$$\begin{aligned}
\delta\cdot k\cdot\mathsf{R}^{1,\mathsf{pub}}_{\varepsilon}(f) &= \delta\cdot k\cdot\mathsf{D}^{1,\mu_1}_{\varepsilon}(f)\\
&= O(\delta\cdot k\cdot\mathsf{rcment}_{\varepsilon/5,\varepsilon/5}(f)) \quad \text{(from Lemma 3.8)}\\
&= O\big(\delta\cdot k\cdot\mathsf{rcment}^{\mu}_{\varepsilon/5,\varepsilon/5}(f)\big)\\
&= O\big(\mathsf{ment}^{\mu^k}_{1-(1-\varepsilon/10)^{\lfloor\delta k\rfloor}}(f^k) + k\big) \quad \text{(from Theorem 3.11)}\\
&= O\big(\mathsf{D}^{1,\mu^k}_{1-(1-\varepsilon/10)^{\lfloor\delta k\rfloor}-2^{-k}}(f^k) + k\big) \quad \text{(from Lemma 3.7)}\\
&= O\big(\mathsf{R}^{1,\mathsf{pub}}_{1-\delta'}(f^k) + k\big).
\end{aligned}$$

Hence, $\mathsf{R}^{1,\mathsf{pub}}_{1-\delta'}(f^k) = \Omega(k\cdot(\delta\cdot\mathsf{R}^{1,\mathsf{pub}}_{\varepsilon}(f) - O(1)))$. □

4. TWO-WAY COMMUNICATION

In this section, we discuss the two-way public-coin model. We begin with the necessary definitions.

Definitions. Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation, $\mu, \lambda \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ be distributions and $\varepsilon > 0$. Let $XY \sim \mu$ and $X_1Y_1 \sim \lambda$ be random variables. Let $S \subseteq \mathcal{Z}$. We abuse the notation and define $f(x,y) \stackrel{\text{def}}{=} \{z \in \mathcal{Z} \mid (x,y,z) \in f\}$.

Definition 4.1 (*Error of a Distribution*). Error of distribution μ with respect to f and answer in S, denoted $\mathsf{err}_{f,S}(\mu)$, is defined as

$$\mathsf{err}_{f,S}(\mu) \stackrel{\text{def}}{=} \min_{z\in S}\left\{\Pr_{(x,y)\leftarrow\mu}[(x,y,z) \notin f]\right\}.$$

Let μ be the distribution of the inputs of Alice and Bob conditioned on a message transcript in a two-way deterministic protocol. Then, if Alice and Bob give an answer in S, they make error on at least $\mathsf{err}_{f,S}(\mu)$ fraction of the inputs.

Definition 4.2 (*Essentialness of an Answer Subset*). Essentialness of answer in S for f with respect to distribution μ, denoted $\mathsf{ess}^{\mu}(f, S)$, is defined as

$$\mathsf{ess}^{\mu}(f, S) \stackrel{\text{def}}{=} \Pr_{(x,y)\leftarrow\mu}[f(x, y) \subseteq S].$$

$\mathsf{ess}^{\mu}(f, S)$ represents the fraction of inputs according to μ for which every correct answer must lie in S. For example, $\mathsf{ess}^{\mu}(f, \mathcal{Z}) = 1$. Our direct product result is eventually stated in terms of the conditional relative entropy bound (which we define subsequently) when the answers lie in S and the essentialness of S.

Here, we define one-way distributions again since we need to define them with respect to both $\mathcal{X}$ and $\mathcal{Y}$. The following two definitions were made in Jain et al. [2008].

Definition 4.3 (*One-Way Distributions*). λ is called one-way for μ with respect to $\mathcal{X}$, if for all (x, y) in the support of λ we have $\mu(y|x) = \lambda(y|x)$. Similarly, λ is called one-way for μ with respect to $\mathcal{Y}$, if for all (x, y) in the support of λ, we have $\mu(x|y) = \lambda(x|y)$.

Definition 4.4 (*SM-Like*). λ is called SM-like (simultaneous-message-like) for μ, if there is a distribution θ on $\mathcal{X} \times \mathcal{Y}$ such that θ is one-way for μ with respect to $\mathcal{X}$ and λ is one-way for θ with respect to $\mathcal{Y}$.

Let the inputs of Alice and Bob be distributed according to μ. Then note that conditioned on any message transcript in a two-way deterministic protocol, the resulting distribution on the inputs will be SM-like for μ.

The following two definitions are new to this work. Again, as in the case of crent, these definitions are helpful in dealing with nonproduct distributions.

Definition 4.5 (*Conditional Relative Entropy*). The $\mathcal{Y}$-conditional relative entropy of λ with respect to μ, denoted $\mathsf{crent}^{\mu}_{\mathcal{Y}}(\lambda)$, is defined as

$$\mathsf{crent}^{\mu}_{\mathcal{Y}}(\lambda) \stackrel{\text{def}}{=} \mathbb{E}_{y\leftarrow Y_1} S((X_1)_y || X_y).$$

Similarly, the $\mathcal{X}$-conditional relative entropy of λ with respect to μ, denoted $\mathsf{crent}^{\mu}_{\mathcal{X}}(\lambda)$, is defined as

$$\mathsf{crent}^{\mu}_{\mathcal{X}}(\lambda) \stackrel{\text{def}}{=} \mathbb{E}_{x\leftarrow X_1} S((Y_1)_x || Y_x).$$

Definition 4.6 (*Conditional Relative Entropy Bound*). The two-way ε-error conditional relative entropy bound of f with answer in S with respect to distribution μ, denoted $\mathsf{crent}^{2,\mu}_{\varepsilon}(f, S)$, is defined as

$$\mathsf{crent}^{2,\mu}_{\varepsilon}(f, S) \stackrel{\text{def}}{=} \min\left\{\mathsf{crent}^{\mu}_{\mathcal{X}}(\lambda) + \mathsf{crent}^{\mu}_{\mathcal{Y}}(\lambda) \mid \lambda \text{ is SM-like for } \mu \text{ and } \mathsf{err}_{f,S}(\lambda) \le \varepsilon\right\}.$$

We use this definition as follows. Let λ be SM-like for μ and $\mathsf{crent}^{\mu}_{\mathcal{X}}(\lambda) + \mathsf{crent}^{\mu}_{\mathcal{Y}}(\lambda) < \mathsf{crent}^{2,\mu}_{\varepsilon}(f, S)$. Then $\mathsf{err}_{f,S}(\lambda) > \varepsilon$.

The following bound is analogous to a bound defined in Jain et al. [2008] where it was referred to as the two-way subdistribution bound. We call it differently here for consistency of nomenclature with the other bounds. Jain et al. [2008] typically considered the cases where $S = \mathcal{Z}$ or S is a singleton set.

Definition 4.7 (*Relative min entropy Bound*). The two-way ε-error relative min entropy bound of f with answer in S with respect to distribution μ, denoted $\mathsf{ment}^{2,\mu}_{\varepsilon}(f, S)$, is defined as

$$\mathsf{ment}^{2,\mu}_{\varepsilon}(f, S) \stackrel{\text{def}}{=} \min\left\{S_{\infty}(\lambda||\mu) | \ \lambda \text{ is SM-like for } \mu \text{ and } \mathsf{err}_{f,S}(\lambda) \le \varepsilon\right\}.$$

The following is easily seen from definitions and Part (2) of Fact 2.1.

LEMMA 4.8.

$$\mathsf{crent}^{\mu}_{\mathcal{X}}(\lambda) + \mathsf{crent}^{\mu}_{\mathcal{Y}}(\lambda) \leq 2 \cdot S_\infty(\lambda||\mu) \quad \textit{and} \quad \mathsf{crent}^{2,\mu}_{\varepsilon}(f, S) \leq 2 \cdot \mathsf{ment}^{2,\mu}_{\varepsilon}(f, S).$$

The following lemma states that, when μ is a product distribution, then ment is upper bounded by crent.

LEMMA 4.9. *Let μ be a product distribution across $\mathcal{X}$ and $\mathcal{Y}$. Then,*

$$\mathsf{ment}^{2,\mu}_{\varepsilon}(f, S) = O\left(\frac{1}{\varepsilon}(\mathsf{crent}^{2,\mu}_{\varepsilon/2}(f, S) + 1)\right).$$

PROOF. Let $\mu = \mu_1 \otimes \mu_2$, where $\mu_1 \in \mathcal{X}$ and $\mu_2 \in \mathcal{Y}$. Let λ be a distribution such that $\mathsf{crent}^{\mu}_{\mathcal{X}}(\lambda) + \mathsf{crent}^{\mu}_{\mathcal{Y}}(\lambda) = \mathsf{crent}^{2,\mu}_{\varepsilon/2}(f, S)$; λ is SM-like for μ and $\mathsf{err}_{f,S}(\lambda) \leq \varepsilon/2$. Since λ is SM-like for μ, it is easily verified that λ is also a product distribution across $\mathcal{X}$ and $\mathcal{Y}$. Let $\lambda = \lambda_1 \otimes \lambda_2$, where $\lambda_1 \in \mathcal{X}$ and $\lambda_2 \in \mathcal{Y}$. Using Part (4), of Fact 2.1, we get that there exists distributions λ'_1 and λ'_2 such that $S_\infty(\lambda'_1||\mu_1) \leq O(\frac{1}{\varepsilon}(S(\lambda_1||\mu_1) + 1))$, $S_\infty(\lambda'_2||\mu_2) \leq O(\frac{1}{\varepsilon}(S(\lambda_2||\mu_2)+1))$, $||\lambda'_1 - \lambda_1||_t \leq \frac{\varepsilon}{4}$ and $||\lambda'_2 - \lambda_2||_t \leq \frac{\varepsilon}{4}$. Let $\lambda' = \lambda'_1 \otimes \lambda'_2$. Note that λ' is SM-like for μ. Since $||\lambda' - \lambda||_t \leq \frac{\varepsilon}{2}$ and $\mathsf{err}_{f,S}(\lambda) \leq \varepsilon/2$, we have $\mathsf{err}_{f,S}(\lambda') \leq \varepsilon$. Now,

$$\begin{aligned} S_\infty(\lambda'||\mu) = S_\infty(\lambda'_1||\mu_1) + S_\infty(\lambda'_2||\mu_2) &= O\left(\frac{1}{\varepsilon}(S(\lambda_1||\mu_1) + S(\lambda_2||\mu_2) + 1)\right) \\ &= O\left(\frac{1}{\varepsilon}(\mathsf{crent}^{\mu}_{\mathcal{X}}(\lambda) + \mathsf{crent}^{\mu}_{\mathcal{Y}}(\lambda) + 1)\right) \quad = O\left(\frac{1}{\varepsilon}(\mathsf{crent}^{2,\mu}_{\varepsilon/2}(f, S) + 1)\right). \end{aligned}$$

Therefore, $\mathsf{ment}^{2,\mu}_{\varepsilon}(f, S) = O(\frac{1}{\varepsilon}(\mathsf{crent}^{2,\mu}_{\varepsilon/2}(f, S) + 1))$. □

4.1. Strong Direct Product

We start with the following theorem.

THEOREM 4.10 (DIRECT PRODUCT IN TERMS OF ment AND crent). *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation, $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ be a distribution and $S \subseteq \mathcal{Z}$. Let $0 < \varepsilon < 1/3$, $0 < 200\delta < 1$ and k be a natural number. Fix $z \in \mathcal{Z}^k$. Let the number of indices $i \in [k]$ with $z_i \in S$ be at least $\delta_1 k$. Assume $\mathsf{crent}^{2,\mu}_{\varepsilon}(f, S) > 2$. Then*

$$\mathsf{ment}^{2,\mu^k}_{1-(1-\varepsilon/2)^{\lfloor \delta\delta_1 k \rfloor}}(f^k, \{z\}) \geq \delta \cdot \delta_1 \cdot k \cdot \mathsf{crent}^{2,\mu}_{\varepsilon}(f, S).$$

PROOF. Let $c = \mathsf{crent}^{2,\mu}_{\varepsilon}(f, S)$. Let $\lambda \in \mathcal{X}^k \times \mathcal{Y}^k$ be a distribution which is SM-like for μ^k and with $S_\infty(\lambda||\mu^k) < \delta\delta_1 ck$. We show that $\mathsf{err}_{f^k,\{z\}}(\lambda) \geq 1 - (1 - \varepsilon/2)^{\lfloor \delta\delta_1 k \rfloor}$. This shows the desired.

We use similar notations as in the previous section. Let $XY \sim \lambda$. For a coordinate i, let the binary random variable $T_i \in \{0, 1\}$, correlated with XY, denote success in the ith coordinate. That is, $T_i = 1$ iff $XY = (x, y)$ such that $(x_i, y_i, z_i) \in f$. Using Claim 4.11, we get a sequence of coordinates $i_1, \ldots, i_{k'}$ such that,

$$\Pr[T_1 \times T_2 \times \cdots \times T_k = 1] \leq \Pr[T_{i_1} \times T_{i_2} \times \cdots \times T_{i_{k'}} = 1] \leq (1 - \varepsilon/2)^{k'}.$$

CLAIM 4.11. *Let $k' = \lfloor \delta\delta_1 k \rfloor$. There exists k' distinct coordinates $i_1, \ldots, i_{k'}$ such that* $\Pr[T_{i_1} = 1] \leq 1 - \varepsilon/2$ *and for each $r < k'$,*

(1) *either* $\Pr[T_{i_1} \times T_{i_2} \times \cdots \times T_{i_r} = 1] \leq (1 - \varepsilon/2)^{k'}$,
(2) *or* $\Pr[T_{i_{r+1}} = 1 | \,(T_{i_1} \times T_{i_2} \times \cdots \times T_{i_r} = 1)] \leq 1 - \varepsilon/2$.

PROOF. Let us say we have identified $r < k'$ coordinates $i_1, \ldots i_r$. Let $C = \{i_1, i_2, \ldots, i_r\}$. Let $T = T_1 \times T_2 \times \cdots \times T_r$. If $\Pr[T = 1] \leq (1 - \varepsilon/2)^{k'}$, then we will be done. So assume that $\Pr[T = 1] > (1 - \varepsilon/2)^{k'} \geq 2^{-\delta\delta_1 k}$. Let $X'Y' \sim \mu$. Let $X^1Y^1 = XY|(T = 1)$. Let D be uniformly distributed in $\{0, 1\}^k$ and independent of X^1Y^1. Let $U_i = X^1_i$ if $D_i = 0$ and $U_i = Y^1_i$ if $D_i = 1$. Let $U = U_1 \cdots U_k$. Here, for any random variable $\tilde{X}\tilde{Y}$, we let $\tilde{X}\tilde{Y}_{d,u}$, represent the random variable obtained by following conditioning on $\tilde{X}\tilde{Y}$: for all i, $\tilde{X}_i = u_i$ if $d_i = 0$, otherwise $\tilde{Y}_i = u_i$ if $d = 1$. Let I be the set of indices i such that $z_i \in S$.

The following are similar to the calculations performed in the one-way case. They help us identify a coordinate i outside C in which the marginal distribution, even conditioned on success on C, is close to μ, this time in terms of the conditional min-entropy bound. This helps us argue that the success in that coordinate, even conditioned on success on C, is bounded away from 1. Consider,

$$\begin{aligned}
&\delta\delta_1 k + \delta\delta_1 ck \\
&> S_\infty(X^1Y^1 || XY) + S_\infty(XY || (X'Y')^{\otimes k}) \\
&\geq S_\infty(X^1Y^1 || (X'Y')^{\otimes k}) \geq S(X^1Y^1 || (X'Y')^{\otimes k}) \\
&\geq \mathbb{E}_{(d,u,x_C,y_C) \leftarrow (DUX^1_C Y^1_C)} S((X^1Y^1)_{d,u,x_C,y_C} || ((X'Y')^{\otimes k})_{d,u,x_C,y_C}) \quad \text{(from Part (2) of Fact 2.1)} \\
&\geq \mathbb{E}_{(d,u,x_C,y_C) \leftarrow (DUX^1_C Y^1_C)} S(X^1_{d,u,x_C,y_C} || X'_{d_1,u_1,x_C,y_C} \otimes \cdots \otimes X'_{d_k,u_k,x_C,y_C}) \quad \text{(from Part (2) of Fact 2.1)} \\
&\geq \mathbb{E}_{(d,u,x_C,y_C) \leftarrow (DUX^1_C Y^1_C)} \sum_{i \notin C, i \in I} S\big((X^1_{d,u,x_C,y_C})_i || X'_{d_i,u_i}\big) \quad \text{(from Part (2) of Fact 2.1)} \\
&= \sum_{i \notin C, i \in I} \mathbb{E}_{(d,u,x_C,y_C) \leftarrow (DUX^1_C Y^1_C)} S\big((X^1_{d,u,x_C,y_C})_i || X'_{d_i,u_i}\big). \qquad (4.1)
\end{aligned}$$

Similarly,

$$\delta\delta_1 k + \delta\delta_1 ck > \sum_{i \notin C, i \in I} \mathbb{E}_{(d,u,x_C,y_C) \leftarrow (DUX^1_C Y^1_C)} S\big((Y^1_{d,u,x_C,y_C})_i || Y'_{d_i,u_i}\big). \qquad (4.2)$$

From Eq. (4.1) and Eq. (4.2) and using Markov's inequality, we get a coordinate j outside of C but in I such that

(1) $\mathbb{E}_{(d,u,x_C,y_C) \leftarrow (DUX^1_C Y^1_C)} S((X^1_{d,u,x_C,y_C})_j || X'_{d_j,u_j}) \leq \frac{2\delta(c+1)}{(1-\delta)} \leq 4\delta c$, and
(2) $\mathbb{E}_{(d,u,x_C,y_C) \leftarrow (DUX^1_C Y^1_C)} S((Y^1_{d,u,x_C,y_C})_j || Y'_{d_j,u_j}) \leq \frac{2\delta(c+1)}{(1-\delta)} \leq 4\delta c$.

Therefore,

$$\begin{aligned}
4\delta c &\geq \mathbb{E}_{(d,u,x_C,y_C) \leftarrow (DUX^1_C Y^1_C)} S\big((X^1_{d,u,x_C,y_C})_j || X'_{d_j,u_j}\big) \\
&= \mathbb{E}_{(d_{-j},u_{-j},x_C,y_C) \leftarrow (D_{-j}U_{-j}X^1_C Y^1_C)} \mathbb{E}_{(d_j,u_j) \leftarrow (D_j U_j)| \; (D_{-j}U_{-j}X^1_C Y^1_C) = (d_{-j},u_{-j},x_C,y_C)} \\
&\quad S((X^1_{d,u,x_C,y_C})_j || X'_{d_j,u_j}).
\end{aligned}$$

And,

$$\begin{aligned}4\delta c &\geq \mathbb{E}_{(d,u,x_C,y_C)\leftarrow(DUX^1_CY^1_C)} S\big(\big(Y^1_{d,u,x_C,y_C}\big)_j || Y'_{d_j,u_j}\big)\\ &= \mathbb{E}_{(d_{-j},u_{-j},x_C,y_C)\leftarrow(D_{-j}U_{-j}X^1_CY^1_C)} \mathbb{E}_{(d_j,u_j)\leftarrow(D_jU_j)|\ (D_{-j}U_{-j}X^1_CY^1_C)=(d_{-j},u_{-j},x_C,y_C)}\\ &\quad S\big(\big(Y^1_{d,u,x_C,y_C}\big)_j || Y'_{d_j,u_j}\big).\end{aligned}$$

Now, using Markov's inequality, there exists set G_1 with $\Pr[D_{-j}U_{-j}X^1_CY^1_C \in G_1] \geq 1-0.2$, such that for all $(d_{-j}, u_{-j}, x_C, y_C) \in G_1$,

$$\mathbb{E}_{(d_j,u_j)\leftarrow(D_jU_j)|\ (D_{-j}U_{-j}X^1_CY^1_C)=(d_{-j},u_{-j},x_C,y_C)} S\big(\big(X^1_{d,u,x_C,y_C}\big)_j || X'_{d_j,u_j}\big) \leq 40\delta c, \tag{4.3}$$

$$\mathbb{E}_{(d_j,u_j)\leftarrow(D_jU_j)|\ (D_{-j}U_{-j}X^1_CY^1_C)=(d_{-j},u_{-j},x_C,y_C)} S\big(\big(Y^1_{d,u,x_C,y_C}\big)_j || Y'_{d_j,u_j}\big) \leq 40\delta c. \tag{4.4}$$

Fix $(d_{-j}, u_{-j}, x_C, y_C) \in G_1$. Conditioning on $D_j = 1$ (which happens with probability 1/2) in Eq. (4.3), we get

$$\mathbb{E}_{y_j\leftarrow Y^1_j|(D_{-j}U_{-j}X^1_CY^1_C)=(d_{-j},u_{-j},x_C,y_C)} S\big(\big(X^1_{d_{-j},u_{-j},y_j,x_C,y_C}\big)_j || X'_{y_j}\big) \leq 80\delta c. \tag{4.5}$$

Conditioning on $D_j = 0$ (which happens with probability 1/2) in Eq. (4.4), we get

$$\mathbb{E}_{x_j\leftarrow X^1_j|(D_{-j}U_{-j}X^1_CY^1_C)=(d_{-j},u_{-j},x_C,y_C)} S\big(\big(Y^1_{d_{-j},u_{-j},x_j,x_C,y_C}\big)_j || Y'_{x_j}\big) \leq 80\delta c. \tag{4.6}$$

Let $X^2Y^2 = ((X^1Y^1)_{d_{-j},u_{-j},x_C,y_C})_j$. Note that conditioned on $(d_{-j}, u_{-j}, x_C, y_C)$, the distribuition of X^1Y^1 is independent across Alice and Bob in all other coordinates except j, so the inputs in coordinates other than j serve as private coins for Alice and Bob. Hence, X^2Y^2 is SM-like for μ. From Eq. (4.5) and Eq. (4.6), we get that

$$\mathsf{crent}^{\mu}_{\mathcal{X}}(X^2Y^2) + \mathsf{crent}^{\mu}_{\mathcal{Y}}(X^2Y^2) < c.$$

Hence, $\mathsf{err}_{f,\{z\}}(((X^1Y^1)_{d_{-j},u_{-j},x_C,y_C})_j) \geq \varepsilon$. This implies,

$$\Pr[T_j = 1 |\ (1, d_{-j}, u_{-j}, x_C, y_C) = (TD_{-j}U_{-j}X_CY_C)] \leq 1 - \varepsilon.$$

Therefore, overall

$$\Pr[T_j = 1 |\ (T = 1)] \leq 0.8(1-\varepsilon) + 0.2 \leq 1 - \varepsilon/2. \quad \square$$

We can now state and prove the main result of this section.

THEOREM 4.12 (DIRECT PRODUCT IN TERMS OF D AND crent). *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation, $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ be a distribution and $S \subseteq \mathcal{Z}$. Let $0 < \varepsilon < 1/3$ and k be a natural number. Let $\delta_2 = \mathsf{ess}^{\mu}(f, S)$. Let $0 < 200\delta < \delta_2$. Let $\delta' = 3(1-\varepsilon/2)^{\lfloor \delta\delta_2 k/2 \rfloor}$. Then,*

$$\mathsf{D}^{2,\mu^k}_{1-\delta'}(f^k) \geq \frac{\delta}{2} \cdot \delta_2 \cdot k \cdot \mathsf{crent}^{2,\mu}_{\varepsilon}(f, S) - k.$$

PROOF. Let $\mathsf{crent}^{2,\mu}_{\varepsilon}(f, S) = c$. For input $(x, y) \in \mathcal{X}^k \times \mathcal{Y}^k$, let $b(x, y)$ be the number of indices i in $[k]$ for which there exists $z_i \notin S$ such that $(x_i, y_i, z_i) \in f$. Let

$$B = \left\{(x, y) \in \mathcal{X}^k \times \mathcal{Y}^k |\ b(x, y) \geq (1 - \delta_2/2)k\right\}.$$

By Chernoff's inequality, we get

$$\Pr_{(x,y)\leftarrow\mu^k}[(x, y) \in B] \leq \exp(-\delta_2^2 k/2).$$

Let $\mathcal{P}$ be a protocol for f^k with inputs $XY \sim \mu^k$ and communication at most $d = (kc\delta\delta_2/2) - k$ bits. Let $M \in \mathcal{M}$ represent the message transcript of $\mathcal{P}$. Let

$$B_1 = \left\{m \in \mathcal{M} |\ \Pr[(XY)_m \in B] \geq \exp\big(-\delta_2^2 k/4\big)\right\}.$$

Then, $\Pr[M \in B_1] \leq \exp(-\delta_2^2 k/4)$. Let

$$B_2 = \{m \in \mathcal{M} |\ \Pr[M = m] \leq 2^{-d-k}\}.$$

Then, $\Pr[M \in B_2] \leq 2^{-k}$. Fix $m \notin B_1 \cup B_2$. Let z_m be the output of $\mathcal{P}$ when $M = m$. Let $b(z_m)$ be the number of indices i such that $z_{m,i} \notin S$. If $b(z_m) \geq (1 - \delta_2/2)k$, then success of $\mathcal{P}$ when $M = m$ is at most $\exp(-\delta_2^2 k/4) \leq (1 - \varepsilon/2)^{\lfloor \delta\delta_2 k/2 \rfloor}$. If $b(z_m) < (1 - \delta_2/2)k$, then from Theorem 4.10 (by setting $z = z_m$ and $\delta_1 = \delta_2/2$), success of $\mathcal{P}$ when $M = m$ is at most $(1 - \varepsilon/2)^{\lfloor \delta\delta_2 k/2 \rfloor}$. Therefore, overall success probability of $\mathcal{P}$ is at most

$$2^{-k} + \exp\left(-\delta_2^2 k/4\right) + \left(1 - 2^{-k} - \exp\left(-\delta_2^2 k/4\right)\right)(1 - \varepsilon/2)^{\lfloor \delta\delta_2 k/2 \rfloor} \leq 3(1 - \varepsilon/2)^{\lfloor \delta\delta_2 k/2 \rfloor}. \quad \square$$

We point that when μ is a product distribution, this result and Lemma 4.9 imply the direct product result, using $\mathsf{ment}_\varepsilon^{2,\mu}(f, S)$, of Jain et al. [2008].

5. STRONG DIRECT PRODUCT FOR SET DISJOINTNESS

In this section, we present an application of Theorem 4.12 to show a strong direct product result for the two-way public-coin communication complexity of the set-disjointness function. For a string $x \in \{0, 1\}^n$, we let x also represent the subset of $[n]$ for which x is the characteristic vector. The set disjointness function $\mathsf{disj}_n : \{0, 1\}^n \times \{0, 1\}^n \to \{0, 1\}$ is defined as $\mathsf{disj}_n(x, y) = 1$ iff the subsets x and y do not intersect.

THEOREM 5.1 (STRONG DIRECT PRODUCT FOR SET DISJOINTNESS). *Let k be a natural number. Then,* $\mathsf{R}^{2,\mathsf{pub}}_{1-2^{-\Omega(k)}}(\mathsf{disj}_n^k) = \Omega(k \cdot n)$.

PROOF. We assume $n = 4l - 1$ (for some integer l) and show our result. This implies our result for all n (since we use Ω in our result). Let $T = (T_1, T_2, I)$ be a uniformly random partition of $[n]$ into three disjoint sets such that $|T_1| = |T_2| = 2l - 1$ and $|I| = 1$. Conditioned on $T = t = (t_1, t_2, \{i\})$, let X be a uniformly random subset of $t_1 \cup \{i\}$ and Y be a uniformly random subset of $t_2 \cup \{i\}$. Note that $X \leftrightarrow T \leftrightarrow Y$ is a Markov chain. It is easily seen that $\mathsf{ess}^{XY}(\mathsf{disj}_n, \{1\}) = 0.75$. Therefore, using Theorem 4.12, Lemma 5.2 and Lemma 2.5 (Yao principle) we conclude the desired,

$$\mathsf{R}^{2,\mathsf{pub}}_{1-2^{-\Omega(k)}}(\mathsf{disj}_n^k) = \Omega(k \cdot n).$$

LEMMA 5.2. $\mathsf{crent}^{2,XY}_{1/20}(\mathsf{disj}_n, \{1\}) = \Omega(n)$.

PROOF. Our proof follows broadly on the lines as the proof of Razborov [1992] showing linear lower bound on the rectangle bound for set-disjointness (see, e.g., Kushilevitz and Nisan [1997, Lemma 4.49]). However, there are important differences. The first difference is that we not only have to argue about distributions conditioned on rectangles but also about the more general SM-like distributions. The second very important difference is that we cannot assume that the distribution (for which we have to show error) has high relative min-entropy with XY, which was the assumption under which Razborov's proof worked. We are working with a much weaker quantity crent and hence we cannot afford to even ignore events with exponentially small probability, which Razborov's proof could. Due to this, the exact technical lemmas that we use and their proofs differ significantly from Razborov's work.

The broad argument works as follows: Consider a distribution λ such that $\mathsf{crent}^{XY}(\lambda)$ is small. We divide λ into a convex combination of several product (across Alice and Bob) distributions. We are able to argue that for most of these ("good") distributions, entropy of Alice's input and Bob's input is still sufficiently high and since these are product distributions this implies that the probability of Alice and Bob's inputs intersecting is sufficiently high. A large part of the technical argument goes in showing that not much probability is lost in “not good” distributions. The details follow.

Let $\delta = 1/(200)^3$. Let $X'Y'$ be such that $\mathsf{crent}^{XY}_{\mathcal{X}}(X'Y') + \mathsf{crent}^{XY}_{\mathcal{Y}}(X'Y') \leq \delta n$ and $X'Y'$ is SM-like for XY. We will show that $\mathsf{err}_{\mathsf{disj}_n,\{1\}}(X'Y') = \Pr[\mathsf{disj}_n(X'Y') = 0] > 1/20$. This will show the desired. We assume that $\Pr[\mathsf{disj}_n(X'Y') = 1] \geq 0.5$ otherwise we are done already. Let $A, B \in \{0, 1\}$ be binary random variables such that $A \leftrightarrow X \leftrightarrow Y \leftrightarrow B$ and $X'Y' = (XY|\ A = B = 1)$. We argue existence of such A, B as follows. Let $X'Y$ be a distribution which is one-way (with respect to $\mathcal{X}$) for XY and $X'Y'$ is one-way (with respect to $\mathcal{Y}$) for $X'Y$. Let $X''Y$ be a distribution on $\mathcal{X} \times \mathcal{Y}$ and $p \in [0, 1]$ be such that $XY = pX'Y + (1 - p)X''Y$. Define correlated random variable $A \in \{0, 1\}$ such that $X'Y = (XY|\ A = 1)$ and $X''Y = (XY|\ A = 0)$. Note that since $X'Y$ is one-way for XY with respect to $\mathcal{X}$, so is $X''Y$. Hence $A \leftrightarrow X \leftrightarrow Y$. Similarly, we can argue the existence of B using the fact that $X'Y'$ is one-way (with respect to $\mathcal{Y}$) for $X'Y$.

Recall that in our notation $X_{t_2} = (X|\ T_2 = t_2)$ and so on. Define

(1) $B^1_x = \{t_2|\ S(X'_{t_2}||X_{t_2}) > 100\delta n\}$, $B^1_y = \{t_1|\ S(Y'_{t_1}||Y_{t_1}) > 100\delta n\}$;
(2) $B^2_x = \{t = (t_1, t_2, i) \mid S((X'_t)_i||(X_t)_i) > 0.01\}$, $B^2_y = \{t = (t_1, t_2, i) \mid S((Y'_t)_i||(Y_t)_i) > 0.01\}$;
(3) $Bad^1_x(T) = 1$ iff $T_2 \in B^1_x$, otherwise 0. $Bad^1_y(T) = 1$ iff $T_1 \in B^1_y$ otherwise 0;
(4) $Bad^2_x(T) = 1$ iff $T \in B^2_x$, otherwise 0. $Bad^2_y(T) = 1$ iff $T \in B^2_y$ otherwise 0.

Note that if $t \notin B^2_y$ then $S((Y'_t)_i||(Y_t)_i) \leq 0.01$. This implies (using Part (3) of Fact 2.1) $||(Y'_t)_i - (Y_t)_i)||_t \leq 0.1$. Recall that $(Y_t)_i$ is uniformly distributed in $\{0, 1\}$. This implies

$$\begin{aligned} 0.45 &\leq \Pr[Y_i = 1|\ B = 1, T = t] \\ \Rightarrow 0.45 \Pr[B = 1|\ T = t] &\leq \Pr[Y_i = 1, B = 1|\ T = t]. \end{aligned} \tag{5.1}$$

Similarly for $t \notin B^2_x$ we have $0.45 \Pr[A = 1|\ T = t] \leq \Pr[X_i = 1, A = 1|\ T = t]$.

The following three claims show that we do not loose much probability when T is bad.

Claim 5.3.

(1) $\Pr[A = B = 1, T_2 \in B^1_x] < \frac{1}{100} \Pr[A = B = 1]$.
(2) $\Pr[A = B = 1, T_1 \in B^1_y] < \frac{1}{100} \Pr[A = B = 1]$.
(3) *Let* $t_2 \notin B^1_x$, *then* $\Pr[T \in B^2_x|\ T_2 = t_2] < \frac{1}{100}$.
(4) *Let* $t_1 \notin B^1_y$, *then* $\Pr[T \in B^2_y|\ T_1 = t_1] < \frac{1}{100}$.

Proof. We show Part (1) and Part (2) follows similarly. Let $T' = (T \mid A = B = 1)$. Here, the second equality follows since $\forall(x, y) :\ (T|\ XY = (x, y))$ is identically distributed as $(T'|\ X'Y' = (x, y)))$ and using Part (2) of Fact 2.1. Consider,

$$\begin{aligned} \delta n &\geq \mathsf{crent}^{XY}_{\mathcal{Y}}(X'Y') = \mathbb{E}_{y \leftarrow Y'} S(X'_y||X_y) \\ &= \mathbb{E}_{y \leftarrow Y'} S((X'T')_y||(XT)_y) \\ &\geq \mathbb{E}_{(y,t) \leftarrow (Y'T')} S(X'_{y,t}||X_{y,t}) \quad \text{(from Part (2) of Fact 2.1)} \\ &= \mathbb{E}_{t \leftarrow T'} S(X'_t||X_t) \quad \text{(since } X \leftrightarrow T \leftrightarrow Y \text{ and } X' \leftrightarrow T' \leftrightarrow Y' \text{ are Markov chains)} \\ &= \mathbb{E}_{t_2 \leftarrow T'_2} S(X'_{t_2}||X_{t_2}). \end{aligned}$$

Here, the last equality follows since for all $t = (t_1, t_2, \{i\})$, X_{t_2} is identically distributed as X_t and similarly X'_{t_2} is identically distributed as X'_t. Therefore, using Markov's inequality,

$$\frac{1}{100} > \Pr[T'_2 \in B^1_x] = \Pr[T_2 \in B^1_x \mid A = B = 1] = \frac{\Pr[T_2 \in B^1_x, A = B = 1]}{\Pr[A = B = 1]}.$$

We show Part (3) and Part (4) follows similarly. Fix $t_2 \notin B^1_x$. Then, using Part (2) of Fact 2.1 (recall that in our notation X_i represents the ith bit of X and so on),

$$100\delta n \geq S(X'_{t_2} || X_{t_2}) \geq \sum_{i \notin t_2} S((X'_{t_2})_i || (X_{t_2})_i).$$

Let $R = \{i \notin t_2 | \; S((X'_{t_2})_i || (X_{t_2})_i) > 0.01\}$. This inequality and the choice of δ implies $|R| \leq n/800$. Now $i \notin R \cup t_2$ implies $t = (t_1, t_2, \{i\}) \notin B^2_x$. Hence,

$$\Pr\left[T \in B^2_x \mid T_2 = t_2\right] \leq \Pr[i \in R \mid T_2 = t_2] = \frac{|R|}{2l} < \frac{1}{100}. \quad \square$$

Claim 5.4.

(1)

$$\mathbb{E}_{t=(t_1,t_2,\{i\})\leftarrow T} \Pr[A = B = 1, Y_i = 0 \mid T = t] Bad^1_x(t) \leq \frac{2}{100} \Pr[A = B = 1, \mathsf{disj}_n(XY) = 1].$$

(2)

$$\mathbb{E}_{t=(t_1,t_2,\{i\})\leftarrow T} \Pr[A = B = 1, X_i = 0 \mid T = t] Bad^1_y(t) \leq \frac{2}{100} \Pr[A = B = 1, \mathsf{disj}_n(XY) = 1].$$

(3) *Fix $t_2 \notin B^1_x$. Let $T_{t_2} = (T \mid T_2 = t_2)$. Then*

$$\begin{aligned}&\mathbb{E}_{t=(t_1,t_2,\{i\})\leftarrow T_{t_2}} \Pr[A = B = 1, Y_i = 0 \mid T = t] Bad^2_x(t)\\ &\leq \frac{1}{100} \mathbb{E}_{t=(t_1,t_2,\{i\})\leftarrow T_{t_2}} \Pr[A = B = 1, \mathsf{disj}_n(XY) = 1 \mid T = t].\end{aligned}$$

(4) *Fix $t_1 \notin B^1_y$. Let $T_{t_1} = (T \mid T_1 = t_1)$. Then*

$$\begin{aligned}&\mathbb{E}_{t=(t_1,t_2,\{i\})\leftarrow T_{t_1}} \Pr[A = B = 1, X_i = 0 \mid T = t] Bad^2_y(t)\\ &\leq \frac{1}{100} \mathbb{E}_{t=(t_1,t_2,\{i\})\leftarrow T_{t_1}} \Pr[A = B = 1, \mathsf{disj}_n(XY) = 1 \mid T = t].\end{aligned}$$

Proof. We show Part (1) and Part (2) follows similarly. Consider,

$$\begin{aligned}&\mathbb{E}_{t=(t_1,t_2,\{i\})\leftarrow T} \Pr[A = 1, B = 1, Y_i = 0 \mid T = t] Bad^1_x(t)\\ &\leq \mathbb{E}_{t=(t_1,t_2,\{i\})\leftarrow T} \Pr[A = B = 1 \mid T = t] Bad^1_x(t)\\ &= \Pr[A = B = 1, T_2 \in B^1_x]\\ &\leq \frac{1}{100} \Pr[A = B = 1] \quad \text{(from Part (1) of Claim 5.3)}\\ &\leq \frac{2}{100} \Pr[A = B = 1, \mathsf{disj}_n(XY) = 1]. \quad (\text{since } \Pr[\mathsf{disj}_n(X'Y') = 1] \geq 0.5).\end{aligned}$$

We show Part (3) and Part (4) follows similarly. Note that

(1) $\Pr[B = 1 \mid Y_i = 0, T = (t_1, t_2, \{i\})]$ is independent of i for fixed t_2. Let us call it $c(t_2)$;
(2) $\Pr[A = 1 \mid T = (t_1, t_2, \{i\})]$ is independent of i for fixed t_2. Let us call it $r(t_2)$.

Fix $t_2 \notin B_x^1$. Consider,

$$\begin{aligned}
&\mathbb{E}_{t=(t_1,t_2,\{i\})\leftarrow T_{t_2}} \Pr[A=1, B=1, Y_i=0 |\ T=t] Bad_x^2(t)\\
&= \mathbb{E}_{t=(t_1,t_2,\{i\})\leftarrow T_{t_2}} \frac{1}{2} \cdot \Pr[A=1 |\ T=t] \Pr[B=1 |\ Y_i=0, T=t] Bad_x^2(t)\\
&= \frac{c(t_2)r(t_2)}{2} \mathbb{E}_{t=(t_1,t_2,\{i\})\leftarrow T_{t_2}} Bad_x^2(t)\\
&\leq \frac{2}{100} c(t_2)r(t_2) \quad \text{(from Part (3) of Claim 5.3)}\\
&= \frac{2}{100} \mathbb{E}_{t=(t_1,t_2,\{i\})\leftarrow T_{t_2}} \Pr[A=1 |\ T=t] \Pr[B=1 |\ Y_i=0, T=t]\\
&= \frac{1}{100} \mathbb{E}_{t=(t_1,t_2,\{i\})\leftarrow T_{t_2}} \Pr[A=1, B=1, Y_i=0 |\ T=t]\\
&\leq \frac{1}{100} \mathbb{E}_{t=(t_1,t_2,\{i\})\leftarrow T_{t_2}} \Pr[A=B=1, \mathsf{disj}_n(XY)=1 |\ T=t]. \quad \square
\end{aligned}$$

In this claim, ∨ represents the logical OR.

CLAIM 5.5. *For any* $t=(t_1,t_2,i)$

$$\begin{aligned}
&\Pr[A=B=1, \mathsf{disj}_n(XY)=1 \mid T=t]\big(Bad_x^2(t) \vee Bad_y^2(t)\big)\\
&\quad \leq 3\Pr[A=B=1, X_i=0 \mid T=t] Bad_y^2(t) + 3\Pr[A=B=1, Y_i=0 \mid T=t] Bad_x^2(t).
\end{aligned}$$

This implies using Claim 5.4.

$$\begin{aligned}
&\mathbb{E}_{t=(t_1,t_2,\{i\})\leftarrow T} \Pr[A=B=1, \mathsf{disj}_n(XY)=1 \mid T=t]\big(Bad_x^2(t) \vee Bad_y^2(t)\big)\\
&\quad \leq \frac{18}{100} \mathbb{E}_{t=(t_1,t_2,\{i\})\leftarrow T} \Pr[A=B=1, \mathsf{disj}_n(XY)=1 \mid T=t].
\end{aligned}$$

PROOF. The claim is obvious when $Bad_x^2(t) = Bad_y^2(t) = 0$. Consider the case when $Bad_x^2(t) = Bad_y^2(t) = 1$. Then,

$$\begin{aligned}
&\Pr[A=B=1, \mathsf{disj}_n(XY)=1 \mid T=t](Bad_x^2(t) \vee Bad_y^2(t))\\
&\leq (\Pr[A=B=1, X_i=0 \mid T=t] + \Pr[A=B=1, Y_i=0 \mid T=t])(Bad_x^2(t) \vee Bad_y^2(t))\\
&= \Pr[A=B=1, X_i=0 \mid T=t] Bad_y^2(t) + \Pr[A=B=1, Y_i=0 \mid T=t] Bad_x^2(t).
\end{aligned}$$

Consider the case $Bad_x^2(t)=1$ and $Bad_y^2(t)=0$ (the case $Bad_x^2(t)=0$ and $Bad_y^2(t)=1$ is similar). In this case, $\Pr[Y_i=0 \mid B=1, T=t] \geq 0.45$. Then,

$$\begin{aligned}
&\Pr[A=B=1, \mathsf{disj}_n(XY)=1 \mid T=t](Bad_x^2(t) \vee Bad_y^2(t))\\
&= \Pr[A=B=1, \mathsf{disj}_n(XY)=1 \mid T=t] Bad_x^2(t)\\
&\leq \Pr[A=B=1 \mid T=t] Bad_x^2(t)\\
&= \Pr[A=1 \mid T=t] \Pr[B=1 \mid T=t] Bad_x^2(t)\\
&\leq \frac{1}{0.45} \Pr[A=1 \mid T=t] \Pr[B=1, Y_i=0 \mid T=t] Bad_x^2(t)\\
&= \frac{1}{0.45} \Pr[A=B=1, Y_i=0 \mid T=t] Bad_x^2(t). \quad \square
\end{aligned}$$

We can now finally prove our lemma. Define $\mathit{Bad}(t) = \mathit{Bad}_x^2(t) \vee \mathit{Bad}_y^2(t)$.

$$
\begin{aligned}
&\Pr[A = B = 1, \mathsf{disj}_n(XY) = 1] \\
&\leq \frac{100}{82} \mathbb{E}_{t=(t_1,t_2,\{i\})\leftarrow T} \Pr[A = B = 1, \mathsf{disj}_n(XY) = 1 \mid T = t](1 - \mathit{Bad}(t)) \quad \text{(using Claim 5.5)} \\
&\leq \frac{100}{82} \mathbb{E}_{t=(t_1,t_2,\{i\})\leftarrow T} \Pr[A = B = 1 \mid T = t](1 - \mathit{Bad}(t)) \\
&= \frac{100}{82} \mathbb{E}_{t=(t_1,t_2,\{i\})\leftarrow T} \Pr[A = 1 \mid T = t] \Pr[B = 1 \mid T = t](1 - \mathit{Bad}(t)) \\
&\leq \frac{100}{82(0.45)^2} \mathbb{E}_{t=(t_1,t_2,\{i\})\leftarrow T} \Pr[A = 1, X_i = 1 \mid T = t] \Pr[B = 1, Y_i = 1 \mid T = t](1 - \mathit{Bad}(t)) \\
&\qquad \text{(using Eq. (5.1) and the statement after it)} \\
&= \frac{100}{82(0.45)^2} \mathbb{E}_{t=(t_1,t_2,\{i\})\leftarrow T} \Pr[A = B = 1, \mathsf{disj}_n(XY) = 0 \mid T = t](1 - \mathit{Bad}(t)) \\
&\leq \frac{100}{82(0.45)^2} \Pr[A = B = 1, \mathsf{disj}_n(XY) = 0].
\end{aligned}
$$

This implies (recall that $\Pr[\mathsf{disj}_n(X'Y') = 1] \geq 0.5$),

$$
\begin{aligned}
\Pr[\mathsf{disj}_n(X'Y') = 0] &= \Pr[\mathsf{disj}_n(XY) = 0 \mid A = B = 1] \\
&= \frac{\Pr[\mathsf{disj}_n(XY) = 0, A = B = 1]}{\Pr[A = B = 1]} \\
&\geq \frac{82(0.45)^2}{100} \cdot \frac{\Pr[\mathsf{disj}_n(XY) = 1, A = B = 1]}{\Pr[A = B = 1]} \\
&= \frac{82(0.45)^2}{100} \cdot \Pr[\mathsf{disj}_n(X'Y') = 1] > \frac{1}{20}.
\end{aligned}
$$

Conclusion. In this work, we present the first generic strong direct product results holding of all relations for the one-way public-coin communication complexity and in terms of a lower bound method for the two-way public-coin communication complexity. As we mentioned in the Introduction, several subsequent works have followed, using and further developing ideas from our work and strengthening our results; the latest being a strong direct product result in terms of (internal) information complexity [Braverman and Weinstein 2014]. We hope that the ideas and techniques from our work are helpful in reaching an important final goal in this research area, that is in settling the strong direct product conjecture for the two-way public-coin communication complexity of all relations. As intermediate open questions, we can ask if a strong direct product result can be shown in terms of other lower bound methods, for example, the *partition bound* [Jain and Klauck 2010] and the *public-coin partition bound* [Jain et al. 2014], the latter has been shown to be quadratically tight for the two-way public-coin communication complexity for all relations.

APPENDIX

A. DEFERRED PROOF

PROOF OF LEMMA 2.4 First we obtain a public-coin protocol $\mathcal{P}'$ from protocol $\mathcal{P}$. In $\mathcal{P}'$, Alice on input x and Bob on input y proceed as follows. Let the set $\mathcal{U}$ be the support of M. Alice and Bob, using public coins, obtain a sequence $\{a_i\}_{i=1}^{\infty} = \{(u_i, p_i)\}_{i=1}^{\infty}$, where each a_i is drawn uniformly and independently from $\mathcal{U} \times [0, 1]$. They, using public coins, also obtain a sequence of random functions $\{h_i : \mathcal{U} \to \{0, 1\}\}_{i=1}^{m}$ ($m = c + \lceil \log \frac{2}{\delta} \rceil$) such that for every $u \neq u'$ $(u, u' \in \mathcal{U})$ and for every $i \in [m]$, we have $\Pr[h_i(u) = h_i(u')] = \frac{1}{2}$. Let $P = (M|\ X = x)$ and $Q = (M|\ Y = y)$. Let

$$\mathcal{A} = \{(u, p)|\ p \le P(u)\} \quad \text{and} \quad \mathcal{B} = \{(u, p)|\ p \le 2^c \cdot Q(u)\},$$

be subsets of $\mathcal{U} \times [0, 1]$. Let a_j be the first point in the sequence $\{a_i\}_{i=1}^{\infty} = \{(u_i, p_i)\}_{i=1}^{\infty}$ such that $a_j \in \mathcal{A}$. Let $k = \lceil \frac{j}{|\mathcal{U}|} \rceil$. If $k > \lceil \log \frac{2}{\delta} \rceil$, then Alice sends 1 to Bob, otherwise she sends the binary encoding of k to Bob. Note that, for any n (assuming $|\mathcal{U}| > 1$),

$$\Pr[k > n] = \Pr[a_i \notin \mathcal{A} \text{ for } i = 1, \ldots, n \cdot |\mathcal{U}|] = (1 - 1/|\mathcal{U}|)^{|\mathcal{U}| \cdot n} < e^{-n}.$$

Therefore,

$$\Pr\left[k > \left\lceil \log \frac{2}{\delta} \right\rceil\right] < e^{-\lceil \log \frac{2}{\delta} \rceil} \le \delta/2.$$

Alice also sends to Bob $h_i(u_j)$ for all $i \in [m]$. Hence, overall communication from Alice to Bob $c + O(\log 1/\delta)$. Bob checks if there is an $a_r = (u_r, p_r)$ (in the sequence $\{a_i\}_{i=1}^{\infty}$) such that

(1) $r \in \{(k-1) \cdot |\mathcal{U}| + 1, \ldots, k \cdot |\mathcal{U}|\}$,
(2) $a_r \in \mathcal{B}$, and
(3) $h_i(u_r) = h_i(u_j)$ for all $i \in [m]$.

If there exists more than one such point, then Bob takes the first such point. If there is no such point then Bob assumes $r = 1$. Bob then proceeds as in $\mathcal{P}$ assuming the message in $\mathcal{P}$ from Alice is u_r. It is easily seen that the distribution of u_j is exactly P. Hence,

$$\begin{aligned} &\Pr[\text{ Bob's output is incorrect in } \mathcal{P}' \text{ on input } (x, y)\ ] \\ &\quad \le \Pr[\text{ Bob's output is incorrect in } \mathcal{P} \text{ on input } (x, y)\ ] + \Pr[r \neq j]. \end{aligned}$$

Note that

$$\Pr\left[r \neq j|\ u_j \in \mathcal{B}, k \le \left\lceil \log \frac{2}{\delta} \right\rceil\right] \le |\mathcal{U}| \cdot \frac{2^c}{|\mathcal{U}|} \cdot 2^{-m} \le \delta/2.$$

Now,

$$\begin{aligned} \Pr[r \neq j] &= \Pr\left[k \le \left\lceil \log \frac{2}{\delta} \right\rceil\right] \cdot \Pr\left[r \neq j|\ k \le \left\lceil \log \frac{2}{\delta} \right\rceil\right] \\ &\quad + \Pr\left[k > \left\lceil \log \frac{2}{\delta} \right\rceil\right] \cdot \Pr\left[r \neq j|\ k > \left\lceil \log \frac{2}{\delta} \right\rceil\right] \\ &\le \delta/2 + \Pr\left[r \neq j|\ k \le \left\lceil \log \frac{2}{\delta} \right\rceil\right] \\ &= \delta/2 + \Pr\left[a_j \in \mathcal{B}|\ k \le \left\lceil \log \frac{2}{\delta} \right\rceil\right] \cdot \Pr\left[r \neq j|\ a_j \in \mathcal{B}, k \le \left\lceil \log \frac{2}{\delta} \right\rceil\right] \\ &\quad + \Pr\left[a_j \notin \mathcal{B}|\ k \le \left\lceil \log \frac{2}{\delta} \right\rceil\right] \cdot \Pr\left[r \neq j|\ a_j \notin \mathcal{B}, k \le \left\lceil \log \frac{2}{\delta} \right\rceil\right] \\ &\le \delta + \Pr\left[a_j \notin \mathcal{B}|\ k \le \left\lceil \log \frac{2}{\delta} \right\rceil\right] \\ &= \delta + \Pr[a_j \notin \mathcal{B}]. \end{aligned}$$

Note that, from Eq. (2.1), expectation over $(x, y) \leftarrow XY$ of $\Pr[a_j \notin \mathcal{B}]$ is at most δ. Hence, overall,

$$\Pr[\text{Bob errs in } \mathcal{P}'] \leq \Pr[\text{Bob errs in } \mathcal{P}] + 2\delta.$$

Here the probability is taken over the inputs and coins used in the protocols. Now by fixing the public-coins in $\mathcal{P}'$ (so that the average error over the inputs is minimized), we get a deterministic one-way protocol $\mathcal{P}_1$ with distributional error (over the inputs) at most $\varepsilon + 2\delta$ and communication at most $c + O(\log \frac{1}{\delta})$. □

ACKNOWLEDGMENTS

We thank Penghui Yao for pointing an important bug in one of the earlier proofs of a theorem. We thank Ashwin Nayak and Shengyu Zhang for helpful discussions. The we thank the anonymous referees for many suggestions that were helpful in greatly improving the presentation of this article.

REFERENCES

B. Barak, M. Braverman, X. Chen, and A. Rao. 2010. How to compress interactive communication. In *Proceedings of the 42nd Annual ACM Symposium on Theory of Computing*.

Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. 2002. An information statistics approach to data stream and communication complexity. In *Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science*. 209–218.

Paul Beame, Toniann Pitassi, Nathan Segerlind, and Avi Wigderson. 2007. A direct sum theorem for corruption and a lower bound for the multiparty communication complexity of set disjointness. *Computat. Complex.*

Avraham Ben-Aroya, Oded Regev, and Ronald de Wolf. 2008. A hypercontractive inequality for matrix-valued functions with applications to quantum computing. In *Proceedings of the 49th IEEE Symposium on Foundations of Computer Science*. 477–486.

Mark Braverman and Anup Rao. 2011. Information equals amortized communication. In *Proceedings of the IEEE 52nd Annual Symposium on Foundations of Computer Science (FOCS'11)*. IEEE Computer Society, Los Alamitos, CA, 748–757. DOI:http://dx.doi.org/10.1109/FOCS.2011.86

Mark Braverman, Anup Rao, Omri Weinstein, and Amir Yehudayoff. 2013a. Direct product via round-preserving compression. In *Proceedings of the 40th International Colloquium on Automata, Languages and Programming (ICALP)*. ECCC-TR13-035.

M. Braverman, A. Rao, O. Weinstein, and A. Yehudayoff. 2013b. Direct Products in communication complexity. In *Proceeding of the IEEE 54th Annual Symposium on Foundations of Computer Science (FOCS),* 746–755. DOI:http://dx.doi.org/10.1109/FOCS.2013.85

Mark Braverman and Omri Weinstein. 2014. An interactive information odometer with applications. ECCC-TR14-047.

Thomas M. Cover and Joy A. Thomas. 1991. *Elements of Information Theory*. Wiley, New York.

Ronald de Wolf. 2005. Random access codes, direct product theorems, and multiparty communication complexity. Unpublished manuscript, incorporated into Ben-Aroya et al. [2008].

Dmitry Gavinsky. 2008. On the role of shared entanglement. *Quant. Inf. Comput.* 8.

Thomas Holenstein. 2007. Parallel repetition: Simplifications and the no-signaling case. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing*. 411–419.

Rahul Jain and Hartmut Klauck. 2009. New results in the simultaneous message passing model via information theoretic techniques. In *Proceeding of the 24th IEEE Conference on Computational Complexity*. 369–378.

R. Jain and H. Klauck. 2010. The partition bound for classical communication complexity and query complexity. In *IEEE 25th Annual Conference on Computational Complexity (CCC)*. 247–258. DOI:http://dx.doi.org/10.1109/CCC.2010.31

Rahul Jain, Hartmut Klauck, and Ashwin Nayak. 2008. Direct product theorems for classical communication complexity via subdistribution bounds. In *Proceedings of the 40th ACM Symposium on Theory of Computing*. 599–608.

Rahul Jain, Troy Lee, and Nisheeth K. Vishnoi. 2014. A quadratically tight partition bound for classical communication complexity and query complexity. arXiv:1401.4512.

Rahul Jain, Attila Pereszlenyi, and Penghui Yao. 2012. A direct product theorem for bounded-round public-coin randomized communication complexity. In *Proceedings of The 53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. 167–176.

Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. 2002. Privacy and interaction in quantum communication complexity and a theorem about the relative entropy of quantum states. In *Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science*. 429–438.

Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. 2003. A direct sum theorem in communication complexity via message compression. In *Proceedings of the 30th International Colloquium on Automata Languages and Programming.* Lecture Notes in Computer Science, vol. 2719. Springer, Berlin, 300–315.

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

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

Hartmut Klauck. 2004. Quantum and classical communication-space tradeoffs from rectangle bounds. In *Proceedings of the 24th Annual IARCS International Conference on Foundations of Software Technology and Theoretical Computer Science.* Lecture Notes in Computer Science, vol. 3328. Springer, Berlin, 384–395.

Hartmut Klauck. 2010. A strong direct product theorem for disjointness. In *Proceedings of the 42nd Annual ACM Symposium on Theory of Computing*. 77–86.

Eyal Kushilevitz and Noam Nisan. 1997. *Communication Complexity*. Cambridge University Press, Cambridge, UK. http://www.cs.technion.ac.il/∼eyalk/book.html.

Troy Lee, Adi Shraibman, and Robert Spalek. 2008. A direct product theorem for discrepancy. In *Proceedings of IEEE Conference on Computational Complexity*. 71–80.

Ilan Newman. 1991. Private vs. common random bits in communication complexity. *Inform. Process. Lett.* 39, 2, 67–71.

Itzhak Parnafes, Ran Raz, and Avi Wigderson. 1997. Direct product results and the GCD problem, in old and new communication models. In *Proceedings of the 29th Annual ACM Symposium on Theory of Computing*. 363–372.

Ran Raz. 1998. A parallel repetition theorem. *SIAM J. Comput.* 27, 3, 763–803.

A. Razborov. 1992. On the distributional complexity of disjointness. *Theoret. Comput. Sci.* 106, 2, 385–390.

Ronen Shaltiel. 2003. Towards proving strong direct product theorems. *Computat. Complex.* 12, 1–2, 1–22.

Alexander A. Sherstov. 2011. Strong direct product theorems for quantum communication and query complexity. In *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing*.

Emanuele Viola and Avi Wigderson. 2008. Norms, XOR lemmas, and lower bounds for polynomials and protocols. *Theory Comput.* 4, 1, 137–168.

Received March 2011; revised May 2013, October 2013, and April 2014; accepted November 2014