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 $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation, $\varepsilon > 0$ and $k \geq 1$ be an integer. We show,

$$\mathrm{R}_{\varepsilon}^{\mathrm{pub}}(f^k) \cdot \log(\mathrm{R}_{\varepsilon}^{\mathrm{pub}}(f^k)) \geq \Omega(k \cdot \mathrm{R}_{\varepsilon}^{\mathrm{pub}}(f)) \ ,$$

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

Given a protocol $\mathcal{P}$ for f^k with communication cost $c \cdot k$ and worst-case error ε, we exhibit a protocol $\mathcal{Q}$ for f with *external-information-cost* $\mathcal{O}(c)$ and worst-error ε. We then use a message compression protocol due to Barak, Braverman, Chen and Rao [2] for *simulating* $\mathcal{Q}$ with communication $\mathcal{O}(c \cdot \log(c \cdot k))$ 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, 1, 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 $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ (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 do local computation, use public and-or private coins, exchange messages between them and at the end output an element $z \in \mathcal{Z}$. They succeed if $(x, y, z) \in f$. For $\varepsilon \in (0, 1)$, let $\mathrm{R}^{\mathrm{pub}}_{\varepsilon}(f)$ be the two-party communication complexity of f with worst case error ε (see Definition 2.7). Let $f^k = f \times \ldots \times f$ (k-times). In a protocol for f^k, Alice receives input from $\mathcal{X}^k$, Bob receives input from $\mathcal{Y}^k$ and the output of the protocol is in $\mathcal{Z}^k$. We show the following.

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

$$\mathrm{R}^{\mathrm{pub}}_{\varepsilon}(f^k) \cdot \log(\mathrm{R}^{\mathrm{pub}}_{\varepsilon}(f^k)/\delta) \geq \Omega\left(\delta^2 \cdot k \cdot \mathrm{R}^{\mathrm{pub}}_{\varepsilon+\delta}(f)\right) \ ,$$

implying (using Fact 2.9),

$$\mathrm{R}^{\mathrm{pub}}_{\varepsilon}(f^k) \cdot \log(\mathrm{R}^{\mathrm{pub}}_{\varepsilon}(f^k)) \geq \Omega\left(k \cdot \mathrm{R}^{\mathrm{pub}}_{\varepsilon}(f)\right) \ .$$

Our techniques

Most previous direct-sum results involved information theoretic arguments and proceeded as follows. Let $\varepsilon, \delta > 0$ and μ be a distribution on $\mathcal{X} \times \mathcal{Y}$ (possibly non-product across $\mathcal{X}$ and $\mathcal{Y}$) such that $\mathrm{R}^{\mathrm{pub}}_{\varepsilon+\delta}(f) = \mathrm{D}^{\mu}_{\varepsilon+\delta}(f) \stackrel{\mathrm{def}}{=} c$ (as guaranteed by Yao's principle, see Fact 2.8). Consider a protocol $\mathcal{P}$ for f^k with $\mathsf{CC}(\mathcal{P}) = o(kc)$ and $\mathsf{err}(\mathcal{P}) = \varepsilon$ (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* [1, 6] $\mathsf{IC}^{\mu}_{\mathsf{INT}}(\mathcal{Q}) = o(c)$ and $\mathsf{err}_{\mathcal{Q}}(f) = \varepsilon$. So the key question that remains is: can one *simulate* $\mathcal{Q}$ with another protocol $\mathcal{Q}'$ such that $\mathsf{CC}(\mathcal{Q}') = \mathcal{O}(\mathsf{IC}^{\mu}_{\mathsf{INT}}(\mathcal{Q}) \cdot \mathrm{polylog}(\mathsf{CC}(\mathcal{Q})))$ and $\mathsf{err}(\mathcal{Q}') = \mathsf{err}(\mathcal{Q}) + \delta$? Compression results are known that introduce dependence on the number of rounds of communication in $\mathcal{Q}$ or heavier (than polylog) dependence on $\mathsf{CC}(\mathcal{Q})$ 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 $\mathsf{CC}(\mathcal{Q}') = \mathcal{O}(\mathsf{IC}^{\mu}_{\mathsf{EXT}}(\mathcal{Q}) \cdot \log(\mathsf{CC}(\mathcal{Q})))$ and $\mathsf{err}^{\mu}_{\mathcal{Q}'}(f) = \mathsf{err}^{\mu}_{\mathcal{Q}}(f) + \delta$, where $\mathsf{IC}^{\mu}_{\mathsf{EXT}}$ represents external-information-cost [10]. So the question then is: can one obtain a protocol $\mathcal{Q}$ such that $\mathsf{IC}^{\mu}_{\mathsf{EXT}}(\mathcal{Q}) = o(c)$ and $\mathsf{err}_{\mathcal{Q}}(f) = \varepsilon$? 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 $-i \stackrel{\text{def}}{=} [k] - \{i\}; \leq i \stackrel{\text{def}}{=} [i]$. 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 $(x_i, x_{-i}) \stackrel{\text{def}}{=} x$. 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 μ. For a random variable X taking values in $\{0,1\}^*$ we define $|X| \stackrel{\text{def}}{=} \max\{n \mid \Pr[X \in \{0,1\}^n] > 0\}$. 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 μ, 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*

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

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

3. *The ℓ_1 distance between μ and λ is defined to be half of the ℓ_1 norm of $\mu - \lambda$; that is,*

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

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

4. *The entropy of X is defined as:* $\mathrm{H}(X) \stackrel{\text{def}}{=} -\sum_x \Pr[X = x] \cdot \log \Pr[X = x]$.

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

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

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

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

7. *The mutual-information between X and Y is defined as*

$$\mathrm{I}(X:Y) \stackrel{\mathrm{def}}{=} \mathrm{H}(X) + \mathrm{H}(Y) - \mathrm{H}(XY) \ .$$

We say that X and Y are independent iff $\mathrm{I}(X:Y) = 0$.

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

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

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

$$g_\mu(x) = \mathop{\mathbb{E}}_{y \leftarrow \mu_x}[g(x,y)] \ ; \ g_\mu(y) = \mathop{\mathbb{E}}_{x \leftarrow \mu_y}[g(x,y)] \ ; \ g_\mu = \mathop{\mathbb{E}}_{(x,y) \leftarrow \mu}[g(x,y)] \ .$$

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)**.**

$$\mathrm{I}(X_1 \dots X_k : M) = \sum_{i=1}^{k} \mathrm{I}(X_i : M | X_{<i}) \ .$$

If $(X_1, \dots, X_k)$ are independent then: $\mathrm{I}(X_1 \dots X_k : M) \geq \sum_{i=1}^{k} \mathrm{I}(X_i : M)$.

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

$$\mathrm{S}\big(p\mu + (1-p)\mu' \big\| p\lambda + (1-p)\lambda'\big) \leq p \cdot \mathrm{S}(\mu \| \lambda) + (1-p) \cdot \mathrm{S}\big(\mu' \big\| \lambda'\big) \ .$$

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

$$\mathrm{S}\big(X'Y' \big\| XY\big) = \mathrm{S}\big(X' \big\| X\big) + \mathop{\mathbb{E}}_{x \leftarrow X'}\big[\mathrm{S}\big(Y'_x \big\| Y_x\big)\big] \ .$$

In particular, using Fact 2.3:

$$\mathrm{S}\big(X'Y' \big\| X \otimes Y\big) = \mathrm{S}\big(X' \big\| X\big) + \mathop{\mathbb{E}}_{x \leftarrow X'}\big[\mathrm{S}\big(Y'_x \big\| Y\big)\big] \geq \mathrm{S}\big(X' \big\| X\big) + \mathrm{S}\big(Y' \big\| Y\big) \ .$$

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

$$\begin{aligned} |X| \geq \mathrm{H}(X) \geq \mathrm{I}(X:Y) &= \mathop{\mathbb{E}}_{y \leftarrow Y}[\mathrm{S}(X_y \| X)] = \mathop{\mathbb{E}}_{x \leftarrow X}[\mathrm{S}(Y_x \| Y)] = \mathrm{S}(XY \| X \otimes Y) \\ &= \min_{X',Y'} \mathrm{S}\big(XY \big\| X' \otimes Y'\big) = \min_{Y'} \mathop{\mathbb{E}}_{x \leftarrow X}\big[\mathrm{S}\big(Y_x \big\| Y'\big)\big] = \min_{X'} \mathop{\mathbb{E}}_{y \leftarrow Y}\big[\mathrm{S}\big(X_y \big\| X'\big)\big] \ . \end{aligned}$$

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 integers. 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 $u_i : \mathcal{A} \to \mathbb{R}$ be a continuous function such that*

$$\forall a = (a_1, \dots, a_k) \in \mathcal{A} \ : \ \text{the set } \{a'_i \in \mathcal{A}_i \ : \ u_i(a'_i, a_{-i}) \geq u_i(a)\} \text{ is convex.}$$

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

$$\forall i \ : \ \max_{a_i \in \mathcal{A}_i} u_i(a_i, a^*_{-i}) \quad = \quad u_i(a^*) \ .$$

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,*

$$
\begin{aligned}
\mathsf{CC}(\mathcal{P}) &= \max_{x,y} |M(x,y)| \quad ; \quad \mathsf{out}_{\mathcal{P}}(x,y) = \textit{output random variable on input } (x,y),\\
\mathsf{err}_{\mathcal{P}}(f,(x,y)) &= \Pr((x,y,\mathsf{out}_{\mathcal{P}}(x,y)) \notin f),\\
\mathsf{err}_{\mathcal{P}}(f) &= \max_{x,y} \mathsf{err}_{\mathcal{P}}(f,(x,y)) \quad ; \quad \mathsf{err}^{\mu}_{\mathcal{P}}(f) = \mathop{\mathbb{E}}_{(x,y)\leftarrow\mu} [\mathsf{err}_{\mathcal{P}}(f,(x,y))]\,,\\
\mathrm{R}^{\mathrm{pub}}_{\varepsilon}(f) &= \min_{\mathcal{P}:\ \mathsf{err}_{\mathcal{P}}(f)\le\varepsilon} \mathsf{CC}(\mathcal{P}) \quad ; \quad \mathrm{D}^{\mu}_{\varepsilon}(f) = \min_{\mathcal{P}:\ \mathsf{err}^{\mu}_{\mathcal{P}}(f)\le\varepsilon} \mathsf{CC}(\mathcal{P}),\\
\mathsf{IC}^{\mu}_{\mathsf{INT}}(\mathcal{P}) &= \mathrm{I}(X:\mathcal{P}|Y) + \mathrm{I}(Y:\mathcal{P}|X) \quad ; \quad \mathsf{IC}^{\mu}_{\mathsf{EXT}}(\mathcal{P}) = \mathrm{I}(XY:\mathcal{P})\,,\\
\mathsf{IC}_{\mathsf{INT}}(\mathcal{P}) &= \max_{\mu} \mathsf{IC}^{\mu}_{\mathsf{INT}}(\mathcal{P}) \quad ; \quad \mathsf{IC}_{\mathsf{EXT}}(\mathcal{P}) = \max_{\mu} \mathsf{IC}^{\mu}_{\mathsf{EXT}}(\mathcal{P}).
\end{aligned}
$$

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])**.** $\mathrm{R}^{\mathrm{pub}}_{\varepsilon}(f) = \max_{\mu} \mathrm{D}^{\mu}_{\varepsilon}(f)$.

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, $\mathrm{R}^{\mathrm{pub}}_{\varepsilon}(f) = \Theta(\mathrm{R}^{\mathrm{pub}}_{\varepsilon'}(f))$.*

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*

$$\forall (x,y) \in \mathcal{X} \times \mathcal{Y} : \ \mathrm{S}(\mathcal{P}(x,y) \| \mathcal{P}_{\mu}) = \mathrm{S}(\mathcal{P}(x,y) \| \mathcal{P}_{\mu}(x)) + \mathrm{S}(\mathcal{P}(x,y) \| \mathcal{P}_{\mu}(y)) \ \ .$$

Proof. Let $M = (M_1 \dots M_t)$ be the transcript of $\mathcal{P}$, correlated with the inputs $XY \sim \mu$ (M_i represents the ith 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,

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

This implies,

$$
\begin{aligned}
\forall i \in [A], m_{<i} : &\quad \mathrm{I}(XM_i : Y | M_{<i} = m_{<i}) = 0 \ \ ,\\
\forall i \in [B], m_{<i} : &\quad \mathrm{I}(YM_i : X | M_{<i} = m_{<i}) = 0 \ \ . \qquad (1)
\end{aligned}
$$

Consider,

$$
\begin{aligned}
&\mathrm{S}(\mathcal{P}(x,y)\|\mathcal{P}_\mu) \\
&= \sum_{i\in A}^{t} \mathop{\mathbb{E}}_{m_{<i}\leftarrow M_{<i}}[\mathrm{S}(M_i(x,y)|m_{<i}\|M_i|m_{<i})] + \sum_{i\in B}^{t} \mathop{\mathbb{E}}_{m_{<i}\leftarrow M_{<i}}[\mathrm{S}(M_i(x,y)|m_{<i}\|M_i|m_{<i})] && \text{(Fact 2.4)} \\
&= \sum_{i\in A}^{t} \mathop{\mathbb{E}}_{m_{<i}\leftarrow M_{<i}}[\mathrm{S}(M_i(x,y)|m_{<i}\|M_i(y)|m_{<i})] + \sum_{i\in B}^{t} \mathop{\mathbb{E}}_{m_{<i}\leftarrow M_{<i}}[\mathrm{S}(M_i(x,y)|m_{<i}\|M_i(x)|m_{<i})] \ . && \text{(Eq. (1))}
\end{aligned}
\tag{2}
$$

Also,

$$
\begin{aligned}
&\mathrm{S}(\mathcal{P}(x,y)\|\mathcal{P}_\mu(x)) \\
&= \sum_{i\in A}^{t} \mathop{\mathbb{E}}_{m_{<i}\leftarrow M_{<i}}[\mathrm{S}(M_i(x,y)|m_{<i}\|M_i(x)|m_{<i})] + \sum_{i\in B}^{t} \mathop{\mathbb{E}}_{m_{<i}\leftarrow M_{<i}}[\mathrm{S}(M_i(x,y)|m_{<i}\|M_i(x)|m_{<i})] && \text{(Fact 2.4)} \\
&= \sum_{i\in B}^{t} \mathop{\mathbb{E}}_{m_{<i}\leftarrow M_{<i}}[\mathrm{S}(M_i(x,y)|m_{<i}\|M_i(x)|m_{<i})] \ . && \text{(Eq. (1))}
\end{aligned}
\tag{3}
$$

Similarly,

$$
\mathrm{S}(\mathcal{P}(x,y)\|\mathcal{P}_\mu(y)) = \sum_{i\in A}^{t} \mathop{\mathbb{E}}_{m_{<i}\leftarrow M_{<i}}[\mathrm{S}(M_i(x,y)|m_{<i}\|M_i(y)|m_{<i})] \ . \tag{4}
$$

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

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

$$
\mathop{\mathbb{E}}_{(x,y)\leftarrow XY}[\|g(\mathcal{Q}(x,y)) - \mathcal{P}(x,y)\|_1] \le \delta \ .
$$

Barak et al. [2] showed that any protocol $\mathcal{P}$ with low external-information-cost can be simulated by a protocol $\mathcal{Q}$ with low communication. A very nice property is that communication in $\mathcal{Q}$ does not depend on the number of rounds of $\mathcal{P}$. 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 $\mathcal{P}$ be a protocol. There exists a protocol $\mathcal{Q}$ that δ-simulates $\mathcal{P}$ and*

$$
\mathsf{CC}(\mathcal{Q}) = \mathcal{O}\left(\frac{1}{\delta^2}\cdot \mathsf{IC}^{\mu}_{\mathsf{EXT}}(\mathcal{P})\cdot \log(\mathsf{CC}(\mathcal{P})/\delta)\right) \ .
$$

3 Chain rules for capacity

Capacity

Let $g : \mathcal{X} \to \mathcal{D}(\mathcal{M})$ be a map (a.k.a *channel*)[1].

Definition 3.1 (Capacity)**.** *The capacity of g is defined as*

$$\mathsf{cap}(g) \stackrel{\text{def}}{=} \max_{\mu \in \mathcal{D}(\mathcal{X})} \mathop{\mathbb{E}}_{x \leftarrow \mu} [\mathrm{S}(g(x) \| g_\mu)] \quad .$$

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

Definition 3.2 (Capacity-dual)**.** *The capacity-dual of g is defined as*

$$\widetilde{\mathsf{cap}}(g) \stackrel{\text{def}}{=} \min_{\gamma \in \mathcal{D}(\mathcal{X})} \max_{x \in \mathcal{X}} \mathrm{S}(g(x) \| g_\gamma) \quad .$$

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])**.**

$$\mathsf{cap}(g) \geq \max_{\mu \in \mathcal{D}(\mathcal{X})} \min_{\gamma \in \mathcal{D}(\mathcal{X})} \mathop{\mathbb{E}}_{x \leftarrow \mu} [\mathrm{S}(g(x) \| g_\gamma)] = \min_{\gamma \in \mathcal{D}(\mathcal{X})} \max_{x \in \mathcal{X}} \mathrm{S}(g(x) \| g_\gamma) = \widetilde{\mathsf{cap}}(g) \quad .$$

We show they are in fact the same.

Lemma 3.4. $\min_{M \in \mathcal{D}(\mathcal{M})} \max_{x \in \mathcal{X}} \mathrm{S}(g(x) \| M) = \mathsf{cap}(g) = \widetilde{\mathsf{cap}}(g)$.

Proof. Consider,

$$\begin{aligned}
\mathsf{cap}(g) &= \max_{\mu \in \mathcal{D}(\mathcal{X})} \mathop{\mathbb{E}}_{x \leftarrow \mu} [\mathrm{S}(g(x) \| g_\mu)] \\
&\leq \min_{M \in \mathcal{D}(\mathcal{M})} \max_{\mu \in \mathcal{D}(\mathcal{X})} \mathop{\mathbb{E}}_{x \leftarrow \mu} [\mathrm{S}(g(x) \| M)] && \text{(Fact 2.5)} \\
&= \min_{M \in \mathcal{D}(\mathcal{M})} \max_{x \in \mathcal{X}} \mathrm{S}(g(x) \| M) \\
&\leq \widetilde{\mathsf{cap}}(g) \quad .
\end{aligned}$$

Combined with Fact 3.3 shows the desired. □

Chain-rules

Let $g : \mathcal{X} \to \mathcal{D}(\mathcal{M})$ be a channel where $\mathcal{X} = (\mathcal{X}_1 \times \ldots \times \mathcal{X}_k)$. For $i \in [k]$ and $\mu \in \mathcal{D}(\mathcal{X})$, define channel $g^i_\mu : \mathcal{X}_i \to \mathcal{D}(\mathcal{M})$ given by $g^i_\mu(x_i) = g_\mu(x_i)$. Let $\mathcal{A} = \mathcal{D}(\mathcal{X}_1) \times \ldots \times \mathcal{D}(\mathcal{X}_k)$.

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

Fact 3.5 (A chain-rule for capacity. Theorem 2.1 [15])**.**

$$\mathsf{cap}(g) \geq \sum_{i=1}^{k} \min_{\mu \in \mathcal{D}(\mathcal{X})} \mathsf{cap}(g^i_\mu) \quad .$$

[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)**.**

$$\begin{aligned}\mathsf{cap}(g) &\geq \min_{(\theta,\gamma)\in\mathcal{A}\times\mathcal{A}} \sum_{i=1}^{k} \max_{x_i} \mathrm{S}\big(g_\theta(x_i)\big\|g_{\theta_{-i},\gamma_i}\big) \\ &= \min_{\theta\in\mathcal{A}} \sum_{i=1}^{k} \mathsf{cap}(g^i_\theta) \ . && \textit{(Lemma 3.4)}\end{aligned}$$

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

$$u_i(\mu) = \min_{\gamma_i\in\mathcal{D}(\mathcal{X}_i)} \mathop{\mathbb{E}}_{x_i\leftarrow\mu_i}\big[\mathrm{S}\big(g_\mu(x_i)\big\|g_{\mu_{-i},\gamma_i}\big)\big] \ .$$

For all $\mu, \mu_i', \mu_i'', p \in [0,1]$,

$$\begin{aligned}&u_i(p\mu_i' + (1-p)\mu_i'', \mu_{-i}) \\ &= \min_{\gamma_i} \mathop{\mathbb{E}}_{x_i\leftarrow p\mu_i'+(1-p)\mu_i''}\big[\mathrm{S}\big(g_\mu(x_i)\big\|g_{\mu_{-i},\gamma_i}\big)\big] \\ &= \min_{\gamma_i}\left(p \mathop{\mathbb{E}}_{x_i\leftarrow\mu_i'}\big[\mathrm{S}\big(g_\mu(x_i)\big\|g_{\mu_{-i},\gamma_i}\big)\big] + (1-p) \mathop{\mathbb{E}}_{x_i\leftarrow\mu_i''}\big[\mathrm{S}\big(g_\mu(x_i)\big\|g_{\mu_{-i},\gamma_i}\big)\big]\right) \\ &\geq p\left(\min_{\gamma_i} \mathop{\mathbb{E}}_{x_i\leftarrow\mu_i'}\big[\mathrm{S}\big(g_\mu(x_i)\big\|g_{\mu_{-i},\gamma_i}\big)\big]\right) + (1-p)\left(\min_{\gamma_i} \mathop{\mathbb{E}}_{x_i\leftarrow\mu_i''}\big[\mathrm{S}\big(g_\mu(x_i)\big\|g_{\mu_{-i},\gamma_i}\big)\big]\right) \\ &= p\cdot u_i(\mu_i', \mu_{-i}) + (1-p)\cdot u_i(\mu_i'', \mu_{-i}) \ . && (5)\end{aligned}$$

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

$$\begin{aligned}\forall i:\ u_i(\theta) &= \max_{\mu_i\in\mathcal{D}(\mathcal{X}_i)} u_i(\mu_i, \theta_{-i}) \\ &= \max_{\mu_i}\min_{\gamma_i} \mathop{\mathbb{E}}_{x_i\leftarrow\mu_i}\big[\mathrm{S}\big(g_\theta(x_i)\big\|g_{\theta_{-i},\gamma_i}\big)\big] \\ &= \min_{\gamma_i}\max_{x_i} \mathrm{S}\big(g_\theta(x_i)\big\|g_{\theta_{-i},\gamma_i}\big) \ . && \text{(Fact 3.3)}\end{aligned}$$

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

$$\begin{aligned}\sum_{i=1}^{k} \min_{\gamma_i}\max_{x_i} \mathrm{S}\big(g_\theta(x_i)\big\|g_{\theta_{-i},\gamma_i}\big) &= \sum_i u_i(\theta) \\ &= \sum_i \min_{\gamma_i} \mathop{\mathbb{E}}_{x_i\leftarrow\theta_i}\big[\mathrm{S}\big(g_\theta(x_i)\big\|g_{\theta_{-i},\gamma_i}\big)\big] \\ &\leq \sum_i \mathop{\mathbb{E}}_{x_i\leftarrow\theta_i}\big[\mathrm{S}\big(g_\theta(x_i)\big\|g_{\theta_{-i},\theta_i}\big)\big] \\ &= \sum_i \mathrm{I}(X_i : M) && \text{(Fact 2.5)} \\ &\leq \mathrm{I}(X : M) && \text{(Fact 2.2)} \\ &\leq \mathsf{cap}(g) \ . && \text{(Definition 3.1)}\end{aligned}$$

This concludes the desired. □

We strengthen the chain rule to allow for conditioning on some events. Let

$$\mathcal{T} = \{(T, x_T) \mid T \subseteq [k], x_T \in \mathcal{X}_T\}.$$

Below whenever $i \in T$, define $\mathrm{S}(\cdot\|\cdot) \stackrel{\text{def}}{=} 0$.

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

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

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

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

For all $\mu, \mu_i', \mu_i'', p \in [0,1]$,

$$\begin{aligned}
u_i(p\mu_i' + (1-p)\mu_i'', \mu_{-i}) &= \min_{\gamma_i} \mathop{\mathbb{E}}_{(T,x_T) \leftarrow \alpha, x_i \leftarrow p\mu_i' + (1-p)\mu_i''} \left[\mathrm{S}\big(g_\mu(x_i, x_T) \big\| g_{\mu_{-i},\gamma_i}(x_T)\big)\right] \\
&= \min_{\gamma_i} \Big(p \mathop{\mathbb{E}}_{(T,x_T) \leftarrow \alpha, x_i \leftarrow \mu_i'} \left[\mathrm{S}\big(g_\mu(x_i, x_T) \big\| g_{\mu_{-i},\gamma_i}(x_T)\big)\right] \\
&\qquad + (1-p) \mathop{\mathbb{E}}_{(T,x_T) \leftarrow \alpha, x_i \leftarrow \mu_i''} \left[\mathrm{S}\big(g_\mu(x_i, x_T) \big\| g_{\mu_{-i},\gamma_i}(x_T)\big)\right] \Big) \\
&\geq p \left(\min_{\gamma_i} \mathop{\mathbb{E}}_{(T,x_T) \leftarrow \alpha, x_i \leftarrow \mu_i'} \left[\mathrm{S}\big(g_\mu(x_i, x_T) \big\| g_{\mu_{-i},\gamma_i}(x_T)\big)\right] \right) \\
&\qquad + (1-p) \left(\min_{\gamma_i} \mathop{\mathbb{E}}_{(T,x_T) \leftarrow \alpha, x_i \leftarrow \mu_i''} \left[\mathrm{S}\big(g_\mu(x_i, x_T) \big\| g_{\mu_{-i},\gamma_i}(x_T)\big)\right] \right) \\
&= p \cdot u_i(\mu_i', \mu_{-i}) + (1-p) \cdot u_i(\mu_i'', \mu_{-i}) \ . \qquad (6)
\end{aligned}$$

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

$$\begin{aligned}
\forall i: \ u_i(\theta) &= \max_{\mu_i \in \mathcal{D}(\mathcal{X}_i)} u_i(\mu_i, \theta_{-i}) \\
&= \max_{\mu_i} \min_{\gamma_i} \mathop{\mathbb{E}}_{(T,x_T) \leftarrow \alpha, x_i \leftarrow \mu_i} \left[\mathrm{S}\big(g_\theta(x_i, x_T) \big\| g_{\theta_{-i},\gamma_i}(x_T)\big)\right] \\
&= \min_{\gamma_i} \max_{x_i} \mathop{\mathbb{E}}_{(T,x_T) \leftarrow \alpha} \left[\mathrm{S}\big(g_\theta(x_i, x_T) \big\| g_{\theta_{-i},\gamma_i}(x_T)\big)\right] \ . \qquad \text{(Fact 2.3 and Fact 2.6)} \qquad (7)
\end{aligned}$$

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

$$\begin{aligned}
\sum_i u_i(\theta) &= \sum_i \min_{\gamma_i} \mathop{\mathbb{E}}_{(T,x_T) \leftarrow \alpha, x_i \leftarrow \theta_i} \left[\mathrm{S}\big(g_\theta(x_i, x_T) \big\| g_{\theta_{-i},\gamma_i}(x_T)\big)\right] \\
&\leq \sum_i \mathop{\mathbb{E}}_{(T,x_T) \leftarrow \alpha, x_i \leftarrow \theta_i} \left[\mathrm{S}\big(g_\theta(x_i, x_T) \big\| g_{\theta_{-i},\theta_i}(x_T)\big)\right] \\
&= \sum_i \mathop{\mathbb{E}}_{(T,x_T) \leftarrow \alpha} [\mathrm{I}(X_i : M | X_T = x_T)] \qquad \text{(Fact 2.5)} \\
&\leq \mathop{\mathbb{E}}_{(T,x_T) \leftarrow \alpha} [\mathrm{I}(X : M | X_T = x_T)] \qquad \text{(Fact 2.2)} \\
&\leq \mathsf{cap}(g) \ . \qquad \text{(Definition 3.1)}
\end{aligned}$$

Combining this with Eq. (7) concludes the desired. □

Following is a strengthening of the above by changing the order of quantifiers.

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

$$\mathsf{cap}(g) \geq \min_{(\theta,\gamma)\in\mathcal{A}\times\mathcal{A}} \max_{T,x_T} \sum_{i\notin T} \max_{x_i} \mathrm{S}\left(g_\theta(x_i,x_T) \middle\| g_{\theta_{-i},\gamma_i}(x_T)\right) \quad .$$

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

$$p\cdot(\beta_1,\dots,\beta_\ell)+(1-p)\cdot(\beta'_1,\dots,\beta'_\ell) = (p\beta_1+(1-p)\beta'_1,\dots,p\beta_\ell+(1-p)\beta'_\ell) \quad .$$

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

$$\begin{aligned}
&\max_{x_i} \mathop{\mathbb{E}}_{(T,x_T)\leftarrow\alpha}\left[\mathrm{S}\left(g_{p\theta+(1-p)\theta'}(x_i,x_T)\middle\| g_{p\theta_{-i}+(1-p)\theta'_{-i},p\gamma_i+(1-p)\gamma'_i}(x_T)\right)\right] \\
&\leq \max_{x_i}\left(p\cdot \mathop{\mathbb{E}}_{(T,x_T)\leftarrow\alpha}\left[\mathrm{S}\left(g_\theta(x_i,x_T)\middle\|g_{\theta_{-i},\gamma_i}(x_T)\right)\right]\right. \\
&\qquad \left. +(1-p)\cdot \mathop{\mathbb{E}}_{(T,x_T)\leftarrow\alpha}\left[\mathrm{S}\left(g_{\theta'}(x_i,x_T)\middle\|g_{\theta'_{-i},\gamma'_i}(x_T)\right)\right]\right) && \text{(Fact 2.3)} \\
&\leq p\left(\max_{x_i} \mathop{\mathbb{E}}_{(T,x_T)\leftarrow\alpha}\left[\mathrm{S}\left(g_\theta(x_i,x_T)\middle\|g_{\theta_{-i},\gamma_i}(x_T)\right)\right]\right) \\
&\qquad +(1-p)\cdot\left(\max_{x_i} \mathop{\mathbb{E}}_{(T,x_T)\leftarrow\alpha}\left[\mathrm{S}\left(g_{\theta'}(x_i,x_T)\middle\|g_{\theta'_{-i},\gamma'_i}(x_T)\right)\right]\right) \quad . && (8)
\end{aligned}$$

Consider,

$$\begin{aligned}
\mathsf{cap}(g) &\geq \max_\alpha \min_{\theta,\gamma} \sum_i \max_{x_i} \mathop{\mathbb{E}}_{(T,x_T)\leftarrow\alpha}\left[\mathrm{S}\left(g_\theta(x_i,x_T)\middle\|g_{\theta_{-i},\gamma_i}(x_T)\right)\right] && \text{(Lemma 3.7)} \\
&= \min_{\theta,\gamma}\max_{T,x_T}\sum_{i\notin T}\max_{x_i}\mathrm{S}\left(g_\theta(x_i,x_T)\middle\|g_{\theta_{-i},\gamma_i}(x_T)\right) \quad . && \text{(Fact 2.6, Eq. (8))}
\end{aligned}$$

□

4 Direct-sum

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

Theorem 4.1. *Let $f\subseteq\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}$ be a relation, $\varepsilon,\delta>0$ and $k\geq1$ be an integer. Then,*

$$\mathrm{R}^{\mathrm{pub}}_\varepsilon(f^k)\cdot\log(\mathrm{R}^{\mathrm{pub}}_\varepsilon(f^k)/\delta) \geq \Omega\left(\delta^2\cdot k\cdot \mathrm{R}^{\mathrm{pub}}_{\varepsilon+\delta}(f)\right) \quad ,$$

implying (using Fact 2.9),

$$\mathrm{R}^{\mathrm{pub}}_\varepsilon(f^k)\cdot\log(\mathrm{R}^{\mathrm{pub}}_\varepsilon(f^k)) \geq \Omega\left(k\cdot \mathrm{R}^{\mathrm{pub}}_{\varepsilon}(f)\right) \quad .$$

Proof. Let $\tilde{\mu} \in \mathcal{D}(\mathcal{X} \times \mathcal{Y})$ be a distribution (guaranteed by Fact 2.8) be such that, $\mathrm{R}^{\mathrm{pub}}_{\varepsilon+\delta}(f) = \mathrm{D}^{\tilde{\mu}}_{\varepsilon+\delta}(f)$. Assume there is a protocol $\mathcal{P} : \mathcal{X}^k \times \mathcal{Y}^k \to \mathcal{D}(\mathcal{M})$ with $\mathsf{CC}(\mathcal{P}) = kc$ and $\mathsf{err}_{\mathcal{P}}(f^k) \leq \varepsilon$, where $\mathcal{M}$ denote the set of transcripts of $\mathcal{P}$.

Let $XY \sim \tilde{\mu}$. Let D be a random variable uniformly distributed in $\{0,1\}^k$. For $d \in \{0,1\}^k$, let $T^d_i = \mathcal{X}_i, S^d_i = \mathcal{Y}_i$ if $d_i = 0$ and $T^d_i = \mathcal{Y}_i, \mathcal{S}^d_i = \mathcal{X}_i$ 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 $\mu^d_i \sim Y$ if $d_i = 0$ and $\mu^d_i \sim X$ if $d_i = 1$. Let $\mu^d = \mu^d_1 \otimes \ldots \mu^d_k$. From Lemma 3.8 (by setting $[k] \leftarrow [2k], \mathcal{X} \leftarrow \mathcal{X}^k \times \mathcal{Y}^k, \mathcal{M} \leftarrow \mathcal{M}, g \leftarrow \mathcal{P}$) we get (θ, γ) such that (below $\theta_i = (\theta^A_i, \theta^B_i)$, similarly $\gamma_i = (\gamma^A_i, \gamma^B_i)$, contains two components, one belonging to Alice and Bob each),

$$\begin{aligned}
kc &= \mathsf{CC}(\mathcal{P}) \geq \mathsf{cap}(\mathcal{P}) && \text{(Fact 2.5)}\\
&\geq \mathop{\mathbb{E}}_{d \leftarrow D, s \leftarrow \mu^d}\left[\sum_{i=1}^{k} \max_{t_i \in T^d_i} \mathrm{S}\left(\mathcal{P}_\theta(t_i, s) \middle\| \mathcal{P}_{\theta_{-i},\gamma_i}(s)\right)\right] && \text{(Lemma 3.8)}\\
&= k \cdot \mathop{\mathbb{E}}_{i \leftarrow [k], d \leftarrow D, s \leftarrow \mu^d}\left[\max_{t_i \in T^d_i} \mathrm{S}\left(\mathcal{P}_\theta(t_i, s) \middle\| \mathcal{P}_{\theta_{-i},\gamma_i}(s)\right)\right]\\
&= \frac{k}{2} \cdot \mathop{\mathbb{E}}_{i \leftarrow [k], d_{-i} \leftarrow D_{-i}, (s_{-i}, y_i) \leftarrow (\mu^{d_{-i}}, Y)}\left[\max_{x_i \in \mathcal{X}_i} \mathrm{S}\left(\mathcal{P}_\theta(x_i, y_i, s_{-i}) \middle\| \mathcal{P}_{\theta_{-i},\gamma_i}(y_i, s_{-i})\right)\right]\\
&\quad + \frac{k}{2} \cdot \mathop{\mathbb{E}}_{i \leftarrow [k], d_{-i} \leftarrow D_{-i}, (s_{-i}, x_i) \leftarrow (\mu^{d_{-i}}, X)}\left[\max_{y_i \in \mathcal{Y}_i} \mathrm{S}\left(\mathcal{P}_\theta(x_i, y_i, s_{-i}) \middle\| \mathcal{P}_{\theta_{-i},\gamma_i}(x_i, s_{-i})\right)\right]\\
&\geq \frac{k}{2} \cdot \mathop{\mathbb{E}}_{i \leftarrow [k], d_{-i} \leftarrow D_{-i}, s_{-i} \leftarrow \mu^{d_{-i}}, (x_i, y_i) \leftarrow \tilde{\mu}}\left[\mathrm{S}\left(\mathcal{P}_\theta(x_i, y_i, s_{-i}) \middle\| \mathcal{P}_{\theta_{-i},\gamma_i}(y_i, s_{-i})\right)\right]\\
&\quad + \frac{k}{2} \cdot \mathop{\mathbb{E}}_{i \leftarrow [k], d_{-i} \leftarrow D_{-i}, s_{-i} \leftarrow \mu^{d_{-i}}, (x_i, y_i) \leftarrow \tilde{\mu}}\left[\mathrm{S}\left(\mathcal{P}_\theta(x_i, y_i, s_{-i}) \middle\| \mathcal{P}_{\theta_{-i},\gamma_i}(x_i, s_{-i})\right)\right]\\
&= \frac{k}{2} \cdot \mathop{\mathbb{E}}_{i \leftarrow [k], d_{-i} \leftarrow D_{-i}, s_{-i} \leftarrow \mu^{d_{-i}}, (x_i, y_i) \leftarrow \tilde{\mu}}\left[\mathrm{S}\left(\mathcal{P}_\theta(x_i, y_i, s_{-i}) \middle\| \mathcal{P}_{\theta_{-i},\gamma_i}(s_{-i})\right)\right] \quad . && \text{(Fact 2.10)}
\end{aligned}$$

Fix (i, d_{-i}, s_{-i}) such that[2],

$$2c \geq \mathop{\mathbb{E}}_{(x_i, y_i) \leftarrow \tilde{\mu}}\left[\mathrm{S}\left(\mathcal{P}_\theta(x_i, y_i, s_{-i}) \middle\| \mathcal{P}_{\theta_{-i},\gamma_i}(s_{-i})\right)\right] \quad . \tag{9}$$

Consider the following protocol $\mathcal{Q}$ for f.

1. Alice gets input $\tilde{x} \in \mathcal{X}$. Bob gets input $\tilde{y} \in \mathcal{Y}$.
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 $\mathcal{P}$.

[2]For Fact 2.10, using standard continuity arguments assume w.l.o.g $\gamma^A_i \otimes \gamma^B_i$ has full support in $\mathcal{X}_i \times \mathcal{Y}_i$.

Note that $\mathsf{CC}(\mathcal{Q}) = \mathsf{CC}(\mathcal{P})$ and $\mathsf{err}_{\mathcal{Q}}(f) = \mathsf{err}_{\mathcal{P}}(f^k)$. We have,

$$\begin{aligned}
2c &\geq \mathop{\mathbb{E}}_{(\tilde{x},\tilde{y})\leftarrow\tilde{\mu}} \left[\mathrm{S}\left(\mathcal{Q}(\tilde{x},\tilde{y}) \middle\| \mathcal{P}_{\theta_{-i},\gamma_i}(s_{-i})\right)\right] && \text{(Eq. (9))} \\
&= \mathrm{S}\left(XY\mathcal{Q} \middle\| XY \otimes \mathcal{P}_{\theta_{-i},\gamma_i}(s_{-i})\right) && \text{(Fact 2.4)} \\
&\geq \mathrm{I}(XY : \mathcal{Q}) \ . && \text{(Fact 2.5)}
\end{aligned}$$

From Fact 2.12 and Definition 2.1, we get a protocol $\mathcal{Q}_1$ that δ-simulates $\mathcal{Q}$ such that

$$\mathsf{CC}(\mathcal{Q}_1) = \mathcal{O}\left(\frac{c}{\delta^2}\log(kc/\delta)\right) \quad \text{and} \quad \mathsf{err}^{\tilde{\mu}}_{\mathcal{Q}_1}(f) \leq \varepsilon + \delta \ ,$$

implying

$$\mathrm{D}^{\tilde{\mu}}_{\varepsilon+\delta}(f) = \mathcal{O}\left(\frac{c}{\delta^2}\log(kc/\delta)\right) \ ,$$

which concludes the desired. □

Open questions

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

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

   Can we get explicit protocols for $\mathsf{CPJ}(\mathsf{C},\mathsf{I})$ 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.