Direct Sum Fails for Zero Error Average Communication

Gillat Kol* Shay Moran† Amir Shpilka* Amir Yehudayoff‡

Abstract

We show that in the model of zero error communication complexity, direct sum fails for average communication complexity as well as for external information cost. Our example also refutes a version of a conjecture by Braverman et al. that in the zero error case amortized communication complexity equals external information cost.

In our examples the underlying distributions do not have full support. One interpretation of a distributions of non full support is as a promise given to the players (the players have a guarantee on their inputs). This brings up the issue of promise versus non-promise problems in this context.

1 Introduction

1.1 Direct Sum for Zero Error Communication

A *direct sum* problem asks whether solving n independent copies of a given task requires n times the amount of resources needed to solve a single copy. This fundamental question has been studied in many computational models. In the context of communication complexity, direct sum theorems in various settings have been the focus of many works [FKNN95, CSWY01, JRS03, HJMR07, BBCR10, Kla10, Jai11]. We study the direct sum problem in the model of communication complexity with zero error.

1.1.1 Average Case Complexity Measures

Since we are interested in computations with zero error, we shall only consider here average case complexity measures. The reason is that worst case communication roughly corresponds to deterministic computation in the zero error case, since by fixing the randomness of a zero error protocol we get a deterministic zero error protocol with the same worst case complexity.

*Department of Computer Science, Technion-IIT, Israel. Emails: gillat.kol@gmail.com, shpilka@cs.technion.ac.il. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement number 257575.

†Max Planck Institute for Informatics, Saarbrücken, Germany. Email: shaymoran1@gmail.com.

‡Department of Mathematics, Technion-IIT, Israel. Email: amir.yehudayoff@gmail.com. Horev fellow – supported by the Taub foundation. Research is also supported by ISF and BSF.

Let π be a communication protocol between two players, one holding an input x and the other y, where (x, y) are drawn from a mutual distribution μ. The protocol π has three average case complexity measures associated with it:

1. *Average communication complexity*, denoted $\mathsf{CC}^{\mathsf{avg}}_{\mu}(\pi)$: The expected number of bits exchanged by the players while executing π.
2. *External information cost*, denoted $\mathsf{IC}^{\mathsf{ext}}_{\mu}(\pi)$: The amount of information an observer who watches the execution of π learns about the players' inputs.
3. *Internal information cost*, denoted $\mathsf{IC}^{\mathsf{int}}_{\mu}(\pi)$: The amount of information the players learn about each other's input while executing π.

We focus on the first two measures, formally defined in Section 2. For a definition of internal information cost see [Bra12a]. The following inequality gives an order between the three measures (see Section 3.2 for the proof).

Claim 1. *For every protocol π,*

$$\mathsf{CC}^{\mathsf{avg}}_{\mu}(\pi) \geq \mathsf{IC}^{\mathsf{ext}}_{\mu}(\pi) \geq \mathsf{IC}^{\mathsf{int}}_{\mu}(\pi).$$

There are several possible definitions for the measures defined above. The options relate to randomized versus distributional complexities. A detailed discussion is given in Section 4. The definitions we use are given in Section 2.

For a given function f, a distribution μ and an error parameter $\epsilon \geq 0$, these three measures on protocols induce three measures on f over μ with error ϵ. For example, $\mathsf{CC}^{\mathsf{avg}}_{\mu}(f, \epsilon)$ is the minimum $\mathsf{CC}^{\mathsf{avg}}_{\mu}(\pi)$ for a protocol π that computes f with error ϵ over μ. We also discuss two options for the definition of "computes" in Section 4.

1.1.2 Direct Sum

We are interested in the complexity of computing n independent copies of a function f, that is, of the function $f^n((x_1, \ldots, x_n), (y_1, \ldots, y_n)) = (f(x_1, y_1), \ldots, f(x_n, y_n))$ over the distribution μ^n.

It was shown in [BBCR10] that direct sum holds for internal information cost: For every f, μ and $\epsilon \geq 0$,

$$\mathsf{IC}^{\mathsf{int}}_{\mu^n}(f^n, \epsilon) = n \cdot \mathsf{IC}^{\mathsf{int}}_{\mu}(f, \epsilon).$$

Here we show (in a strong sense) that direct sum does *not* hold for average communication and for external information in the zero error case. For this part, we focus on average communication (external information is considered later on). A direct sum theorem for this model would say that if $\mathsf{CC}^{\mathsf{avg}}_{\mu}(f, 0) \geq C$ then $\mathsf{CC}^{\mathsf{avg}}_{\mu^n}(f^n, 0)$ is at least Cn or $Cn/100$ or even[1] $(C-1)n/100$. We show that this is far from true: For every $C > 0$ there exist a boolean

[1] This is because in any non trivial case the communication complexity is at least one.

function f and a distribution μ such that $\mathsf{CC}^{\mathsf{avg}}_{\mu}(f,0) \geq C$, but for every n it holds that $\mathsf{CC}^{\mathsf{avg}}_{\mu^n}(f^n,0)$ is at most roughly $C + C2^{-C}n \leq C + n$.

Observe that for "reasonable" computational models, if a function f requires an amount of C resources then f^n requires $\Omega(C+n)$ resources: The reason is that solving n copies is at least as hard as solving a single copy, and outputting the results of the n independent tasks requires at least $\Omega(n)$ resources.

1.2 Zero Error Amortized Cost Versus External Information

Another implication of our construction is the following. The amortized (average case) communication complexity of function f over μ with error $\epsilon \geq 0$ is the per-copy average communication cost, when we solve $n \to \infty$ copies of f simultaneously:

$$\mathsf{AC}^{\mathsf{avg}}_{\mu}(f,\epsilon) \triangleq \lim_{n\to\infty} \frac{\mathsf{CC}^{\mathsf{avg}}_{\mu^n}(f^n,\epsilon)}{n}.$$

Braveman et al. [Bra12a, BGPW13] conjecture that in the zero error regime, the amortized communication complexity of a function is exactly captured by its external information complexity.

Conjecture 2 ([Bra12a, BGPW13])**.** *For every function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ and distribution μ over $\mathcal{X} \times \mathcal{Y}$, it holds that*

$$\mathsf{AC}^{\mathsf{avg}}_{\mu}(f,0) = \mathsf{IC}^{\mathsf{ext}}_{\mu}(f,0).$$

Conjecture 2 was proposed as a zero error analog of the fact that, for error $\epsilon > 0$, the internal information cost of a function is equal to its amortized (worst case) communication complexity: $\mathsf{AC}_{\mu}(f,\epsilon) = \mathsf{IC}^{\mathsf{int}}_{\mu}(f,\epsilon)$ [BR10]. This connection gives an "operational" meaning to the internal information of a function, but fails for $\epsilon = 0$.

The "$\leq$" direction of Conjecture 2 is known to hold [BGPW13, HJMR07]. The conjecture is true for product distributions $\mu = \mu_x \times \mu_y$, since then the external and internal information costs coincide. Furthermore, the conjecture was shown to hold for well studied functions, such as the message transmission function $f(x,y) = x$, the bit equality function, and the AND function. Further discussion of the motivation for the conjecture can be found in [Bra12a, BGPW13].

The conjecture has several interpretations, depending on the meaning of "compute" and "average complexity" as discussed in detail in Section 4. With the definitions we use, our example refutes the conjecture.

1.3 Promise Versus Non-Promise

In our example, the measure μ does not have full support (i.e. $\mathrm{supp}(\mu) \neq \mathcal{X} \times \mathcal{Y}$). This seems to be a crucial property of μ in our proof. It, therefore, can still be the case that both direct sum and Conjecture 2 hold for measures of full support.

One way to think of a measure that does not have full support is as a "promise" that is given to the players about their inputs (the inputs are always from support). In this respect, our examples correspond to promise problems, whereas full support measures correspond to non-promise problems. This highlights a (possible) difference between promise and non-promise problems in this context that is worth a further investigation.

1.4 Results

Our main result is a construction of a function f and distribution μ on inputs with high external information and low amortized cost. The function f is the equality function on k-bit strings. The distribution μ puts almost all its weight on the diagonal (i.e., on input pairs with $x = y$). The remaining weight is on input pairs with $x < y$ (when we interpret both x and y as integers).

Theorem 3 (Main)**.** *For every $k \in \mathbb{N}$, there exist a function $f : \{0,1\}^k \times \{0,1\}^k \to \{0,1\}$ and a distribution μ over $\{0,1\}^k \times \{0,1\}^k$ such that for every $n \in \mathbb{N}$,*

$$\mathsf{IC}^{\mathsf{ext}}_{\mu}(f,0) \geq 0.99k, \text{ while } \mathsf{CC}^{\mathsf{avg}}_{\mu^n}(f^n,0) \leq 5k + 10k2^{-k}n.$$

The following two corollaries show two cases where direct sum does not hold. Both corollaries follow from Theorem 3 using Claim 1.

Corollary 4. *For every $k \in \mathbb{N}$, there exist a function $f : \{0,1\}^k \times \{0,1\}^k \to \{0,1\}$ and a distribution μ over $\{0,1\}^k \times \{0,1\}^k$ such that for every $n \in \mathbb{N}$,*

$$\mathsf{CC}^{\mathsf{avg}}_{\mu}(f,0) \geq 0.99k, \text{ while } \mathsf{CC}^{\mathsf{avg}}_{\mu^n}(f^n,0) \leq 5k + 10k2^{-k}n.$$

Corollary 5. *For every $k \in \mathbb{N}$, there exist a function $f : \{0,1\}^k \times \{0,1\}^k \to \{0,1\}$ and a distribution μ over $\{0,1\}^k \times \{0,1\}^k$ such that for every $n \in \mathbb{N}$,*

$$\mathsf{IC}^{\mathsf{ext}}_{\mu}(f,0) \geq 0.99k, \text{ while } \mathsf{IC}^{\mathsf{ext}}_{\mu^n}(f^n,0) \leq 5k + 10k2^{-k}n.$$

Another corollary of Theorem 3 refutes a version of Conjecture 2.

Corollary 6. *For every $\alpha \in (0,1)$, there exist a function $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ and a distribution μ over $\mathcal{X} \times \mathcal{Y}$ such that*

$$\mathsf{AC}^{\mathsf{avg}}_{\mu}(f,0) < \alpha \cdot \mathsf{IC}^{\mathsf{ext}}_{\mu}(f,0).$$

We complement Theorem 3 and Corollaries 4 and 5 by the following Claim 7, which provides a matching lower bound.

Claim 7. *Let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ be a function and μ a distribution over $\mathcal{X} \times \mathcal{Y}$. Assume that f is not constant on the support of μ, that is, there exist $(x,y) \neq (x',y')$ so that $\mu(x,y), \mu(x',y') > 0$ and $f(x,y) \neq f(x',y')$. Then, there exists a constant $\beta > 0$ so that for*

every $n \in \mathbb{N}$,

$$\mathsf{CC}^{\mathsf{avg}}_{\mu^n}(f^n, 0) \geq \tfrac{1}{2}\mathsf{CC}^{\mathsf{avg}}_{\mu}(f, 0) + \beta n \quad \textit{and} \quad \mathsf{IC}^{\mathsf{ext}}_{\mu^n}(f^n, 0) \geq \tfrac{1}{2}\mathsf{IC}^{\mathsf{ext}}_{\mu}(f, 0) + \beta n.$$

The constant β in the claim above depends on f and μ. It is roughly the entropy of $f(x, y)$ for $(x, y) \sim \mu$. This lower bound almost matches the upper bound given in Theorem 3 where the $k2^{-k}$ term is roughly the entropy of the distribution given in the theorem.

2 Preliminaries and Definitions

2.1 Communication Complexity

We use the definitions of protocols from [Bra12b].

Private coin protocols. Let $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ be finite sets. A *private coin protocol* π between two players, Alice and Bob, over inputs in $\mathcal{X} \times \mathcal{Y}$ and outputs in $\mathcal{Z}$, is a rooted tree with the following structure:

- Each non-leaf node is owned by either Alice or Bob.
- Each non-leaf node owned by a particular player has a set of children that are owned by the other player. Each of these children is labeled by a binary string. This labeling is prefix-free: No child has a label that is a prefix of a different child.
- Each non-leaf node owned by Alice is associated with a function mapping $\mathcal{X}$ to distributions on children of the node, and each non-leaf node owned by Bob is associated with a function mapping $\mathcal{Y}$ to distributions on children of the node.
- The leaves of the protocol are labeled by values from $\mathcal{Z}$.

On input $(x, y) \in \mathcal{X} \times \mathcal{Y}$, the protocol π is executed as in Algorithm 1.

Algorithm 1 (Generic Communication Protocol)

1. Alice gets x and Bob gets y.
2. Set v to be the root of the protocol tree.
3. If v is a leaf, the protocol ends and outputs[2] $\pi(x, y)$ which is the value labeling v. Otherwise, the player owning v samples a child of v according to the distribution associated with v and her input, and sends the child's label to the other player.
4. Set v to be the newly sampled node and return to the previous step.

[2]Observe that even for fixed x, y the value $\pi(x, y)$ may be random.

Public coin protocols. A *public coin protocol* is a distribution on private coin protocols. A public coin protocol is executed by first using the public randomness R to sample a private coin protocol π_R, and then running π_R. Every private coin protocol is thus also a public coin protocol.

Notation. In all that follows, we assume that f is a function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$, π is a public coin protocol over inputs in $\mathcal{X} \times \mathcal{Y}$, μ is a distribution over $\mathcal{X} \times \mathcal{Y}$, and $\epsilon \geq 0$ is an error parameter.

We denote by X, Y the (random) pair of inputs given to the players (i.e., they are distributed according to μ). We denote by T the random variable that is the *transcript* of the protocol π with respect to μ. That is, T is the concatenation of all the messages that are sent during the execution of π. Observe that $\mathrm{supp}(T)$ is a prefix-free code. Let $|T|$ denote the bit-length of T. We denote by R the public randomness of π.

Computing a function. We say that π *computes* f with error ϵ with respect to μ if there exists a function d of the transcript T and public randomness R so that

$$\Pr[d(T, R) \neq f(x, y)] \leq \epsilon,$$

where the probability is over $(x, y) \sim \mu$ and the randomness of π. We denote by $\Pi_{\epsilon,\mu}(f)$ the set of all protocols that compute f with error ϵ with respect to μ.

Definition 8 (**Average Communication Complexity**)**.** *The* average communication complexity *of a protocol π with respect to a distribution μ is defined as*

$$\mathsf{CC}^{\mathsf{avg}}_{\mu}(\pi) = \mathbf{E}\left[|T|\right].$$

The average communication complexity *of a function f with error ϵ with respect to a distribution μ is defined as*

$$\mathsf{CC}^{\mathsf{avg}}_{\mu}(f, \epsilon) = \inf\left\{\mathsf{CC}^{\mathsf{avg}}_{\mu}(\pi) : \pi \in \Pi_{\epsilon,\mu}(f)\right\}.$$

Definition 9 (**Amortized Communication**)**.** *The* (average case) amortized communication complexity *of a function f with error ϵ with respect to a distribution μ is defined as*

$$\mathsf{AC}^{\mathsf{avg}}_{\mu}(f, \epsilon) = \lim_{n\to\infty} \frac{\mathsf{CC}^{\mathsf{avg}}_{\mu^n}(f^n, \epsilon)}{n}.$$

We consider only average case amortized communication complexity since we focus on the zero error case where average case is more meaningful than worst case.

2.2 Information Cost

Information cost of protocols is defined using information theory notions (see e.g. [BBCR10]). Let A be a random variable taking values in a set $\mathcal{A}$. We denote by $\mathbf{H}(A)$ the *Shannon entropy*

of the distribution over $\mathcal{A}$ induced by A. The logarithms in this work (including the log in the entropy function) are taken with base 2.

Definition 10. *Let A, B, C be random variables. The* conditional entropy *of A given B is defined as*

$$\mathbf{H}(A|B) = \mathbf{H}(A, B) - \mathbf{H}(B).$$

The conditional mutual information *of A and B given C is*

$$\mathbf{I}(A; B|C) = \mathbf{H}(A|C) - \mathbf{H}(A|BC).$$

Definition 11 (**External Information**). *The* external information cost *of a public coin protocol π with respect to a distribution μ is defined as*

$$\mathsf{IC}^{\mathsf{ext}}_{\mu}(\pi) = \mathbf{I}(XY; T|R),$$

where, as before, T is the transcript of π and R is the public randomness. The external information cost *of a function f with error ϵ with respect to a distribution μ is defined as*

$$\mathsf{IC}^{\mathsf{ext}}_{\mu}(f, \epsilon) = \inf \left\{ \mathsf{IC}^{\mathsf{ext}}_{\mu}(\pi) : \pi \in \Pi_{\epsilon,\mu}(f) \right\}.$$

3 Proofs

3.1 An Example Violating Direct Sum

In this section we prove Theorem 3. The assertion of Theorem 3 follows directly from Lemma 12 (found in Section 3.1.2), Lemma 13 (found in Section 3.1.3), and Lemma 14 (found in Section 3.1.4). For the rest of the section we assume to be given $k, n \in \mathbb{N}$ as in the statement of Theorem 3.

3.1.1 The Example

The equality function. Let $EQ : \{0,1\}^k \times \{0,1\}^k \to \{0,1\}$ be the equality function on k-bit strings. That is, $EQ(x, y) = 1$ if and only if $x = y$.

The input distribution. We think of $x \in \{0,1\}^k$ both as a binary string and as an integer between 0 and $2^k - 1$. For $x, y \in \{0,1\}^k$, e.g., we write $x < y$ if x is smaller than y as integers. Let $\mu_{x=y}$ denote the uniform distribution on the set $\{(x, y) \mid x = y\} \subseteq \{0,1\}^k \times \{0,1\}^k$ and $\mu_{x<y}$ denote the uniform distribution on the set $\{(x, y) \mid x < y\} \subseteq \{0,1\}^k \times \{0,1\}^k$. Set

$$\delta = \frac{1}{25k2^{2k}}.$$

Define

$$\mu = (1 - \delta) \cdot \mu_{x=y} + \delta \cdot \mu_{x<y}.$$

3.1.2 The External Information of EQ

The following lemma gives a lower bound of $(1-\delta)k$ on the zero error external information cost of EQ with respect to μ. Informally speaking, the reason is that in the case that $x = y = a$, the protocol must reveal a to make sure that x and y are equal.

Lemma 12. $\mathsf{IC}^{\mathsf{ext}}_{\mu}(EQ, 0) \geq (1-\delta)k \geq 0.99k$.

Proof. It suffices to prove the lemma for private coin protocols (as external information is defined as an average over the public randomness). Let π be a zero error private coin protocol for EQ.

Let t be a possible transcript of π. Denote by $A(t)$ the set of all input pairs (x, y) so that $\mu(x, y) > 0$ and[3] $\Pr[T(x, y) = t] > 0$. Since π has zero error, $A(t)$ is a *μ-monochromatic rectangle*. That is, there exist subsets $\mathcal{X}', \mathcal{Y}' \subseteq \{0,1\}^k$ such that $A(t) = (\mathcal{X}' \times \mathcal{Y}') \cap \mathrm{supp}(\mu)$,[4] and for every $(x, y), (x', y') \in A(t)$ it holds that $EQ(x, y) = EQ(x', y')$.

Let $a \in \{0,1\}^k$. We claim that if $(a, a) \in A(t)$ then $A(t) = \{(a, a)\}$: By definition of μ, if $x > y$ then $(x, y) \notin A(t)$ (as it is not in $\mathrm{supp}(\mu)$). Since $A(t)$ is a μ-monochromatic rectangle, it cannot contain any (x, y) with $x < y$ as well. Finally, if $(a', a') \in A(t)$ for some $a' > a$, then since $A(t)$ is a rectangle it is also the case that $(a, a') \in A(t)$, which is impossible (a similar argument works for $a' < a$).

Let T denote the transcript of π and think of $A = A(T)$ as a random variable. Therefore,

$$\mathbf{H}(A \mid X = Y, XY) = 0.$$

Furthermore, conditioned on $X = Y$, the random variable $A(T)$ is uniform over $\{0,1\}^k$, and so

$$\mathbf{H}(A \mid X = Y) = k.$$

It follows that

$$\mathbf{I}(XY; A \mid X = Y) = \mathbf{H}(A \mid X = Y) - \mathbf{H}(A \mid X = Y, XY) = k.$$

Let E be indicator random variable of the event $X = Y$. Finally,

$$\begin{aligned}
\mathsf{IC}^{\mathsf{ext}}_{\mu}(\pi) &= \mathbf{I}(XY; T) \\
&\geq \mathbf{I}(XY; A) && (T \text{ determines } A) \\
&= \mathbf{I}(XYE; A) && (XY \text{ determines } E) \\
&= \mathbf{I}(E; A) + \Pr[E = 1] \cdot \mathbf{I}(XY; A \mid E = 1) \\
&\qquad + \Pr[E = 0] \cdot \mathbf{I}(XY; A \mid E = 0) && (\text{the chain rule}) \\
&\geq \Pr[X = Y] \cdot \mathbf{I}(XY; A \mid X = Y) \\
&= (1-\delta)k.
\end{aligned}$$

[3] Even for fixed (x, y) the transcript $T(x, y)$ is a random variable that depends on the private coins.

[4] $\mathrm{supp}(\mu)$ is the set of all $(x, y) \in \mathcal{X}' \times \mathcal{Y}'$ such that $\mu(x, y) > 0$.

□

3.1.3 Efficiently Computing EQ^n for $n < 2^k$

Consider the zero error protocol π_n for EQ^n described in Algorithm 2:

Algorithm 2 (A zero error protocol π_n for EQ^n over μ^n, $n < 2^k$)

1. Alice computes $x = \sum_{i=1}^{n} x_i$ and Bob computes $y = \sum_{i=1}^{n} y_i$, where x_i, y_i are viewed as integers, and "+" is integer addition[5]. So, the bit-length of x, y is most $k + \log n + 1$.
2. Alice sends x to Bob, and Bob sends y to Alice. If $x = y$ the protocol terminates and outputs: "$\forall i \in [n] : \ EQ(x_i, y_i) = 1$".
3. Otherwise, Alice sends $x_1, \ldots, x_n$ to Bob, and Bob sends $y_1, \ldots, y_n$ to Alice. The protocol trivially outputs the true value of EQ^n.

The protocol π_n works. By choice of μ, for every $i \in [n]$, it holds that $x_i \leq y_i$. If strict inequality $x_i < y_i$ holds for some $i \in [n]$, then $x < y$. Therefore, $x = y$ if and only if $x_i = y_i$ for every $i \in [n]$.

Lemma 13. *For $n < 2^k$ it holds that $\mathsf{CC}^{\mathsf{avg}}_{\mu^n}(\pi_n) \leq 5k$.*

Proof. Denote by p the probability that there exists $i \in [n]$ such that $x_i \neq y_i$. It holds that

$$p = 1 - (1 - \delta)^n \leq 1 - (1 - \delta n) = \delta n.$$

Step 2 of the protocol π_n exchanges at most $2(k + \log n + 1)$ bits. If $x = y$, the protocol ends after Step 2. Otherwise, if $x \neq y$, at most $2kn$ additional bits are sent in Step 3. We conclude that

$$\mathsf{CC}^{\mathsf{avg}}_{\mu^n}(\pi_n) \leq (1 - p) \cdot 2(k + \log n + 1) + p \cdot 2kn \leq 2(k + \log n + 1) + 2\delta kn^2 \leq 5k,$$

where in the last inquality we used the fact that $\delta = \frac{1}{25k2^{2k}}$ and $n < 2^k$. □

3.1.4 Efficiently Computing EQ^n for $n \geq 2^k$

Consider the zero error protocol π'_n for EQ^n described in Algorithm 3:

[5] For example, $01 + 11 = 100$.

Algorithm 3 (A zero error protocol π'_n for EQ^n over μ^n, $n \geq 2^k$)

Express n as $n = d2^k + k'$ for $d \in \mathbb{N}$ and $k' \in \{0, \ldots, 2^k - 1\}$. Partition the input pairs into d disjoint sets of 2^k pairs each, and one set of k' pairs. Run the previous protocol (for small n) on each of these $d+1$ sets. That is, run

$$\pi_{2^k}\left((x_1, \ldots, x_{2^k}), (y_1, \ldots, y_{2^k})\right),$$
$$\pi_{2^k}\left((x_{2^k+1}, \ldots, x_{2\cdot 2^k}), (y_{2^k+1}, \ldots, y_{2\cdot 2^k})\right), \ldots,$$
$$\pi_{k'}\left((x_{d2^k+1}, \ldots, x_n), (y_{d2^k+1}, \ldots, y_n)\right).$$

Lemma 14. *For $n \geq 2^k$ it holds that $\mathsf{CC}^{\mathsf{avg}}_{\mu^n}(\pi'_n) \leq 10k2^{-k}n$.*

Proof. By Lemma 13,

$$\mathsf{CC}^{\mathsf{avg}}_{\mu^n}(\pi'_n) \leq d \cdot \mathsf{CC}^{\mathsf{avg}}_{\mu^k}(\pi_k) + \mathsf{CC}^{\mathsf{avg}}_{\mu^{k'}}(\pi_{k'}) \leq (d+1)5k \leq 10k2^{-k}n.$$

□

3.2 Order on Measures

Here we prove Claim 1. The inequality $\mathsf{IC}^{\mathsf{ext}}_{\mu}(\pi) \geq \mathsf{IC}^{\mathsf{int}}_{\mu}(\pi)$ is known to hold, see e.g. [BBCR10]. We show $\mathsf{CC}^{\mathsf{avg}}_{\mu}(\pi) \geq \mathsf{IC}^{\mathsf{ext}}_{\mu}(\pi)$. Intuitively, the proof uses the fact that each bit sent by the players can contain at most one bit of information about the inputs.

Formally, let π be a public coin protocol over an input space $\mathcal{X} \times \mathcal{Y}$, and let μ be a distribution over $\mathcal{X} \times \mathcal{Y}$. Denote by π_R the private coin protocol induced by public randomness R. Let T be the transcript of π and T_R be the transcript of π_R.

For every value r that R may attain, we know that T_r is a prefix-free code. So, since the expected length of a binary prefix-free code is at least its entropy, we know $\mathbf{E}[|T_r|] \geq \mathbf{H}(T_r)$. Therefore,

$$\mathsf{CC}^{\mathsf{avg}}_{\mu}(\pi) = \mathbf{E}[|T|] = \mathbf{E}[\mathbf{E}[|T| \mid R]] \geq \mathbf{E}[\mathbf{H}(T)|R] = \mathbf{H}(T|R) \geq \mathbf{I}(XY;T|R) = \mathsf{IC}^{\mathsf{ext}}_{\mu}(\pi).$$

□

3.3 Amortized Cost is Non Zero

Here we prove Claim 7. Informally speaking, the proof uses the fact that a protocol for f^n needs to encode the results of n independent evaluations of f, and each such result has constant entropy (since f is not constant on $\mathrm{supp}(\mu)$).

Denote by X, Y the (random) pair of inputs given to the players, distributed according to μ. Choose

$$\beta = \tfrac{1}{2}\min\{\mathbf{H}(f(X,Y)), 1\}.$$

Since f is not a constant function on the support of μ, it holds that $\beta > 0$.

Now, let π be a public coin protocol computing f^n with zero error. Denote by X^n, Y^n the n (random) pairs of inputs given to the players that are distributed according to μ^n. Let T denote the transcript of the protocol π. It hold that

$$\begin{aligned}
\mathsf{IC}^{\mathsf{ext}}_{\mu^n}(\pi) &= \mathbf{I}\left(X^nY^n; T|R\right) \\
&\geq \mathbf{I}\left(X^nY^n; f^n(X^n,Y^n)|R\right) && (T \text{ determines } f^n \text{ as } \pi \text{ computes } f^n \text{ with zero error}) \\
&= \mathbf{H}\left(f^n(X^n,Y^n)|R\right) && (f^n \text{ is a deterministic function of } X^nY^n) \\
&= \mathbf{H}\left(f^n(X^n,Y^n)\right) && ((X^n,Y^n) \text{ is independent of } R) \\
&\geq 2\beta n. && (\text{the input pairs are independently selected})
\end{aligned}$$

Claim 1, therefore, implies

$$\mathsf{CC}^{\mathsf{avg}}_{\mu^n}(f^n, 0) \geq \mathsf{IC}^{\mathsf{ext}}_{\mu^n}(f^n, 0) \geq 2\beta n.$$

Clearly,

$$\mathsf{CC}^{\mathsf{avg}}_{\mu^n}(f^n, 0) \geq \mathsf{CC}^{\mathsf{avg}}_{\mu}(f, 0) \text{ and } \mathsf{IC}^{\mathsf{ext}}_{\mu^n}(f^n, 0) \geq \mathsf{IC}^{\mathsf{ext}}_{\mu}(f, 0).$$

Hence,

$$\mathsf{CC}^{\mathsf{avg}}_{\mu^n}(f^n, 0) \geq \frac{1}{2} \cdot 2\beta n + \frac{1}{2}\mathsf{CC}^{\mathsf{avg}}_{\mu}(f, 0).$$

Similarly for $\mathsf{IC}^{\mathsf{ext}}_{\mu^n}(f^n, 0)$. □

4 Discussion of Definitions

We now discuss the definitions of "compute" and "average complexity" with a focus on zero error communication complexity and implications to Conjecture 2. The difference between the possible definitions relate to randomized computation versus distributional computation. We first explain two possible definitions for each notion and then discuss them in more detail.

We start by discussing the notion of "compute." Let f be a function defined over $\mathcal{X} \times \mathcal{Y}$ and let μ be a distribution over $\mathcal{X} \times \mathcal{Y}$. We wish to define when does a protocol π computes f with error at most $\epsilon \geq 0$. Two possible definitions are:

(1) For every $(x, y) \in \mathcal{X} \times \mathcal{Y}$, we have[6] $\Pr_r[\pi(x, y, r) = f(x, y)] \geq 1 - \epsilon$ where r is the randomness of the protocol.

(2) $\Pr_{(x,y)\sim\mu,r}[\pi(x, y, r) = f(x, y)] \geq 1 - \epsilon$.

What does "average complexity" mean? Let π be a protocol with transcript T. We wish to define when does the average complexity of π is at most C. Two possible definitions are:

(a) For every $(x, y) \in \mathcal{X} \times \mathcal{Y}$, we have $\mathbf{E}_r[|T(x, y, r)|] \leq C$.

[6] By $\pi(x, y, r) = f(x, y)$ we mean that knowledge of the transcript and the public randomness yields knowledge of f as in Section 2.1.

(b) $\mathbf{E}_{(x,y)\sim\mu,r}[|T(x,y,r)|] \leq C$.

Definitions (1) and (a) do not depend on μ at all, whereas Definitions (2) and (b) do. Definitions (1) and (a) are standard for randomized protocols when there is no distribution on inputs, see e.g. [KN97]. Definitions (2) and (b) are usually used in the distributional setting, for example, when studying information complexity, see e.g. [BBCR10].

There are therefore four options for defining $\mathsf{CC}^{\mathsf{avg}}_{\mu}(f,\epsilon)$: One that does not depend on μ at all (Definitions (1), (a) are used), and three options that depend on μ in different ways. Observe that the definition of "information complexity" of a protocol does not make sense in the randomized framework, and is defined in the distributional framework.

How many options to interpret Conjecture 2 are there? In any interpretation it seems reasonable to define "compute" in the same fashion for amortized cost and information cost. This leaves us with four interpretations:

The first two interpretations correspond to the definition of "average complexity" as per Definition (a). The two choices left are for the meaning of "compute." We claim that in these two interpretations the conjecture does not hold. The reason is that with these definitions, if μ has full support then $\mathsf{CC}^{\mathsf{avg}}_{\mu}(f,0)$ does not depend on μ at all, whereas $\mathsf{IC}^{\mathsf{ext}}_{\mu}(f,0)$ does. For example, if μ is $1-\delta$ times $\mu_{x=y}$ and δ times the uniform measure on $\mathcal{X}\times\mathcal{Y}$, then $\mathsf{IC}^{\mathsf{ext}}_{\mu}(EQ,0)$ depends on δ. (We use the fact that in the zero error case, Definitions (1) and (2) have the same meaning.)

In the other two interpretations "average complexity" is given by Definition (b). Our example refutes the conjecture if "compute" is as in Definition (2). The case where "compute" is given by Definition (1) remains open. One drawback of this interpretation is that for $\epsilon > 0$ the standard way to define "compute" in the context of information complexity is as in Definition (2). So, this option leads to that $\mathsf{IC}^{\mathsf{ext}}_{\mu}(f,\epsilon)$ is not continuous at $\epsilon = 0$, for example, $\mathsf{IC}^{\mathsf{ext}}_{\mu_{x=y}}(EQ,\epsilon) = 0$ for $\epsilon > 0$ but $\mathsf{IC}^{\mathsf{ext}}_{\mu_{x=y}}(EQ,0) > 0$. We note that for measures μ of full support, external information cost is still continuous (which fits well with the possibility that Conjecture 2 is true for measures of full support).

References

[BBCR10] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. In *STOC*, pages 67–76, 2010. 1, 2, 6, 10, 12

[BGPW13] Mark Braverman, Ankit Garg, Denis Pankratov, and Omri Weinstein. From information to exact communication. In *STOC (to appear)*, 2013. 3

[BR10] Mark Braverman and Anup Rao. Towards coding for maximum errors in interactive communication. *Electronic Colloquium on Computational Complexity (ECCC)*, 17:166, 2010. 3

[Bra12a] Mark Braverman. Coding for interactive computation: Progress and challenges. In *Allerton*, 2012. 2, 3

[Bra12b] Mark Braverman. Interactive information complexity. In *STOC*, pages 505–524, 2012. 5

[CSWY01] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *In Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science*, pages 270–278, 2001. 1

[FKNN95] Tomás Feder, Eyal Kushilevitz, Moni Naor, and Noam Nisan. Amortized communication complexity. *SIAM J. Comput.*, 24(4):736–750, 1995. 1

[HJMR07] Prahladh Harsha, Rahul Jain, David A. McAllester, and Jaikumar Radhakrishnan. The communication complexity of correlation. In *IEEE Conference on Computational Complexity*, pages 10–23, 2007. 1, 3

[Jai11] Rahul Jain. New strong direct product results in communication complexity. *Electronic Colloquium on Computational Complexity (ECCC)*, 18:24, 2011. 1

[JRS03] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A direct sum theorem in communication complexity via message compression. In *ICALP*, pages 300–315, 2003. 1

[Kla10] Hartmut Klauck. A strong direct product theorem for disjointness. In *STOC*, pages 77–86, 2010. 1

[KN97] Eyal Kushilevitz and Noam Nisan. Communication complexity. *Cambridge University Press*, 1997. 12