A direct product theorem for the two-party bounded-round public-coin communication complexity

Rahul Jain
Centre for Quantum Technologies
and Department of Computer Science
National University of Singapore
Singapore
rahul@comp.nus.edu.sg

Attila Pereszlényi
Centre for Quantum Technologies,
National University of Singapore
Singapore
attila.pereszlenyi@gmail.com

Penghui Yao
Centre for Quantum Technologies,
National University of Singapore
Singapore
pyao@nus.edu.sg

***Abstract*—A strong direct product theorem for a problem in a given model of computation states that, in order to compute k instances of the problem, if we provide resource which is less than k times the resource required for computing one instance of the problem with constant success probability, then the probability of correctly computing all the k instances together, is exponentially small in k. In this paper, we consider the model of two-party bounded-round public-coin randomized communication complexity. We show a direct product theorem for the communication complexity of any relation in this model. In particular, our result implies a strong direct product theorem for the two-party constant-message public-coin randomized communication complexity of all relations. As an immediate application of our result, we get a strong direct product theorem for the pointer chasing problem. This problem has been well studied for understanding round v/s communication trade-offs in both classical and quantum communication protocols [27], [18], [29], [20], [14]. Our result generalizes the result of Jain [11] which can be regarded as the special case when $t = 1$. Our result can be considered as an important progress towards settling the strong direct product conjecture for the two-party public-coin communication complexity, a major open question in this area.**

We show our result using information theoretic arguments. Our arguments and techniques build on the ones used in Jain [11]. One key tool used in our work and also in Jain [11] is a message compression technique due to Braverman and Rao [5], who used it to show a *direct sum* theorem in the same model of communication complexity as considered by us. Another important tool that we use is a correlated sampling protocol, which for example, has been used in Holenstein [9] for proving a parallel repetition theorem for two-prover games.

***Index Terms*—Communication complexity, information theory, direct product, bounded rounds.**

I. Introduction

A fundamental question in complexity theory is how much resource is needed to solve k independent instances of a problem compared to the resource required to solve one instance. More specifically, suppose for solving one instance of a problem with probability of correctness p, we require c units of some resource in a given model of computation. A natural way to solve k independent instances of the same problem is to solve them independently, which needs $k \cdot c$ units of resource and the overall success probability is p^k. A *strong direct product* theorem for this problem would state that any algorithm, which solves k independent instances of this problem with $o(k \cdot c)$ units of the resource, can only compute all the k instances correctly with probability at most $p^{-\Omega(k)}$.

In this work, we are concerned with the model of communication complexity which was introduced by Yao [35]. In this model there are different parties who wish to compute a joint relation of their inputs. They do local computation, use public/private coins, and communicate between them to achieve this task. The resource that is counted is the number of bits communicated. The text by Kushilevitz and Nisan [23] is an excellent reference for this model. Direct product questions and the weaker *direct sum* questions have been extensively investigated in different sub-models of communication complexity. A direct sum theorem states that in order to compute k independent instances of a problem, if we provide resource less than k times the resource required to compute one instance of the problem with a constant success probability $p < 1$, then the success probability for computing all the k instances correctly is at most a constant $q < 1$. Some examples of known direct product theorems are: Parnafes, Raz and Wigderson's [28] theorem for *forests* of communication protocols; Shaltiel's [32] theorem for the *discrepancy bound* (which is a lower bound on the *distributional* communication complexity) under the uniform distribution; extended to arbitrary distributions by Lee, Shraibman and Špalek [25]; extended to the multiparty case by Viola and Wigderson [34]; extended to the generalized discrepancy bound by Sherstov [33]; Jain, Klauck and Nayak's [13] theorem for the *subdistribution bound*; Klauck, Špalek, de Wolf's [21] theorem for the *quantum* communication complexity of the *set disjointness* problem; Klauck's [19] theorem for the public-coin communication complexity of the set-disjointness problem (which was re-proven using very different arguments in Jain [11]); Ben-Aroya, Regev, and de Wolf's [4] theorem for the one-way quantum communication complexity of the *index* function problem; Jain's [11] theorem for randomized one-way communication complexity and Jain's [11] theorem for *conditional relative min-entropy bound* (which is a lower bound on the public-coin communication complexity). Direct sum theorems have been shown in the public-coin one-way model [15], public-coin simultaneous message passing model [15], entanglement-assisted quantum one-way communication model [17], private-coin simultaneous message passing model [12] and constant-round public-coin two-way model [5]. On the other hand, strong direct product conjectures have been shown to be false by Shaltiel [32] in some models of distributional communication complexity (and of *query complexity* and *circuit complexity*) under specific choices for the error parameter.

Examples of direct product theorems in others models of computation include Yao's *XOR lemma* [36], Raz's [30] theorem for two-prover games; Shaltiel's [32] theorem for *fair decision trees*; Nisan, Rudich and Saks' [26] theorem for *decision forests*; Drucker's [8] theorem for randomized query complexity; Sherstov's [33] theorem for *approximated polynomial degree* and Lee and Roland's [24] theorem for quantum query complexity. Besides their inherent importance, direct product theorems have had various important applications such as in *Probabilistically checkable proofs* [30]; in circuit complexity [36] and in showing time-space tradeoffs [22], [1], [19].

In this paper, we show a direct product theorem for the two-party bounded-round public-coin randomized communication complexity. In this model, for computing a relation $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ ($\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ are finite sets), one party, say Alice, is given an input $x \in \mathcal{X}$ and the other party, say Bob, is given an input $y \in Y$. They are supposed to do local computations using public coins shared between them, communicate a fixed number of messages between them and at the end, output an element $z \in Z$. They are said to succeed if $(x, y, z) \in f$. For a natural number $t \geq 1$ and $\varepsilon \in (0, 1)$, let $\mathrm{R}_{\varepsilon}^{(t),\mathrm{pub}}(f)$ denote the two-party t-message public-coin communication complexity of f with worst case error ε, that is the communication of the best public-coin protocol between Alice and Bob with t messages exchanged between them, and the error (over the public coins) on any input (x, y) being at most ε. We show the following.

Theorem I.1. *Let $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$ be finite sets, $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ a relation, $\varepsilon > 0$ and $k, t \geq 1$ be integers. There exists a constant $\kappa \geq 0$ such that,*

$$\mathrm{R}_{1-(1-\varepsilon/2)^{\Omega(k\varepsilon^2/t^2)}}^{(t),\mathrm{pub}}(f^k) = \Omega\left(\frac{\varepsilon \cdot k}{t} \cdot \left(\mathrm{R}_{\varepsilon}^{(t),\mathrm{pub}}(f) - \frac{\kappa t^2}{\varepsilon}\right)\right).$$

In particular, it implies a strong direct product theorem for the two-party constant-message public-coin randomized communication complexity of all relations f.[1] Our result generalizes the result of Jain [11] which can be regarded as the special case when $t = 1$. Our result can be considered as an important progress towards settling the strong direct product conjecture for the two-party public-coin communication complexity, a major open question in this area.

As a direct consequence of our result we get a direct product theorem for the *pointer chasing* problem defined as follows. Let $n, t \geq 1$ be integers. Alice and Bob are given functions $F_A : [n] \to [n]$ and $F_B : [n] \to [n]$, respectively. Let F^t represent alternate composition of F_A and F_B done t times, starting with F_A. The parties are supposed to communicate and determine $F^t(1)$. In the bit version of the problem, the players are supposed to output the least significant bit of $F^t(1)$. We refer to the t-pointer chasing problem as FP_t and the bit version as BP_t. The pointer chasing problem naturally captures the trade-off between number of messages exchanged and the communication used. There is a straightforward t-message deterministic protocol with $t \cdot \log n$ bits of communication for both FP_t and BP_t. However if only $t-1$ messages are allowed to be exchanged between the parties, exponentially more communication is required. The communication complexity of this problem has been very well studied both in the classical and quantum models of communication complexity [27], [18], [29], [20], [14]. Some tight lower bounds that we know so far

[1] When $\mathrm{R}_{\varepsilon}^{(t),\mathrm{pub}}(f)$ is a constant, a direct product result can be shown via direct arguments as for example in [11].

are as follows (below $\mathrm{Q}^{(t)}(\cdot)$ represents the t-message quantum communication complexity).

Theorem I.2. *For integer $t \geq 1$,*

1) *[29]* $\mathrm{R}^{(t-1),\mathrm{pub}}_{1/3}(\mathrm{FP}_t) \geq \Omega(n \log^{(t-1)} n)$;
$\mathrm{R}^{(t-1),\mathrm{pub}}_{1/3}(\mathrm{BP}_t) \geq \Omega(n)$.
2) *[14]* $\mathrm{Q}^{(t-1)}_{1/3}(\mathrm{FP}_t) \geq \Omega(n \log^{(t-1)} n)$.

As a consequence of Theorem I.1 we get strong direct product results for this problem. Note that in the descriptions of FP_t and BP_t, t is a fixed constant, not dependent on the input size.

Corollary I.3. *For integers $t, k \geq 1$,*

1) $\mathrm{R}^{(t-1),\mathrm{pub}}_{1-2^{-\Omega(k/t^2)}}(\mathrm{FP}^k_t) \geq \Omega\left(\frac{k}{t} \cdot n \log^{(t-1)} n\right)$;
2) $\mathrm{R}^{(t-1),\mathrm{pub}}_{1-2^{-\Omega(k/t^2)}}(\mathrm{BP}^k_t) \geq \Omega\left(\frac{k}{t} \cdot n\right)$.

Our techniques

We prove our direct product result using information theoretic arguments. Information theory is a versatile tool in communication complexity, especially in proving lower bounds and direct sum and direct product theorems [6], [2], [15], [16], [17], [12], [3], [5], [11]. The broad argument that we use is as follows. For a given relation f, let the communication required for computing one instance with t messages and constant success be c. Let us consider a protocol for computing f^k with t messages and communication cost $o(kc)$. Let us condition on success on some l 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 j-th coordinate, even conditioned on the success in the l coordinates, is bounded away from 1. This way the overall success keeps going down and becomes exponentially small (in k) eventually. 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 is then related to the worst case setting by the well known Yao's principle [35].

More concretely, let μ be a distribution on $\mathcal{X} \times \mathcal{Y}$, possibly non-product across $\mathcal{X}$ and $\mathcal{Y}$. Let c be the minimum communication required for computing f with t-message protocols having error at most ε averaged over μ. Let us consider the inputs for f^k drawn from the distribution μ^k (k independent copies of μ). Consider a t-message protocol $\mathcal{P}$ for f^k with communication $o(kc)$ and for the rest of the argument condition on success on a set C of coordinates. If the success probability of this event is as small as we desire then we are done. Otherwise we exhibit a new coordinate $j \notin C$ satisfying the following conditions: the distribution of inputs $X_j Y_j$ (of Alice and Bob respectively) in the j-th coordinate is quite close to μ and the joint distribution $X_j Y_j M$ (where M is the message transcript of $\mathcal{P}$) can be approximated very well by Alice and Bob using a t message protocol for f, when they are given input according to μ, using communication less than c. This shows that success in the j-th coordinate must be bounded away from one.

To simulate the transcript, we adopt the message compression protocol due to Braverman and Rao [5], where they used the protocol to show a direct sum theorem for the same communication model we are considering. Informally, the protocol can be stated as follows.

Braverman-Rao protocol (informal). *Given a Markov chain $X \leftrightarrow Y \leftrightarrow M$, there exists a public-coin protocol between Alice and Bob, with input X, Y, respectively, with a single message from Alice to Bob of $\mathcal{O}(\mathrm{I}(X : M | Y))$ bits, such that at the end of the protocol, Alice and Bob both possess a random variable M', close to M in ℓ_1 distance.*

Consider the situation after conditioning on the success in the set C as above, and let $X_j Y_j$ represent the input in the jth coordinate. The Braverman-Rao compression protocol cannot be directly applied at this stage. Take the first message M_1 sent by Alice, for instance. It is easily seen that $Y_j X_j M_1$ is not necessarily a Markov chain. However, we are able to show that $Y_j X_j M_1$ is 'close' to being a Markov chain by further conditioning on appropriate sub-events. We then use a more 'robust' Braverman-Rao compression protocol (along the lines of the original), where by being 'robust', we mean that the communication cost and the error does not vary much even for XYM which is close to being a Markov chain (similar arguments were used in Jain [11]). We then apply such a robust message compression protocol to each successive message. We accumulate some errors for each of these messages. Thus in order to keep the overall error bounded, we are able to make our argument for protocols with a bounded number of message exchanges.

Another difficulty that is faced in this argument is that since μ may be a non-product distribution, Alice and Bob may obtain information about each other's input in the j-th coordinate via their inputs in other coordinates. This is overcome by splitting the distribution μ into a convex combination of several product distributions. This idea of splitting a non-product distribution into convex combination of product distributions has been used in several previous works to handle non-product distributions in different settings [31], [30], [2], [9], [3], [5], [11]. This splitting of non-product distribution leads us to use another important tool namely the *correlated sampling* protocol, that was also used for example by Holenstein [9] while arguing a strong direct product result for the two-prover one-round games.

As mentioned previously, we build on the arguments used in Jain [11]. Jain shows a new characterization of the two-party one-way public-coin communication complexity and uses it to show a strong direct product result for all relations in this model. We are unable to arrive at such a characterization for protocols with more than one message and use a more direct approach, as outlined above, to arrive at our direct product result.

Organization: The rest of the paper is organized as follows. In Section II, we present some background on information theory and communication complexity. In Section III, we prove our main result Theorem I.1, starting with some lemmas that are helpful in building the proof. Some proofs are deferred to Appendix A.

II. Preliminaries

Information theory

For integer $n \geq 1$, let $[n]$ represent the set $\{1, 2, \ldots, n\}$. Let $\mathcal{X}$, $\mathcal{Y}$ be finite sets and k be a natural number. Let $\mathcal{X}^k$ be the set $\mathcal{X} \times \cdots \times \mathcal{X}$, the cross product of $\mathcal{X}$, k times. Let μ be a (probability) distribution on $\mathcal{X}$. Let $\mu(x)$ represent the probability of $x \in \mathcal{X}$ according to μ. Let X be a random variable distributed according to μ, which we denote by $X \sim \mu$. We use the same symbol to represent a random variable and its distribution whenever it is clear from the context. The expectation value of function f on $\mathcal{X}$ is denoted as $\mathbb{E}_{x \leftarrow X}[f(x)] \overset{\mathrm{def}}{=} \sum_{x \in \mathcal{X}} \Pr[X = x] \cdot f(x)$. The entropy of X is defined as $\mathrm{H}(X) \overset{\mathrm{def}}{=} -\sum_x \mu(x) \cdot \log \mu(x)$. For two distributions μ, λ on $\mathcal{X}$, the distribution $\mu \otimes \lambda$ is defined as $(\mu \otimes \lambda)(x_1, x_2) \overset{\mathrm{def}}{=} \mu(x_1) \cdot \lambda(x_2)$. Let $\mu^k \overset{\mathrm{def}}{=} \mu \otimes \cdots \otimes \mu$, k times. The ℓ_1 distance between μ and λ is defined to be half of the ℓ_1 norm of $\mu - \lambda$; that is, $\|\lambda - \mu\|_1 \overset{\mathrm{def}}{=} \frac{1}{2}\sum_x |\lambda(x) - \mu(x)| = \max_{S \subseteq \mathcal{X}} |\lambda_S - \mu_S|$, where $\lambda_S \overset{\mathrm{def}}{=} \sum_{x \in S} \lambda(x)$. We say that λ is ε-close to μ if $\|\lambda - \mu\|_1 \leq \varepsilon$. The relative entropy between distributions X and Y on $\mathcal{X}$ is defined as $\mathrm{S}(X\|Y) \overset{\mathrm{def}}{=} \mathbb{E}_{x \leftarrow X}\left[\log \frac{\Pr[X=x]}{\Pr[Y=x]}\right]$. The relative min-entropy between them is defined as $\mathrm{S}_\infty(X\|Y) \overset{\mathrm{def}}{=} \max_{x \in \mathcal{X}}\left\{\log \frac{\Pr[X=x]}{\Pr[Y=x]}\right\}$. It is easy to see that $\mathrm{S}(X\|Y) \leq \mathrm{S}_\infty(X\|Y)$. Let X, Y, Z be jointly distributed random variables. Let Y_x denote the distribution of Y conditioned on $X = x$. The conditional entropy of Y conditioned on X is defined as $\mathrm{H}(Y|X) \overset{\mathrm{def}}{=} \mathbb{E}_{x \leftarrow X}[\mathrm{H}(Y_x)] = \mathrm{H}(XY) - \mathrm{H}(X)$. The mutual information between X and Y is defined as: $\mathrm{I}(X : Y) \overset{\mathrm{def}}{=} \mathrm{H}(X) + \mathrm{H}(Y) - \mathrm{H}(XY) = \mathbb{E}_{y \leftarrow Y}[\mathrm{S}(X_y\|X)] = \mathbb{E}_{x \leftarrow X}[\mathrm{S}(Y_x\|Y)]$. It is easily seen that $\mathrm{I}(X : Y) = \mathrm{S}(XY\|X \otimes Y)$. We say that X and Y are independent iff $\mathrm{I}(X : Y) = 0$. The conditional mutual information between X and Y, conditioned on Z, is defined as: $\mathrm{I}(X : Y|Z) \overset{\mathrm{def}}{=} \mathbb{E}_{z \leftarrow Z}[\mathrm{I}(X : Y|Z = z)] = \mathrm{H}(X|Z) + \mathrm{H}(Y|Z) - \mathrm{H}(XY|Z)$. The following *chain rule* for mutual information is easily seen : $\mathrm{I}(X : YZ) = \mathrm{I}(X : Z) + \mathrm{I}(X : Y|Z)$. Let X, X', Y, Z be jointly distributed random variables. We define the joint distribution of $(X'Z)(Y|X)$ by: $\Pr[(X'Z)(Y|X) = x, z, y] \overset{\mathrm{def}}{=} \Pr[X' = x, Z = z] \cdot \Pr[Y = y|X = x]$. We say that X, Y, Z is a Markov chain iff $XYZ = (XY)(Z|Y)$ and we denote it by $X \leftrightarrow Y \leftrightarrow Z$. It is easy to see that X, Y, Z is a Markov chain if and only if $\mathrm{I}(X : Z|Y) = 0$. Ibinson, Linden and Winter [10] showed that if $\mathrm{I}(X : Y|Z)$ is small then XYZ is close to being a Markov chain.

Lemma II.1 ([10])**.** *For any random variables X, Y and Z, it holds that*

$$\mathrm{I}(X : Z|Y) = \min\left\{\mathrm{S}(XYZ\|X'Y'Z') : X' \leftrightarrow Y' \leftrightarrow Z'\right\}.$$

The minimum is achieved by distribution $X'Y'Z' = (XY)(Z|Y)$.

We will need the following basic facts. A very good text for reference on information theory is [7].

Fact II.2. Relative entropy is jointly convex in its arguments. That is, for distributions $\mu, \mu^1, \lambda, \lambda^1 \in \mathcal{X}$ and $p \in [0,1]$: $\mathrm{S}\left(p\mu + (1-p)\mu^1 \middle\| \lambda + (1-p)\lambda^1\right) \leq p \cdot \mathrm{S}(\mu\|\lambda) + (1-p) \cdot \mathrm{S}\left(\mu^1\middle\|\lambda^1\right)$.

Fact II.3. Relative entropy satisfies the following chain rule. Let XY and X^1Y^1 be random variables on $\mathcal{X} \times \mathcal{Y}$. It holds that: $\mathrm{S}\left(X^1Y^1\middle\|XY\right) = \mathrm{S}\left(X^1\middle\|X\right) + \mathbb{E}_{x \leftarrow X^1}\left[\mathrm{S}\left(Y^1_x\middle\|Y_x\right)\right]$. In particular, using Fact II.2: $\mathrm{S}\left(X^1Y^1\middle\|X \otimes Y\right) = \mathrm{S}\left(X^1\middle\|X\right) + \mathbb{E}_{x \leftarrow X^1}\left[\mathrm{S}\left(Y^1_x\middle\|Y\right)\right] \geq \mathrm{S}\left(X^1\middle\|X\right) + \mathrm{S}\left(Y^1\middle\|Y\right)$.

Fact II.4. Let XY and X^1Y^1 be random variables on $\mathcal{X} \times \mathcal{Y}$. It holds that

$$\mathrm{S}\left(X^1Y^1\middle\|X \otimes Y\right) \geq \mathrm{S}\left(X^1Y^1\middle\|X^1 \otimes Y^1\right) = \mathrm{I}\left(X^1 : Y^1\right).$$

Fact II.5. For distributions λ and μ: $0 \leq \|\lambda - \mu\|_1 \leq \sqrt{\mathrm{S}(\lambda\|\mu)}$.

Fact II.6. Let λ and μ be distributions on $\mathcal{X}$. For any subset $\mathcal{S} \subseteq \mathcal{X}$, it holds that: $\sum_{x \in \mathcal{S}} \lambda(x) \cdot \log \frac{\lambda(x)}{\mu(x)} \geq -1$.

Fact II.7. The ℓ_1 distance and relative entropy are monotone non-increasing when subsystems are considered. Let XY and X^1Y^1 be random variables on $\mathcal{X} \times \mathcal{Y}$, then

$$\left\|XY - X^1Y^1\right\|_1 \geq \left\|X - X^1\right\|_1 \quad \text{and}$$
$$\mathrm{S}\left(XY\middle\|X^1Y^1\right) \geq \mathrm{S}\left(X\middle\|X^1\right).$$

Fact II.8. For function $f : \mathcal{X} \times \mathcal{R} \to \mathcal{Y}$ and random variables X, X_1 on $\mathcal{X}$ and R on $\mathcal{R}$, such that R is independent of (XX_1), it holds that: $\|Xf(X,R) - X_1f(X_1,R)\|_1 = \|X - X_1\|_1$.

The following definition was introduced by Holenstein [9]. It plays a critical role in his proof of a parallel repetition theorem for two-prover games.

Definition II.9 ([9])**.** For two distributions (X_0Y_0) and (X_1SY_1T), we say that (X_0, Y_0) is $(1-\varepsilon)$-embeddable in (X_1S, Y_1T) if there exists a probability distribution R over a set $\mathcal{R}$, which is independent of X_0Y_0 and functions $f_A : \mathcal{X} \times \mathcal{R} \to \mathcal{S}$, $f_B : \mathcal{Y} \times \mathcal{R} \to \mathcal{T}$, such that

$$\|X_0Y_0f_A(X_0,R)f_B(Y_0,R) - X_1Y_1ST\|_1 \leq \varepsilon.$$

The following lemma was shown by Holenstein [9] using a correlated sampling protocol.

Lemma II.10 ([9])**.** *For random variables S, X and Y, if*

$$\|XYS - (XY)(S|X)\|_1 \leq \varepsilon \quad \textit{and}$$
$$\|XYS - (XY)(S|Y)\|_1 \leq \varepsilon,$$

then (X,Y) is $(1-4\varepsilon)$-embeddable in (XS, YS).

We will need the following generalization of the previous lemma. Its proof appears in Appendix A.

Lemma II.11. *For joint random variables (A', B', C') and (A, B), satisfying*

$$\mathrm{S}(A'B'\|AB) \leq \varepsilon. \tag{1}$$
$$\mathbb{E}_{(a,c)\leftarrow A',C'}\left[\mathrm{S}\left(B'_{a,c}\middle\|B_a\right)\right] \leq \varepsilon, \tag{2}$$
$$\mathbb{E}_{(b,c)\leftarrow B',C'}\left[\mathrm{S}\left(A'_{b,c}\middle\|A_b\right)\right] \leq \varepsilon, \tag{3}$$

it holds that (A,B) is $(1-5\sqrt{\varepsilon})$-embeddable in $(A'C', B'C')$.

Communication complexity

Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation, $t \geq 1$ be an integer and $\varepsilon \in (0,1)$. In this work we only consider *complete* relations, that is for every $(x,y) \in \mathcal{X} \times \mathcal{Y}$, there is some $z \in \mathcal{Z}$ such that $(x,y,z) \in f$. In the two-party t-message public-coin model of communication, Alice with input $x \in \mathcal{X}$ and Bob with input $y \in \mathcal{Y}$, do local computation using public coins shared between them and exchange t messages, with Alice sending the first message. At the end of their protocol the party receiving the t-th message outputs some $z \in \mathcal{Z}$. The output is declared correct if $(x,y,z) \in f$ and wrong otherwise. Let $\mathrm{R}_\varepsilon^{(t),\mathrm{pub}}(f)$ represent the two-party t-message public-coin communication complexity of f with worst case error ε, i.e., the communication of the best two-party t-message public-coin protocol for f with error for each input (x,y) being at most ε. We similarly consider two-party t-message deterministic protocols where there are no public coins used by Alice and Bob. Let $\mu \in \mathcal{X} \times \mathcal{Y}$ be a distribution. We let $\mathrm{D}_\varepsilon^{(t),\mu}(f)$ represent the two-party t-message distributional communication complexity of f under μ with expected error ε, i.e., the communication of the best two-party t-message deterministic protocol for f, with distributional error (average error over the inputs) at most ε under μ. The following is a consequence of the min-max theorem in game theory, see e.g., [23, Theorem 3.20, page 36].

Lemma II.12 (Yao's principle, [35])**.** $\mathrm{R}_\varepsilon^{(t),\mathrm{pub}}(f) = \max_\mu \mathrm{D}_\varepsilon^{(t),\mu}(f)$.

The following fact about communication protocols can be verified easily.

Fact II.13. Let there be t messages $M_1, \ldots, M_t$ in a deterministic communication protocol between Alice and Bob with inputs X, Y respectively where X and Y are independent. Then for any $s \in [t]$, X and Y are independent even conditioned on $M_1, \ldots, M_s$.

Let $f^k \subseteq \mathcal{X}^k \times \mathcal{Y}^k \times \mathcal{Z}^k$ be defined to be cross product of f with itself k times. In a protocol for computing f^k, Alice will receive input in $\mathcal{X}^k$, Bob will receive input in $\mathcal{Y}^k$ and the output of the protocol will be in $\mathcal{Z}^k$.

III. Proof of Theorem I.1

We start by showing a few lemmas which are helpful in the proof of the main result. The following lemma was shown in Jain [11] and follows primarily from a message compression argument due to Braverman and Rao [5].

Theorem III.1. *Let $\delta > 0, c \geq 0$. Let X', Y', N be random variables for which $Y' \leftrightarrow X' \leftrightarrow N$ is a Markov chain and the following holds,*

$$\Pr_{(x,y,m)\leftarrow X',Y',N}\left[\log\frac{\Pr[N=m|X'=x]}{\Pr[N=m|Y'=y]} > c\right] \leq \delta. \tag{4}$$

There exists a public-coin protocol between Alice and Bob, with inputs X', Y' respectively, with a single message from Alice to Bob of $c + \mathcal{O}(\log(1/\delta))$ bits, such that *at the end of the protocol, both Alice and Bob possess a random variable M satisfying $\|X'Y'N - X'Y'M\|_1 \leq 2\delta$.*

Remark III.2. In [5], the condition $\mathrm{I}(X' : N|Y') \leq c$ is used instead of (4). It is changed to the current one in Jain [11]. By Markov's inequality, $\mathrm{I}(X' : N|Y') \leq c$ implies

$$\Pr_{(x,y,m)\leftarrow X',Y',N}\left[\log \frac{\Pr[N=m|X'=x]}{\Pr[N=m|Y'=y]} > \frac{c+1}{\delta}\right] \leq \delta.$$

This modification is essential in our arguments since the condition (4) is robust when the underlying joint distribution is perturbed slightly, while $\mathrm{I}\left(X' : N|Y'\right)$ may change a lot with such a perturbation.

As mentioned in Section 1, we will have to work with approximate Markov chains in our arguments for the direct product. The following lemma makes Theorem I.1 more robust to deal with approximate Markov chains. Its proof appears in the Appendix A.

Lemma III.3. *Let $c \geq 0, 1 > \varepsilon > 0, \varepsilon' > 0$. Let X', Y', M' be random variables for which the following holds,*

$$\mathrm{I}(X' : M'|Y') \leq c \text{ and } \mathrm{I}(Y' : M'|X') \leq \varepsilon.$$

There exists a public-coin protocol between Alice and Bob, with inputs X', Y' respectively, with a single message from Alice to Bob of $\frac{c+5}{\varepsilon'} + \mathcal{O}(\log \frac{1}{\varepsilon'})$ bits, such that at the end of the protocol, both Alice and Bob possess a random variable M satisfying $\|X'Y'M' - X'Y'M\|_1 \leq 3\sqrt{\varepsilon} + 6\varepsilon'$.

The following lemma generalizes the lemma above to deal with multiple messages, as needed for our purposes. Its proof appears in the Appendix A.

Lemma III.4. *Let $t \geq 1$ be an integer. Let $\varepsilon' > 0$, $c_s \geq 0, 1 > \varepsilon_s > 0$ for each $1 \leq s \leq t$. Let $R', X', Y', M'_1, \ldots, M'_t$, be random variables for which the following holds (below $M'_{<s} \overset{\mathrm{def}}{=} M'_1 \cdots M'_{s-1}$),*

$$\mathrm{I}(X' : M'_s|Y'R'M'_{<s}) \leq c_s, \mathrm{I}(Y' : M'_s|X'R'M'_{<s}) \leq \varepsilon_s, \quad (5)$$

for odd s, and

$$\mathrm{I}(Y' : M'_s|X'R'M'_{<s}) \leq c_s, \mathrm{I}(X' : M'_s|Y'R'M'_{<s}) \leq \varepsilon_s,$$

for even s.

There exists a public-coin t-message protocol $\mathcal{P}_t$ between Alice, with input $X'R'$, and Bob, with input $Y'R'$, with Alice sending the first message. The total communication of $\mathcal{P}_t$ is $\frac{\sum_{s=1}^{t} c_s + 5t}{\varepsilon'} + \mathcal{O}\left(t \log \frac{1}{\varepsilon'}\right)$, and at end of the protocol, both Alice and Bob possess random variables $M_1, \ldots, M_t$, satisfying: $\|R'X'Y'M_1 \cdots M_t - R'X'Y'M'_1 \cdots M'_t\|_1 \leq 3\sum_{s=1}^{t} \sqrt{\varepsilon_s} + 6\varepsilon' t$.

In the lemma above, Alice and Bob shared an input R' (potentially correlated with $X'Y'$). Eventually we will need Alice and Bob to generate this shared part themselves using correlated sampling. The following lemma, obtained from the lemma above, is the one that we will finally use in the proof of our main result. Its proof appears in the Appendix A.

Lemma III.5. *Let random variables $R', X', Y', M'_1, \ldots, M'_t$ and numbers $\varepsilon', c_s, \varepsilon_s$ satisfy all the conditions in Lemma III.4. Let $\tau > 0$ and let random variables (X, Y) be $(1-\tau)$-embeddable in $(X'R', Y'R')$. There exists a public-coin t-message protocol $\mathcal{Q}_t$ between Alice, with input X, and Bob, with input Y, with Alice sending the first message, and total communication $\frac{\sum_{s=1}^{t} c_s + 5t}{\varepsilon'} + \mathcal{O}\left(t \log \frac{1}{\varepsilon'}\right)$ bits, such that at the end Alice possesses $R_A M_1 \cdots M_t$ and Bob possesses $R_B M_1 \cdots M_t$, such that: $\|XYR_AR_BM_1 \cdots M_t - X'Y'R'R'M'_1 \cdots M'_t\|_1 \leq \tau + 3\sum_{s=1}^{t} \sqrt{\varepsilon_s} + 6\varepsilon' t$.*

We are now ready to prove our main result, Theorem I.1. We restate it here for convenience.

Theorem I.1. *Let $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$ be finite sets, $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ a relation, $\varepsilon > 0$ and $k, t \geq 1$ be integers. There exists a constant $\kappa \geq 0$ such that,*

$$\mathrm{R}^{(t),\mathrm{pub}}_{1-(1-\varepsilon/2)^{\Omega(k\varepsilon^2/t^2)}}(f^k) = \Omega\left(\frac{\varepsilon \cdot k}{t} \cdot \left(\mathrm{R}^{(t),\mathrm{pub}}_{\varepsilon}(f) - \frac{\kappa t^2}{\varepsilon}\right)\right).$$

Proof of Theorem I.1: Let $\delta \overset{\mathrm{def}}{=} \frac{\varepsilon^2}{7500t^2}$ and $\delta_1 = \frac{\varepsilon}{3000t}$. From Yao's principle, Lemma II.12, it suffices to prove that for any distribution μ on $\mathcal{X} \times \mathcal{Y}$, $\mathrm{D}^{(t),\mu^k}_{1-(1-\varepsilon/2)^{\lfloor \delta k \rfloor}}(f^k) \geq \delta_1 kc$, where $c \overset{\mathrm{def}}{=} \mathrm{D}^{(t),\mu}_{\varepsilon}(f) - \frac{\kappa t^2}{\varepsilon}$, for constant κ to be chosen later. Let $XY \sim \mu^k$. Let $\mathcal{Q}$ be a t-message deterministic protocol between Alice, with input X, and Bob, with input Y, that computes f^k, with Alice sending the first message and total communication $\delta_1 kc$ bits. We assume t is odd for the rest of the argument and Bob makes the final output (the case when t is even follows similarly). The following Claim III.6 implies that the success of $\mathcal{Q}$ is at most $(1-\varepsilon/2)^{\lfloor \delta k \rfloor}$ and this shows the desired. □

Claim III.6. For each $i \in [k]$, define a binary random variable $T_i \in \{0, 1\}$, which represents the success of $\mathcal{Q}$ (that is Bob's output being correct) on the i-th instance. That is, $T_i = 1$ if the protocol $\mathcal{Q}$ computes the i-th instance of f correctly, and $T_i = 0$ otherwise. Let $k' \overset{\text{def}}{=} \lfloor \delta k \rfloor$. There exist k' coordinates $\{i_1, \ldots, i_{k'}\}$ such that for each $1 \leq r \leq k'-1$,

$$\text{either} \quad \Pr\left[T^{(r)} = 1\right] \leq (1-\varepsilon/2)^{k'} \quad \text{or}$$

$$\Pr\left[T_{i_{r+1}} = 1 \middle| T^{(r)} = 1\right] \leq 1 - \varepsilon/2,$$

where $T^{(r)} \overset{\text{def}}{=} \prod_{j=1}^{r} T_{i_j}$.

Proof of Claim III.6: For $s \in [t]$, denote the s-th message of $\mathcal{Q}$ by M_s. Define $M \overset{\text{def}}{=} M_1 \cdots M_t$. In the following we assume $1 \leq r < k'$, however same arguments also work when $r = 0$, that is for identifying the first coordinate, which we skip for the sake of avoiding repetition. Suppose we have already identified r coordinates $i_1, \ldots, i_r$ satisfying that $\Pr[T_{i_1} = 1] \leq 1 - \varepsilon/2$ and $\Pr[T_{i_{j+1}} = 1 | T^{(j)} = 1] \leq 1 - \varepsilon/2$ for $1 \leq j \leq r-1$. If $\Pr\left[T^{(r)} = 1\right] \leq (1-\varepsilon/2)^{k'}$, we are done. So from now on, assume $\Pr\left[T^{(r)} = 1\right] > (1-\varepsilon/2)^{k'} \geq 2^{-\delta k}$.

Let D be a random variable uniformly distributed in $\{0,1\}^k$ and independent of XY. Let $U_i = X_i$ if $D_i = 0$, and $U_i = Y_i$ if $D_i = 1$. For any random variable L, let us introduce the notation: $L^1 \overset{\text{def}}{=} (L | T^{(r)} = 1)$. For example, $X^1Y^1 = (XY | T^{(r)} = 1)$. If $L = L_1 \cdots L_k$, define $L_{-i} \overset{\text{def}}{=} L_1 \cdots L_{i-1} L_{i+1} \cdots L_k$, and $L_{<i} \overset{\text{def}}{=} L_1 \cdots L_{i-1}$. Random variable $L_{\leq i}$ is defined analogously. Let $C \overset{\text{def}}{=} \{i_1, \ldots, i_r\}$. Define $R_i \overset{\text{def}}{=} D_{-i} U_{-i} X_{C \cup [i-1]} Y_{C \cup [i-1]}$ for $i \in [k]$. We denote an element from the range of R_i by r_i. [2]

To prove the claim, we will show that there exists a coordinate $j \notin C$ such that,

1) $(X_j Y_j)$ can be embedded well in $(X_j^1 R_j^1, Y_j^1 R_j^1)$ (with appropriate parameters as required in Lemma II.11.)
2) Random variables $R_j^1, X_j^1, Y_j^1, M_1^1, \ldots, M_t^1$ satisfy the conditions of Lemma III.4 with appropriate parameters.

The following calculations are helpful for achieving condition (1) of Lemma II.11.

$$\begin{aligned} \delta k &> \mathrm{S}_\infty\left(X^1Y^1 \middle\| XY\right) \geq \mathrm{S}\left(X^1Y^1 \middle\| XY\right) \\ &\geq \sum_{i \notin C} \mathrm{S}\left(X_i^1 Y_i^1 \middle\| X_i Y_i\right), \end{aligned} \tag{6}$$

where first inequality follows from the assumption that $\Pr\left[T^{(r)} = 1\right] > 2^{-\delta k}$, and the last inequality follows from Fact II.3. The following calculations are helpful for achieving conditions (2) and (3) of Lemma II.11.

$$\begin{aligned} \delta k &> \mathrm{S}_\infty\left(X^1Y^1D^1U^1 \middle\| XYDU\right) \\ &\geq \mathrm{S}\left(X^1Y^1D^1U^1 \middle\| XYDU\right) \\ &\geq \mathop{\mathbb{E}}_{\substack{(d,u,x_C,y_C) \\ \leftarrow D^1,U^1,X_C^1,Y_C^1}} \left[\mathrm{S}\left(\left(X^1Y^1\right)_{d,u,x_C,y_C} \middle\| (XY)_{d,u,x_C,y_C}\right)\right] \end{aligned} \tag{7}$$

$$\begin{aligned} &= \sum_{i \notin C} \mathop{\mathbb{E}}_{\substack{(d,u,x_{C\cup[i-1]},y_{C\cup[i-1]}) \\ \leftarrow D^1,U^1,X^1_{C\cup[i-1]},Y^1_{C\cup[i-1]}}} \\ &\left[\mathrm{S}\left((X_i^1Y_i^1)_{\substack{d,u,x_{C\cup[i-1]}, \\ y_{C\cup[i-1]}}} \middle\| (X_iY_i)_{\substack{d,u,x_{C\cup[i-1]}, \\ y_{C\cup[i-1]}}} \right)\right] \end{aligned} \tag{8}$$

$$= \sum_{i \notin C} \mathop{\mathbb{E}}_{\substack{(d_i,u_i,r_i) \\ \leftarrow D_i^1,U_i^1,R_i^1}} \left[\mathrm{S}\left((X_i^1Y_i^1)_{d_i,u_i,r_i} \middle\| (X_iY_i)_{d_i,u_i,r_i}\right)\right] \tag{9}$$

$$\begin{aligned} &= \frac{1}{2} \sum_{i \notin C} \mathop{\mathbb{E}}_{(r_i,x_i) \leftarrow R_i^1,X_i^1} \left[\mathrm{S}\left(\left(Y_i^1\right)_{r_i,x_i} \middle\| \left(Y_i\right)_{x_i}\right)\right] + \\ &\frac{1}{2} \sum_{i \notin C} \mathop{\mathbb{E}}_{(r_i,y_i) \leftarrow R_i^1,Y_i^1} \left[\mathrm{S}\left(\left(X_i^1\right)_{r_i,y_i} \middle\| \left(X_i\right)_{y_i}\right)\right]. \end{aligned} \tag{10}$$

Above, Eq. (7) and Eq. (8) follow from Fact II.3; Eq. (9) is from the definition of R_i. Eq. (10) follows since D_i^1 is independent of R_i^1 and with probability half D_i^1 is 0, in which case $U_i^1 = X_i^1$ and with probability half D_i^1 is 1 in which case $U_i^1 = Y_i^1$.

The following calculations are useful for achieving partly the conditions in (5) exhibiting that the information carried by messages about sender's input is small.

$$\begin{aligned} \delta_1 ck &\geq \left|M^1\right| \geq \mathrm{I}\left(X^1Y^1 : M^1 \middle| D^1U^1X_C^1Y_C^1\right) \\ &= \sum_{i \notin C} \mathrm{I}\left(X_i^1Y_i^1 : M^1 \middle| D^1U^1X^1_{C\cup[i-1]}Y^1_{C\cup[i-1]}\right) \\ &= \sum_{i \notin C} \sum_{s=1}^{t} \mathrm{I}\left(X_i^1Y_i^1 : M_s^1 \middle| D^1U^1X^1_{C\cup[i-1]}Y^1_{C\cup[i-1]}M^1_{<s}\right) \\ &= \sum_{i \notin C} \sum_{s=1}^{t} \mathrm{I}\left(X_i^1Y_i^1 : M_s^1 \middle| D_i^1U_i^1R_i^1M^1_{<s}\right) \end{aligned} \tag{11}$$

[2] We justify here the composition of R_i. Random variables $D_{-i}U_{-i}$ are useful since conditioning on them makes the distribution of inputs product across Alice and Bob (for fixed values of X_iY_i) and is helpful in our arguments later. Random variables X_CY_C are helpful since conditioning on them ensures that the inputs become product even conditioned on success on C. Random variables $X_{[i-1]}Y_{[i-1]}$ are helpful since the following chain rule is used to draw a new coordinate outside C with low information content:

$$\mathrm{I}(XY : M) = \sum_i \mathrm{I}\left(X_iY_i : M \middle| X_{[i-1]}Y_{[i-1]}\right).$$

$$= \sum_{i \notin C} \left(\left(\sum_{s \text{ odd}} + \sum_{s \text{ even}} \right) \mathrm{I}\big(X_i^1 Y_i^1 : M_s^1 \big| D_i^1 U_i^1 R_i^1 M_{<s}^1\big) \right)$$

$$\geq \frac{1}{2} \sum_{i \notin C} \big(\sum_{s \text{ odd}} \mathrm{I}\big(X_i^1 : M_s^1 \big| R_i^1 Y_i^1 M_{<s}^1\big) + \sum_{s\text{even}} \mathrm{I}\big(Y_i^1 : M_s^1 \big| R_i^1 X_i^1 M_{<s}^1\big) \big). \tag{12}$$

Above we have used the chain rule for mutual information several times. Last inequality follows since D_i^1 is independent of $(X_i^1 Y_i^1 R_i^1 M^1)$ and with probability half D_i^1 is 0, in which case $U_i^1 = X_i^1$ and with probability half D_i^1 is 1 in which case $U_i^1 = Y_i^1$.

The following calculations are useful for achieving partly the conditions in (5) exhibiting that the information carried by messages about receiver's input is very small. Here we are only able to argue round by round and hence pay a factor proportional to the number of messages in the final result. Let $s \in [t]$ be odd.

$$\delta k \geq \mathrm{S}_\infty\big(D^1 U^1 X^1 Y^1 M_{\leq s}^1 \big\| DUXYM_{\leq s}\big)$$

$$\geq \mathrm{S}\big(D^1 U^1 X^1 Y^1 M_{\leq s}^1 \big\| DUXYM_{\leq s}\big)$$

$$\geq \mathbb{E}_{(d,u,x_C,y_C,m_{\leq s}) \leftarrow D^1, U^1, X_C^1, Y_C^1, M_{\leq s}^1} \Big[\mathrm{S}\big((X^1Y^1)_{d,u,x_C,y_C,m_{\leq s}} \big\| (XY)_{d,u,x_C,y_C,m_{\leq s}}\big)\Big]$$

$$= \sum_{i \notin C} \mathop{\mathbb{E}}_{\substack{(d,u,x_{C\cup[i-1]},y_{C\cup[i-1]},m_{\leq s}) \\ \leftarrow D^1, U^1, X^1_{C\cup[i-1]}, Y^1_{C\cup[i-1]}, M^1_{\leq s}}} \left[\mathrm{S}\left((X_i^1 Y_i^1)_{\substack{d,u,x_{C\cup[i-1]}, \\ y_{C\cup[i-1]},m_{\leq s}}} \middle\| (X_iY_i)_{\substack{d,u,x_{C\cup[i-1]}, \\ y_{C\cup[i-1]},m_{\leq s}}}\right)\right]$$

$$= \sum_{i \notin C} \mathop{\mathbb{E}}_{\substack{(d_i,u_i,r_i,m_{\leq s}) \\ \leftarrow D_i^1, U_i^1, R_i^1, M_{\leq s}^1}} \Big[\mathrm{S}\big((X_i^1 Y_i^1)_{d_i,u_i,r_i,m_{\leq s}} \big\| (X_iY_i)_{d_i,u_i,r_i,m_{\leq s}}\big)\Big] \tag{13}$$

$$\geq \frac{1}{2} \sum_{i \notin C} \mathop{\mathbb{E}}_{\substack{(x_i,r_i,m_{\leq s}) \\ \leftarrow X_i^1, R_i^1, M_{\leq s}^1}} \Big[\mathrm{S}\big((Y_i^1)_{x_i,r_i,m_{\leq s}} \big\| (Y_i)_{x_i,r_i,m_{\leq s}}\big)\Big]$$

$$= \frac{1}{2} \sum_{i \notin C} \mathop{\mathbb{E}}_{\substack{(x_i,r_i,m_{\leq s}) \\ \leftarrow X_i^1, R_i^1, M_{\leq s}^1}} \Big[\mathrm{S}\big((Y_i^1)_{x_i,r_i,m_{\leq s}} \big\| (Y_i)_{x_i,r_i,m_{< s}}\big)\Big] \tag{14}$$

$$= \frac{1}{2} \sum_{i \notin C} \mathbb{E}_{(x_i,r_i,m_{<s}) \leftarrow X_i^1, R_i^1, M_{<s}^1} \Big[\mathrm{S}\big((Y_i^1 M_s^1)_{x_i,r_i,m_{<s}} \big\| (Y_i)_{x_i,r_i,m_{<s}} \otimes (M_s^1)_{x_i,r_i,m_{<s}}\big)\Big]$$

$$\geq \frac{1}{2} \sum_{i \notin C} \mathop{\mathbb{E}}_{\substack{(x_i,r_i,m_{<s}) \\ \leftarrow X_i^1, R_i^1, M_{<s}^1}} \Big[\mathrm{I}\big((Y_i^1)_{x_i,r_i,m_{<s}} : (M_s^1)_{x_i,r_i,m_{<s}}\big)\Big] \tag{15}$$

$$= \frac{1}{2} \sum_{i \notin C} \mathrm{I}\big(Y_i^1 : M_s^1 \big| X_i^1 R_i^1 M_{<s}^1\big). \tag{16}$$

Above we have used Fact II.3 several times. Eq. (13) follows from the definition of R_i; Eq. (14) follows from the fact that $Y \leftrightarrow X_i R_i M_{<s} \leftrightarrow M_s$ for any i, whenever s is odd; Eq. (15) follows from Fact II.4. From a symmetric argument, we can show that when $s \in [t]$ is even, $\frac{1}{2}\sum_{i \notin C} \mathrm{I}\big(X_i^1 : M_s^1 \big| Y_i^1 R_i^1 M_{<s}^1\big) \leq \delta k$. This and Eq. (16) together imply

$$\sum_{i \notin C} \big(\sum_{s \text{ odd}} \mathrm{I}\big(Y_i^1 : M_s^1 \big| R_i^1 X_i^1 M_{<s}^1\big) + \sum_{s \text{ even}} \mathrm{I}\big(X_i^1 : M_s^1 \big| R_i^1 Y_i^1 M_{<s}^1\big) \big) \leq 2\delta kt. \tag{17}$$

Combining Equations (6)(10)(12)(17), and making standard use of Markov's inequality, we can get a coordinate $j \notin C$ such that

$$\mathrm{S}\big(X_j^1 Y_j^1 \big\| X_j Y_j\big) \leq 12\delta, \tag{18}$$

$$\mathop{\mathbb{E}}_{\substack{(r_j,x_j) \\ \leftarrow R_j^1, X_j^1}} \Big[\mathrm{S}\Big(\big(Y_j^1\big)_{r_j,x_j} \Big\| \big(Y_j\big)_{x_j}\Big)\Big] \leq 12\delta, \tag{19}$$

$$\mathop{\mathbb{E}}_{\substack{(r_j,y_j) \\ \leftarrow R_j^1, Y_j^1}} \Big[\mathrm{S}\Big(\big(X_j^1\big)_{r_j,y_j} \Big\| \big(X_j\big)_{y_j}\Big)\Big] \leq 12\delta, \tag{20}$$

$$\sum_{s \text{ odd}} \mathrm{I}\big(X_j^1 : M_s^1 \big| R_j^1 Y_j^1 M_{<s}^1\big) + \sum_{s \text{ even}} \mathrm{I}\big(Y_j^1 : M_s^1 \big| R_j^1 X_j^1 M_{<s}^1\big) \leq 12\delta_1 c, \tag{21}$$

$$\sum_{s \text{ odd}} \mathrm{I}\big(Y_j^1 : M_s^1 \big| R_j^1 X_j^1 M_{<s}^1\big) + \sum_{s \text{ even}} \mathrm{I}\big(X_j^1 : M_s^1 \big| R_j^1 Y_j^1 M_{<s}^1\big) \leq 12\delta t. \tag{22}$$

Set $\varepsilon' \overset{\text{def}}{=} \frac{\varepsilon}{125t}$, and

$$\varepsilon_s \overset{\text{def}}{=} \begin{cases} \mathrm{I}\big(Y_j^1 : M_s^1 \big| R_j^1 X_j^1 M_{<s}^1\big) & s \in [t] \text{ odd}, \\ \mathrm{I}\big(X_j^1 : M_s^1 \big| R_j^1 Y_j^1 M_{<s}^1\big) & s \in [t] \text{ even}. \end{cases};$$

$$c_s \overset{\text{def}}{=} \begin{cases} \mathrm{I}\big(Y_j^1 : M_s^1 \big| R_j^1 X_j^1 M_{<s}^1\big) & s \in [t] \text{ even}, \\ \mathrm{I}\big(X_j^1 : M_s^1 \big| R_j^1 Y_j^1 M_{<s}^1\big) & s \in [t] \text{ odd}. \end{cases}$$

By (22), $\sum_{s=1}^t \sqrt{\varepsilon_s} \leq \sqrt{12\delta t}$. From Equations (18)(19)(20) and Lemma II.11 we can infer that (X_jY_j) is $(1 - 10\sqrt{3\delta})$-embeddable in $(X_j^1 R_j^1; Y_j^1 R_j^1)$. This, combined with Equations (21)(22) and Lemma III.5 (take $\varepsilon', \varepsilon_s, c_s$ in the lemma to be as defined above and take $XYX'Y'R'M_1' \cdots M_t'$ in the lemma to be $X_jY_jX_j^1Y_j^1R_j^1M_1^1 \cdots M_t^1$) imply the following (for appropriate constant κ). There exists a public-coin t-message protocol $\mathcal{Q}^1$ between Alice, with input X_j, and Bob, with input Y_j, with Alice sending the first message and total communication, $\frac{12\delta_1 c + 5t}{\varepsilon'} +$

$\mathcal{O}(t \log \frac{1}{\varepsilon'}) < \mathrm{D}_{\varepsilon}^{(t),\mu}(f)$, such that at the end Alice possesses $R_A M_1 \cdots M_t$ and Bob possesses $R_B M_1 \cdots M_t$, satisfying

$$\begin{aligned}&\left\| X_j Y_j R_A R_B M_1 \cdots M_t - X_j^1 Y_j^1 R_j^1 R_j^1 M_1^1 \cdots M_t^1 \right\|_1 \\ &\leq 10\sqrt{3\delta} + 3\sqrt{12\delta}t + 6\varepsilon' t < \varepsilon/2.\end{aligned}$$

Assume for contradiction that $\Pr\left[T_j = 1 \middle| T^{(r)} = 1\right] > 1 - \varepsilon/2$. Consider a protocol $\mathcal{Q}^2$ (with no communication) for f between Alice, with input $X_j^1 R_j^1 M_1^1 \cdots M_t^1$, and Bob, with input $Y_j^1 R_j^1 M_1^1 \cdots M_t^1$, as follows. Bob generates the rest of the random variables present in Y^1 (not present in his input) himself since, conditioned on his input, those other random variables are independent of Alice's input (here we use Fact II.13). Bob then generates the output for the j-th coordinate in $\mathcal{Q}$, and makes it the output of $\mathcal{Q}^2$. This ensures that the success probability of Bob in $\mathcal{Q}^2$ is $\Pr\left[T_j = 1 \middle| T^{(r)} = 1\right] > 1 - \varepsilon/2$. Now consider protocol $\mathcal{Q}^3$ for f, with Alice's input X_j and Bob's input Y_j, which is a composition of $\mathcal{Q}^1$ followed by $\mathcal{Q}^2$. This ensures, using Fact II.8, that success probability of Bob (averaged over public coins and the inputs $X_j Y_j$) in $\mathcal{Q}^3$ is larger than $1 - \varepsilon$. Finally by fixing the public coins of $\mathcal{Q}^3$, we get a deterministic protocol $\mathcal{Q}^4$ for f with Alice's input X_j and Bob's input Y_j such that the communication of $\mathcal{Q}^4$ is less than $\mathrm{D}_{\varepsilon}^{(t),\mu}(f)$ and Bob's success probability (averaged over the inputs $X_j Y_j$) in $\mathcal{Q}^4$ is larger than $1 - \varepsilon$. This is a contradiction to the definition of $\mathrm{D}_{\varepsilon}^{(t),\mu}(f)$ (recall that $X_j Y_j$ are distributed according to μ). Hence it must be that $\Pr\left[T_j = 1 \middle| T^{(r)} = 1\right] \leq 1 - \varepsilon/2$. The claim now follows by setting $i_{r+1} = j$. □

Open problems

Some natural questions that arise from this work are:

1) Can the dependence on t in our direct product theorem be improved?
2) Can these techniques be extended to show direct product theorems for bounded-round quantum communication complexity?

References

[1] Andris Ambainis, Robert Špalek, and Ronald de Wolf. A new quantum lower bound method, with applications to direct product theorems and time-space tradeoffs. *Algorithmica*, 55:422–461, 2009. 10.1007/s00453-007-9022-9.

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

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

[4] Avraham Ben-Aroya, Oded Regev, and Ronald de Wolf. A hypercontractive inequality for matrix-valued functions with applications to quantum computing. In *Proceedings of the 52nd Symposium on Foundations of Computer Science*, FOCS '08, pages 477–486, Washington, DC, USA, 2008. IEEE Computer Society.

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

[6] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Chi-Chih Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *Proceedings of the 42nd IEEE symposium on Foundations of Computer Science*, FOCS '01, pages 270–278, Washington, DC, USA, 2001. IEEE Computer Society.

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

[8] Andrew Drucker. Improved direct product theorems for randomized query complexity. In *Proceedings of the 2011 IEEE 26th Annual Conference on Computational Complexity*, CCC '11, pages 1–11, Washington, DC, USA, 2011. IEEE Computer Society.

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

[10] Ben Ibinson, Noah Linden, and Andreas Winter. Robustness of quantum markov chains. *Communications in Mathematical Physics*, 277:289–304, 2008. 10.1007/s00220-007-0362-8.

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

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

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

[14] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. The quantum communication complexity of the pointer chasing problem: The bit version. In *Proceedings of the 22nd Conference Kanpur on Foundations of Software Technology and Theoretical Computer Science*, FST TCS '02, pages 218–229, London, UK, 2002. Springer-Verlag.

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

[16] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A lower bound for the bounded round quantum communication complexity of set disjointness. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '03, pages 220–229, Washington, DC, USA, 2003. IEEE Computer Society.

[17] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. Prior entanglement, message compression and privacy in quantum communication. In *Proceedings of the 20th Annual IEEE Confer-*

ence on Computational Complexity, pages 285–296, Washington, DC, USA, 2005. IEEE Computer Society.

[18] Hartmut Klauck. On quantum and probabilistic communication: Las vegas and one-way protocols. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing*, STOC '00, pages 644–651, New York, NY, USA, 2000. ACM.

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

[20] Hartmut Klauck, Ashwin Nayak, Amnon Ta-Shma, and David Zuckerman. Interaction in quantum communication and the complexity of set disjointness. In *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, STOC '01, pages 124–133, New York, NY, USA, 2001. ACM.

[21] Hartmut Klauck, Robert Špalek, and Ronald de Wolf. Quantum and classical strong direct product theorems and optimal time-space tradeoffs. In *Proceedings of the 45th Symposium on Foundations of Computer Science*, FOCS '04, pages 12–21, Rome, ITALY, 2004. IEEE Computer Society.

[22] Hartmut Klauck, Robert Špalek, and Ronald de Wolf. Quantum and classical strong direct product theorems and optimal time-space tradeoffs. In *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*, pages 12–21, Washington, DC, USA, 2004. IEEE Computer Society.

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

[24] Troy Lee and Jérémie Roland. A strong direct product theorem for quantum query complexity. In *Proceedings of the 27th IEEE Conference on Computational Complexity*, CCC '12, pages 236–246, Washington, DC, USA, 2012. IEEE Computer Society.

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

[26] Noam Nisan, Steven Rudich, and Michael Saks. Products and help bits in decision trees. *SIAM J. Comput.*, 28:1035–1050, February 1999.

[27] Noam Nisan and Avi Widgerson. Rounds in communication complexity revisited. In *Proceedings of the twenty-third annual ACM symposium on Theory of computing*, STOC '91, pages 419–429, New York, NY, USA, 1991. ACM.

[28] Itzhak Parnafes, Ran Raz, and Avi Wigderson. Direct product results and the gcd problem, in old and new communication models. In *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, STOC '97, pages 363–372, New York, NY, USA, 1997. ACM.

[29] Stephen J. Ponzio, Jaikumar Radhakrishnan, and S. Venkatesh. The communication complexity of pointer chasing. *J. Comput. Syst. Sci.*, 62:323–355, March 2001.

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

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

[32] Ronen Shaltiel. Towards proving strong direct product theorems. *Comput. Complex.*, 12:1–22, July 2004.

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

[34] Emanuele Viola and Avi Wigderson. Norms, xor lemmas, and lower bounds for poly- nomials and protocols. *Theory of Computing*, 4(1):137–168, 2008.

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

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

Appendix

Proof of Lemma II.11: Using the definition of the relative entropy, we have the following.

$$\begin{aligned}
&\mathop{\mathbb{E}}_{(a,c)\leftarrow A',C'}\left[\mathrm{S}\left(B'_{a,c}\middle\|B_a\right)\right] - \mathop{\mathbb{E}}_{(a,c)\leftarrow A',C'}\left[\mathrm{S}\left(B'_{a,c}\middle\|B'_a\right)\right]\\
&= \mathop{\mathbb{E}}_{(a,b,c)\leftarrow A',B',C'}\left[\log\frac{\Pr[B'=b|A'=a]}{\Pr[B=b|A=a]}\right]\\
&= \mathop{\mathbb{E}}_{a\leftarrow A'}\left[\mathrm{S}(B'_a\|B_a)\right] \quad \geq \quad 0.
\end{aligned}$$

This means that

$$\begin{aligned}
&\mathop{\mathbb{E}}_{(a,c)\leftarrow A',C'}\left[\mathrm{S}\left(B'_{a,c}\middle\|B'_a\right)\right] && (23)\\
&\leq \mathop{\mathbb{E}}_{(a,c)\leftarrow A',C'}\left[\mathrm{S}\left(B'_{a,c}\middle\|B_a\right)\right] \leq \varepsilon. && (24)
\end{aligned}$$

Then

$$\begin{aligned}
&\mathop{\mathbb{E}}_{(a,c)\leftarrow A',C'}\left[\mathrm{S}\left(B'_{a,c}\middle\|B'_a\right)\right]\\
&= \mathrm{S}(A'C'B'\|(A'C')\,(B'|A')) && (25)\\
&= \mathrm{S}(A'B'C'\|(A'B')\,(C'|A')) && (26)\\
&\geq \left\|A'B'C' - (A'B')\,(C'|A')\right\|_1^2. && (27)
\end{aligned}$$

Above, Eq. (25) follows from the chain rule for the relative entropy, Eq. (26) follows because $(A'C')\,(B'|A')$ and $(A'B')\,(C'|A')$ are identically distributed, and Eq. (27) follows from Fact II.5. Now from Equations (27) and (24) we get

$$\left\|A'B'C' - (A'B')\,(C'|A')\right\|_1 \leq \sqrt{\varepsilon}.$$

By similar arguments we get

$$\left\|A'B'C' - (A'B')\,(C'|B')\right\|_1 \leq \sqrt{\varepsilon}.$$

The inequalities above and Lemma II.10 imply that (A', B') is $(1-4\sqrt{\varepsilon})$-embeddable in $(A'C', B'C')$. Furthermore from Fact II.5 and $\mathrm{S}(A'B'\|AB) \leq \varepsilon$ we get

$$\left\|A'B' - AB\right\|_1 \leq \sqrt{\varepsilon}.$$

Finally using the inequality above, Fact II.8 and the triangle inequality for the ℓ_1 norm, we get that (A, B) is $(1-5\sqrt{\varepsilon})$-embeddable in $(A'C', B'C')$. □

Proof of Lemma III.3: Let us introduce a new random variable N with joint distribution $X'Y'N \overset{\mathrm{def}}{=}$

$(X'Y')(M'|X')$. Note that $Y' \leftrightarrow X' \leftrightarrow N$ is a Markov chain. Using Lemma II.1, we have

$$\mathrm{S}(X'Y'M'\|X'Y'N) = \mathrm{I}(Y' : M'|X') \le \varepsilon. \tag{28}$$

Applying Fact II.5, we get $\|X'Y'M' - X'Y'N\|_1 \le \sqrt{\varepsilon}$. Theorem III.1 and the following claim together imply that there exists a public-coin protocol between Alice and Bob, with input X', Y', respectively, with a single message from Alice to Bob of $\frac{c+5}{\varepsilon'} + \mathcal{O}(\log \frac{1}{\varepsilon'})$ bits, at the end of which both Alice and Bob possess a random variable N' satisfying $\|X'Y'N' - X'Y'N\|_1 \le 2\sqrt{\varepsilon} + 6\varepsilon'$. Finally using the triangle inequality for the ℓ_1 norm we conclude the desired. □

Claim A.1.

$$\Pr_{(m,x,y)\leftarrow N,X',Y'}\left[\log \frac{\Pr[N=m|X'=x]}{\Pr[N=m|Y'=y]} \ge \frac{c+5}{\varepsilon'}\right] \le 3\varepsilon' + \sqrt{\varepsilon}.$$

Proof: For any m, x, y it holds that

$$\begin{aligned}
&\log \frac{\Pr[N=m|X'=x]}{\Pr[N=m|Y'=y]} \\
&= \log \frac{\Pr[N=m|X'=x,Y'=y]}{\Pr[N=m|Y'=y]} \\
&= \log \frac{\Pr[N=m|X'=x,Y'=y]}{\Pr[M'=m|X'=x,Y'=y]} \\
&\quad + \log \frac{\Pr[M'=m|X'=x,Y'=y]}{\Pr[M'=m|Y'=y]} \\
&\quad + \log \frac{\Pr[M'=m,Y'=y]}{\Pr[N=m,Y'=y]}.
\end{aligned} \tag{29}$$

From union bound and above we get (recall $1 > \varepsilon > 0$),

$$\begin{aligned}
&\Pr_{\substack{(m,x,y)\\ \leftarrow M',X',Y'}}\left[\log \frac{\Pr[N=m|X'=x]}{\Pr[N=m|Y'=y]} \ge \frac{c+5}{\varepsilon'}\right] \\
&= \Pr_{\substack{(m,x,y)\\ \leftarrow M',X',Y'}}\left[\log \frac{\Pr[N=m|X'=x,Y'=y]}{\Pr[N=m|Y'=y]} \ge \frac{c+5}{\varepsilon'}\right] \\
&\le \Pr_{\substack{(m,x,y)\\ \leftarrow M',X',Y'}}\left[\log \frac{\Pr[N=m|X'=x,Y'=y]}{\Pr[M'=m|X'=x,Y'=y]} \ge \frac{\varepsilon+1}{\varepsilon'}\right] \\
&+ \Pr_{\substack{(m,x,y)\\ \leftarrow M',X',Y'}}\left[\log \frac{\Pr[M'=m|X'=x,Y'=y]}{\Pr[M'=m|Y'=y]} \ge \frac{c+1}{\varepsilon'}\right] \\
&+ \Pr_{\substack{(m,x,y)\\ \leftarrow M',X',Y'}}\left[\log \frac{\Pr[M'=m,Y'=y]}{\Pr[N=m,Y'=y]} \ge \frac{\varepsilon+1}{\varepsilon'}\right].
\end{aligned} \tag{30}$$

We bound each term above separately. For the first one, let us define the set

$$G_1 \overset{\text{def}}{=} \left\{(m,x,y) : \log \frac{\Pr[N=m|X'=x,Y'=y]}{\Pr[M'=m|X'=x,Y'=y]} < \frac{\varepsilon+1}{\varepsilon'}\right\}.$$

Consider,

$$\begin{aligned}
0 &\ge - \mathop{\mathbb{E}}_{(x,y)\leftarrow X',Y'}\left[\mathrm{S}\left(M'_{xy}\|N_{xy}\right)\right] \\
&= \mathop{\mathbb{E}}_{(m,x,y)\leftarrow M',X',Y'}\left[\log \frac{\Pr[N=m|X'=x,Y'=y]}{\Pr[M'=m|X'=x,Y'=y]}\right] \qquad (31) \\
&= \sum_{(m,x,y)\in G_1} \Big(\Pr[M'=m,X'=x,Y'=y] \cdot \\
&\qquad \log \frac{\Pr[N=m|X'=x,Y'=y]}{\Pr[M'=m|X'=x,Y'=y]}\Big) \\
&\quad + \sum_{(m,x,y)\notin G_1} \Big(\Pr[M'=m,X'=x,Y'=y] \cdot \\
&\qquad \log \frac{\Pr[N=m|X'=x,Y'=y]}{\Pr[M'=m|X'=x,Y'=y]}\Big) \\
&\ge \sum_{(m,x,y)\in G_1} \Big(\Pr[M'=m,X'=x,Y'=y] \cdot \\
&\qquad \log \frac{\Pr[N=m|X'=x,Y'=y]}{\Pr[M'=m|X'=x,Y'=y]}\Big) \\
&\quad + \Pr[(M',X',Y') \notin G_1] \cdot \frac{\varepsilon+1}{\varepsilon'} \qquad (32) \\
&= \sum_{(m,x,y)\notin G_1} \Big(\Pr[M'=m,X'=x,Y'=y] \cdot \\
&\qquad \log \frac{\Pr[M'=m|X'=x,Y'=y]}{\Pr[N=m|X'=x,Y'=y]}\Big) \\
&\quad - \mathrm{S}(M'X'Y'\|NX'Y') \\
&\quad + \Pr[(M',X',Y') \notin G_1] \cdot \frac{\varepsilon+1}{\varepsilon'} \qquad (33) \\
&\ge -1 - \varepsilon + \Pr[(M',X',Y') \notin G_1] \cdot \frac{\varepsilon+1}{\varepsilon'}. \qquad (34)
\end{aligned}$$

Above, Eq. (31) and Eq. (33) follow from the definition of the relative entropy, and Eq. (32) follows from the definition of G_1. To get Eq. (34), we use Fact II.6 and Eq. (28). Eq. (34) implies that $\Pr[(M',X',Y') \notin G_1] \le \varepsilon'$.

To upper bound the second term let us define

$$G_2 \overset{\text{def}}{=} \left\{(m,x,y) : \log \frac{\Pr[M'=m|X'=x,Y'=y]}{\Pr[M'=m|Y'=y]} < \frac{c+1}{\varepsilon'}\right\}.$$

Consider,

$$c \geq \mathrm{I}\big(M' : X' \big| Y'\big) \tag{35}$$

$$= \mathop{\mathbb{E}}_{(m,x,y)\leftarrow M',X',Y'} \left[\log \frac{\Pr[M' = m | X' = x, Y' = y]}{\Pr[M' = m | Y' = y]} \right] \tag{36}$$

$$\tag{37}$$

$$= \sum_{(m,x,y)\in G_2} \Big(\Pr\big[M' = m, X' = x, Y' = y\big] \cdot \log \frac{\Pr[M' = m | X' = x, Y' = y]}{\Pr[M' = m | Y' = y]} \Big) + \sum_{(m,x,y)\notin G_2} \Big(\Pr\big[M' = m, X' = x, Y' = y\big] \cdot \log \frac{\Pr[M' = m | X' = x, Y' = y]}{\Pr[M' = m | Y' = y]} \Big)$$

$$\geq \frac{c+1}{\varepsilon'} \cdot \Pr\big[\big(M', X', Y'\big) \notin G_2\big] - 1. \tag{38}$$

Above Eq. (35) is one of the assumptions in the lemma; Eq. (36) follows from the definition of the conditional mutual information; Eq. (38) follows from the definition of G_2 and Fact II.6. Eq. (38) implies that $\Pr\left[(M', X', Y') \notin G_2\right] \leq \varepsilon'$.

To bound the last term define

$$G_3 \overset{\text{def}}{=} \left\{ (m,x,y) : \log \frac{\Pr[M' = m, Y' = y]}{\Pr[N = m, Y' = y]} < \frac{\varepsilon + 1}{\varepsilon'} \right\}.$$

Consider,

$$\varepsilon \geq \mathrm{S}\big(X'Y'M' \big\| X'Y'N\big)$$

$$\geq \mathrm{S}\big(Y'M' \big\| Y'N\big) \tag{39}$$

$$= \mathop{\mathbb{E}}_{(m,x,y)\leftarrow M',X',Y'} \left[\log \frac{\Pr[M' = m, Y' = y]}{\Pr[N = m, Y' = y]} \right]$$

$$= \sum_{(m,x,y)\in G_3} \Big(\Pr\big[M' = m, X' = x, Y' = y\big] \cdot \log \frac{\Pr[M' = m, Y' = y]}{\Pr[N = m, Y' = y]} \Big) + \sum_{(m,x,y)\notin G_3} \Big(\Pr\big[M' = m, X' = x, Y' = y\big] \cdot \log \frac{\Pr[M' = m, Y' = y]}{\Pr[N = m, Y' = y]} \Big)$$

$$\geq -1 + \Pr\big[\big(M', X', Y'\big) \notin G_3\big] \cdot \frac{\varepsilon + 1}{\varepsilon'}. \tag{40}$$

Above Eq. (39) follows from Fact II.7 and Eq. (40) follows from definition of G_3. This implies $\Pr[(M', X', Y') \notin G_3] \leq \varepsilon'$.

Combining the bounds for the three terms we get

$$\Pr_{(m,x,y)\leftarrow M',X',Y'} \left[\log \frac{\Pr[N = m | X' = x]}{\Pr[N = m | Y' = y]} \geq \frac{c+5}{\varepsilon'} \right] \leq 3\varepsilon'.$$

Using $\|X'Y'M' - X'Y'N\|_1 \leq \sqrt{\varepsilon}$ (as was shown previously), we finally have,

$$\Pr_{(m,x,y)\leftarrow N,X',Y'} \left[\log \frac{\Pr[N = m | X' = x]}{\Pr[N = m | Y' = y]} \geq \frac{c+5}{\varepsilon'} \right] \leq 3\varepsilon' + \sqrt{\varepsilon}.$$

■

Proof of Lemma III.4: We prove the lemma by induction on t. For the base case $t = 1$, note that

$$\mathrm{I}(X'R' : M'_1 | Y'R') = \mathrm{I}(X' : M'_1 | Y'R') \leq c_1$$

and

$$\mathrm{I}(Y'R' : M'_1 | X'R') = \mathrm{I}(Y' : M'_1 | X'R') \leq \varepsilon_1.$$

Lemma III.3 implies (by taking $X', Y,' M'$ in Lemma III.3 to be $X'R', Y'R', M'_1$ respectively) that Alice, with input $X'R'$, and Bob, with input $Y'R'$, can run a public-coin protocol with a single message from Alice to Bob of

$$\frac{c_1 + 5}{\varepsilon'} + \mathcal{O}(\log \frac{1}{\varepsilon'})$$

bits and generate a new random variable M_1 satisfying

$$\|R'X'Y'M'_1 - R'X'Y'M_1\|_1 \leq 3\sqrt{\varepsilon_1} + 6\varepsilon'.$$

Now let $t > 1$. Assume t is odd, for even t a similar argument will follow. From the induction hypothesis there exists a public-coin $t - 1$ message protocol $\mathcal{P}_{t-1}$ between Alice, with input $X'R'$, and Bob, with input $Y'R'$, with Alice sending the first message, and total communication

$$\frac{\sum_{s=1}^{t-1} c_s + 5(t-1)}{\varepsilon'} + \mathcal{O}\left((t-1) \log \frac{1}{\varepsilon'} \right), \tag{41}$$

such that at the end both Alice and Bob possess random variables $M_1, \ldots, M_{t-1}$ satisfying

$$\|R'X'Y'M_1 \cdots M_{t-1} - R'X'Y'M'_1 \cdots M'_{t-1}\|_1 \leq 3 \sum_{s=1}^{t-1} \sqrt{\varepsilon_s} + 6\varepsilon'(t-1). \tag{42}$$

Note that

$$\mathrm{I}(Y'R'M'_{<t} : M'_t | X'R'M'_{<t}) = \mathrm{I}(Y' : M'_t | X'R'M'_{<t}) \leq c_t$$

and

$$\mathrm{I}(X'R'M'_{<t} : M'_t | Y'R'M'_{<t}) = \mathrm{I}(X' : M'_t | Y'R'M'_{<t}) \leq \varepsilon_t.$$

Therefore Lemma III.3 implies (by taking $X', Y,' M'$ in Lemma III.3 to be $X'R'M'_{<t}, Y'R'M'_{<t}, M'_t$ respectively) that Alice, with input $X'R'M'_{<t}$, and Bob, with input $Y'R'M'_{<t}$, can run a public coin protocol $\mathcal{P}$ with a single message from Alice to Bob of

$$\frac{c_t + 5}{\varepsilon'} + \mathcal{O}\left(\log \frac{1}{\varepsilon'} \right) \tag{43}$$

bits and generate a new random variable M''_t satisfying

$$\left\|R'X'Y'M'_1\cdots M'_{t-1}M'_t - R'X'Y'M'_1\cdots M'_{t-1}M''_t\right\|_1 \\ \leq 3\sqrt{\varepsilon_t} + 6\varepsilon'. \tag{44}$$

Fact II.8 and Eq. (42) imply that Alice, on input $X'R'M_{<t}$ and Bob on input $Y'R'M_{<t}$, on running the same protocol $\mathcal{P}$ will generate a new random variable M_t satisfying

$$\|R'X'Y'M_1\cdots M_{t-1}M_t - R'X'Y'M'_1\cdots M'_{t-1}M''_t\|_1 \\ = \|R'X'Y'M_1\cdots M_{t-1} - R'X'Y'M'_1\cdots M'_{t-1}\|_1 \\ \leq 3\sum_{s=1}^{t-1}\sqrt{\varepsilon_s} + 6\varepsilon'(t-1). \tag{45}$$

Therefore by composing protocol $\mathcal{P}_{t-1}$ and protocol $\mathcal{P}$, using Equations (41), (43), (44), (45) and the triangle inequality for the ℓ_1 norm, we get a public-coin t-message protocol $\mathcal{P}_t$ between Alice, with input $X'R'$, and Bob, with input $Y'R'$, with Alice sending the first message, and total communication

$$\frac{\sum_{s=1}^{t} c_s + 5t}{\varepsilon'} + \mathcal{O}\left(t\log\frac{1}{\varepsilon'}\right),$$

such that at the end Alice and Bob both possess random variables $M_1,\ldots,M_t$ satisfying

$$\|R'X'Y'M_1\cdots M_t - R'X'Y'M'_1\cdots M'_t\|_1 \\ \leq 3\sum_{s=1}^{t}\sqrt{\varepsilon_s} + 6\varepsilon' t.$$

□

Proof of Lemma III.5: In $\mathcal{Q}_t$, Alice and Bob, using public coins and no communication first generate R_A, R_B such that $\left\|XYR_AR_B - X'Y'R'R'\right\|_1 \leq \tau$. They can do this from the Definition II.9 of embedding. Now they will run protocol $\mathcal{P}_t$ (as in Lemma III.4) with Alice's input being XR_A and Bob's input being YR_B and at the end both possess $M_1,\ldots,M_t$. From Lemma III.4, the communication of $\mathcal{Q}_t$ is as desired. Now from Fact II.8, Lemma III.4 and the triangle inequality for the ℓ_1 norm,

$$\|XYR_AR_BM_1\cdots M_t - X'Y'R'R'M'_1\cdots M'_t\|_1 \\ \leq \tau + 3\sum_{s=1}^{t}\sqrt{\varepsilon_s} + 6\varepsilon' t.$$

□