Amortized Communication Complexity of Distributions

Jérémie Roland and Mario Szegedy

[1] NEC Laboratories America
jroland@nec-labs.com
[2] Rutgers University
szegedy@cs.rutgers.edu

Abstract. Consider the following general communication problem: Alice and Bob have to simulate a probabilistic function $\mathbf{p}$, that with every $(x, y) \in \mathcal{X} \times \mathcal{Y}$ associates a probability distribution on $\mathcal{A} \times \mathcal{B}$. The two parties, upon receiving inputs x and y, need to output $a \in \mathcal{A}$, $b \in \mathcal{B}$ in such a manner that the (a, b) pair is distributed according to $\mathbf{p}(x, y)$. They share randomness, and have access to a channel that allows two-way communication. Our main focus is an instance of the above problem coming from the well known EPR experiment in quantum physics. In this paper, we are concerned with the amount of communication required to simulate the EPR experiment when it is repeated in parallel a large number of times, giving rise to a notion of amortized communication complexity.

In the 3-dimensional case, Toner and Bacon showed that this problem could be solved using on average 0.85 bits of communication per repetition [1]. We show that their approach cannot go below 0.414 bits, and we give a fundamentally different technique, relying on the reverse Shannon theorem, which allows us to reduce the amortized communication to 0.28 bits for dimension 3, and 0.410 bits for arbitrary dimension. We also give a lower bound of 0.13 bits for this problem (valid for one-way protocols), and conjecture that this could be improved to match the upper bounds. In our investigation we find interesting connections to a number of different problems in communication complexity, in particular to [2]. The results contained herein are entirely classical and no knowledge of the quantum phenomenon is assumed.

1 Communication Complexity of Distributions

Communication complexity has been an amazingly potent tool for studying lower bounds for circuits, branching programs, VLSI and streaming data. Lately it is also used to quantify non-local nature of quantum systems.

Recall that in the original version of the model [3] Alice and Bob jointly evaluate a Boolean predicate $f(x, y)$ ($x \in \mathcal{X},\ y \in \mathcal{Y}$) through exchanging messages. Throughout, we will be concerned with the following generalization of the model:

Let $\mathcal{X}$ and $\mathcal{A}$ be the sets of inputs and possible outputs for Alice, and $\mathcal{Y}$ and $\mathcal{B}$ be the sets of inputs and possible outputs for Bob.

Task: A task $\mathbf{p}$ is specified by a function $\mathbf{p} : \mathcal{X} \times \mathcal{Y} \to \mathrm{Distrib}(\mathcal{A} \times \mathcal{B})$, where $\mathrm{Distrib}(\mathcal{A} \times \mathcal{B})$ is the set of all probability distributions on $\mathcal{A} \times \mathcal{B}$.

S. Albers et al. (Eds.): ICALP 2009, Part I, LNCS 5555, pp. 738–749, 2009.
© Springer-Verlag Berlin Heidelberg 2009

Alice and Bob meet the specification $\mathbf{p}$ if upon receiving $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ their output pair (a, b) is distributed according to $\mathbf{p}(x, y)$. Task $\mathbf{p}$ is completely described by the probabilities

$$\mathbf{p}(a, b||x, y) \stackrel{def}{=} \text{the probability of } (a, b) \text{ under distribution } \mathbf{p}(x, y).$$

Alice and Bob share randomness from a common source Λ, i.e. in addition to their input they both receive λ, where $\lambda \in \Lambda$ is picked randomly.

Even with unlimited computational power Alice and Bob usually need to communicate to produce the desired output. The exact rules concerning the communication are critical for our analysis of very low communication problems. Under the wrong definition, Alice may signal to Bob simply by her choice of sending or not sending a bit. To exclude this, we postulate that Alice and Bob are either in send-mode or in receive-mode or in output mode. The communication runs in rounds. After each round the players get into a new mode, which is a function of the player's input, the shared random string λ, and the messages received so far by the player. *A protocol must satisfy that in each round either of two cases happens:* 1. one player is in send-mode and the other is in receive mode; 2. both players are in output mode. No other combination is permitted. Note that if the parties needed to make random choices, we could add them to the shared randomness, Λ. Thus we assume that the protocol is deterministic for any fixed λ.

Protocol: A protocol P for a given simulation task $\mathbf{p}$ is a probability distribution $p(\lambda)$ over deterministic protocols P_λ, each solving a task $\mathbf{p}_\lambda$, such that $\mathbf{p} = \sum_\lambda p(\lambda)\mathbf{p}_\lambda$.

Note that since any $\lambda \in \Lambda$ corresponds to a deterministic protocol, we may extend Λ to the set of all possible deterministic communication protocols with inputs in $\mathcal{X} \times \mathcal{Y}$ and outputs in $\mathcal{A} \times \mathcal{B}$ (we would just set $p(\lambda) = 0$ for all deterministic protocols that never occur when executing the shared randomness protocol P).

LHV: Let Λ_0 be the set of all deterministic protocols that do not use any communication. A task $\mathbf{p}$ is in LHV if it may be simulated using a distribution over protocols in Λ_0 only, that is, if there is a zero (classical) communication protocol for it.

LHV stands for Local Hidden Variable referring to $\lambda \in \Lambda$, which is the only source of correlation between Alice and Bob. Note that these correlations do not violate locality because we assume that the parties receive the "hidden" λ when they are not yet spatially separated.

Fix the input and output sets $\mathcal{X}, \mathcal{Y}, \mathcal{A}, \mathcal{B}$ for the rest of this paragraph.

Bell inequality: A Bell inequality is an inequality of the form

$$\sum_{x,y,a,b} B_{xyab}\, \mathbf{p}(a, b||x, y) \leq B_0, \tag{1}$$

which holds for all $\mathbf{p} \in$ LHV.

Notice that the left hand side of Eq. (1) is a linear functional, which we will shortly denote as $B(\mathbf{p})$. A task $\mathbf{p}$ is in LHV if and only if it satisfies all Bell inequalities.

In other words, LHV is a convex set. We are now interested in tasks outside LHV, which may be identified by the fact that they violate some Bell inequality. This means that such a task $\mathbf{p}$ may not be simulated using shared randomness only, and that some additional communication is required. Let P be a communication protocol simulating $\mathbf{p}(a,b||x,y)$ using shared randomness Λ, $M(x,y,\lambda)$ be the transcript of the messages on input x and y when the shared randomness is fixed to λ, and $|M|$ be the length in bits of this transcript. We define the worst-case cost $C_w(P)$ as the maximal number of bits communicated between Alice and Bob in any particular execution of the protocol, that is, $C_w(P) = \max_{x,y,\lambda} |M(x,y,\lambda)|$, where the maximum is over inputs $(x,y) \in \mathcal{X} \times \mathcal{Y}$ and shared randomness $\lambda \in \Lambda$ such that $p(\lambda) \neq 0$. We then define the worst-case communication complexity as $C_w(\mathbf{p}) = \min_P C_w(P)$. In this paper, we are more interested in the average cost:

Average cost: Given a distribution D on $\mathcal{X} \times \mathcal{Y}$, the average cost $C^D(P)$ is the expected number of bits communicated between Alice and Bob, where the expectation is taken over the shared randomness $\lambda \in \Lambda$ and the inputs $(x,y) \in D$,

$$C^D(P) = \sum_{\lambda \in \Lambda} p(\lambda) \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} D(x,y)\, |M(x,y,\lambda)|. \tag{2}$$

Average communication complexity: $C(\mathbf{p}) = \max_D C^D(\mathbf{p})$, where $C^D(\mathbf{p}) = \min_P C^D(P)$ is the *distributional* average communication complexity for fixed input distribution D, the minimum being taken over all protocols P implementing $\mathbf{p}$, and the maximum over all distributions D on $\mathcal{X} \times \mathcal{Y}$.

We emphasize that even when we are concerned with the average case complexity, P needs to meet the specification for every input pair $(x,y) \in \mathcal{X} \times \mathcal{Y}$.

Example. The CHSH correlations ($\mathbf{p}_\mu$): Let us define the task $\mathbf{p}_\mu$ for $0 \leq \mu \leq 1$ as follows: $\mathcal{X} = \mathcal{Y} = \{0,1\}$, $\mathcal{A} = \mathcal{B} = \{1,-1\}$ and

$$\mathbf{p}_\mu(a,b||x,y) = \frac{1 + \mu\, ab\, (-1)^{x \cdot y}}{4}.$$

The task is defined in such a way that for all $(x,y) \in \mathcal{X} \times \mathcal{Y}$, the relation $ab = (-1)^{x \cdot y}$ between the inputs and the outputs has to be satisfied with probability $\frac{1+\mu}{2}$. It is not hard to show that $\mathbf{p}_\mu$ can be implemented classically with zero communication only for $0 \leq \mu \leq 1/2$. In particular, for $\mu > 1/2$, $\mathbf{p}_\mu$ violates the so-called CHSH Bell inequality [4]:

$$\sum_{x,y,a,b} ab\, (-1)^{x \cdot y} \mathbf{p}(a,b||x,y) \leq 2,$$

so that in a classical world, this task requires communication to be implemented. However, if Alice and Bob are separated in space, but they share a pair of entangled qubits, in the quantum world they can solve $\mathbf{p}_{1/\sqrt{2}}$ *with no communication* whatsoever. This is

because quantum correlations may violate Bell inequalities, and therefore have a non-local character, as was first shown by Bell [5].

The EPR-Bohm experiment ($\mathbf{p}_{\dim=d}$): The inputs to Alice and Bob are unit vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ from the d-dimensional sphere $\mathbb{S}_{d-1}$. The output is again an element of $\{1,-1\}$, a for Alice, and b for Bob, with the specification

$$\mathbf{p}_{\dim=d}(a,b||\boldsymbol{x},\boldsymbol{y}) = \frac{1-ab\,\boldsymbol{x}\cdot\boldsymbol{y}}{4}. \tag{3}$$

$\mathbf{p}_{\dim=d}$ arises from the EPR-Bohm experiment [6,7], and can be solved in the quantum world with zero communication.

The communication complexity of $\mathbf{p}_{\dim=d}$, and quantum distributions in general, has been studied in a series of paper [8,9,10,11]. $\mathbf{p}_{\dim=3}$ is particularly interesting because it corresponds to Bohm's original version of the experiment, involving a maximally entangled qubit pair, the most simple quantum system that captures the essential properties of entanglement. The best known protocol for $\mathbf{p}_{\dim=3}$ was presented by Toner and Bacon, and uses one bit of communication [1]. This was shown to be optimal, even for the average complexity, by Barrett, Kent and Pironio [12], hence, $C_w(\mathbf{p}_{\dim=3}) = C(\mathbf{p}_{\dim=3}) = 1$ bit.

The higher dimensional case $\mathbf{p} = \mathbf{p}_{\dim=d}$ has been studied by Degorre, Laplante and Roland [13], who proved that the average communication complexity scaled at most as $C(\mathbf{p}_{\dim=d}) = O(\log d)$. This was significantly improved by Regev and Toner [14], who showed that bounded worst-case communication was sufficient, by providing an explicit 2-bit protocol, so that $C_w(\mathbf{p}_{\dim=d}) \leq 2$. When considering average communication, they could improve their protocol to 1.82 bits.

The amortized communication cost of simulating $\mathbf{p}_{\dim=d}$ (and its powers) will be the focus of this paper, and will be described in more details in the next section.

Finiteness: In this example $\mathcal{X}$, $\mathcal{Y}$, and probability space Λ are infinite, equipped with some measure. The communication still needs to be bounded. Note that as in the finite case, each bit communicated in the protocol by a given party can be described by a measurable function that goes from this party's input, the shared randomness and the bits communicated so far into $\{0,1\}$. Even though the domain of these functions is infinite, the parties can compute them for free because they are computationally unbounded. We also need to modify formula (2) by replacing the sum with an integral.

2 Amortization

We now consider the task $\mathbf{p}^{\otimes n}$, given by the n-fold parallelization of $\mathbf{p}$,

$$\mathbf{p}^{\otimes n}(\boldsymbol{a},\boldsymbol{b}||\boldsymbol{x},\boldsymbol{y}) = \prod_{i=1}^{n} \mathbf{p}(a_i,b_i||x_i,y_i).$$

We then define the following communication complexity.

Amortized communication complexity: $C_\infty(\mathbf{p}) = \max_D C^D_\infty(\mathbf{p})$, where $C^D_\infty(\mathbf{p}) = \lim_{n\to\infty} C^{D^{\otimes n}}(\mathbf{p}^{\otimes n})/n$ is the *distributional* amortized communication complexity, and $D^{\otimes n}(\boldsymbol{x},\boldsymbol{y}) = \prod_{i=1}^{n} D(x_i,y_i)$.

2.1 Entropic Complexity

Let P be a communication protocol simulating $\mathbf{p}(a,b||x,y)$ using shared randomness Λ, and let D be the input distribution.

Entropic cost $C_H^D(P)$: Conditional entropy $H(M|\Lambda)$ of the transcript M of the messages communicated between Alice and Bob, given the shared randomness $\lambda \in \Lambda$.

We also define the corresponding (distributional and non-distributional) entropic communication complexities for a task $\mathbf{p}$ as $C_H^D(\mathbf{p}) = \min_P C_H^D(P)$ and $C_H(\mathbf{p}) = \max_D C_H^D(\mathbf{p})$, where the minimum is taken over all protocols P implementing $\mathbf{p}$, and the maximum is taken over all distributions D on $\mathcal{X} \times \mathcal{Y}$.

2.2 The Input Distribution

As first observed by Yao [15], von Neumann's minmax principle [16] implies the following statement.

Theorem 1. *Let C_* be any of C, C_∞, C_H. We have $C_*(\mathbf{p}) = \min_P \max_D C_*^D(P)$.*

Note that for a fixed protocol P, the maximum over distributions D is achieved for a given input couple $(x,y) \in \mathcal{X} \times \mathcal{Y}$.

For specific tasks, symmetries allow to make assumptions on the hardest distribution, which attains $C_*(\mathbf{p}) = \max_D \min_P C_*^D(P)$. In particular, for the CHSH problem $\mathbf{p}_\mu$, we can show that the uniform distribution is the hardest distribution.

Claim 1. *Let C_* be any of C, C_∞, C_H. Then, $C_*(\mathbf{p}_\mu) = C_*^U(\mathbf{p}_\mu)$, where U is the uniform distribution on $\{0,1\}^2$.*

Similarly, for the EPR-Bohm problem $\mathbf{p}_{\dim=d}$, we may assume that the hardest distribution has uniform marginals (this observation is due to Toner and Bacon [1]). For an input distribution D, we will denote D_A and D_B the marginal distributions of x and y, respectively.

Claim 2. *Let C_* be any of C, C_∞, C_H. Then, there exists a distribution U on $\mathbb{S}_{d-1} \times \mathbb{S}_{d-1}$ with uniform marginals U_A and U_B such that $C_*^U(\mathbf{p}_{\dim=d}) = \max_D C_*^D(\mathbf{p}_{\dim=d})$.*

Note that for $\mathbf{p}_{\dim=d}$, we can only show that the marginals of the hardest distribution are uniform, not that the hardest distribution itself is uniform. However, let us note that when restricting to one-way communication protocols, the communication complexity only depends on the marginal distribution for the player sending the messages. For any notion of communication complexity, we add a superscript $\rightarrow$ when we only consider protocols restricted to one-way communication.

Claim 3. *Let C_* be any of C, C_∞, C_H, C_I, and let D, D' be two distributions on $\mathcal{X} \times \mathcal{Y}$ having the same marginal distributions for x, that is, $D_A = D'_A$. Then, $C_*^{\rightarrow,D}(\mathbf{p}) = C_*^{\rightarrow,D'}(\mathbf{p})$.*

2.3 Relation between the Communication Complexities

We will use as an intermediate step the following cost for communication protocols using *private* randomness only, first introduced by Chakrabarti *et al.* [17]:

Information cost $C_I^D(P)$: Mutual information $I(XY : M)$ between the inputs X, Y and the transcript M of the messages communicated between Alice and Bob.

As for the other complexities, we also define the information complexities $C_I^D(\mathbf{p}) = \min_P C_I^D(P)$ and $C_I(\mathbf{p}) = \min_P \max_D C_I^D(P)$, where the minimum is taken over all *private* randomness protocols P implementing $\mathbf{p}$ (note that in the presence of shared randomness, there always exists a protocol P such that $C_I^D(P) = 0$, so this quantity would not be relevant). Chakrabarti *et al.* have shown that this complexity satisfies the following direct sum property.

Theorem 2 ([17,2]). *If $D = D_A \otimes D_B$ is a product distribution, then $C_I^{D^{\otimes n}}(\mathbf{p}^{\otimes n}) = n\, C_I^D(\mathbf{p})$.*

In the case of one-way communication, the complexity only depends on the marginal distribution, so we have the following corollary.

Corollary 3. *For one-way communication, $C_I^{\rightarrow, D^{\otimes n}}(\mathbf{p}^{\otimes n}) = n\, C_I^{\rightarrow, D}(\mathbf{p})$.*

Finally, we will also use the reverse Shannon theorem [18], which in our notations may be stated as follows:

Theorem 4 ([18]). *Let $\mathbf{p} : \mathcal{X} \rightarrow \mathrm{Distrib}(\mathcal{B})$ be a simulation task with no output on Alice's side, and no input on Bob's side. Then, $C_\infty^{\rightarrow, D}(\mathbf{p}) \leq I(X : B)$.*

Our statement is slightly different from [18] but may be proved using the same construction. Informally, it says that a communication channel $X \rightarrow B$ may be simulated, in the limit of a large number of repetitions, using (free) shared randomness and one-way communication at most $I(X : B)$ per repetition.

Proposition 5. *The communication complexities satisfy the following relations: $C_\infty^D(\mathbf{p}) \leq C_I^D(\mathbf{p}) \leq C_H^D(\mathbf{p}) \leq C^D(\mathbf{p}) \leq C_w(\mathbf{p})$. For a product input distribution $D = D_A \otimes D_B$, we also have $C_\infty^{D_A \otimes D_B}(\mathbf{p}) = C_I^{D_A \otimes D_B}(\mathbf{p})$. Similarly, for any input distribution D but restricting to one-way communication protocols, $C_\infty^{\rightarrow, D}(\mathbf{p}) = C_I^{\rightarrow, D}(\mathbf{p})$.*

Proof (Sketch). These relations are based on fundamental propositions of information theory, such as Shannon's source coding theorem. Let us focus on the less obvious relations, involving $C_I(\mathbf{p})$.

[$C_I^D(\mathbf{p}) \leq C_H^D(\mathbf{p})$]. Let $C_H^D(\mathbf{p}) = H(M|\Lambda)$ be achieved by a protocol P with shared randomness Λ, where M is the transcript of the messages communicated during the protocol. Let us build a protocol P' using private randomness only. In this protocol, only Alice knows the random string Λ, and her first action is to send Λ to Bob. From there, the players proceed as in protocol P. Since the transcript of P' is the concatenation of Λ and M, we have $C_I^D(P') = I(XY : M\Lambda)$. From the facts that $I(XY : \Lambda) = 0$ (since the shared randomness is independent from the inputs), and $H(M|XY\Lambda) = 0$ (since the messages depend deterministically on the inputs and the randomness), it is straightforward to check that $I(XY : M\Lambda) = H(M|\Lambda)$.

[$C^D_\infty(\mathbf{p}) \leq C^D_I(\mathbf{p})$]. Let $C^D_I(\mathbf{p}) = I(XY : M)$ be achieved by a protocol P without shared randomness, where M is the transcript of the messages communicated during the protocol P. These messages are alternatingly sent by Alice to Bob and vice-versa. Let us denote by M_k the k^{th} message and $M_{[k]}$ the restriction of the transcript to the first k messages. We may express the information complexity of $\mathbf{p}$ as:

$$C^D_I(\mathbf{p}) = \sum_{k=1}^{t} I(XY : M_k | M_{[k-1]}),$$

where t is the maximal number of rounds of the protocol (possibly infinite). Let us focus on the k^{th} message M_k, and suppose it is sent by Alice to Bob. In a particular execution of the protocol, the partial transcript $M_{[k-1]}$ will be fixed to some string m, which is at this point known to both Alice and Bob, so that

$$I(XY : M_k | M_{[k-1]}) = \sum_m p(m)\, I(X : M_k | M_{[k-1]} = m),$$

where $p(m) = \Pr[M_{[k-1]} = m]$, and we have used the fact that M_k only depends on X and $M_{[k-1]}$, and not on Y. Alice now needs to send M_k to Bob, which only depends on X when we condition on $M_{[k-1]}$, so she actually needs to simulate a communication channel $X \longrightarrow M_k$. Since the partial transcript $M_{[k-1]} = m$ happens with probability $p(m)$, this particular channel will have to be simulated on average $n \cdot p(m)$ times when repeating the protocol n times, and the reverse Shannon theorem (Theorem 4) ensures that as n goes to infinity, this simulation may be achieved using shared randomness and communication $I(X : M_k | M_{[k-1]} = m)$ per repetition. By compressing similarly each successive message, and averaging over all possible transcripts, we get that $C^D_\infty(\mathbf{p}) \leq I(XY : M) = C^D_I(\mathbf{p})$.

[$C^{D_A \otimes D_B}_I(\mathbf{p}) \leq C^{D_A \otimes D_B}_\infty(\mathbf{p})$]. For a product input distribution $D = D_A \otimes D_B$, Theorem 2 implies that $C^D_I(\mathbf{p}) = C^{D^{\otimes n}}_I(\mathbf{p}^{\otimes n})/n$. Moreover, since $C^{D^{\otimes n}}_I(\mathbf{p}^{\otimes n}) \leq C^{D^{\otimes n}}(\mathbf{p}^{\otimes n})$, we obtain $C^D_I(\mathbf{p}) \leq C^{D^{\otimes n}}(\mathbf{p}^{\otimes n})/n$ and, in the limit $n \to \infty$, $C^D_I(\mathbf{p}) \leq C^D_\infty(\mathbf{p})$. Similarly, Corollary 3 implies that for any input distribution but one-way communication, $C^{\rightarrow,D}_I(\mathbf{p}) \leq C^{\rightarrow,D}_\infty(\mathbf{p})$.

3 Lower Bound on the Entropic Complexity

The previous best upper bound on the amortized communication complexity of $\mathbf{p}_{\dim=3}$ is due to Toner and Bacon, who proved that $C_\infty(\mathbf{p}_{\dim=3}) \leq \text{Si}(\pi)/(\pi \ln 2) \approx 0.85$ bits [1], where $\text{Si}(x)$ is the sine integral function. Indeed, they showed that in their one-bit protocol, the conditional entropy of the messages given the shared randomness (what we defined as the entropic cost) is only 0.85 bits, so that one can use Shannon's source coding theorem to compress the communication from 1 bit to 0.85 bits. In this section we prove that the entropic complexity of $\mathbf{p}_{\dim=d}$ is at least 0.414 bits for any $d \geq 2$, which shows that to reduce the communication further, a new technique is required. Such a technique will be presented in the next section.

We prove the lower bound on the entropic complexity by adapting a method proposed by Pironio for lower bounds on the average communication complexity [19]. The idea behind the following theorem is that a task $\mathbf{p}$ outside LHV may violate a Bell inequality, so that it will require to use deterministic protocols P_λ for $\lambda \notin \Lambda_0$, simulating tasks $\mathbf{p}_\lambda$ outside LHV, with some probability. More precisely, we consider for each deterministic protocol a "violation per entropy" ratio. To achieve the same violation as the task $\mathbf{p}$ using as little communication as possible (where the communication is counted as the entropy of the messages), one should use a distribution over deterministic protocols that have a large violation per entropy ratio. In particular, the deterministic protocol having the largest ratio gives a lower bound on the entropic communication complexity of $\mathbf{p}$.

Theorem 6. *Let B be a linear functional over the set of tasks, which defines a Bell inequality $B(\mathbf{p}_\lambda) \leq B_0$ satisfied for all $\lambda \in \Lambda_0$, but violated by a simulation task $\mathbf{p}$, that is, $B(\mathbf{p}) > B_0$. Then, the entropic communication complexity of $\mathbf{p}$ is lower bounded as follows:*

$$C_H^D(\mathbf{p}) \geq \frac{B(\mathbf{p}) - B_0}{B(\mathbf{p}_{\lambda^*}) - B_0} C_H^D(P_{\lambda^*}),$$

where P_{λ_*} is a deterministic protocol for a task $\mathbf{p}_{\lambda^*}$ such that

$$\frac{B(\mathbf{p}_{\lambda^*}) - B_0}{C_H^D(P_{\lambda^*})} = \max_{\lambda \notin \Lambda_0} \frac{B(\mathbf{p}_\lambda) - B_0}{C_H^D(P_\lambda)}.$$

This may be proved along the lines of the proof of Proposition 1 in [19], which gives a similar statement for the average communication complexity. We may now completely determine the entropic communication complexity of $\mathbf{p}_\mu$:

Theorem 7. *For any $1/2 \leq \mu \leq 1$ we have, $C_H(\mathbf{p}_\mu) = 2\mu - 1$.*

Note that for $0 \leq \mu \leq 1/2$, we trivially have $C_H(\mathbf{p}_\mu) = 0$.

Proof. The lower bound comes from the previous theorem. This is then shown to be tight by giving an explicit protocol. Since we have shown in Claim 1 that $C_H(\mathbf{p}_\mu) = C_H^U(\mathbf{p}_\mu)$, it suffices to consider the uniform input distribution.

[$C_H^U(\mathbf{p}_\mu) \geq 2\mu - 1$]. We use the CHSH inequality [4], which is defined by a linear functional B acting on a task $\mathbf{p}$ as:

$$B(\mathbf{p}) = \sum_{x,y,a,b} ab\ (-1)^{x \cdot y} \mathbf{p}(a, b||x, y).$$

It is straightforward to check that $B(\mathbf{p}) \leq 2$ for all $\mathbf{p}$ in LHV, so we set $B_0 = 2$. For the simulation task $\mathbf{p}_\mu$, we have $B(\mathbf{p}_\mu) = 4\mu$, so that the inequality is violated as soon as $\mu > 1/2$. Moreover, $\max_{P_\lambda}(B(\mathbf{p}_\lambda) - B_0)/C_H^U(P_\lambda)$ is attained by a protocol P_{λ^*} where one player sends his input to the other, such that $C_H^U(P_{\lambda^*}) = 1$ and $B(\mathbf{p}_{\lambda^*}) = 4$. We then obtain

$$C_H^U(\mathbf{p}_\mu) \geq \frac{4\mu - 2}{4 - 2} 1 = 2\mu - 1.$$

$[C_H^U(\mathbf{p}_\mu) \leq 2\mu - 1]$. Let us consider the extreme cases $\mu = 1/2$ and $\mu = 1$. For $\mu = 1/2$, there exists a shared randomness protocol $P_{1/2}$ without any communication ($\mathbf{p}_{1/2}$ is in LHV), therefore satisfying $C_H^U(P_{1/2}) = 0$. On the other hand, for $\mu = 1$, there exists a protocol P_1 with one bit of communication (one of the player sends his input to the other), that is, $C_H^U(P_1) = 1$. It is also straightforward to show that $\mathbf{p}_\mu = (2 - 2\mu)\mathbf{p}_{1/2} + (2\mu - 1)\mathbf{p}_1$, so that for implementing $\mathbf{p}_\mu$, it suffices to use the protocol $P_{1/2}$ with probability $(2 - 2\mu)$ and the protocol P_1 with probability $(2\mu - 1)$. By linearity, the obtained protocol has entropic cost $2\mu - 1$.

Using a reduction from $\mathbf{p}_{1/\sqrt{2}}$ to $\mathbf{p}_{\dim=d}$, Theorem 7 implies as a corollary a lower bound on the entropic complexity of $\mathbf{p}_{\dim=d}$.

Claim 4. *Let C_* be any of C, C_∞, C_H. Then, $C_*(\mathbf{p}_{\dim=d}) \geq C_*(\mathbf{p}_{1/\sqrt{2}})$ for any $d \geq 2$.*

Proof. The key observation is that the task $\mathbf{p}_{1/\sqrt{2}}$ for uniformly distributed inputs is equivalent to the task $\mathbf{p}_{\dim=d}$ for a special distribution $\tilde{D}$, where the inputs are uniform over two vectors $\{\boldsymbol{x}_0, \boldsymbol{x}_1\}$ for Alice and two vectors $\{\boldsymbol{y}_0, \boldsymbol{y}_1\}$ for Bob, laid out such that $\boldsymbol{x}_i \cdot \boldsymbol{y}_j = (-1)^{i \cdot j}/\sqrt{2}$. We then have $C_*(\mathbf{p}_{\dim=d}) \geq C_*^{\tilde{D}}(\mathbf{p}_{\dim=d}) = C_*^U(\mathbf{p}_{1/\sqrt{2}})$, which concludes the proof since we have shown that $C_*^U(\mathbf{p}_\mu) = C_*(\mathbf{p}_\mu)$.

Corollary 8. $C_H(\mathbf{p}_{\dim=d}) \geq \sqrt{2} - 1 \approx 0.414$ *bits.*

This lower bound means that for parallel repetitions of the problem, if we simply compress the messages using Shannon's source coding theorem, we may not reduce the communication further than 0.414 bits. We show in the next section that we can beat this lower bound by using another technique, based on the reverse Shannon theorem.

4 A New Protocol

In this section, we show how to reduce the communication for parallel repetitions of the problem of simulating $\mathbf{p}_{\dim=d}$, beating the lower bound on the entropic communication complexity derived in the previous section. We use a result due to Degorre *et al.*, which shows that the problem reduces to a distributed sampling task:

Theorem 9 ([20,13]). *Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be Alice's and Bob's inputs. If Alice and Bob share a random variable $\boldsymbol{v} \in \mathbb{S}_{d-1}$ distributed according to a probability measure*

$$\rho(\boldsymbol{v}||\boldsymbol{x}) = \frac{|\boldsymbol{x} \cdot \boldsymbol{v}|}{\mathcal{R}_d},$$

where $\mathcal{R}_d = \int_{\mathbb{S}_{d-1}} |\boldsymbol{x} \cdot \boldsymbol{v}| \, d\boldsymbol{v}$, then they are able to simulate $\mathbf{p}_{\dim=d}$ without any further resource.

This observation leads to an apparently very bad communication protocol for $\mathbf{p}_{\dim=d}$ with private randomness only: using her input and private randomness, Alice locally samples $\boldsymbol{v}$ according to the distribution $\rho(\boldsymbol{v}||\boldsymbol{x})$, and then communicates $\boldsymbol{v}$ to Bob. This

would require infinite communication, but the point is that the information cost of this protocol would actually be not only finite, but also rather low, so that for parallel repetitions of the problem, and with the help of shared randomness, we may significantly reduce the communication using the reverse Shannon theorem. In particular, we prove the following upper bound:

Theorem 10. *The amortized communication complexity of* $\mathbf{p}_{\dim=d}$ *satisfies*

$$C_\infty(\mathbf{p}_{\dim=d}) \leq \begin{cases} \frac{1}{\ln 2}\left[\ln \frac{(d-1)A_d}{A_{d-1}} - \sum_{k=0}^{\frac{d}{2}-1} \frac{1}{2k+1}\right] & \text{for } d \text{ even}, \\ \frac{1}{\ln 2}\left[\ln \frac{(d-1)A_d}{2A_{d-1}} - \sum_{k=1}^{\frac{d-1}{2}} \frac{1}{2k}\right] & \text{for } d \text{ odd}, \end{cases} \tag{4}$$

where $A_d = \int_{\mathbb{S}_{d-1}} d\boldsymbol{v}$ *is the surface area of the* d*-dimensional sphere.*

In particular, we have

1. $C_\infty(\mathbf{p}_{\dim=2}) \leq \frac{1}{\ln 2}(\ln \pi - 1) \approx 0.21$ bits,
2. $C_\infty(\mathbf{p}_{\dim=3}) \leq 1 - \frac{1}{2\ln 2} \approx 0.28$ bits,
3. $C_\infty(\mathbf{p}_{\dim=d}) \leq \frac{1}{2\ln 2}(\ln \pi - \gamma) \approx 0.410$ bits for arbitrary d, where γ is the Euler-Mascheroni constant.

Proof. Let P be the following private randomness protocol for $\mathbf{p}_{\dim=d}$: using her input together with private randomness, Alice samples a random variable V according to the distribution $\rho(\boldsymbol{v}||\boldsymbol{x})$ defined above and communicates the obtained sample $\boldsymbol{v}$ to Bob. By Theorem 9, they are then able to solve task $\mathbf{p}_{\dim=d}$. More precisely, it suffices for the players to set their respective outputs as $a = \mathrm{sgn}(\boldsymbol{x} \cdot \boldsymbol{v})$ and $b = \mathrm{sgn}(\boldsymbol{y} \cdot \boldsymbol{v})$, where $\mathrm{sgn}(x) = 1$ if $x \geq 0$, and -1 otherwise [20,13].

We have shown in the previous section that the reverse Shannon theorem implies that $C_\infty^D(\mathbf{p}_{\dim=d}) \leq C_I^D(\mathbf{p}_{\dim=d})$ (Proposition 5) and also that the hardest distribution for $\mathbf{p}_{\dim=d}$ has uniform marginals (Claim 2), so it suffices to compute the information cost $C_I^D(P) = I(X:V)$ for a distribution D with uniformly distributed $\boldsymbol{x}$. The computation of $I(X:V)$ will be given in the full version of the paper, and yields Eq. (4).

For completeness, let us note that using the same technique, we can prove the following upper bound on the amortized communication complexity of $\mathbf{p}_\mu$:

Theorem 11. *For any* $1/2 \leq \mu \leq 1$*, we have* $C_\infty(\mathbf{p}_\mu) \leq 1 - H[\mu]$*, where* $H[\mu] = \mu \log \frac{1}{\mu} + (1-\mu)\log \frac{1}{1-\mu}$.

Proof. Let v be a random bit correlated with x, such that $p(x = v) = \mu$. The channel defined by the Markov process $X \to V$ is then a binary symmetric channel, with channel capacity $1 - H[\mu]$. It is straightforward to show that if Alice may use such a channel to communicate information about her input x to Bob, it is sufficient to simulate $\mathbf{p}_\mu$. Indeed, it suffices for Alice and Bob to output

$$a = (-1)^{(x \oplus v \oplus 1)\cdot \lambda_0}\ (-1)^{(x \oplus v)\cdot \lambda_1},$$
$$b = (-1)^{\lambda_0}\ (-1)^{y \cdot v},$$

where λ_0, λ_1 are shared unbiased random bits. The reverse Shannon theorem then ensures that asymptotically, the channel $X \to V$ may be simulated using on average $1 - H[\mu]$ bits per repetition.

In the next section, we will show that this protocol is optimal, at least when the players are restricted to one-way communication.

5 The Difference Method

Amortized lower bounds are notoriously hard to prove. Examples include the Shannon capacity of graphs [21] and the parallel repetition theorem of Raz [22]. In some lucky cases the situation is better. Quantum values of XOR games [23] and the communication complexity of correlation [2] are examples, where mathematics seems to be in our favor. Incidentally, both topics have relevance to lower bounding amortized communication complexity. We develop a new method we call the *difference method*, which so far we could apply only in the one-way communication context. Note, however, that all efficient protocols we know for this problem are one-way.

Theorem 12. *For any* $1/2 \leq \mu \leq 1$*, we have* $C^{\rightarrow}_{\infty}(\mathbf{p}_\mu) \geq 1 - H[\mu]$.

This matches the upper bound of Theorem 11, showing that the above protocol is optimal, at least for one-way communication.

Proof. Since the hardest distribution is the uniform distribution U (Claim 1), we have $C^{\rightarrow}_{\infty}(\mathbf{p}_\mu) = C^{\rightarrow,U}_{\infty}(\mathbf{p}_\mu)$. Moreover, Proposition 5 implies that $C^{\rightarrow,U}_{\infty}(\mathbf{p}_\mu) = C^{\rightarrow,U}_{I}(\mathbf{p}_\mu)$, so it suffices to show that for any protocol for $\mathbf{p}_\mu$, $I(X : M) \geq 1 - H[\mu]$, where x is an unbiased random bit. The idea of the proof is to reduce the problem to the communication complexity problem of a correlation *à la* Harsha *et al.* [2] and then, following their approach, use the mutual information between Alice's input and Bob's output to bound the communication. To reduce to [2], 1. We have to get rid of Bob's input; 2. We have to get rid of Alice's output. If we fix Bob's input and omit Alice's output, we get nothing. Nevertheless, since the communication is one-way, when Bob receives Alice's message, he can just compute the output on any input he wants to. We show that if we run the protocol with a random input x on Alice's side, take both $y = 0$ and $y' = 1$ as inputs on Bob's side, and receive outputs b and b', respectively from Bob, then the product $b \cdot b'$ will contain a lot of information about Alice's input, x.

Observe that $(a \cdot b) \cdot (a \cdot b') = b \cdot b'$. The specification of $\mathbf{p}_\mu$ tells us that $a \cdot b$ should take 1 with probability $(1+\mu)/2$ and -1 with probability $(1-\mu)/2$. Also, $a \cdot b'$ should take $(-1)^x$ with probability $(1+\mu)/2$ and $(-1)^{x+1}$ with probability $(1-\mu)/2$. The union bound gives that the probability that $b \cdot b' = (-1)^x$ is at least μ. This shows that the mutual information $I(X : E)$ between x and $e = b \cdot b'$ is at least $1 - H[\mu]$, where we have used the fact that $H(X) = 1$ (since X is an unbiased random bit). The data processing inequality on the Markov chain $X \rightarrow M \rightarrow E$ then implies that $I(X : M) \geq I(X : E)$.

For $\mu = 1/\sqrt{2}$, we have $C^{\rightarrow}_{\infty}(\mathbf{p}_{1/\sqrt{2}}) \geq 1 - H[1/\sqrt{2}]$, and in turn, using the reduction from Claim 4, we obtain the following lower bound on $C^{\rightarrow}_{\infty}(\mathbf{p}_{\dim=d})$ as a corollary:

Corollary 13. $C^{\rightarrow}_{\infty}(\mathbf{p}_{\dim=d}) \geq 1 - H[1/\sqrt{2}] \approx 0.13$ *bits for any* $d \geq 2$.

Acknowledgements

J. Roland would like to thank Julien Degorre and Sophie Laplante for many interesting discussions. Part of this work was done while J. Roland was affiliated with U.C. Berkeley and FNRS Belgium.

References

1. Toner, B.F., Bacon, D.: Communication Cost of Simulating Bell Correlations. Phys. Rev. Lett. 91, 187904 (2003)
2. Harsha, P., Jain, R., McAllester, D., Radhakrishnan, J.: The communication complexity of correlation. In: Proc. 22nd CCC, pp. 10–23 (2007)
3. Yao, A.C.C.: Some complexity questions related to distributive computing. In: Proc. 11th STOC, pp. 209–213 (1979)
4. Clauser, J.F., Horne, M.A., Shimony, A., Holt, R.A.: Proposed Experiment to Test Local Hidden-Variable Theories. Phys. Rev. Lett. 23, 880–884 (1969)
5. Bell, J.S.: On the Einstein Podolsky Rosen paradox. Physics 1, 195 (1964)
6. Einstein, A., Podolsky, B., Rosen, N.: Can quantum-mechanical description of physical reality be considered complete? Phys. Rev. 47, 777–780 (1935)
7. Bohm, D., Aharonov, Y.: Discussion of Experimental Proof for the Paradox of Einstein, Rosen, and Podolsky. Phys. Rev. 108, 1070–1076 (1957)
8. Maudlin, T.: Bell's inequality, information transmission, and prism models. In: Biennal Meeting of the Philosophy of Science Association, pp. 404–417 (1992)
9. Brassard, G., Cleve, R., Tapp, A.: Cost of Exactly Simulating Quantum Entanglement with Classical Communication. Phys. Rev. Lett. 83, 1874–1877 (1999)
10. Steiner, M.: Towards quantifying non-local information transfer: finite-bit non-locality. Phys. Lett. A 270, 239–244 (2000)
11. Cerf, N.J., Gisin, N., Massar, S.: Classical Teleportation of a Quantum Bit. Phys. Rev. Lett. 84, 2521–2524 (2000)
12. Barrett, J., Kent, A., Pironio, S.: Maximally nonlocal and monogamous quantum correlations. Phys. Rev. Lett. 97(17), 170409 (2006)
13. Degorre, J., Laplante, S., Roland, J.: Classical simulation of traceless binary observables on any bipartite quantum state. Phys. Rev. A 75, 012309 (2006)
14. Regev, O., Toner, B.: Simulating quantum correlations with finite communication. In: Proc. 48th FOCS, pp. 384–394 (2007)
15. Yao, A.C.C.: Probabilistic computations: Toward a unified measure of complexity. In: Proc. 18th FOCS, pp. 222–227 (1977)
16. von Neumann, J.: Zur Theorie der Gesellschaftsspiele. Math. Ann. 100(1), 295–320 (1928)
17. Chakrabarti, A., Shi, Y., Wirth, A., Yao, A.: Informational complexity and the direct sum problem for simultaneous message complexity. In: Proc. 42nd FOCS, pp. 270–278 (2001)
18. Bennett, C., Shor, P., Smolin, J., Thapliyal, A.: Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem. IEEE Trans. Inf. Theor. 48(10), 26–37 (2002)
19. Pironio, S.: Violations of Bell inequalities as lower bounds on the communication cost of nonlocal correlations. Phys. Rev. A 68(6), 062102 (2003)
20. Degorre, J., Laplante, S., Roland, J.: Simulating quantum correlations as a distributed sampling problem. Phys. Rev. A 72, 062314 (2005)
21. Lovász, L.: On the Shannon capacity of a graph. IEEE Trans. Inf. Theor. 25(1), 1–7 (1979)
22. Raz, R.: A parallel repetition theorem. SIAM J. Comput. 27(3), 763–803 (1998)
23. Cleve, R., Slofstra, W., Unger, F., Upadhyay, S.: Perfect parallel repetition theorem for quantum XOR proof systems. In: Proc. 22nd CCC, pp. 109–114 (2007)