Algorithmica (2016) 76:720–748
DOI 10.1007/s00453-015-0100-0

A Direct Product Theorem for Two-Party Bounded-Round Public-Coin Communication Complexity

Rahul Jain[1,2] · **Attila Pereszlényi**[3] · **Penghui Yao**[3]

Received: 25 September 2014 / Accepted: 10 December 2015 / Published online: 18 December 2015

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 complete relation in this model. In particular, our result implies a strong direct product theorem for the two-party constant-round public-coin randomized communication complexity of all complete 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. Our result generalizes the

A preliminary version of this article has appeared in the Proceedings of the 53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012.

✉ Penghui Yao
phyao1985@gmail.com

Rahul Jain
rahul@comp.nus.edu.sg

Attila Pereszlényi
attila.pereszlenyi@gmail.com

1 Department of Computer Science, Centre for Quantum Technologies, National University of Singapore, Singapore, Singapore

2 MajuLab, CNRS-UNS-NUS-NTU International Joint Research Unit, UMI 3654, Singapore, Singapore

3 Centre for Quantum Technologies, National University of Singapore, Singapore, Singapore

result of Jain which can be regarded as the special case when the number of messages is one. Our result can be considered as an important progress towards settling the strong direct product conjecture for 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 by Jain. One key tool used in our work and also by Jain is a message compression technique due to Braverman and Rao, 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 by Holenstein for proving a parallel repetition theorem for two-prover games.

Keywords Communication complexity · Information theory · Strong direct product theorem

Mathematics Subject Classification 68Q10 · 68Q17

1 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 that 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 $p^{-\Omega(k)}$. The weaker *direct sum* theorems state that in order to compute k independent instances of a problem, if we provide $o(k \cdot c)$ units of resource, then the success probability for computing all the k instances correctly is at most a constant $q < 1$.

In this work, we are concerned with the model of communication complexity which was introduced by Yao [40]. In this model there are different parties who wish to compute a joint relation of their inputs. They do local computation, use public or private coins, and communicate to achieve this task. The resource that is counted is the number of bits communicated. The text by Kushilevitz and Nisan [27] is an excellent reference for this model.

Direct product questions and direct sum questions have been extensively investigated in different sub-models of communication complexity. Some examples of known direct product theorems are Parnafes et al. [32] theorem for *forests* of communication protocols, Shaltiel's [36] 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 et al. [29], extended to the multi-party case by Viola and Wigderson [39], extended to the generalized discrepancy bound by Sherstov [38]. Jain et al. [16] proved direct product theorem for the *subdistribu-*

tion bound. Klauck et al. [26] proved it for the *quantum* communication complexity of the *set disjointness* problem and Klauck [24] proved it for the public-coin communication complexity of the set disjointness problem (which was re-proven using different arguments by Jain [14]). Ben-Aroya et al. [4] showed it for the one-way quantum communication complexity of the *index* function problem. Jain showed it for randomized one-way communication complexity and for the *conditional relative min-entropy bound* [14], which is a lower bound on public-coin communication complexity. Recently, Jain and Yao [22] showed a strong direct product theorem in terms of the *smooth rectangle bound*. Later, Braverman and Weinstein [8] strengthened the result by showing a strong direct product theorem in terms of the (internal) *information cost*. Direct sum theorems were shown in the public-coin one-way model [18], in the public-coin simultaneous message passing model [18], in the entanglement-assisted quantum one-way communication model [20], in the private-coin simultaneous message passing model [15], in the constant-round public-coin two-way model [5] and in the general two-way model [3]. On the other hand, strong direct product conjectures have been shown to be false by Shaltiel [36] 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* [41], Raz's [34] theorem for two-prover games, Shaltiel's [36] theorem for *fair decision trees*, Nisan et al. [30] theorem for *decision forests*, Drucker's [11] theorem for randomized query complexity, Sherstov's [38] theorem for *approximated polynomial degree*, and Lee and Roland's [28] theorem for quantum query complexity. Besides their inherent importance, direct product theorems had various important applications such as in *probabilistically checkable proofs* [34], in circuit complexity [41], and in showing time-space trade-offs [1,24,26].

In this paper, we show a direct product theorem for two-party bounded-round public-coin randomized communication complexity. In this model, for computing a relation $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ (where $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ 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 $y \in \mathcal{Y}$. They are supposed to do local computations using public coins shared between them, communicate a fixed number of rounds and at the end, output an element $z \in \mathcal{Z}$. We only consider complete relations so there exists a z. They succeed if $(x, y, z) \in f$. For a natural number $t \geq 1$ and $\varepsilon \in (0, 1)$, let $\mathrm{R}_{\varepsilon}^{(t),\mathrm{pub}}(f)$ be the two-party t-round public-coin communication complexity of f with worst case error ε (see Definition 2.13).

We show the following.

Theorem 1.1 *Let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ be finite sets, $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ a complete relation, $\varepsilon > 0$, and $k, t \geq 1$ integers. There exists a constant $\kappa \geq 0$ such that*

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

In particular, it implies a strong direct product theorem for the two-party constant-round public-coin randomized communication complexity of all complete

relations.[1] Our result generalizes the result of Jain [14] which can be regarded as the special case when $t = 1$. Prior to our result, randomized one-way communication complexity is the only model whose strong direct product theorem was established [14]. Hence our result can be considered as an important progress towards settling the strong direct product conjecture for two-party public-coin communication complexity, a major open question in this area. Recently, our result was improved by Braverman et al. [6] with better dependence on the number of rounds, using a new sampling technique introduced in Ref. [7].

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-round deterministic protocol with $t \cdot \log n$ bits of communication for both FP_t and BP_t. However, if only $t - 1$ rounds are allowed to be exchanged between the parties, exponentially more communication is required, treating t as a fixed constant. The communication complexity of this problem has been very well studied in both the classical and the quantum models [17,23,25,31,33]. Some tight lower bounds that we know so far are as follows.

Theorem 1.2 [33] *For any integer $t \geq 1$,*

$$\mathrm{R}_{1/3}^{(t-1),\mathrm{pub}}(\mathrm{FP}_t) \geq \Omega\left(n \log^{(t-1)} n\right)$$
$$\mathrm{R}_{1/3}^{(t-1),\mathrm{pub}}(\mathrm{BP}_t) \geq \Omega(n)$$

As a consequence of Theorem 1.1 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, independent of the input size.

Corollary 1.3 *For integers $t, k \geq 1$,*

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

1.1 Our Techniques

We prove our direct product result using information theoretic arguments. Information theory is a versatile tool in communication complexity, especially in proving

[1] When $\mathrm{R}_{\varepsilon}^{(t),\mathrm{pub}}(f)$ is a constant, all the lower bounds are constants as well. It is known that several lower bounds satisfy a strong direct product theorem, such as *conditional relative entropy* [14]. Thus in this case a strong direct product result for the model we concerns follows directly.

lower bounds and direct sum and direct product theorems [2,3,5,9,14,15,18–20]. The similar information theoretic arguments have been used to prove *parallel repetition theorems* for *two-prover one-round games* as well [12,34]. The broad argument that we use is as follows. For a given relation f, let the communication required for computing one instance with t rounds and constant success be c. Let us consider a protocol for computing f^k with t rounds and communication $o(kc)$. 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. Otherwise, we exhibit another coordinate j outside of these ℓ coordinates such that success in the j-th coordinate, even when conditioned on success in these 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 [40].

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-round protocols having error at most ε averaged over μ. Let the inputs for f^k be drawn from distribution μ^k (k independent copies of μ). Consider a t-round protocol $\mathcal{P}$ for f^k with communication $o(kc)$ and for the rest of the argument condition on success on a set of coordinates C. 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 conditioning on success on all coordinates in C. The distribution of Alice's and Bob's input in the j-th coordinate $(X_j Y_j)$ is quite close to μ. Here use the same symbol to represent a random variable and its distribution. 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-round 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 sample the transcript, we adopt the message compression protocol of 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 $Y \leftrightarrow X \leftrightarrow M$ (see Definition 2.1), there exists a public-coin protocol between Alice and Bob, with inputs X and Y, with a single message from Alice to Bob of $O\left(\mathrm{I}(X : M \mid Y) + \sqrt{\mathrm{I}(X : M \mid Y)}\right) + 1$ bits, such that at the end of the protocol, Alice and Bob both possess a random variable M' which is close to M in the ℓ_1 distance.*

Consider the situation after conditioning on success in all the coordinates in C, as above, and let $X_j Y_j$ represent the input in the j-th coordinate. The Braverman–Rao compression protocol cannot be directly applied at this point. Take the first message M_1, sent by Alice, for instance. $Y_j X_j M_1$ doesn't necessarily form a Markov chain. For example, M_1 is the message in which Alice tries to guess Bob's input Y_j

and the event of success is Alice succeeds in doing so. Then it is easy to see that $Y_j X_j M_1$ is not a Markov chain conditioning on success. 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 by Jain in Ref. [14].) We then apply such a robust message compression protocol to each successive message. Conditioning on success in C incurs a small statistical loss for each message. Thus, the overall error is bounded as the number of messages exchanged is bounded in our model. Recently, Braverman et.al. introduced in Ref. [7] a new simulation whose statistical error is independent of the number of messages. Using this simulation, Braverman et al. [6] strengthened our result with better dependence on the number of rounds.

Another difficulty in this argument is that since μ may be a non-product distribution, the input of Alice and Bob in other coordinates may be correlated with each other's input in the j-th coordinate when conditioned on success in C. We overcome this by introducing new random variables DU conditioning on which Alice's input is independent of Bob's input. Namely, DU split μ^k into a convex combination of product distributions.

This idea of splitting a non-product distribution into convex combination of product distributions has been used in several previous works [2,3,5,12,14,34,35]. $D_{-j}U_{-j}$ is independent of $X_j Y_j$ without conditioning on success in all coordinates in C. This fact is sufficient for several direct sum results [2,5]. However, after conditioning on success in all coordinates in C, $D_{-j}U_{-j}$ is correlated with $X_j Y_j$. This lead us to use another important tool namely the *correlated sampling* protocol, that was also used for example by Holenstein [12] in his proof of a strong direct product theorem for two-prover one-round games. We prove that $D_{-j}U_{-j}$ can be correlatedly sampled by Alice and Bob. Conditioning on $D_{-j}U_{-j}$ and their own inputs, Alice and Bob are able to complete the remaining XY.

As mentioned previously, we build on the arguments used by Jain [14]. He showed a new characterization of two-party one-way public-coin communication complexity and used that characterization to show a strong direct product result for all relations in this model. We are unable to arrive at such characterization for protocols with more than one messages so we use a more direct approach, as outlined above, to prove our direct product result.

1.2 Organization

The rest of the paper is organized as follows. In Sect. 2, we present some background on information theory and communication complexity. In Sect. 3, we prove our main result, Theorem 1.1, starting with some lemmas that are helpful in building the proof. Some proofs are deferred to Sect. 4.

2 Preliminaries

2.1 Information Theory

For integer $n \geq 1$, let $[n]$ represent the set $\{1, 2, \ldots, n\}$ and let $[0]$ be the empty set. Let $\mathcal{X}$ and $\mathcal{Y}$ 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. We use the same symbol to represent a random variable and its distribution whenever it is clear from the context. We use letters in lower-case such as x, y, z to represent the elements in the supports of X, Y, Z, respectively. The expectation of function f on $\mathcal{X}$ is defined as

$$\mathop{\mathbb{E}}_{x \leftarrow X}[f(x)] \stackrel{\text{def}}{=} \sum_{x \in \mathcal{X}} \mu(x) \cdot f(x).$$

The entropy of X is defined by Shannon in [37] as

$$\mathrm{H}(X) \stackrel{\text{def}}{=} -\sum_{x \in \mathcal{X}} \mu(x) \cdot \log \mu(x).$$

For two distributions μ and λ on $\mathcal{X}$, the distribution $\mu \otimes \lambda$ is defined as $(\mu \otimes \lambda)(x_1, x_2) \stackrel{\text{def}}{=} \mu(x_1) \cdot \lambda(x_2)$. Define μ^k to be $\mu \otimes \cdots \otimes \mu$ with k times. If $L = L_1 \cdots L_k$, we define $L_{-i} \stackrel{\text{def}}{=} L_1 \cdots L_{i-1} L_{i+1} \cdots L_k$ and $L_{<i} \stackrel{\text{def}}{=} L_1 \cdots L_{i-1}$. The random variable $L_{\leq i}$ is defined analogously. The total variance distance between μ and λ is defined to be half of the ℓ_1 norm of $\mu - \lambda$, i.e.,

$$\|\lambda - \mu\|_1 \stackrel{\text{def}}{=} \frac{1}{2} \sum_x |\lambda(x) - \mu(x)| = \max_{S \subseteq \mathcal{X}} |\lambda_S - \mu_S|$$

where $\lambda_S \stackrel{\text{def}}{=} \sum_{x \in S} \lambda(x)$. 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

$$\mathrm{D}\big(X \big\| Y\big) \stackrel{\text{def}}{=} \mathop{\mathbb{E}}_{x \leftarrow X}\left[\log \frac{\Pr[X = x]}{\Pr[Y = x]}\right].$$

The relative min-entropy between them is defined as

$$\mathrm{S}_\infty\big(X \big\| Y\big) \stackrel{\text{def}}{=} \max_{x \in \mathcal{X}} \left\{\log \frac{\Pr[X = x]}{\Pr[Y = x]}\right\}.$$

It is easy to see that $\mathrm{D}\big(X \big\| Y\big) \leq \mathrm{S}_\infty\big(X \big\| Y\big)$. Let X, Y, and Z be jointly distributed random variables. We often write XY as a shorthand for the pair (X, Y). With slight abuse of notations, we write XX for a joint distribution XX' where X and X' are

always equal, i.e., $\Pr[X = X'] = 1$ Let Y_x denote the distribution of Y conditioned on $X = x$. The conditional entropy of Y conditioned on X is defined as

$$\mathrm{H}(Y|X) \stackrel{\text{def}}{=} \mathop{\mathbb{E}}_{x \leftarrow X}[\mathrm{H}(Y_x)] = \mathrm{H}(XY) - \mathrm{H}(X).$$

The mutual information between X and Y is defined as

$$\begin{aligned}\mathrm{I}(X : Y) &\stackrel{\text{def}}{=} \mathrm{H}(X) + \mathrm{H}(Y) - \mathrm{H}(XY) \\ &= \mathop{\mathbb{E}}_{y \leftarrow Y}\left[\mathrm{D}\left(X_y \middle\| X\right)\right] \\ &= \mathop{\mathbb{E}}_{x \leftarrow X}\left[\mathrm{D}\left(Y_x \middle\| Y\right)\right].\end{aligned}$$

It is easily seen that $\mathrm{I}(X : Y) = \mathrm{D}\left(XY \middle\| X \otimes Y\right)$. We say that X and Y are independent if $\mathrm{I}(X : Y) = 0$. The conditional mutual information between X and Y, conditioned on Z, is defined as

$$\begin{aligned}\mathrm{I}\left(X : Y \middle| Z\right) &\stackrel{\text{def}}{=} \mathop{\mathbb{E}}_{z \leftarrow Z}\left[\mathrm{I}\left(X : Y \middle| Z = z\right)\right] \\ &= \mathrm{H}\left(X|Z\right) + \mathrm{H}\left(Y|Z\right) - \mathrm{H}\left(XY|Z\right).\end{aligned}$$

The following *chain rule* for mutual information can be proved easily

$$\mathrm{I}(X : YZ) = \mathrm{I}(X : Z) + \mathrm{I}\left(X : Y \middle| Z\right).$$

Definition 2.1 Let X, X', Y, and Z be jointly distributed random variables. We define the joint distribution of $(X'Z)(Y|X)$ by

$$\Pr[(X'Z)(Y|X) = (x, z, y)] \stackrel{\text{def}}{=} \Pr[X' = x, Z = z] \cdot \Pr[Y = y|X = x].$$

We say that X, Y, and Z is a Markov chain if $XYZ = (XY)(Z|Y)$ 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 $\mathrm{I}\left(X : Z \middle| Y\right) = 0$. Ibinson et al. [13] showed that if $\mathrm{I}\left(X : Z \middle| Y\right)$ is small then XYZ is close to being a Markov chain.

Lemma 2.2 ([13]) *For any random variables X, Y, and Z, it holds that*

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

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

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

Fact 2.3 ([10, page32]) The relative entropy is jointly convex in its arguments. That is, for distributions $\mu, \mu^1, \lambda, \lambda^1 \in \mathscr{X}$ and $p \in [0, 1]$,

$$\mathrm{D}\Big(p\mu + (1-p)\mu^1 \Big\| p\lambda + (1-p)\lambda^1\Big) \leq p \cdot \mathrm{D}\big(\mu\|\lambda\big) + (1-p) \cdot \mathrm{D}\Big(\mu^1\|\lambda^1\Big).$$

Fact 2.4 ([10, page24]) The relative entropy satisfies the following chain rule. Let XY and X^1Y^1 be random variables on $\mathscr{X} \times \mathscr{Y}$. It holds that

$$\mathrm{D}\Big(X^1Y^1\Big\|XY\Big) = \mathrm{D}\Big(X^1\Big\|X\Big) + \mathop{\mathbb{E}}_{x \leftarrow X^1}\Big[\mathrm{D}\Big(Y^1_x\Big\|Y_x\Big)\Big].$$

In particular,

$$\begin{aligned}\mathrm{D}\Big(X^1Y^1\Big\|X \otimes Y\Big) &= \mathrm{D}\Big(X^1\Big\|X\Big) + \mathop{\mathbb{E}}_{x \leftarrow X^1}\Big[\mathrm{D}\Big(Y^1_x\Big\|Y\Big)\Big] \\ &\geq \mathrm{D}\Big(X^1\Big\|X\Big) + \mathrm{D}\Big(Y^1\Big\|Y\Big),\end{aligned}$$

where the inequality is from Fact 2.3.

Note that there is no conditioning on x in Y at the end of the first line as in the second argument of the relative entropy X and Y are independent. The following fact proves that the minimizer of the relative entropy is the product of the marginals.

Fact 2.5 Let XY and X^1Y^1 be random variables on $\mathscr{X} \times \mathscr{Y}$. It holds that

$$\mathrm{D}\Big(X^1Y^1\Big\|X \otimes Y\Big) \geq \mathrm{D}\Big(X^1Y^1\Big\|X^1 \otimes Y^1\Big) = \mathrm{I}\Big(X^1 : Y^1\Big).$$

Proof From the definition of the relative entropy, we have

$$\begin{aligned}\mathrm{D}\Big(X^1Y^1\Big\|X \otimes Y\Big) &= \sum_{xy} \Pr\Big[X^1Y^1 = xy\Big] \log \frac{\Pr\big[X^1Y^1 = xy\big]}{\Pr[X = x]\Pr[Y = y]} \\ &= \sum_{xy} \Pr\Big[X^1Y^1 = xy\Big]\Bigg(\log \frac{\Pr\big[X^1Y^1 = xy\big]}{\Pr\big[X^1 = x\big]\Pr\big[Y^1 = y\big]} \\ &\qquad + \log \frac{\Pr\big[X^1 = x\big]\Pr\big[Y^1 = y\big]}{\Pr[X = x]\Pr[Y = y]}\Bigg) \\ &= \mathrm{D}\Big(X^1Y^1\Big\|X^1 \otimes Y^1\Big) + \mathrm{D}\Big(X^1\Big\|X\Big) + \mathrm{D}\Big(Y^1\Big\|Y\Big) \\ &\geq \mathrm{D}\Big(X^1Y^1\Big\|X^1 \otimes Y^1\Big).\end{aligned}$$

The equality $\mathrm{D}\big(X^1Y^1\big\|X^1 \otimes Y^1\big) = \mathrm{I}\big(X^1 : Y^1\big)$ can easily be verified from the definitions.

Fact 2.6 (Pinsker's inequality, [10, page370]) For distributions λ and μ,

$$0 \leq \|\lambda - \mu\|_1 \leq \sqrt{\mathrm{D}(\lambda\|\mu)}.$$

The following fact gives a lower bound on each term in the summation in the definition of the relative entropy.

Fact 2.7 ([21]) Let λ and μ be distributions on $\mathscr{X}$. For any subset $\mathscr{S} \subseteq \mathscr{X}$, it holds that

$$\sum_{x \in \mathscr{S}} \lambda(x) \cdot \log \frac{\lambda(x)}{\mu(x)} \geq -1.$$

Hence, for any $r > 0, c > 0$, if $\mathrm{D}(\lambda\|\mu) \leq c$, then it holds that

$$\Pr_{x \leftarrow \lambda}\left[\log \frac{\lambda(x)}{\mu(x)} \leq \frac{c+1}{r}\right] \leq r.$$

The following fact easily follows from the triangle inequality and Fact 2.4.

Fact 2.8 The ℓ_1 distance and the relative entropy are monotone non-increasing when subsystems are considered. Let XY and X^1Y^1 be random variables on $\mathscr{X} \times \mathscr{Y}$, then

$$\left\|XY - X^1Y^1\right\|_1 \geq \left\|X - X^1\right\|_1$$

and

$$\mathrm{D}\left(XY\|X^1Y^1\right) \geq \mathrm{D}\left(X\|X^1\right).$$

Fact 2.9 For any function $f : \mathscr{X} \times \mathscr{R} \to \mathscr{Y}$ and random variables X, X_1 on $\mathscr{X}$ and R on $\mathscr{R}$, such that R is independent of XX_1, it holds that

$$\|Xf(X,R) - X_1 f(X_1,R)\|_1 = \|X - X_1\|_1.$$

Proof

$$\begin{aligned}
&\|Xf(X,R) - X_1 f(X_1,R)\|_1 \\
&= \frac{1}{2}\sum_{xy}\left|\Pr[Xf(X,R) = xy] - \Pr\left[X^1 f(X^1,R) = xy\right]\right| \\
&= \frac{1}{2}\sum_{x}\left|\Pr[X = x] - \Pr\left[X^1 = x\right]\right| \cdot \sum_{y}\Pr[f(x,R) = y] \\
&= \frac{1}{2}\sum_{x}\left|\Pr[X = x] - \Pr\left[X^1 = x\right]\right| = \left\|X - X^1\right\|_1.
\end{aligned}$$

□

The following definition was introduced by Holenstein [12]. It plays a critical role in his proof of a parallel repetition theorem for two-prover games.

Definition 2.10 ([12]) 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 $\mathscr{R}$, which is independent of X_0Y_0 and functions $f_A : \mathscr{X} \times \mathscr{R} \to \mathscr{S}$, $f_B : \mathscr{Y} \times \mathscr{R} \to \mathscr{T}$, such that

$$\|X_0Y_0 f_A(X_0, R) f_B(Y_0, R) - X_1Y_1ST\|_1 \leq \varepsilon.$$

The following lemma was shown by Holenstein [12] using a correlated sampling protocol.

Lemma 2.11 (Corollary 5.3 in [12]) *For random variables S, X, and Y, if*

$$\|XYS - (XY)(S|X)\|_1 \leq \varepsilon$$

and

$$\|XYS - (XY)(S|Y)\|_1 \leq \varepsilon$$

then (X, Y) is $(1-5\varepsilon)$-embeddable in (XS, YS).

We will need the following generalization of the previous lemma.

Lemma 2.12 *For joint random variables (A', B', C') and (A, B), satisfying*

$$\mathrm{D}\big(A'B' \big\| AB\big) \leq \varepsilon \tag{1}$$

$$\mathop{\mathbb{E}}_{(a,c)\leftarrow A'C'} \big[\mathrm{D}\big(B'_{a,c} \big\| B_a\big)\big] \leq \varepsilon \tag{2}$$

$$\mathop{\mathbb{E}}_{(b,c)\leftarrow B'C'} \big[\mathrm{D}\big(A'_{b,c} \big\| A_b\big)\big] \leq \varepsilon \tag{3}$$

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

Proof Using the definition of the relative entropy, we get the following.

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

This means that

$$\mathop{\mathbb{E}}_{(a,c)\leftarrow A'C'} \big[\mathrm{D}\big(B'_{a,c} \big\| B'_a\big)\big] \leq \mathop{\mathbb{E}}_{(a,c)\leftarrow A'C'} \big[\mathrm{D}\big(B'_{a,c} \big\| B_a\big)\big] \leq \varepsilon. \tag{4}$$

Furthermore,

$$\mathop{\mathbb{E}}_{(a,c)\leftarrow A'C'}\left[\mathrm{D}\left(B'_{a,c}\middle\|B'_a\right)\right] = \mathrm{D}\left(A'C'B'\middle\|\left(A'C'\right)\left(B'|A'\right)\right) \tag{5}$$

$$= \mathrm{D}\left(A'B'C'\middle\|\left(A'B'\right)\left(C'|A'\right)\right) \tag{6}$$

$$\geq \left\|A'B'C' - \left(A'B'\right)\left(C'|A'\right)\right\|_1^2. \tag{7}$$

Above, Eq. (5) follows from the chain rule for the relative entropy, Eq. (6) is because the distributions $\left(A'C'\right)\left(B'|A'\right)$ and $\left(A'B'\right)\left(C'|A'\right)$ are same by Definition 2.1, and Eq. (7) follows from Fact 2.6. Now from Eqs. (4) and (7) we get

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

By similar arguments we get

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

The two inequalities above (using Lemma 2.11) imply that $\left(A', B'\right)$ is $\left(1-5\sqrt{\varepsilon}\right)$-embeddable in $\left(A'C', B'C'\right)$. Namely, there exist functions f_1 and f_2 and random variable R independent of $A'B'$ such that $\left\|A'B'f_1\left(A', R\right)f_2\left(B', R\right) - A'B'C'C'\right\|_1 \leq 5\sqrt{\varepsilon}$. Furthermore, from Fact 2.6 and Eq. (1) we get that

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

Finally,

$$\begin{aligned}
&\left\|ABf\left(A, R\right)f\left(B, R\right) - A'B'C'C'\right\|_1 \\
&\quad\leq \left\|ABf_1\left(A, R\right)f_2\left(B, R\right) - A'B'f_1\left(A', R\right)f_2\left(B', R\right)\right\|_1 \\
&\qquad + \left\|A'B'f_1\left(A', R\right)f_2\left(B', R\right) - A'B'C'C'\right\|_1 \\
&\quad= \left\|AB - A'B'\right\|_1 + \left\|A'B'f_1\left(A', R\right)f_2\left(B', R\right) - A'B'C'C'\right\|_1 \leq 6\sqrt{\varepsilon},
\end{aligned}$$

where the equality is from Fact 2.9. Thus
we get that (A, B) is $\left(1-6\sqrt{\varepsilon}\right)$-embeddable in $(A'C', B'C')$. □

2.2 Communication Complexity

Let $f \subseteq \mathscr{X} \times \mathscr{Y} \times \mathscr{Z}$ be a relation, $t \geq 1$ an integer, and $\varepsilon \in (0, 1)$. In this work we only consider *complete* relations, i.e., for every $(x, y) \in \mathscr{X} \times \mathscr{Y}$, there is some $z \in \mathscr{Z}$ such that $(x, y, z) \in f$. In the two-party t-round public-coin model of communication, Alice, with input $x \in \mathscr{X}$, and Bob, with input $y \in \mathscr{Y}$, do local computation using public coins shared between them and exchange t messages, with Alice sending the first message. At the end of the protocol, the party receiving the t-th message outputs some $z \in \mathscr{Z}$. The output is declared correct if $(x, y, z) \in f$ and wrong otherwise.

Definition 2.13 Let $\mathrm{R}_{\varepsilon}^{(t),\mathrm{pub}}(f)$ represent the two-party t-round public-coin communication complexity of f with worst case error ε, i.e., the minimum number of bits that Alice and Bob need to exchange in a t-round public-coin protocol that the output for each input (x, y) is correct with probability at least $1-\varepsilon$. We similarly consider two-party t-round deterministic protocols where there are no public coins used by Alice and Bob. Let $\mu \in \mathscr{X} \times \mathscr{Y}$ be a distribution. We let $\mathrm{D}_{\varepsilon}^{(t),\mu}(f)$ represent the two-party t-round distributional communication complexity of f under μ with expected error ε, i.e., the minimum number of bits Alice and Bob need to exchange in a two-party t-round 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., [27, Theorem 3.20].

Lemma 2.14 (Yao's principle, [40]) $\mathrm{R}_{\varepsilon}^{(t),\mathrm{pub}}(f) = \max_{\mu} \mathrm{D}_{\varepsilon}^{(t),\mu}(f)$.

The following fact about communication protocols can be verified easily.

Fact 2.15 Let $M_1, \ldots, M_t$ be t messages in a deterministic communication protocol between Alice and Bob with inputs X and Y, 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 $f^k \subseteq \mathscr{X}^k \times \mathscr{Y}^k \times \mathscr{Z}^k$ be the cross product of f with itself k times. In a protocol for computing f^k, Alice will receive input in $\mathscr{X}^k$, Bob will receive input in $\mathscr{Y}^k$ and the output of the protocol will be in $\mathscr{Z}^k$.

3 Proof of Theorem 1.1

We start by showing a few lemmas which are helpful in the proof of the main result. The following theorem was shown by Jain [14] and follows primarily from a message compression argument due to Braverman and Rao [5].

Theorem 3.1 (Lemma 3.8 in [14]) *Let $\delta > 0$ and $c \geq 0$. Let X', Y', and N be random variables for which $Y' \leftrightarrow X' \leftrightarrow N$ is a Markov chain and the following holds.*

$$\Pr_{(x,y,m)\leftarrow X',Y',N}\left[\log \frac{\Pr\left[N=m|X'=x\right]}{\Pr[N=m|Y'=y]} > c\right] \leq \delta. \tag{8}$$

There exists a public-coin protocol between Alice and Bob, with inputs X' and Y', with a single message from Alice to Bob of at most $c + O(\log(1/\delta))$ bits, such that at the end of the protocol, Alice and Bob possess random variables M_A and M_B, respectively, satisfying $\left\| X'Y'NN - X'Y'M_AM_B \right\|_1 \leq 2\delta$.

Remark 3.2 In Ref. [5], the condition $\mathrm{I}\left(X' : N \middle| Y'\right) \leq c$ is used instead of Eq. (8). It is changed to the current one in Ref. [14]. By the equality $\mathrm{I}\left(X' : N \middle| Y\right) = \mathrm{D}\left(X'Y'N \middle\| X'Y'\left(N|Y'\right)\right)$ and Fact 2.7, $\mathrm{I}\left(X' : N \middle| Y'\right) \leq c$ 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 argument since the condition in Eq. (8) is robust when the underlying joint distribution is perturbed slightly, while $\mathrm{I}\left(X' : N \middle| Y'\right)$ may change a lot with such a perturbation.

As mentioned in Sect. 1, we will have to work with approximate Markov chains in our argument for the direct product. The following lemma makes Theorem 3.1 more robust to deal with approximate Markov chains. Its proof appears in Sect. 4.

Lemma 3.3 *Let $c \geq 0$, $1 > \varepsilon > 0$, and $\varepsilon' > 0$. Let X', Y', and M' be random variables for which the following holds,*

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

There exists a public-coin protocol between Alice and Bob, with inputs X' and Y', with a single message from Alice to Bob of at most $\frac{c+5}{\varepsilon'} + O\left(\log\frac{1}{\varepsilon'}\right)$ bits, such that at the end of the protocol, Alice and Bob possess a random variable M_A and M_B, respectively, satisfying

$$\left\| X'Y'M'M' - X'Y'M_AM_B \right\|_1 \leq 3\sqrt{\varepsilon} + 6\varepsilon'.$$

The following lemma generalizes the above lemma to deal with multiple messages. Its proof appears in Sect. 4.

Lemma 3.4 *Let $t \geq 1$ be an integer. Let $\varepsilon' > 0$, $c_s \geq 0$, and $1 > \varepsilon_s > 0$ for all $1 \leq s \leq t$. Let R', X', Y', $M'_1, \ldots, M'_t$ be random variables for which the following holds. (Below $M'_{<s} = M'_1 \cdots M'_{s-1}$ by definition.)*

$$\mathrm{I}\left(X' : M'_s \middle| Y'R'M'_{<s}\right) \leq c_s \tag{9}$$
$$\mathrm{I}\left(Y' : M'_s \middle| X'R'M'_{<s}\right) \leq \varepsilon_s \tag{10}$$

for odd s and

$$\mathrm{I}\left(Y' : M'_s \middle| X'R'M'_{<s}\right) \leq c_s \tag{11}$$
$$\mathrm{I}\left(X' : M'_s \middle| Y'R'M'_{<s}\right) \leq \varepsilon_s \tag{12}$$

for even s. There exists a public-coin t-round protocol $\mathscr{P}_t$ between Alice, with input $X'R'$, and Bob, with input $Y'R'$, with Alice sending the first message. The total communication of $\mathscr{P}_t$ is at most

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

At end of the protocol, Alice and Bob possess random variables $M'_{A,1}, \ldots, M'_{A,t}$ and $M'_{B,1}, \ldots, M'_{B,t}$, respectively, satisfying

$$\left\| R'X'Y'M'_1M'_1 \cdots M'_tM'_t - R'X'Y'M'_{A,1}M'_{B,1} \cdots M'_{A,t}M'_{B,t} \right\|_1 \leq 3\sum_{s=1}^{t} \sqrt{\varepsilon_s} + 6\varepsilon' t.$$

In the above lemma, 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 Sect. 4.

Lemma 3.5 *Let random variables R', X', Y', and $M'_1, \ldots, M'_t$ and numbers ε', c_s, and ε_s satisfy all the conditions in Lemma 3.4. Let $\tau > 0$ and let random variables (X, Y) be $(1-\tau)$-embeddable in $(X'R', Y'R')$. There exists a public-coin t-round protocol $\mathscr{Q}_t$ between Alice, with input X, and Bob, with input Y, with Alice sending the first message, and total communication at most*

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

At the end of the protocol, Alice possesses $R_A M_{A,1} \cdots M_{A,t}$ and Bob possesses $R_B M_{B,1} \cdots M_{B,t}$, such that

$$\begin{aligned} &\left\| XYR_AR_BM_1M_1 \cdots M_tM_t - X'Y'R'R'M'_{A,1}M'_{B,1} \cdots M'_{A,t}M'_{B,t} \right\|_1 \\ &\leq \tau + 3\sum_{s=1}^{t} \sqrt{\varepsilon_s} + 6\varepsilon' t. \end{aligned}$$

We are now ready to prove our main result, Theorem 1.1. We restate it here for convenience.

Theorem 1.1 *Let $\mathscr{X}$, $\mathscr{Y}$, and $\mathscr{Z}$ be finite sets, $f \subseteq \mathscr{X} \times \mathscr{Y} \times \mathscr{Z}$ a complete relation, $\varepsilon > 0$, and $k, t \geq 1$ integers. There exists a constant $\kappa \geq 0$ such that*

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

Proof of Theorem 1.1 Let $\delta \stackrel{\mathrm{def}}{=} \frac{\varepsilon^2}{7500t^2}$ and $\delta_1 = \frac{\varepsilon}{3000t}$. From Yao's principle (Lemma 2.14) it suffices to prove that for any distribution μ on $\mathscr{X} \times \mathscr{Y}$,

$$\mathrm{D}^{(t),\mu^k}_{1-(1-\varepsilon/2)^{\lfloor \delta k \rfloor}}\left(f^k\right) \geq \delta_1 kc$$

where $c \stackrel{\mathrm{def}}{=} \mathrm{D}^{(t),\mu}_{\varepsilon}(f) - \frac{\kappa t^2}{\varepsilon}$, for constant κ to be chosen later. Let XY be distributed according to μ^k. Let $\mathscr{Q}$ be a t-round 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 that t is odd for the rest of the argument and so Bob makes the final output. (The case when t is even follows similarly.) The following claim implies that the success of $\mathcal{Q}$ is at most $(1-\varepsilon/2)^{\lfloor \delta k \rfloor}$ and this shows the desired. □

Claim 3.6 For each $i \in [k]$, let us define a binary random variable $T_i \in \{0,1\}$, which represents the success of $\mathcal{Q}$, i.e., Bob's output being correct, on the i-th instance. So $T_i = 1$ if the protocol $\mathcal{Q}$ computes the i-th instance of f correctly and $T_i = 0$ otherwise. Let $k' \stackrel{\text{def}}{=} \lfloor \delta k \rfloor$. There exist k' coordinates $\{i_1, \ldots, i_{k'}\}$ such that for each $1 \le r \le k'-1$, either

$$\Pr\left[T^{(r)} = 1\right] \le (1 - \frac{\varepsilon}{2})^{k'}$$

or

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

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

Proof For $s \in [t]$, we denote the s-th message of $\mathcal{Q}$ by M_s. Let $M \stackrel{\text{def}}{=} M_1 \cdots M_t$. In the following, we assume that $1 \le r < k'$. However, the same argument also works when $r = 0$, i.e., for identifying the first coordinate, which we skip for the sake of avoiding repetition. Suppose that we have already identified r coordinates $i_1, \ldots, i_r$ satisfying that

$$\Pr\left[T_{i_1} = 1\right] \le 1 - \frac{\varepsilon}{2}$$

and

$$\Pr\left[T_{i_{j+1}} = 1 \middle| T^{(j)} = 1\right] \le 1 - \frac{\varepsilon}{2}$$

for $1 \le j \le r-1$. If $\Pr\left[T^{(r)} = 1\right] \le (1 - \frac{\varepsilon}{2})^{k'}$ then we are done. So from now on, assume that

$$\Pr\left[T^{(r)} = 1\right] > (1 - \frac{\varepsilon}{2})^{k'} \ge 2^{-\delta k}.$$

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

$$L^1 \stackrel{\text{def}}{=} (L | T^{(r)} = 1).$$

For example, $X^1Y^1 = (XY|T^{(r)} = 1)$. Let $C \stackrel{\text{def}}{=} \{i_1, \ldots, i_r\}$ and

$$R_i \stackrel{\text{def}}{=} D_{-i}U_{-i}X_{C\cup[i-1]}Y_{C\cup[i-1]}$$

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. $\left(X_j, Y_j\right)$ can be embedded well in $\left(X_j^1R_j^1, Y_j^1R_j^1\right)$ (with appropriate parameters as required by Lemma 2.12.)
2. Random variables R_j^1, X_j^1, Y_j^1, and $M_1^1, \ldots, M_t^1$ satisfy the conditions of Lemma 3.4 with appropriate parameters.

The following calculations are helpful for achieving the condition in Eq. (1) in Lemma 2.12. That is, $X_j^1Y_j^1$ is close to μ.

$$\begin{aligned}\delta k &> \mathrm{S}_\infty\left(X^1Y^1 \middle\| XY\right)\\ &\geq \mathrm{D}\left(X^1Y^1 \middle\| XY\right)\\ &\geq \sum_{i\notin C}\mathrm{D}\left(X_i^1Y_i^1 \middle\| X_iY_i\right)\end{aligned} \tag{13}$$

where the first inequality follows from the assumption that $\Pr\left[T^{(r)} = 1\right] > 2^{-\delta k}$ and the last inequality follows from Fact 2.4. The following calculations are helpful for achieving the conditions in Eqs. (2) and (3) in Lemma 2.12 that $\left(X_j^1|R_j^1Y_j^1\right) \approx \left(X_j|Y_j\right)$ and $\left(Y_j^1|R_j^1X_j^1\right) \approx \left(Y_j|X_j\right)$. It implies that Alice and Bob are able to sample R_j^1 correlatedly with inputs $X_j^1Y_j^1$.

$$\delta k > \mathrm{S}_\infty\left(X^1Y^1D^1U^1 \middle\| XYDU\right) \geq \mathrm{D}\left(X^1Y^1D^1U^1 \middle\| XYDU\right)$$

$$\geq \mathop{\mathbb{E}}_{(d,u,x_C,y_C)\leftarrow D^1U^1X_C^1Y_C^1}\left[\mathrm{D}\left(\left(X^1Y^1\right)_{d,u,x_C,y_C} \middle\| (XY)_{d,u,x_C,y_C}\right)\right] \tag{14}$$

$$= \sum_{i\notin C}\mathop{\mathbb{E}}_{\substack{(d,u,x_{C\cup[i-1]},y_{C\cup[i-1]})\\ \leftarrow D^1U^1X^1_{C\cup[i-1]}Y^1_{C\cup[i-1]}}}\left[\mathrm{D}\left(\left(X_i^1Y_i^1\right)_{\substack{d,u,x_{C\cup[i-1]},\\ y_{C\cup[i-1]}}} \middle\| (X_iY_i)_{\substack{d,u,x_{C\cup[i-1]},\\ y_{C\cup[i-1]}}}\right)\right] \tag{15}$$

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

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

$$= \sum_{i \notin C} \mathop{\mathbb{E}}_{(d_i,u_i,r_i) \leftarrow D_i^1 U_i^1 R_i^1} \left[\mathrm{D}\left((X_i^1 Y_i^1)_{d_i,u_i,r_i} \,\middle\|\, (X_i Y_i)_{d_i,u_i,r_i} \right) \right] \tag{16}$$

$$= \frac{1}{2} \sum_{i \notin C} \mathop{\mathbb{E}}_{(r_i,x_i) \leftarrow R_i^1 X_i^1} \left[\mathrm{D}\left(\left(Y_i^1\right)_{r_i,x_i} \,\middle\|\, (Y_i)_{x_i} \right) \right]$$

$$+ \frac{1}{2} \sum_{i \notin C} \mathop{\mathbb{E}}_{(r_i,y_i) \leftarrow R_i^1 Y_i^1} \left[\mathrm{D}\left(\left(X_i^1\right)_{r_i,y_i} \,\middle\|\, (X_i)_{y_i} \right) \right]. \tag{17}$$

Above, Eqs. (14) and (15) follow from Fact 2.4. Equation (16) is because (d_i, u_i, r_i) and $\left(d, u, x_{C\cup[i-1]}, y_{C\cup[i-1]}\right)$ are same up to the order. Equation (17) follows because 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 calculation is useful for achieving the conditions of Eqs. (9) and (11), exhibiting that the information carried by the messages about the sender's input is small.

$$\begin{aligned}
\delta_1 ck &\geq \left|M^1\right| \quad (|M^1| \text{represents the length of} M^1) \\
&\geq \mathrm{I}\left(X^1 Y^1 : M^1 \,\middle|\, D^1 U^1 X_C^1 Y_C^1\right) \\
&= \sum_{i \notin C} \mathrm{I}\left(X_i^1 Y_i^1 : M^1 \,\middle|\, D^1 U^1 X_{C\cup[i-1]}^1 Y_{C\cup[i-1]}^1\right) \\
&= \sum_{i \notin C} \sum_{s=1}^{t} \mathrm{I}\left(X_i^1 Y_i^1 : M_s^1 \,\middle|\, D^1 U^1 X_{C\cup[i-1]}^1 Y_{C\cup[i-1]}^1 M_{<s}^1\right) \\
&= \sum_{i \notin C} \sum_{s=1}^{t} \mathrm{I}\left(X_i^1 Y_i^1 : M_s^1 \,\middle|\, D_i^1 U_i^1 R_i^1 M_{<s}^1\right) \\
&= \sum_{i \notin C} \left(\sum_{s\text{odd}} \mathrm{I}\left(X_i^1 Y_i^1 : M_s^1 \,\middle|\, D_i^1 U_i^1 R_i^1 M_{<s}^1\right) + \sum_{s\text{even}} \mathrm{I}\left(X_i^1 Y_i^1 : M_s^1 \,\middle|\, D_i^1 U_i^1 R_i^1 M_{<s}^1\right) \right) \\
&= \frac{1}{2} \sum_{i \notin C} \left(\sum_{s\text{odd}} \mathrm{I}\left(X_i^1 : M_s^1 \,\middle|\, R_i^1 Y_i^1 M_{<s}^1\right) + \sum_{s\text{even}} \mathrm{I}\left(Y_i^1 : M_s^1 \,\middle|\, R_i^1 X_i^1 M_{<s}^1\right) \right)
\end{aligned} \tag{18}$$

Above, we have used the chain rule for the mutual information in the first two equalities. The last inequality follows because 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 calculation is useful for achieving the conditions of Eqs. (10) and (12), exhibiting that the information carried by the messages about the 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 for now.

$$\begin{aligned}
\delta k &\geq \mathrm{S}_\infty\left(D^1 U^1 X^1 Y^1 M_{\leq s}^1 \,\middle\|\, DUXYM_{\leq s}\right) \\
&\geq \mathrm{D}\left(D^1 U^1 X^1 Y^1 M_{\leq s}^1 \,\middle\|\, DUXYM_{\leq s}\right)
\end{aligned}$$

$$\geq \mathop{\mathbb{E}}_{(d,u,x_C,y_C,m_{\leq s})\leftarrow D^1U^1X^1_CY^1_CM^1_{\leq s}} \left[\mathrm{D}\left((X^1Y^1)_{d,u,x_C,y_C,m_{\leq s}} \middle\| (XY)_{d,u,x_C,y_C,m_{\leq s}}\right)\right]$$

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

$$= \sum_{i\notin C} \mathop{\mathbb{E}}_{(d_i,u_i,r_i,m_{\leq s})\leftarrow D^1_iU^1_iR^1_iM^1_{\leq s}} \left[\mathrm{D}\left((X^1_iY^1_i)_{d_i,u_i,r_i,m_{\leq s}} \middle\| (X_iY_i)_{d_i,u_i,r_i,m_{\leq s}}\right)\right] \tag{19}$$

$$\geq \frac{1}{2}\sum_{i\notin C} \mathop{\mathbb{E}}_{(x_i,r_i,m_{\leq s})\leftarrow X^1_iR^1_iM^1_{\leq s}} \left[\mathrm{D}\left((Y^1_i)_{x_i,r_i,m_{\leq s}} \middle\| (Y_i)_{x_i,r_i,m_{\leq s}}\right)\right]$$

$$= \frac{1}{2}\sum_{i\notin C} \mathop{\mathbb{E}}_{(x_i,r_i,m_{\leq s})\leftarrow X^1_iR^1_iM^1_{\leq s}} \left[\mathrm{D}\left((Y^1_i)_{x_i,r_i,m_{\leq s}} \middle\| (Y_i)_{x_i,r_i,m_{<s}}\right)\right] \tag{20}$$

$$= \frac{1}{2}\sum_{i\notin C} \mathop{\mathbb{E}}_{(x_i,r_i,m_{<s})\leftarrow X^1_iR^1_iM^1_{<s}} \left[\mathrm{D}\left((Y^1_iM^1_s)_{x_i,r_i,m_{<s}} \middle\| (Y_i)_{x_i,r_i,m_{<s}} \otimes (M^1_s)_{x_i,r_i,m_{<s}}\right)\right]$$

$$\geq \frac{1}{2}\sum_{i\notin C} \mathop{\mathbb{E}}_{(x_i,r_i,m_{<s})\leftarrow X^1_iR^1_iM^1_{<s}} \left[\mathrm{I}\left((Y^1_i)_{x_i,r_i,m_{<s}} : (M^1_s)_{x_i,r_i,m_{<s}}\right)\right] \tag{21}$$

$$= \frac{1}{2}\sum_{i\notin C} \mathrm{I}\left(Y^1_i : M^1_s \middle| X^1_iR^1_iM^1_{<s}\right) \tag{22}$$

Above, we have used Fact 2.4 several times. Equation (19) follows from the definition of R_i, Eq. (20) follows from the fact that $Y \leftrightarrow X_iR_iM_{<s} \leftrightarrow M_s$ for any i, when s is odd, and Eq. (21) follows from Fact 2.5. From a symmetric argument, we can show that when $s \in [t]$ is even then

$$\frac{1}{2}\sum_{i\notin C} \mathrm{I}\left(X^1_i : M^1_s \middle| Y^1_iR^1_iM^1_{<s}\right) \leq \delta k.$$

This and Eq. (22) together imply that

$$\sum_{i\notin C}\left(\sum_{s\,\mathrm{odd}} \mathrm{I}\left(Y^1_i : M^1_s \middle| R^1_iX^1_iM^1_{<s}\right) + \sum_{s\,\mathrm{even}} \mathrm{I}\left(X^1_i : M^1_s \middle| R^1_iY^1_iM^1_{<s}\right)\right) \leq 2\delta kt. \tag{23}$$

Note that in a true protocol, the LHS in the above inequality is 0. Here we prove that conditioning on success on all coordinates in C, it is still small.

Combining Eqs. (13), (17), (18) and (23), and making standard use of Markov's inequality, we can get a coordinate $j \notin C$ such that

$$\mathrm{D}\left(X^1_jY^1_j \middle\| X_jY_j\right) \leq 12\delta \tag{24}$$

$$\mathop{\mathbb{E}}_{(r_j,x_j)\leftarrow R^1_jX^1_j}\left[\mathrm{D}\left(\left(Y^1_j\right)_{r_j,x_j} \middle\| \left(Y_j\right)_{x_j}\right)\right] \leq 12\delta \tag{25}$$

$$\mathop{\mathbb{E}}_{(r_j,y_j)\leftarrow R_j^1 Y_j^1}\left[\mathrm{D}\left(\left(X_j^1\right)_{r_j,y_j}\,\middle\|\,\left(X_j\right)_{y_j}\right)\right] \le 12\delta \tag{26}$$

$$\sum_{s\,\mathrm{odd}} \mathrm{I}\left(X_j^1 : M_s^1 \,\middle|\, R_j^1 Y_j^1 M_{<s}^1\right) + \sum_{s\,\mathrm{even}} \mathrm{I}\left(Y_j^1 : M_s^1 \,\middle|\, R_j^1 X_j^1 M_{<s}^1\right) \le 12\delta_1 c \tag{27}$$

$$\sum_{s\,\mathrm{odd}} \mathrm{I}\left(Y_j^1 : M_s^1 \,\middle|\, R_j^1 X_j^1 M_{<s}^1\right) + \sum_{s\,\mathrm{even}} \mathrm{I}\left(X_j^1 : M_s^1 \,\middle|\, R_j^1 Y_j^1 M_{<s}^1\right) \le 12\delta t. \tag{28}$$

Let

$$\varepsilon' \stackrel{\mathrm{def}}{=} \frac{\varepsilon}{125t} \tag{29}$$

$$\varepsilon_s \stackrel{\mathrm{def}}{=} \begin{cases} \mathrm{I}\left(Y_j^1 : M_s^1 \,\middle|\, R_j^1 X_j^1 M_{<s}^1\right) & \text{if } s \in [t] \text{ is odd} \\ \mathrm{I}\left(X_j^1 : M_s^1 \,\middle|\, R_j^1 Y_j^1 M_{<s}^1\right) & \text{if } s \in [t] \text{ is even} \end{cases} \tag{30}$$

$$c_s \stackrel{\mathrm{def}}{=} \begin{cases} \mathrm{I}\left(Y_j^1 : M_s^1 \,\middle|\, R_j^1 X_j^1 M_{<s}^1\right) & \text{if } s \in [t] \text{ is even} \\ \mathrm{I}\left(X_j^1 : M_s^1 \,\middle|\, R_j^1 Y_j^1 M_{<s}^1\right) & \text{if } s \in [t] \text{ is odd.} \end{cases} \tag{31}$$

From Eq. (28) and Cauchy–Schwartz inequality, we have that $\sum_{s=1}^{t} \sqrt{\varepsilon_s} \le \sqrt{12\delta}t$. From Eqs. (24) to (26) and Lemma 2.12, we can infer that $\left(X_j, Y_j\right)$ is $(1 - 10\sqrt{3\delta})$-embeddable in $\left(X_j^1 R_j^1, Y_j^1 R_j^1\right)$. This, combined with Eqs. (27) and (28) and Lemma 3.5, (by taking ε', ε_s, and c_s in Lemma 3.5 to be Eqs. (29) and (30) and Eq. (31), respectively, and taking $XYX'Y'R'M_1' \cdots M_t'$ to be $X_j Y_j X_j^1 Y_j^1 R_j^1 M_1^1 \cdots M_t^1$) imply the following. There exists a public-coin t-round protocol $\mathscr{Q}^1$ between Alice, with input X_j, and Bob, with input Y_j, with Alice sending the first message and total communication

$$\frac{12\delta_1 c + 5t}{\varepsilon'} + O\left(t \log \frac{1}{\varepsilon'}\right) < \mathrm{D}_{\varepsilon}^{(t),\mu}(f)$$

such that at the end Alice possesses $R_A M_{A,1} \cdots M_{A,t}$ and Bob possesses $R_B M_{B,1} \cdots M_{B,t}$, satisfying

$$\left\| X_j Y_j R_A R_B M_{A,1} M_{B,1} \cdots M_{A,t} M_{B,t} - X_j^1 Y_j^1 R_j^1 R_j^1 M_1^1 M_1^1 \cdots M_t^1 M_t^1 \right\|_1$$
$$\le 10\sqrt{3\delta} + 3\sqrt{12\delta}t + 6\varepsilon' t < \frac{\varepsilon}{2}.$$

Assume towards contradiction that $\Pr\left[T_j = 1 \middle| T^{(r)} = 1\right] > 1 - \frac{\varepsilon}{2}$. Consider a protocol $\mathscr{Q}^2$ (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. He can do this because, conditioned on his input, those other random variables are independent of Alice's input. (Here we used Fact 2.15.) He 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 $\mathcal{Q}^2$ is $\Pr\left[T_j = 1 \middle| T^{(r)} = 1\right] > 1 - \frac{\varepsilon}{2}$. Now consider a 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 2.9, that the probability of success (averaged over the public coins and inputs X_j and Y_j) of $\mathcal{Q}^3$ 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 the success probability (averaged over the inputs X_j and Y_j) of $\mathcal{Q}^4$ is larger than $1-\varepsilon$. This is a contradiction to the definition of $\mathrm{D}_{\varepsilon}^{(t),\mu}(f)$. (Recall that X_jY_j are distributed according to μ.) So it must be that $\Pr\left[T_j = 1 \middle| T^{(r)} = 1\right] \leq 1 - \frac{\varepsilon}{2}$. The claim now follows by setting $i_{r+1} = j$.

4 Deferred Proofs

Proof of Lemma 3.3 Let us introduce a new random variable N with joint distribution

$$X'Y'N \stackrel{\text{def}}{=} (X'Y')(M'|X').$$

Note that $Y' \leftrightarrow X' \leftrightarrow N$ is a Markov chain. Using Lemma 2.2, we have that

$$\mathrm{D}\left(X'Y'M' \middle\| X'Y'N\right) = \mathrm{I}\left(Y' : M' \middle| X'\right) \leq \varepsilon. \tag{32}$$

Applying Fact 2.6, we get $\left\| X'Y'M' - X'Y'N \right\|_1 \leq \sqrt{\varepsilon}$. Theorem 3.1 and Claim 4.1 below together imply that there exists a public-coin protocol between Alice and Bob, with inputs X' and Y', with a single message from Alice to Bob of $\frac{c+5}{\varepsilon'} + O\left(\log \frac{1}{\varepsilon+\varepsilon'}\right) = \frac{c+5}{\varepsilon'} + O\left(\log \frac{1}{\varepsilon'}\right)$ bits, at the end of which Alice and Bob possess random variables N'_A and N'_B, respectively, satisfying

$$\left\| X'Y'N'_AN'_B - X'Y'NN \right\|_1 \leq 2\sqrt{\varepsilon} + 6\varepsilon'.$$

Finally, using the triangle inequality for the ℓ_1 norm we conclude the desired.

Claim 4.1 Let random variables X', Y', M', and N and numbers c, ε, and ε' be the same as in the statement and the proof of Lemma 3.3. It holds that

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

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

$$\begin{aligned}\log \frac{\Pr\left[N = m|X' = x\right]}{\Pr[N = m|Y' = y]} &= \log \frac{\Pr\left[N = m|X' = x, Y' = y\right]}{\Pr[N = m|Y' = y]} \\ &= \log \frac{\Pr\left[N = m|X' = x, Y' = y\right]}{\Pr[M' = m|X' = x, Y' = y]}\end{aligned}$$

$$+ \log \frac{\Pr\left[M' = m | X' = x, Y' = y\right]}{\Pr[M' = m | Y' = y]} + \log \frac{\Pr\left[M' = m, Y' = y\right]}{\Pr[N = m, Y' = y]}.$$

From the union bound, the above calculation, and using that $1 > \varepsilon > 0$, we get

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

We bound each of the above term separately, starting with the first one. Let us define the set

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

The following calculation gives a bound on the first term.

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

$$= \sum_{(m,x,y)\notin G_1} \left(\Pr\big[M' = m, X' = x, Y' = y\big] \cdot \log \frac{\Pr\big[M' = m | X' = x, Y' = y\big]}{\Pr[N = m | X' = x, Y' = y]} \right)$$

$$- \mathrm{D}\big(M'X'Y' \big\| NX'Y'\big) + \Pr\big[\big(M', X', Y'\big) \notin G_1\big] \cdot \frac{\varepsilon + 1}{\varepsilon'} \tag{35}$$

$$\geq -1 - \varepsilon + \Pr\big[\big(M', X', Y'\big) \notin G_1\big] \cdot \frac{\varepsilon + 1}{\varepsilon'} \tag{36}$$

Above, Eqs. (33) and (35) follow from the definition of the relative entropy and Eq. (34) follows from the definition of G_1. To get Eq. (36), we use Fact 2.7 and Eq. (32). Equation (36) implies that

$$\Pr\big[\big(M', X', Y'\big) \notin G_1\big] \leq \varepsilon'.$$

To upper bound the second term, let us define

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

The following calculation gives a bound on the second term.

$$c \geq \mathrm{I}\big(M' : X' \big| Y'\big) \tag{37}$$

$$= \mathop{\mathbb{E}}_{(m,x,y)\leftarrow M'X'Y'} \left[\log \frac{\Pr\big[M' = m | X' = x, Y' = y\big]}{\Pr[M' = m | Y' = y]} \right] \tag{38}$$

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

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

$$\geq -1 + \frac{c + 1}{\varepsilon'} \cdot \Pr\big[\big(M', X', Y'\big) \notin G_2\big] \tag{39}$$

Above, Eq. (37) is one of the assumptions in Lemma 3.3. Equation (38) follows from the definition of the conditional mutual information and Eq. (39) follows from the definition of G_2 and Fact 2.7. Equation (39) implies that

$$\Pr\big[\big(M', X', Y'\big) \notin G_2\big] \leq \varepsilon'.$$

To bound the last term, we define

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

The following calculation gives a bound on the third term.

$$\begin{aligned}
\varepsilon &\geq \mathrm{D}\big(X'Y'M' \big\| X'Y'N\big) \\
&\geq \mathrm{D}\big(Y'M' \big\| Y'N\big) \qquad (40)\\
&= \mathop{\mathbb{E}}_{(m,x,y)\leftarrow M'X'Y'} \left[\log \frac{\Pr\big[M'=m, Y'=y\big]}{\Pr[N=m, Y'=y]}\right] \\
&= \sum_{(m,x,y)\in G_3} \left(\Pr\big[M'=m, X'=x, Y'=y\big] \cdot \log \frac{\Pr\big[M'=m, Y'=y\big]}{\Pr[N=m, Y'=y]}\right) \\
&\quad + \sum_{(m,x,y)\notin G_3} \left(\Pr\big[M'=m, X'=x, Y'=y\big] \cdot \log \frac{\Pr\big[M'=m, Y'=y\big]}{\Pr[N=m, Y'=y]}\right) \\
&\geq -1 + \Pr\big[\big(M', X', Y'\big) \notin G_3\big] \cdot \frac{\varepsilon+1}{\varepsilon'} \qquad (41)
\end{aligned}$$

Above, Eq. (40) follows from Fact 2.8 and Eq. (41) follows from the definition of G_3 and Fact 2.7. Equation (41) implies that

$$\Pr\big[\big(M', X', Y'\big) \notin G_3\big] \leq \varepsilon'.$$

Combining the bounds for the three terms we get

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

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

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

□

Proof of Lemma 3.4 We prove the lemma by induction on t. For the base case $t = 1$, note that

$$\mathrm{I}\big(X'R' : M_1' \big| Y'R'\big) = \mathrm{I}\big(X' : M_1' \big| Y'R'\big) \leq c_1$$

and

$$\mathrm{I}\big(Y'R' : M_1' \big| X'R'\big) = \mathrm{I}\big(Y' : M_1' \big| X'R'\big) \leq \varepsilon_1.$$

Lemma 3.3 implies (by taking X', Y', and M' in Lemma 3.3 to be $X'R'$, $Y'R'$, and 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

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

bits and generate random variables $M'_{A,1}$ and $M'_{B,1}$, respectively, satisfying

$$\left\|R'X'Y'M'_1M'_1-R'X'Y'M'_{A,1}M'_{B,1}\right\|_1\le 3\sqrt{\varepsilon_1}+6\varepsilon'.$$

Now let $t>1$. We assume that t is odd. For even t, a similar argument follows. From the induction hypothesis, there exists a public-coin $t-1$ round protocol $\mathscr{P}_{t-1}$ between Alice, with input $X'R'$, and Bob, with input $Y'R'$, with Alice sending the first message, and total communication

$$\frac{\sum_{s=1}^{t-1}c_s+5(t-1)}{\varepsilon'}+O\left((t-1)\log\frac{1}{\varepsilon'}\right) \tag{42}$$

such that at the end both Alice and Bob possess random variables $M'_{A,1},\ldots,M'_{A,t-1}$ and $M'_{B,1},\ldots,M'_{B,t-1}$, satisfying

$$\begin{aligned}&\left\|R'X'Y'M'_{A,1}M'_{B,1}\cdots M'_{A,t-1}M'_{B,t-1}-R'X'Y'M'_1M'_1\cdots M'_{t-1}M'_{t-1}\right\|_1\\&\le 3\sum_{s=1}^{t-1}\sqrt{\varepsilon_s}+6\varepsilon'(t-1).\end{aligned} \tag{43}$$

Note that

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

and

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

Therefore, Lemma 3.3 implies (by taking X', Y', and M' in Lemma 3.3 to be $X'R'M'_{<t}$, $Y'R'M'_{<t}$, and 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 $\mathscr{P}$ with a single message from Alice to Bob of

$$\frac{c_t+5}{\varepsilon'}+O\left(\log\frac{1}{\varepsilon'}\right) \tag{44}$$

bits and generate new random variable $M''_{A,t}$ and $M''_{B,t}$, respectively, satisfying

$$\left\|R'X'Y'M'_1\cdots M'_{t-1}M'_tM'_t-R'X'Y'M'_1\cdots M'_{t-1}M''_{A,t}M''_{B,t}\right\|_1\le 3\sqrt{\varepsilon_t}+6\varepsilon'. \tag{45}$$

Fact 2.9 and Eq. (43) imply that

Thus, Alice, with input $X'R'M'_{A,<t}$, and Bob, with input $Y'R'M'_{B,<t}$, on running protocol $\mathscr{P}$ will generate new random variables $M'_{A,t}$ and $M'_{B,t}$, respectively, satisfying

$$\begin{aligned}&\left\|R'X'Y'M'_{A,1}M'_{B,1}\cdots M'_{A,t-1}M'_{B,t-1}M'_{A,t}M'_{B,t}-R'X'Y'M'_1M'_1\cdots M'_{t-1}M'_{t-1}M''_{A,t}M''_{B,t}\right\|_1\\&\quad=\left\|R'X'Y'M'_{A,1}M'_{B,1}\cdots M'_{A,t-1}M'_{B,t-1}-R'X'Y'M'_1M'_1\cdots M'_{t-1}M'_{t-1}\right\|_1\\&\quad\le 3\sum_{s=1}^{t-1}\sqrt{\varepsilon_s}+6\varepsilon'(t-1).\end{aligned}\tag{46}$$

where the equality follows from Fact 2.9 because $M'_{A,t}$ and $M''_{A,t}$ can be obtained by applying a same function (protocol) on $XR'M'_{A,<t}$ and $Y'R'M'_{B,<t}$, respectively. Same for $M'_{B,t}$ and $M''_{B,t}$. The equality is from Eq. (43). Therefore, by composing protocol $\mathscr{P}_{t-1}$ and protocol $\mathscr{P}$, using Eqs. (42) and (44)–(46) and the triangle inequality for the ℓ_1 norm, we get a public-coin t-round protocol $\mathscr{P}_t$ between Alice, with input $X'R'$, and Bob, with input $Y'R'$, with Alice sending the first message, and total communication

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

such that at the end Alice and Bob possess random variables $M'_{A,1},\ldots,M'_{A,t}$ and $M'_{B,1},\ldots,M'_{B,t}$, respectively, satisfying

$$\left\|R'X'Y'M'_1M'_1\cdots M'_tM'_t-R'X'Y'M'_{A,1}M'_{B,1}\cdots M'_{A,t}M'_{B,t}\right\|_1\le 3\sum_{s=1}^{t}\sqrt{\varepsilon_s}+6\varepsilon't.$$

□

Proof of Lemma 3.5 In $\mathscr{Q}_t$, Alice and Bob, using public coins and no communication, first generate R_A and R_B such that $\left\|XYR_AR_B-X'Y'R'R'\right\|_1\le\tau$. They can do this because (X,Y) is $(1-\tau)$-embeddable in $\left(X'R',Y'R'\right)$. Now they will run protocol $\mathscr{P}_t$ (given by Lemma 3.4) with Alice's input being XR_A and Bob's input being YR_B and at the end Alice and Bob possess $M_{A,1},\ldots,M_{A,t}$ and $M_{B,1},\ldots,M_{B,t}$, respectively. From Lemma 3.4, the communication of $\mathscr{Q}_t$ is as desired. Now, from Fact 2.9, Lemma 3.4, and the triangle inequality for the ℓ_1 norm, we get

$$\begin{aligned}&\left\|XYR_AR_BM_{A,1}M_{B,1}\cdots M_{A,t}M_{B,t}-X'Y'R'R'M'_1M'_1\cdots M'_tM'_t\right\|_1\\&\le\tau+3\sum_{s=1}^{t}\sqrt{\varepsilon_s}+6\varepsilon't.\end{aligned}$$

□

5 Open Problems

Some natural questions that arise from this work are:

1. Recently Braverman et al. [6] improved our result by showing that

$$\mathrm{R}^{(t),\mathrm{pub}}_{1-2^{-\Omega(\varepsilon^2k)}}\left(f^k\right)=\Omega\left(\varepsilon^2\cdot k\cdot\left(\mathrm{R}^{(7t),\mathrm{pub}}_{\varepsilon}(f)-\kappa\left(\frac{t\log t}{\varepsilon}-\frac{t}{\varepsilon^2}\right)\right)\right),$$

for some constant κ. Can the dependence on t be improved further?
2. Direct product conjectures for quantum communication complexity are still widely open. Can these techniques be extended to show direct product theorems for bounded-round quantum communication complexity?

Acknowledgments This work is funded by the Singapore Ministry of Education, partly through the Academic Research Fund Tier3 MOE2012-T3-1-009 and partly through the internal Grants of the Center for Quantum Technologies, Singapore. We thank the referees for their helpful comments.

References

1. Ambainis, A., Špalek, R., de Wolf, R.: A new quantum lower bound method, with applications to direct product theorems and time-space tradeoffs. Algorithmica **55**(3), 422–461 (2009). doi:10.1007/s00453-007-9022-9
2. Bar-Yossef, Z., Jayram, T., Kumar, R., Sivakumar, D.: An information statistics approach to data stream and communication complexity. In: Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science, FOCS '02, pp. 209–218 (2002). doi:10.1109/SFCS.2002.1181944
3. Barak, B., Braverman, M., Chen, X., Rao, A.: How to compress interactive communication. SIAM J. Comput. **42**(3), 1327–1363 (2013). doi:10.1137/100811969
4. Ben-Aroya, A., Regev, O., de Wolf, R.: A hypercontractive inequality for matrix-valued functions with applications to quantum computing and LDCs. In: Proceedings of the 49rd Annual IEEE Symposium on Foundations of Computer Science, FOCS '08, pp. 477–486 (2008). doi:10.1109/FOCS.2008.45
5. Braverman, M., Rao, A.: Information equals amortized communication. IEEE Trans. Inf. Theory **60**(10), 6058–6069 (2014). doi:10.1109/TIT.2014.2347282
6. Braverman, M., Rao, A., Weinstein, O., Yehudayoff, A.: Direct product via round-preserving compression. In: Automata, Languages, and Programming, Lecture Notes in Computer Science, vol. 7965, pp. 232–243. Springer, Berlin (2013). doi:10.1007/978-3-642-39206-1_20
7. Braverman, M., Rao, A., Weinstein, O., Yehudayoff, A.: Direct products in communication complexity. In: Proceedings of the 54rd Annual IEEE Symposium on Foundations of Computer Science, FOCS '13, pp. 746–755 (2013). doi:10.1109/FOCS.2013.85
8. Braverman, M., Weinstein, O.: An interactive information odometer with applications. In: Proceedings of the 47th Annual ACM Symposium on Theory of Computing, STOC '15, pp. 341–350 (2000). doi:10.1145/2746539.2746548
9. Chakrabarti, A., Shi, Y., Wirth, A., Yao, A.: Informational complexity and the direct sum problem for simultaneous message complexity. In: Proceedings of the 42rd Annual IEEE Symposium on Foundations of Computer Science, FOCS '01, pp. 270–278 (2001). doi:10.1109/SFCS.2001.959901
10. Cover, T.M., Thomas, J.A.: Elements of Information Theory, 2nd edn. Wiley, London (2006)
11. Drucker, A.: Improved direct product theorems for randomized query complexity. Comput. Complex. **21**(2), 197–244 (2012). doi:10.1007/s00037-012-0043-7
12. Holenstein, T.: Parallel repetition: simplification and the no-signaling case. Theory Comput. **5**(8), 141–172 (2009). doi:10.4086/toc.2009.v005a008. http://www.theoryofcomputing.org/articles/v005a008
13. Ibinson, B., Linden, N., Winter, A.: Robustness of quantum Markov chains. Commun. Math. Phys. **277**(2), 289–304 (2008). doi:10.1007/s00220-007-0362-8
14. Jain, R.: New strong direct product results in communication complexity. Journal of the ACM (JACM), 62(3), (2015). doi:10.1145/2699432
15. Jain, R., Klauck, H.: New results in the simultaneous message passing model via information theoretic techniques. In: Proceedings of the 24rd Annual IEEE Conference on Computational Complexity, CCC '09, pp. 369–378 (2009). doi:10.1109/CCC.2009.28
16. Jain, R., Klauck, H., Nayak, A.: Direct product theorems for classical communication complexity via subdistribution bounds: extended abstract. In: Proceedings of the 40rd Annual ACM Symposium on Theory of Computing, STOC '08, pp. 599–608 (2008). doi:10.1145/1374376.1374462
17. Jain, R., Radhakrishnan, J., Sen, P.: The quantum communication complexity of the pointer chasing problem: the bit version. In: FST TCS 2002: Foundations of Software Technology and Theoretical Computer Science, Lecture Notes in Computer Science, vol. 2556, pp. 218–229. Springer, Berlin (2002). doi:10.1007/3-540-36206-1_20

18. Jain, R., Radhakrishnan, J., Sen, P.: A direct sum theorem in communication complexity via message compression. In: Automata, Languages and Programming, Lecture Notes in Computer Science, vol. 2719, pp. 300–315. Springer, Berlin (2003). doi:10.1007/3-540-45061-0_26
19. Jain, R., Radhakrishnan, J., Sen, P.: A lower bound for the bounded round quantum communication complexity of set disjointness. In: Proceedings of the 44rd Annual IEEE Symposium on Foundations of Computer Science, FOCS '03, pp. 220–229 (2003). doi:10.1109/SFCS.2003.1238196
20. Jain, R., Radhakrishnan, J., Sen, P.: Prior entanglement, message compression and privacy in quantum communication. In: Proceedings of the 20rd Annual IEEE Conference on Computational Complexity, CCC '05, pp. 285–296 (2005). doi:10.1109/CCC.2005.24
21. Jain, R., Sen, P., Radhakrishnan, J.: Optimal direct sum and privacy trade-off results for quantum and classical communication complexity (2008). arXiv:0807.1267
22. Jain, R., Yao, P.: A strong direct product theorem in terms of the smooth rectangle bound (2012). arXiv:1209.0263
23. Klauck, H.: On quantum and probabilistic communication: Las Vegas and one-way protocols. In: Proceedings of the 32rd Annual ACM Symposium on Theory of Computing, STOC '00, pp. 644–651 (2000). doi:10.1145/335305.335396
24. Klauck, H.: A strong direct product theorem for disjointness. In: Proceedings of the 42rd ACM Symposium on Theory of Computing, STOC '10, pp. 77–86 (2010). doi:10.1145/1806689.1806702
25. Klauck, H., Nayak, A., Ta-Shma, A., Zuckerman, D.: Interaction in quantum communication and the complexity of set disjointness. In: Proceedings of the 33rd Annual ACM Symposium on Theory of Computing, STOC '01, pp. 124–133 (2001). doi:10.1145/380752.380786
26. Klauck, H., Špalek, R., de Wolf, R.: Quantum and classical strong direct product theorems and optimal time-space tradeoffs. SIAM J. Comput. **36**(5), 1472–1493 (2007). doi:10.1137/05063235X
27. Kushilevitz, E., Nisan, N.: Communication Complexity. Cambridge University Press, Cambridge (1996)
28. Lee, T., Roland, J.: A strong direct product theorem for quantum query complexity. Comput. Complex. **22**(2), 429–462 (2013). doi:10.1007/s00037-013-0066-8
29. Lee, T., Shraibman, A., Špalek, R.: A direct product theorem for discrepancy. In: Proceedings of the 23rd Annual IEEE Conference on Computational Complexity, CCC '08, pp. 71–80 (2008). doi:10.1109/CCC.2008.25
30. Nisan, N., Rudich, S., Saks, M.: Products and help bits in decision trees. SIAM J. Comput. **28**(3), 1035–1050 (1999). doi:10.1137/S0097539795282444
31. Nisan, N., Widgerson, A.: Rounds in communication complexity revisited. In: Proceedings of the 23rd Annual ACM Symposium on Theory of Computing, STOC '91, pp. 419–429 (1991). doi:10.1145/103418.103463
32. Parnafes, I., Raz, R., Wigderson, A.: Direct product results and the GCD problem, in old and new communication models. In: Proceedings of the 29rd Annual ACM Symposium on Theory of Computing, STOC '97, pp. 363–372 (1997). doi:10.1145/258533.258620
33. Ponzio, S.J., Radhakrishnan, J., Venkatesh, S.: The communication complexity of pointer chasing. J. Comput. Syst. Sci. **62**(2), 323–355 (2001). doi:10.1006/jcss.2000.1731. http://www.sciencedirect.com/science/article/pii/S0022000000917318
34. Raz, R.: A parallel repetition theorem. SIAM J. Comput. **27**(3), 763–803 (1998). doi:10.1137/S0097539795280895
35. Razborov, A.: On the distributional complexity of disjointness. Theor. Comput. Sci. **106**(2), 385–390 (1992). doi:10.1016/0304-3975(92)90260-M. http://www.sciencedirect.com/science/article/pii/030439759290260M
36. Shaltiel, R.: Towards proving strong direct product theorems. Comput. Complex. **12**(1–2), 1–22 (2003). doi:10.1007/s00037-003-0175-x
37. Shannon, C.E.: A mathematical theory of communication. SIGMOBILE Mob. Comput. Commun. Rev. **5**(1), 3–55 (2001). doi:10.1145/584091.584093
38. Sherstov, A.A.: Strong direct product theorems for quantum communication and query complexity. SIAM J. Comput. **41**(5), 1122–1165 (2012). doi:10.1137/110842661
39. Viola, E., Wigderson, A.: Norms, XOR lemmas, and lower bounds for polynomials and protocols. Theory Comput. **4**(7), 137–168 (2008). doi:10.4086/toc.2008.v004a007. http://www.theoryofcomputing.org/articles/v004a007

40. Yao, A.C.C.: Some complexity questions related to distributive computing (preliminary report). In: Proceedings of the 11rd Annual ACM Symposium on Theory of Computing, STOC '79, pp. 209–213 (1979). doi:10.1145/800135.804414
41. Yao, A.C.C.: Theory and application of trapdoor functions. In: Proceedings of the 23rd Annual IEEE Symposium on Foundations of Computer Science, FOCS '82, pp. 80–91 (1982). doi:10.1109/SFCS.1982.45