A near-optimal direct-sum theorem for communication complexity

Rahul Jain∗

Abstract

We show a near optimal direct-sum theorem for the two-party randomized communication complexity. Let \( f \subseteq X \times Y \times Z \) be a relation, \( \varepsilon > 0 \) and \( k \geq 1 \) be an integer. We show,

\[
R_{\varepsilon}^{\text{pub}}(f^k) \cdot \log(R_{\varepsilon}^{\text{pub}}(f^k)) \geq \Omega(k \cdot R_{\varepsilon}^{\text{pub}}(f)),
\]

where \( f^k = f \times \ldots \times f \) (\( k \)-times) and \( R_{\varepsilon}^{\text{pub}}(\cdot) \) represents the public-coin randomized communication complexity with worst-case error \( \varepsilon \).

Given a protocol \( P \) for \( f^k \) with communication cost \( c \cdot k \) and worst-case error \( \varepsilon \), we exhibit a protocol \( Q \) for \( f \) with external-information-cost \( O(c) \) and worst-error \( \varepsilon \). We then use a message compression protocol due to Barak, Braverman, Chen and Rao [2] for simulating \( Q \) with communication \( O(c \cdot \log(c \cdot k)) \) to arrive at our result.

To show this reduction we show some new chain-rules for capacity, the maximum information that can be transmitted by a communication channel. We use the powerful concept of Nash-Equilibrium in game-theory, and its existence in suitably defined games, to arrive at the chain-rules for capacity. These chain-rules are of independent interest.

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 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 direct-product (a.k.a. parallel-repetition) 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 [35]. In this model there are different parties who wish to compute a joint relation of their inputs. They do local computation, use public and-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 [26] is an excellent reference for this model.

Direct-product and direct-sum questions have been extensively investigated in different sub-models of communication complexity, a partial list includes [30, 29, 10, 11, 31, 20, 14, 21, 24, 27, 34, 18, 12, 23, 17, 3, 22, 32, 9, 13, 4, 2, 5, 8, 6, 19, 25, 7, 33].

∗Centre for Quantum Technologies, Department of Computer Science, National University of Singapore. MajuLab, UMI 3654, Singapore. Email: rahul@comp.nus.edu.sg
Our result

In this paper, we show a direct-sum theorem for the two-party randomized communication complexity. In this model, for computing a relation \( f \subseteq X \times Y \times Z \) (where \( X, Y, \) and \( Z \) are finite sets), one party, say Alice, is given an input \( x \in X \) and the other party, say Bob, is given an input \( y \in Y \). They do local computation, use public and-or private coins, exchange messages between them and at the end output an element \( z \in Z \). They succeed if \((x,y,z) \in f\). For \( \varepsilon \in (0,1) \), let \( R^\text{pub}_\varepsilon(f) \) be the two-party communication complexity of \( f \) with worst case error \( \varepsilon \) (see Definition 2.7). Let \( f^k = f \times \ldots \times f \) \((k\text{-times})\). In a protocol for \( f^k \), Alice receives input from \( X^k \), Bob receives input from \( Y^k \) and the output of the protocol is in \( Z^k \). We show the following.

**Theorem 1.1.** Let \( f \subseteq X \times Y \times Z \) be a relation, \( \varepsilon, \delta > 0 \) and \( k \geq 1 \) be an integer. Then,

\[
R^\text{pub}_\varepsilon(f^k) \cdot \log(R^\text{pub}_\varepsilon(f^k)/\delta) \geq \Omega \left( \delta^2 \cdot k \cdot R^\text{pub}_\varepsilon(f) \right),
\]

implying (using Fact 2.4),

\[
R^\text{pub}_\varepsilon(f^k) \cdot \log(R^\text{pub}_\varepsilon(f^k)) \geq \Omega \left( k \cdot R^\text{pub}_\varepsilon(f) \right).
\]

Our techniques

Most previous direct-sum results involved information theoretic arguments and proceeded as follows. Let \( \varepsilon, \delta > 0 \) and \( \mu \) be a distribution on \( X \times Y \) (possibly non-product across \( X \) and \( Y \)) such that \( R^\text{pub}_{\varepsilon+\delta}(f) = D^\mu_{\varepsilon+\delta}(f) \) \( \overset{\text{def}}{=} c \) (as guaranteed by Yao’s principle, see Fact 2.8). Consider a protocol \( \mathcal{P} \) for \( f^k \) with \( \text{CC}(\mathcal{P}) = o(kc) \) and \( \text{err}(\mathcal{P}) = \varepsilon \) (see Definition 2.7). Using chain-rule for mutual-information and use of correlation-breaking random variables one is able to obtain a protocol \( \mathcal{Q} \) for \( f \) such that the internal-information-cost \( I^\mu_{\text{INT}}(\mathcal{Q}) = o(c) \) and \( \text{err}_\mathcal{Q}(f) = \varepsilon \). So the key question that remains is: can one simulate \( \mathcal{Q} \) with another protocol \( \mathcal{Q}' \) such that \( \text{CC}(\mathcal{Q}') = O(I^\mu_{\text{INT}}(\mathcal{Q}) \cdot \text{polylog}((\text{CC}(\mathcal{Q}))) \) and \( \text{err}(\mathcal{Q}') = \text{err}(\mathcal{Q}) + \delta \)? Compression results are known that introduce dependence on the number of rounds of communication in \( \mathcal{Q} \) or heavier (than polylog) dependence on \( \text{CC}(\mathcal{Q}) \) implying various direct-sum results [2.4].

On the other hand it is known [2] that \( \mathcal{Q} \) can be simulated with another protocol \( \mathcal{Q}' \) such that \( \text{CC}(\mathcal{Q}') = O(I^\mu_{\text{EXT}}(\mathcal{Q}) \cdot \log((\text{CC}(\mathcal{Q}))) \) and \( \text{err}^\mu_{\mathcal{Q}'}(f) = \text{err}^\mu_{\mathcal{Q}}(f) + \delta \), where \( I^\mu_{\text{EXT}} \) represents external-information-cost [10]. So the question then is: can one obtain a protocol \( \mathcal{Q} \) such that \( I^\mu_{\text{EXT}}(\mathcal{Q}) = o(c) \) and \( \text{err}_\mathcal{Q}(f) = \varepsilon \)? We answer this in the affirmative. To obtain this reduction (from \( \mathcal{P} \) to \( \mathcal{Q} \)), we show some new chain-rules for capacity, the maximum information that can be transferred by a communication channel. Chain-rules for capacity (instead of chain-rules for information) facilitate bounds on external-information-cost instead of bounds on internal-information-cost. We use the powerful concept of Nash-Equilibrium in game-theory, and its existence in suitably defined games, to arrive at the chain-rules for capacity. These chain-rules are of independent interest.

Use of chain-rules for capacity to obtain a direct-sum result has been done previously by Jain and Klauck [13] to obtain an optimal direct-sum result for the private-coin classical and entanglement-unassisted quantum *Simultaneous-Message-Passing* (SMP) models. They used a chain-rule for capacity due to Jain [15] (see Fact 3.5).
Organization

In Section 2 we present some background on information theory and communication complexity. In Section 3 we prove chain-rules for capacity. In Section 4 we present the proof of the direct-sum result.

2 Preliminaries

Information theory

For natural number $k$, let $[k]$ represent the set $\{1, 2, \ldots, k\}$. For $i \in [k]$ let $-i \equiv [k] - \{i\} = \{i\}$. Similarly define $\geq i; < i; > i$. For string $x = (x_1, \ldots, x_k)$ and $T \subseteq [k]$, let $x_T$ be sub-string of $x$ with indices in $T$. For all $i$, define $(x_i, x_{-i}) \equiv x$. For a random variable $X = (X_1, \ldots, X_k)$, similarly define $X_T, X_{-i}, X_{<i}$ and so on.

Let $\mathcal{X}, \mathcal{Y}, \mathcal{M}$ be finite sets (we only consider finite sets in this work unless otherwise specified). Let $\mathcal{D}(\mathcal{X})$ be the set of probability distributions supported on $\mathcal{X}$. For $\mu \in \mathcal{D}(\mathcal{X})$, let $\mu(x)$ represent the probability of $x \in \mathcal{X}$ according to $\mu$. For a random variable $X$ taking values in $\{0, 1\}^*$ we define $|X| \equiv \max\{n \mid \Pr[X \in \{0, 1\}^n] > 0\}$. We use the same symbol to represent a random variable and its distribution whenever it is clear from the context. For jointly distributed random variables $XY$ distributed according to $\mu$, denoted $XY \sim \mu$, let $(Y|X = x) = Y_x \sim \mu_x$.

Definition 2.1. 1. The expectation value of function $f$ is denoted as

$$\mathbb{E}_{x \in \mathcal{X}}[f(x)] \equiv \sum_{x \in \mathcal{X}} \Pr[X = x] \cdot f(x).$$

2. For $\mu, \lambda \in \mathcal{D}(\mathcal{X})$, the distribution $\mu \otimes \lambda$ is defined as $(\mu \otimes \lambda)(x_1, x_2) \equiv \mu(x_1) \cdot \lambda(x_2)$. We sometimes use $(\mu, \lambda)$ to represent $\mu \otimes \lambda$ when it is clear from the context. Let $\mu^k \equiv \mu \otimes \cdots \otimes \mu$, $k$ times.

3. The $\ell_1$ distance between $\mu$ and $\lambda$ is defined to be half of the $\ell_1$ norm of $\mu - \lambda$; that is,

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

where $\lambda_S \equiv \sum_{x \in S} \lambda(x)$.

4. The entropy of $X$ is defined as: $H(X) \equiv -\sum_x \Pr[X = x] \cdot \log \Pr[X = x]$.

5. The conditional-entropy of $Y$ conditioned on $X$ is defined as

$$H(Y|X) \equiv \mathbb{E}_{x \in \mathcal{X}}[H(Y_x)] = H(XY) - H(X).$$

6. The relative-entropy between $X$ and $Y$ is defined as

$$S(X||Y) \equiv \mathbb{E}_{x \in \mathcal{X}} \left[ \log \frac{\Pr[X = x]}{\Pr[Y = x]} \right].$$
7. The mutual-information between $X$ and $Y$ is defined as

\[ I(X : Y) \overset{\text{def}}{=} H(X) + H(Y) - H(XY) \, . \]

We say that $X$ and $Y$ are independent if and only if $I(X : Y) = 0$.

8. The conditional-mutual-information between $X$ and $Y$, conditioned on $Z$, is defined as:

\[ I(X : Y \mid Z) \overset{\text{def}}{=} E_{z \rightarrow Z} [I(X : Y \mid Z = z)] = H(X \mid Z) + H(Y \mid Z) - H(XY \mid Z) \, . \]

9. Let $g : X \times Y \rightarrow \mathcal{D}(M)$ be a map (a.k.a channel). For distribution $\mu \in \mathcal{D}(X \times Y)$, define

\[ g_\mu(x) = E_{y \rightarrow x} [g(x, y)] \, ; \quad g_\mu(y) = E_{x \rightarrow y} [g(x, y)] \, ; \quad g_\mu = E_{(x,y) \rightarrow \mu} [g(x, y)] \, . \]

We will need the following basic facts. A very good text for reference on information theory is [11].

**Fact 2.2** (Chain-rule for mutual-information).

\[ I(X_1 \ldots X_k : M) = \sum_{i=1}^k I(X_i : M \mid X_{<i}) \, . \]

If $(X_1, \ldots, X_k)$ are independent then:

\[ I(X_1 \ldots X_k : M) \geq \sum_{i=1}^k I(X_i : M) \, . \]

**Fact 2.3** (Joint-convexity for relative-entropy). For all $\mu, \mu', \lambda, \lambda'$ and $p \in [0, 1]$,

\[ S(p\mu + (1-p)\mu' \| p\lambda + (1-p)\lambda') \leq p \cdot S(\mu \| \lambda) + (1-p) \cdot S(\mu' \| \lambda') \, . \]

**Fact 2.4** (Chain-rule for relative-entropy). For random variables $XY$ and $X'Y'$,

\[ S(X'Y' \| XY) = S(X' \| X) + E_{x \rightarrow X'} [S(Y'_x \| Y_x)] \, . \]

In particular, using Fact 2.3

\[ S(X'Y' \| X \otimes Y) = S(X' \| X) + E_{x \rightarrow X'} [S(Y'_x \| Y_x)] \geq S(X' \| X) + S(Y' \| Y) \, . \]

**Fact 2.5** (see e.g Fact 2.5 [19]).

\[ |X| \geq H(X) \geq I(X : Y) = E_{y \rightarrow Y} [S(X_y \| X)] = E_{x \rightarrow X} [S(Y_x \| Y)] = S(XY \| X \otimes Y) \]

\[ = \min_{X',Y'} [S(XY \| X' \otimes Y')] = \min_{Y'} E_{x \rightarrow X} [S(Y_x \| Y')] = \min_{X'} E_{y \rightarrow Y} [S(X_y \| X')] \, . \]

### Game theory

This work relies on the following powerful theorem from game theory, which is a consequence of the Kakutani fixed-point theorem in real analysis.

**Fact 2.6** (Nash-Equilibrium, Proposition 20.3 [28]). Let $k, n$ be a positive integer. Let $\mathcal{A} = \mathcal{A}_1 \times \ldots \times \mathcal{A}_k$, where each $\mathcal{A}_i$ is a non-empty, convex and compact subset of $\mathbb{R}^n$. For each $i \in [k]$, let $u_i : \mathcal{A} \rightarrow \mathbb{R}$ be a continuous function such that

∀a = (a_1, \ldots, a_k) ∈ A : the set \{a_i ∈ A_i : u_i(a_i, a_{-i}) ≥ u_i(a)\} is convex.

There is an equilibrium point $a^* \in A$ such that

\[ \forall i : \max_{a_i \in A_i} u_i(a_i, a^*_{-i}) = u_i(a^*) \, . \]
Communication complexity

Let \( f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \) be a relation 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 a two-party communication protocol (or just a protocol) \( \mathcal{P} \) for \( f \), Alice with input \( x \in \mathcal{X} \) and Bob with input \( y \in \mathcal{Y} \), do local computation, use public and-or private coins and exchange messages. The last message consists of output \( z \in \mathcal{Z} \). Let \( XY \) represent the inputs, \( M \) the messages exchanged and \( R \) the public-coin used in \( \mathcal{P} \). We call messages and public-coin together as transcript of \( \mathcal{P} \). We use \( \mathcal{P} \) to present the transcript random variable of \( \mathcal{P} \) and also the map \( \mathcal{P} : \mathcal{X} \times \mathcal{Y} \to \mathcal{D}(\mathcal{M}) \), where \( \mathcal{M} \) is the set of transcripts of \( \mathcal{P} \).

**Definition 2.7.** Let \( \mathcal{P} \) be a protocol, \( \mu \in \mathcal{D}(\mathcal{X} \times \mathcal{Y}) \) and \( XY \sim \mu \). Define,

\[
\text{CC}(\mathcal{P}) = \max_{x,y} |M(x, y)| ; \quad \text{out}_\mathcal{P}(x, y) = \text{output random variable on input } (x, y),
\]

\[
\text{err}_\mathcal{P}(f, (x, y)) = \Pr((x, y, \text{out}_\mathcal{P}(x, y)) \notin f),
\]

\[
\text{err}_\mathcal{P}(f) = \max_{x,y} \text{err}_\mathcal{P}(f, (x, y)) ; \quad \text{err}^\mu_\mathcal{P}(f) = \mathbb{E}_{(x,y)\sim\mu} [\text{err}_\mathcal{P}(f, (x, y))],
\]

\[
R^\mu_{\epsilon}(f) = \min_{\mathcal{P} : \text{err}^\mu_\mathcal{P}(f) \leq \epsilon} \text{CC}(\mathcal{P}) ; \quad D^\mu_\epsilon(f) = \min_{\mathcal{P} : \text{err}^\mu_\mathcal{P}(f) \leq \epsilon} \text{CC}(\mathcal{P}),
\]

\[
|\text{IC}^\mu_{\text{INT}}(\mathcal{P})| = I(X : \mathcal{P}|Y) + I(Y : \mathcal{P}|X) ; \quad |\text{IC}^\mu_{\text{EXT}}(\mathcal{P})| = I(XY : \mathcal{P}),
\]

\[
|\text{IC}^\mu_{\text{INT}}(\mathcal{P})| = \max_{\mu} |\text{IC}^\mu_{\text{INT}}(\mathcal{P})| ; \quad |\text{IC}^\mu_{\text{EXT}}(\mathcal{P})| = \max_{\mu} |\text{IC}^\mu_{\text{EXT}}(\mathcal{P})|.
\]

The following is a consequence of the min-max theorem in game theory which in turn is a consequence of Fact 2.6.

**Fact 2.8** (Yao’s principle [35]). \( R^\mu_{\epsilon}(f) = \Theta(R^\mu_{\epsilon'}(f)) \).

Success in randomized protocols can be boosted by the standard repetition and taking majority arguments.

**Fact 2.9.** Let \( \varepsilon, \varepsilon' > 0 \) be constants, then, \( R^\mu_{\varepsilon}(f) = \Theta(R^\mu_{\varepsilon'}(f)) \).

Following fact is known in previous works, we provide a proof for completeness.

**Fact 2.10.** Let \( \mathcal{P} \) be protocol and \( \mu = \mu_A \otimes \mu_B \) have full support in \( \mathcal{X} \times \mathcal{Y} \). Then

\[
\forall (x, y) \in \mathcal{X} \times \mathcal{Y} : \quad S(\mathcal{P}(x, y)||P_\mu) = S(\mathcal{P}(x, y)||P_\mu(x)) + S(\mathcal{P}(x, y)||P_\mu(y))
\]

**Proof.** Let \( M = (M_1 \ldots M_t) \) be the transcript of \( \mathcal{P} \), correlated with the inputs \( XY \sim \mu \) (\( M_i \) represents the \( i \)th bit in the transcript). Let \( A \subseteq [t] \) be the set of bits transmitted by Alice and \( B \subseteq [t] \) be the set of bits transmitted by Bob. Note that,

\[
\forall i \in [t], m_{<i} : \quad I(X : Y | M_{<i} = m_{<i}) = 0.
\]

This implies,

\[
\forall i \in [A], m_{<i} : \quad I(X M_i : Y | M_{<i} = m_{<i}) = 0,
\]

\[
\forall i \in [B], m_{<i} : \quad I(Y M_i : X | M_{<i} = m_{<i}) = 0. \quad (1)
\]

5
Consider, 

\[ S(P(x,y)\|\mathcal{P}_\mu) = \sum_{i \in A} t_{m_{<i} \leftarrow M_{<i}} \mathbb{E}_{m_{<i}} \left[ S(M_i(x,y)|m_{<i}|M_i|m_{<i}) \right] + \sum_{i \in B} t_{m_{<i} \leftarrow M_{<i}} \mathbb{E}_{m_{<i}} \left[ S(M_i(x,y)|m_{<i}|M_i|m_{<i}) \right] \]  

(Fact 2.4) \hspace{1cm} (2)

Also, 

\[ S(P(x,y)\|\mathcal{P}_\mu(x)) = \sum_{i \in A} t_{m_{<i} \leftarrow M_{<i}} \mathbb{E}_{m_{<i}} \left[ S(M_i(x,y)|m_{<i}|M_i(x)|m_{<i}) \right] + \sum_{i \in B} t_{m_{<i} \leftarrow M_{<i}} \mathbb{E}_{m_{<i}} \left[ S(M_i(x,y)|m_{<i}|M_i(x)|m_{<i}) \right] \]  

(Eq. (1)) \hspace{1cm} (3)

Similarly, 

\[ S(P(x,y)\|\mathcal{P}_\mu(y)) = \sum_{i \in A} t_{m_{<i} \leftarrow M_{<i}} \mathbb{E}_{m_{<i}} \left[ S(M_i(x,y)|m_{<i}|M_i(y)|m_{<i}) \right] \]  

(Eq. (4)) \hspace{1cm} (4)

Combining Eq. (2), (3), (4) we get the desired. \qed

**Definition 2.11** (Simulation of a protocol). Let \( \delta > 0 \). We say a protocol \( Q \), \( \delta \)-simulates a protocol \( P \) with inputs \( XY \), if there exists a function \( g \) such that:

\[ \mathbb{E}_{(x,y) \leftarrow XY} \left[ \|g(Q(x,y)) - P(x,y)\|_1 \right] \leq \delta . \]

Barak et al. [2] showed that any protocol \( P \) with low external-information-cost can be simulated by a protocol \( Q \) with low communication. A very nice property is that communication in \( Q \) does not depend on the number of rounds of \( P \). We use the version as stated in Theorem 10 in [5] where it is credited to [2].

**Fact 2.12** (Compression to external-information [2]). Let \( \delta > 0, \mu \in \mathcal{D}(\mathcal{X} \times \mathcal{Y}) \) and \( P \) be a protocol. There exists a protocol \( Q \) that \( \delta \)-simulates \( P \) and 

\[ \text{CC}(Q) = \mathcal{O} \left( \frac{1}{\delta^2} \cdot \text{IC}^\mu_{\text{EXT}}(P) \cdot \log(\text{CC}(P)/\delta) \right) . \]
3 Chain rules for capacity

Capacity

Let \( g : X \to D(M) \) be a map (a.k.a. channel)\(^1\).

**Definition 3.1** (Capacity). The capacity of \( g \) is defined as
\[
\text{cap}(g) \overset{\text{def}}{=} \max_{\mu \in D(X)} \mathbb{E}_{x \sim \mu} [S(g(x) \| g_\mu)] .
\]

Following notion of a capacity-dual was considered by Jain [16].

**Definition 3.2** (Capacity-dual). The capacity-dual of \( g \) is defined as
\[
\tilde{\text{cap}}(g) \overset{\text{def}}{=} \min_{\gamma \in D(X)} \max_{x \in X} S(g(x) \| g_\gamma) .
\]

Using Fact 2.3 and Fact 2.6, Jain [16] showed that capacity is lower bounded by capacity-dual.

**Fact 3.3** (Lemma 2. [16]).
\[
\text{cap}(g) \geq \max_{\mu \in D(X)} \min_{\gamma \in D(X)} \mathbb{E}_{x \sim \mu} [S(g(x) \| g_\gamma)] = \min_{\gamma \in D(X)} \max_{x \in X} S(g(x) \| g_\gamma) = \tilde{\text{cap}}(g) .
\]

We show they are in fact the same.

**Lemma 3.4.** \( \min_{M \in D(M)} \max_{x \in X} S(g(x) \| M) = \text{cap}(g) = \tilde{\text{cap}}(g) \).

**Proof.** Consider,
\[
\text{cap}(g) = \max_{\mu \in D(X)} \mathbb{E}_{x \sim \mu} [S(g(x) \| g_\mu)] \\
\leq \min_{M \in D(M)} \max_{\mu \in D(X)} \mathbb{E}_{x \sim \mu} [S(g(x) \| M)] \overset{\text{Fact 2.5}}{=} \min_{M \in D(M)} \max_{x \in X} S(g(x) \| M) \\
\leq \tilde{\text{cap}}(g) .
\]

Combined with Fact 3.3 shows the desired. \( \square \)

Chain-rules

Let \( g : X \to D(M) \) be a channel where \( X = (X_1 \times \ldots \times X_k) \). For \( i \in [k] \) and \( \mu \in D(X) \), define channel \( g^i_\mu : X_i \to D(M) \) given by \( g^i_\mu(x_i) = g_\mu(x_i) \). Let \( A = D(X_1) \times \ldots \times D(X_k) \).

Following chain-rule for capacity was shown by Jain [15].

**Fact 3.5** (A chain-rule for capacity. Theorem 2.1 [15]).
\[
\text{cap}(g) \geq \sum_{i=1}^{k} \min_{\mu \in D(X_i)} \text{cap}(g^i_\mu) .
\]

\(^1\)All the results in this section also hold for c-q channels, mapping classical inputs to quantum states.
We show a stronger chain-rule.

**Lemma 3.6 (A chain-rule for capacity).**

\[
\text{cap}(g) \geq \min_{(\theta, \gamma) \in \mathcal{A} \times \mathcal{A}} \sum_{i=1}^{k} \max_{x_i} S(g_\theta(x_i) \| g_{\theta_{-i}, \gamma_i})
\]

\[
= \min_{\theta \in \mathcal{A}} \sum_{i=1}^{k} \text{cap}(g_\theta^i) \ .
\]

**(Lemma 3.4)**

**Proof.** For all \( i \in [k], \mu = (\mu_1, \ldots, \mu_k) \in \mathcal{A} \), define

\[
u_i(\mu) = \min_{\gamma_i \in \mathcal{D}(X_i)} \mathbb{E}_{x_i \leftarrow \mu_i} \left[ S(g_\mu(x_i) \| g_{\mu_{-i}, \gamma_i}) \right] .
\]

For all \( \mu, \mu_i', \mu_i'' \in [0, 1], \)

\[
u_i(p \mu_i' + (1 - p) \mu_i'', \mu_{-i})
\]

\[
= \min_{\gamma_i} \mathbb{E}_{x_i \leftarrow \mu_i' + (1 - p) \mu_i''} \left[ S(g_\mu(x_i) \| g_{\mu_{-i}, \gamma_i}) \right]
\]

\[
= \min_{\gamma_i} \left( p \mathbb{E}_{x_i \leftarrow \mu_i'} \left[ S(g_\mu(x_i) \| g_{\mu_{-i}, \gamma_i}) \right] + (1 - p) \mathbb{E}_{x_i \leftarrow \mu_i''} \left[ S(g_\mu(x_i) \| g_{\mu_{-i}, \gamma_i}) \right] \right)
\]

\[
\geq p \left( \min_{\gamma_i} \mathbb{E}_{x_i \leftarrow \mu_i'} \left[ S(g_\mu(x_i) \| g_{\mu_{-i}, \gamma_i}) \right] \right) + (1 - p) \left( \min_{\gamma_i} \mathbb{E}_{x_i \leftarrow \mu_i''} \left[ S(g_\mu(x_i) \| g_{\mu_{-i}, \gamma_i}) \right] \right)
\]

\[
= p \cdot \nu_i(\mu_i', \mu_{-i}) + (1 - p) \cdot \nu_i(\mu_i'', \mu_{-i}) .
\]

(5)

From Eq. (5) and Fact 2.6 (by letting \( \forall i : (A_i, u_i) \leftarrow (\mathcal{D}(X_i), u_i) \)), we get \( \theta = (\theta_1, \ldots, \theta_k) \in \mathcal{A} \) such that,

\[
\forall i : u_i(\theta) = \max_{\mu_i \in \mathcal{D}(X_i)} \nu_i(\mu_i, \theta_{-i})
\]

\[
= \max_{\mu_i} \min_{\gamma_i} \mathbb{E}_{x_i \leftarrow \mu_i} \left[ S(g_\theta(x_i) \| g_{\theta_{-i}, \gamma_i}) \right]
\]

\[
= \min_{\gamma_i} \max_{\mu_i} \mathbb{E}_{x_i \leftarrow \mu_i} \left[ S(g_\theta(x_i) \| g_{\theta_{-i}, \gamma_i}) \right] .
\]

(Fact 3.3)

Let \( X = (X_1 \ldots X_k) \sim \theta \) and \( \forall x \in \mathcal{X} : (M \mid X = x) \sim g(x) \). Consider,

\[
\sum_{i=1}^{k} \min_{\gamma_i} \mathbb{E}_{x_i \leftarrow \theta_i} \left[ S(g_\theta(x_i) \| g_{\theta_{-i}, \gamma_i}) \right] = \sum_{i} u_i(\theta)
\]

\[
= \sum_{i} \min_{\gamma_i} \mathbb{E}_{x_i \leftarrow \theta_i} \left[ S(g_\theta(x_i) \| g_{\theta_{-i}, \gamma_i}) \right]
\]

\[
\leq \sum_{i} \mathbb{E}_{x_i \leftarrow \theta_i} \left[ S(g_\theta(x_i) \| g_{\theta_{-i}, \theta_{i}}) \right]
\]

\[
= \sum_{i} I(X_i : M) \ .
\]

(Fact 2.5)

\[
\leq I(X : M) \ .
\]

(Fact 2.2)

\[
\leq \text{cap}(g) .
\]

(Definition 3.1)

This concludes the desired. \( \square \)
We strengthen the chain rule to allow for conditioning on some events. Let 
\[ T = \{ (T, x_T) \mid T \subseteq [k], x_T \in \mathcal{X} \}. \]
Below whenever \( i \in T \), define \( S(\cdot \mid \cdot) \overset{\text{def}}{=} 0. \)

**Lemma 3.7** (A chain-rule for capacity).

\[ \cap(g) \geq \max_{\alpha \in \mathcal{D}(T)} \min_{\gamma_i \in \mathcal{A} \times \mathcal{A}} \sum_{t=1}^{k} \max_{x_i} \mathbb{E}_{(T, x_T) \leftarrow \alpha, x_t \leftarrow \mu_i} \left[ S(g_\theta(x_i, x_T)\|g_{\theta_{T-i}, \gamma_t}(x_T)) \right]. \]

Proof. Let \( \alpha \in \mathcal{D}(T) \). For all \( i \in [k], \mu = (\mu_1, \ldots, \mu_k) \in \mathcal{A} \), define,

\[ u_i(\mu) = \min_{\gamma_i \in \mathcal{D}(\mathcal{X})} \mathbb{E}_{(T, x_T) \leftarrow \alpha, x_t \leftarrow \mu_i} \left[ S(g_\mu(x_i, x_T)\|g_{\mu_{T-i}, \gamma_t}(x_T)) \right]. \]


For all \( \mu, \mu_i', \mu''_i, p \in [0, 1], \)

\[ u_i(p_\mu' + (1-p)\mu''_i, \mu_{T-i}) = \\min_{\gamma_i \in \mathcal{D}(\mathcal{X})} \mathbb{E}_{(T, x_T) \leftarrow \alpha, x_t \leftarrow \mu_i'} \left[ S(g_\mu(x_i, x_T)\|g_{\mu_{T-i}, \gamma_t}(x_T)) \right] \]

\[ + (1-p) \mathbb{E}_{(T, x_T) \leftarrow \alpha, x_t \leftarrow \mu''_i} \left[ S(g_\mu(x_i, x_T)\|g_{\mu_{T-i}, \gamma_t}(x_T)) \right] \]

\[ \geq p \left( \min_{\gamma_i \in \mathcal{D}(\mathcal{X})} \mathbb{E}_{(T, x_T) \leftarrow \alpha, x_t \leftarrow \mu_i'} \left[ S(g_\mu(x_i, x_T)\|g_{\mu_{T-i}, \gamma_t}(x_T)) \right] \right) \]

\[ + (1-p) \left( \min_{\gamma_i \in \mathcal{D}(\mathcal{X})} \mathbb{E}_{(T, x_T) \leftarrow \alpha, x_t \leftarrow \mu''_i} \left[ S(g_\mu(x_i, x_T)\|g_{\mu_{T-i}, \gamma_t}(x_T)) \right] \right) \]

\[ = p \cdot u_i(\mu_i', \mu_{T-i}) + (1-p) \cdot u_i(\mu''_i, \mu_{T-i}). \] (6)

From Eq. (6) and Fact 2.6 (by letting \( \forall i : (A_i, u_i) \leftarrow (\mathcal{D}(\mathcal{X}), u_i)) \), we get \( \theta = (\theta_1, \ldots, \theta_k) \in \mathcal{A} \) such that,

\[ \forall i : u_i(\theta) = \max_{\mu_i \in \mathcal{D}(\mathcal{X})} u_i(\mu_i', \theta_{T-i}) \]

\[ = \max_{\mu_i} \min_{\gamma_i \in \mathcal{D}(\mathcal{X})} \mathbb{E}_{(T, x_T) \leftarrow \alpha, x_t \leftarrow \mu_i} \left[ S(g_\theta(x_i, x_T)\|g_{\theta_{T-i}, \gamma_t}(x_T)) \right] \]

\[ = \min_{\gamma_i} \max_{x_i} \mathbb{E}_{(T, x_T) \leftarrow \alpha} \left[ S(g_\theta(x_i, x_T)\|g_{\theta_{T-i}, \gamma_t}(x_T)) \right]. \] (Fact 2.3 and Fact 2.6) (7)

Let \( X = (X_1 \ldots X_k) \sim \theta \) and \( \forall x \in \mathcal{X} : (M | X = x) \sim g(x) \). Consider,

\[ \sum_i u_i(\theta) = \sum_i \min_{\gamma_i \in \mathcal{D}(\mathcal{X})} \mathbb{E}_{(T, x_T) \leftarrow \alpha, x_t \leftarrow \theta_i} \left[ S(g_\theta(x_i, x_T)\|g_{\theta_{T-i}, \gamma_t}(x_T)) \right] \]

\[ \leq \sum_i \mathbb{E}_{(T, x_T) \leftarrow \alpha, x_t \leftarrow \theta_i} \left[ I(X_i : M | X_T = x_T) \right] \]

\[ = \sum_i \mathbb{E}_{(T, x_T) \leftarrow \alpha} \left[ I(X_i : M | X_T = x_T) \right] \]

\[ \leq \mathbb{E}_{(T, x_T) \leftarrow \alpha} \left[ I(X : M | X_T = x_T) \right] \]

\[ \leq \cap(g). \] (Definition 3.1)

Combining this with Eq. (7) concludes the desired. \( \square \)
Following is a strengthening of the above by changing the order of quantifiers.

**Lemma 3.8** (A chain-rule for capacity).

\[
\text{cap}(g) \geq \min_{(\theta, \gamma) \in A \times A} \max_{x_i} \max_{i \notin T} \sum_{x_i} \max_{i \notin T} \left[ S \left( g_{\theta}(x_i, x_T) \bigg| \bigg| g_{\theta_i, \gamma_i}(x_T) \right) \right] .
\]

**Proof.** For tuples \((\beta_1, \ldots, \beta_\ell), (\beta'_1, \ldots, \beta'_\ell)\) and \(p \in [0, 1]\), define the convex combination,

\[
p \cdot (\beta_1, \ldots, \beta_\ell) + (1-p) \cdot (\beta'_1, \ldots, \beta'_\ell) = (p\beta_1 + (1-p)\beta'_1, \ldots, p\beta_\ell + (1-p)\beta'_\ell) .
\]

For all \(\alpha \in \mathcal{D}(T), i \in [k], (\theta, \gamma), (\theta', \gamma'), p \in [0, 1]::

\[
\begin{align*}
\max_{x_i} & \quad \mathbb{E}_{(T,x_T) \leftarrow \alpha} \left[ S \left( g_{p\theta + (1-p)\theta'}(x_i, x_T) \bigg| \bigg| g_{p\theta_i, \gamma_i}(x_T) \right) \right] \\
& \leq \max_{x_i} \left( p \cdot \mathbb{E}_{(T,x_T) \leftarrow \alpha} \left[ S \left( g_{\theta}(x_i, x_T) \bigg| \bigg| g_{\theta_i, \gamma_i}(x_T) \right) \right] + (1-p) \cdot \mathbb{E}_{(T,x_T) \leftarrow \alpha} \left[ S \left( g_{\theta'}(x_i, x_T) \bigg| \bigg| g_{\theta'_i, \gamma'_i}(x_T) \right) \right] \right) \\
& \leq p \left( \max_{x_i} \mathbb{E}_{(T,x_T) \leftarrow \alpha} \left[ S \left( g_{\theta}(x_i, x_T) \bigg| \bigg| g_{\theta_i, \gamma_i}(x_T) \right) \right] \right) \\
& \quad + (1-p) \cdot \left( \max_{x_i} \mathbb{E}_{(T,x_T) \leftarrow \alpha} \left[ S \left( g_{\theta'}(x_i, x_T) \bigg| \bigg| g_{\theta'_i, \gamma'_i}(x_T) \right) \right] \right) .
\end{align*}
\]

Consider,

\[
\text{cap}(g) \geq \max_{\alpha} \min_{\theta, \gamma} \sum_i \max_{(T,x_T) \leftarrow \alpha} \mathbb{E} \left[ S \left( g_{\theta}(x_i, x_T) \bigg| \bigg| g_{\theta_i, \gamma_i}(x_T) \right) \right] \quad \text{(Lemma 3.7)}
\]

\[
\begin{align*}
& = \min_{\theta, \gamma} \max_{T,x_T} \sum_i \max_{x_i} \mathbb{E} \left[ S \left( g_{\theta}(x_i, x_T) \bigg| \bigg| g_{\theta_i, \gamma_i}(x_T) \right) \right] .
\end{align*}
\]

\[
\text{(Fact 2.6 Eq. 3)}
\]

\[\square\]

### 4 Direct-sum

We are now ready to prove the direct-sum result.

**Theorem 4.1.** Let \(f \subseteq X \times Y \times Z\) be a relation, \(\varepsilon, \delta > 0\) and \(k \geq 1\) be an integer. Then,

\[
R_{\text{pub}}^\varepsilon(f^k) \cdot \log(R_{\text{pub}}^\varepsilon(f^k)/\delta) \geq \Omega \left( \delta^2 \cdot k \cdot R_{\text{pub}}^{\varepsilon+\delta}(f) \right) ,
\]

implying (using Fact 2.4),

\[
R_{\text{pub}}^\varepsilon(f^k) \cdot \log(R_{\text{pub}}^\varepsilon(f^k)) \geq \Omega \left( k \cdot R_{\text{pub}}^\varepsilon(f) \right) .
\]
Proof. Let $\hat{\mu} \in D(X \times Y)$ be a distribution (guaranteed by Fact 2.8) be such that, $R_{\varepsilon+\delta}^{\text{ph}}(f) = D_{\varepsilon+\delta}$. Assume there is a protocol $P : X^k \times Y^k \rightarrow D(M)$ with CC-$P = k_{c}$ and err-$P(f^k) \leq \varepsilon$, where $M$ denote the set of transcripts of $P$.

Let $XY \sim \hat{\mu}$. Let $D$ be a random variable uniformly distributed in $\{0,1\}^k$. For $d \in \{0,1\}^k$, let $T_d = X_i, S_d = Y_i$ if $d_i = 0$ and $T_d = Y_i, S_d = X_i$ if $d_i = 1$. Let $T_d = T_d^1 \times \ldots \times T_d^k, S_d = S_d^1 \times \ldots \times S_d^k$. Let $\mu_d^i \sim X$ if $d_i = 0$ and $\mu_d^i \sim Y$ if $d_i = 1$. Let $\mu^d = \mu_d^1 \otimes \ldots \mu_d^d$. From Lemma 3.8 (by setting $[k] \leftarrow [2k], X \leftarrow X^k \times Y^k, M \leftarrow \mathcal{M}, g \leftarrow P$) we get $(\theta, \gamma)$ such that (below $\theta_i = (\theta_i^A, \theta_i^B)$, similarly $\gamma_i = (\gamma_i^A, \gamma_i^B)$), contains two components, one belonging to Alice and Bob each,

$$kc = \text{CC}(P) \geq \text{cap}(P)$$  \hspace{1cm} \text{(Fact 2.5)}

$$\geq \mathbb{E}_{d \leftarrow D, s \leftarrow \mu^d} \left[ \sum_{i=1}^{k} \max_{t_i \in T_d^i} S(P_\theta(t_i, s) || P_{\theta_{-i}, \gamma_i}(s)) \right]$$  \hspace{1cm} \text{(Lemma 3.8)}

$$= k \cdot \mathbb{E}_{i \leftarrow [k], d \leftarrow D, s \leftarrow \mu^d} \left[ \max_{t_i \in T_d^i} S(P_\theta(t_i, s) || P_{\theta_{-i}, \gamma_i}(s)) \right]$$

$$= \frac{k}{2} \cdot \mathbb{E}_{i \leftarrow [k], d \leftarrow D, s \leftarrow \mu^d} \left[ \max_{(x_i, y_i) \in (\mu^d)^i, Y} S(P_\theta(x_i, y_i, s_{-i}) || P_{\theta_{-i}, \gamma_i}(y_i, s_{-i})) \right]$$

$$+ \frac{k}{2} \cdot \mathbb{E}_{i \leftarrow [k], d \leftarrow D, s \leftarrow \mu^d} \left[ \max_{(x_i, s_{-i}) \in \mu^d, X} S(P_\theta(x_i, y_i, s_{-i}) || P_{\theta_{-i}, \gamma_i}(x_i, s_{-i})) \right]$$

$$\geq \frac{k}{2} \cdot \mathbb{E}_{i \leftarrow [k], d \leftarrow D, s \leftarrow \mu^d} \left[ S(P_\theta(x_i, y_i, s_{-i}) || P_{\theta_{-i}, \gamma_i}(y_i, s_{-i})) \right]$$

$$+ \frac{k}{2} \cdot \mathbb{E}_{i \leftarrow [k], d \leftarrow D, s \leftarrow \mu^d} \left[ S(P_\theta(x_i, y_i, s_{-i}) || P_{\theta_{-i}, \gamma_i}(x_i, s_{-i})) \right]$$

$$= \frac{k}{2} \cdot \mathbb{E}_{i \leftarrow [k], d \leftarrow D, s \leftarrow \mu^d} \left[ S(P_\theta(x_i, y_i, s_{-i}) || P_{\theta_{-i}, \gamma_i}(s_{-i})) \right].$$  \hspace{1cm} \text{(Fact 2.10)}

Fix $(i, s_{-i})$ such that $\mu^d_i \neq 0$.

$$2c \geq \mathbb{E}_{(x_i, y_i) \leftarrow \mu^d} \left[ S(P_\theta(x_i, y_i, s_{-i}) || P_{\theta_{-i}, \gamma_i}(s_{-i})) \right].$$  \hspace{1cm} \text{(9)}$$

Consider the following protocol $Q$ for $f$.

1. Alice gets input $\tilde{x} \in X$. Bob gets input $\tilde{y} \in Y$.
2. They set $(x_i, y_i) = (\tilde{x}, \tilde{y})$.
3. They set $s_{-i}$ in $S^{d_{-i}}$.
4. They generate $t_{-i} \leftarrow \theta_{T_{d_{-i}}}$ using private-coin and set in $T^{d_{-i}}$.
5. They run $P$. 

\footnote{For Fact 2.10 using standard continuity arguments assume w.l.o.g $\gamma_i^A \otimes \gamma_i^B$ has full support in $X_i \times Y_i$.}
Note that $\text{CC}(Q) = \text{CC}(P)$ and $\text{err}_Q(f) = \text{err}_P(f^k)$. We have,

$$
2c \geq \mathbb{E}_{(\tilde{x},\tilde{y}) \sim \mu} [S(Q(\tilde{x}, \tilde{y}) \| P_{\theta_{-i},\gamma_i}(s_{-i}))]
$$

(Eq. (9))

$$
= S(XYQ\|XY \otimes P_{\theta_{-i},\gamma_i}(s_{-i}))
$$

(Fact 2.4)

$$
\geq I(XY : Q).
$$

(Fact 2.5)

From Fact 2.12 and Definition 2.1, we get a protocol $Q_1$ that $\delta$-simulates $Q$ such that

$$
\text{CC}(Q_1) = O\left(\frac{c}{\delta^2} \log(kc/\delta)\right)
$$

and $\text{err}_{Q_1}(f) \leq \varepsilon + \delta$,

implying

$$
D_{\varepsilon+\delta}(f) = O\left(\frac{c}{\delta^2} \log(kc/\delta)\right),
$$

which concludes the desired.

\[ \square \]

Open questions

1. Braverman and Rao \[4\] defined a correlated-pointer-jumping promise-problem $\text{CPJ}(C, I)$ and showed that it is in a sense complete for the direct-sum question. Our result shows

$$
R^{\text{pub}}(\text{CPJ}(C, I)) = O(I \log C).
$$

Can we get explicit protocols for $\text{CPJ}(C, I)$ with similar communication?

2. Can our arguments be extended to show near optimal direct-product results for communication complexity?

Acknowledgment

This work is supported by the Singapore Ministry of Education and the National Research Foundation (NRF) and in part by the NRF2017-NRF-ANR004 VanQuTe Grant.

References

[1] Z. Bar-Yossef, T.S. Jayram, R. Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. In Proceedings of the 43th Annual IEEE Symposium on Foundations of Computer Science, FOCS ’02, pages 209–218, 2002.

[2] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. SIAM Journal on Computing, 42(3):1327–1363, 2013.

[3] A. Ben-Aroya, O. Regev, and R. de Wolf. A hypercontractive inequality for matrix-valued functions with applications to quantum computing and LDCs. In Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science, FOCS ’08, pages 477–486, Oct 2008.

[4] M. Braverman and A. Rao. Information equals amortized communication. IEEE Transactions on Information Theory, 60(10):6058–6069, Oct 2014.
[5] M. Braverman, A. Rao, O. Weinstein, and A. Yehudayoff. Direct products in communication complexity. In Proceedings of the 54th Annual IEEE Symposium on Foundations of Computer Science, FOCS ’13, pages 746–755, Oct 2013.

[6] Mark Braverman. Interactive information complexity. SIAM Journal on Computing, 44(6):1698–1739, 2015.

[7] Mark Braverman and Gillat Kol. Interactive compression to external information. In Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, page 964977, New York, NY, USA, 2018. Association for Computing Machinery.

[8] Mark Braverman, Anup Rao, Omri Weinstein, and Amir Yehudayoff. Direct product via round-preserving compression. In Automata, Languages, and Programming, volume 7965 of Lecture Notes in Computer Science, pages 232–243. Springer Berlin Heidelberg, 2013.

[9] Mark Braverman and Omri Weinstein. An interactive information odometer and applications. In Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing, STOC 15, page 341350, New York, NY, USA, 2015. Association for Computing Machinery.

[10] A. Chakrabarti, Yaoyun Shi, A. Wirth, and A. Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science, FOCS ’01, pages 270–278, Oct 2001.

[11] Thomas M. Cover and Joy A. Thomas. Elements of Information Theory. Wiley-Interscience, 2nd edition, 2006.

[12] Prahladh Harsha, Rahul Jain, David McAllester, and Jaikumar Radhakrishnan. The Communication Complexity of Correlation. IEEE Transactions on Information Theory, 56(1):438–449, 2010.

[13] R. Jain and H. Klauck. New results in the simultaneous message passing model via information theoretic techniques. In Proceedings of the 24th Annual IEEE Conference on Computational Complexity, CCC ’09, pages 369–378, July 2009.

[14] R. Jain, J. Radhakrishnan, and P. 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, Oct 2003.

[15] Rahul Jain. A super-additivity inequality for channel capacity of classical-quantum channels, 2005.

[16] Rahul Jain. Communication complexity of remote state preparation with entanglement. Quantum Info. Comput., 6(4):461464, July 2006.

[17] Rahul Jain. New strong direct product results in communication complexity. J. ACM, 62(3), June 2015.

[18] 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, 2008.
[19] Rahul Jain, Attila Pereszlényi, and Penghui Yao. A direct product theorem for two-party bounded-round public-coin communication complexity. *Algorithmica*, 76(3):720748, November 2016.

[20] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A direct sum theorem in communication complexity via message compression. In *Automata, Languages and Programming*, volume 2719 of *Lecture Notes in Computer Science*, pages 300–315. Springer Berlin Heidelberg, 2003.

[21] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. Prior entanglement, message compression and privacy in quantum communication. In *20th Annual IEEE Conference on Computational Complexity (CCC 2005), 11-15 June 2005, San Jose, CA, USA*, pages 285–296. IEEE Computer Society, 2005.

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

[23] 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, 2010.

[24] Hartmut Klauck, Robert Špalek, and Ronald de Wolf. Quantum and classical strong direct product theorems and optimal time-space tradeoffs. *SIAM Journal on Computing*, 36(5):1472–1493, 2007.

[25] Gillat Kol. Interactive compression for product distributions. In *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing*, STOC 16, page 987998, New York, NY, USA, 2016. Association for Computing Machinery.

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

[27] T. Lee, A. Shraibman, and R. Špalek. A direct product theorem for discrepancy. In *Proceedings of the 23rd Annual IEEE Conference on Computational Complexity, CCC ’08*, pages 71–80, June 2008.

[28] M. Osborne and A. Rubinstein. *A course in game theory*. MIT Press, 1994.

[29] Itzhak Parnafes, Ran Raz, and Avi Wigderson. Direct product results and the GCD problem, in old and new communication models. In *Proceedings of the 29th Annual ACM Symposium on Theory of Computing*, STOC ’97, pages 363–372, 1997.

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

[31] Ronen Shaltiel. Towards proving strong direct product theorems. *Computational Complexity*, 12(1-2):1–22, 2003.

[32] Alexander A. Sherstov. Strong direct product theorems for quantum communication and query complexity. *SIAM Journal on Computing*, 41(5):1122–1165, 2012.

[33] Alexander A. Sherstov. Compressing interactive communication under product distributions. *SIAM Journal on Computing*, 47(2):367–419, 2018.
[34] Emanuele Viola and Avi Wigderson. Norms, XOR lemmas, and lower bounds for polynomials and protocols. *Theory of Computing, 4*(7):137–168, 2008.

[35] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing (preliminary report). In *Proceedings of the 11th Annual ACM Symposium on Theory of Computing, STOC ’79*, pages 209–213, 1979.