

# On the communication complexity of distributions

JÉRÉMIE ROLAND <sup>\*</sup>      MARIO SZEGEDY <sup>†</sup>

January 24, 2008

## 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 (this is their only source of randomness), and have access to a channel that allows two-ways communication. Our main focus is an instance of the above problem coming from the well known EPR experiment in quantum physics, but we also present some more general facts about this rather young and promising complexity measure. The results contained herein are entirely classical and no knowledge of the quantum phenomenon is assumed.

Different notions of complexity may be defined for this problem. Due to an upper bound by Toner and Bacon [TB03], and a matching lower bound by Barrett, Kent and Pironio [BKP06], the average and worst-case communication complexities are known to be both one bit. In this paper, we focus on the *asymptotic* communication complexity, for which only an upper bound of 0.85 bits was presented by Toner and Bacon. We show that the technique they use to compress the communication is limited by yet another notion of complexity, which we call the *entropic* communication, but that it is possible to reduce the communication even further to 0.28 bits. This is somewhat of a surprise, because some conjectured that the earlier bound was optimal. In our investigation we find interesting connections to a number of different problems in communication complexity, in particular to [HJMR06].

---

<sup>\*</sup>University of California, Berkeley. [jroland@berkeley.edu](mailto:jroland@berkeley.edu)

<sup>†</sup>Rutgers University. [szegedy@cs.rutgers.edu](mailto:szegedy@cs.rutgers.edu)

# 1 Introduction

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 [Yao79] 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  $\mathbf{p}$  is specified by a function  $\mathbf{p} : \mathcal{X} \times \mathcal{Y} \rightarrow \text{Distrib}(\mathcal{A} \times \mathcal{B})$ , where  $\text{Distrib}(\mathcal{A} \times \mathcal{B})$  is the set of all probability distributions on  $\mathcal{A} \times \mathcal{B}$ . 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  $\Lambda$ , i.e. in addition to their input they both receive  $\lambda$ , 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,  $\lambda$ , and the messages received so far by the player. *A protocol must satisfy that in each round either of two cases may happen:* 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.

If the parties needed to make random choices, we could add them to the shared randomness,  $\Lambda$ , thus making the running of the protocol deterministic for any fixed  $\lambda$ . More specifically, let us consider a communication protocol  $P$  for a given simulation task  $\mathbf{p}$ , using shared randomness  $\Lambda$ . For any fixed value  $\lambda \in \Lambda$  of the shared randomness, the players execute a deterministic protocol  $P_\lambda$ , which alone would solve another simulation task  $\mathbf{p}_\lambda$ . The protocol for  $\mathbf{p}$  therefore consists in running different protocols  $P_\lambda$  with probability  $p(\lambda)$ , where  $p(\lambda)$  defines a probability distribution over  $\Lambda$ , 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  $\Lambda$  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  $P$ ).

**Local Hidden Variable Model (LHV):** Let us consider the set  $\Lambda_0$  of all deterministic protocols that do not use any communication. We say that a task  $\mathbf{p}$  is in LHV if it may be simulated using a distribution over protocols in  $\Lambda_0$  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”  $\lambda$  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. A *Bell-type* inequality is a linear inequality involving parameters  $\{\mathbf{p}(a, b|x, y)\}_{x, y, a, b}$ , that is valid for all  $\mathbf{p} \in \text{LHV}$ . The parameters corresponding to all LHVs form a convex set. The fact that the subclass LHV alone has sparked mathematical interest shows that our model is highly non-trivial.

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  $\Lambda$ ,  $M(x, y, \lambda)$  be the transcript

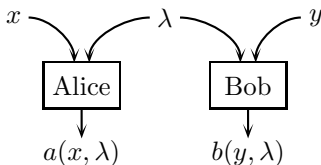


Figure 1: The special case of our model, where Alice and Bob do not communicate at all is called the *Local Hidden Variable* model (LHV). Bell-type inequalities imply that in this simple model most quantum correlations cannot be simulated.

of the messages on input  $x$  and  $y$  when the shared randomness is fixed to  $\lambda$ , and  $|M|$  be the length in bits of this transcript.. Given a distribution  $D$  on  $\mathcal{X} \times \mathcal{Y}$ , we define the following communication costs:

**Worst-case cost**  $C_w(P)$ : the maximal number of bits communicated between Alice and Bob in any particular execution of the protocol,

$$C_w(P) = \max_{(x,y) \in D, \lambda \in \Lambda} |M(x, y, \lambda)|. \quad (1)$$

**Average cost**  $\bar{C}(P)$ : 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  $(x, y) \in D$ ,

$$\bar{C}(P) = \sum_{\lambda \in \Lambda} p(\lambda) \sum_{(x,y) \in D} p(x, y) |M(x, y, \lambda)|. \quad (2)$$

The corresponding worst-case and average communication complexities of a task  $\mathbf{p}$  are then defined as  $C_w(\mathbf{p}) = \min_P C_w(P)$  and  $\bar{C}(\mathbf{p}) = \min_P \bar{C}(P)$ , where the minimum is taken over all protocols  $P$  implementing  $\mathbf{p}$ . 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

$$\mathbf{p}_\mu(a, b|x, y) = \frac{1 + \mu ab (-1)^{x \cdot y}}{4}. \quad (3)$$

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 [CHSH69]:

$$\sum_{x,y,a,b} ab (-1)^{x \cdot y} \mathbf{p}(a, b|x, y) \leq 2 \quad (4)$$

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 [Bel64].

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

$$\mathbf{p}_{\dim=d}(a, b | \vec{x}, \vec{y}) = \frac{1 - ab \vec{x} \cdot \vec{y}}{4}. \quad (5)$$

$\mathbf{p}_{\dim=d}$  arises from the EPR-Bohm experiment [EPR35, BA57], and can be solved in the quantum world with zero communication. Already for  $d = 2$  the problem  $\mathbf{p}_{\dim=d}$  is more general than  $\mathbf{P}_{\frac{1}{\sqrt{2}}}$ .

In the effort of closing the gap between  $\mathbf{p}_{\dim=d}$  and LHV, one may investigate for what  $0 \leq \mu \leq 1$  does the task described by

$$\mathbf{p}_{\dim=d,\mu}(a, b | \vec{x}, \vec{y}) = \frac{1 - \mu ab \vec{x} \cdot \vec{y}}{4} \quad (6)$$

belong to LHV. The question is shown by Toner *et al* [AGT06] for general  $d$  to be equivalent to the well-known open problem of finding the exact value for the Groethendieck constant! The asymptotic communication cost of simulating  $\mathbf{p}_{\dim=3}$  (and its powers) will be the focus of this paper, where points of a sphere are chosen according to the Haar measure.

**Finiteness:** In the above example  $\mathcal{X}$ ,  $\mathcal{Y}$ , and probability space  $\Lambda$  are infinite. This is permitted in our model as long as the communication is bounded. When  $\mathcal{X}$  and/or  $\mathcal{Y}$  are infinite, we need to modify formula (2) by replacing the sum with an integral, where  $D$  is now a probability measure on  $\mathcal{X} \times \mathcal{Y}$ .

## 2 History, Related Works, Our Results

Even though the EPR-Bohm experiment may not be simulated by a local hidden variable model, we know that this becomes possible when the local hidden variable model is supplemented by additional resources, such as communication [Mau92, BCT99, Ste00, CGM00], post-selection [GG99], or non-local boxes [CGMP05]. These papers mostly focus on  $\mathbf{p}_{\dim=3}$ , which is of special interest 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. As a consequence, this case has been intensively studied in the literature.

In the case of communication, the best known protocol was presented by Toner and Bacon, and uses one bit of communication [TB03]. This was shown to be optimal, even for the average complexity, by Barrett, Kent and Pironio [BKP06]. In this article, we focus on the *asymptotic* communication complexity, which is defined as follows:

**Asymptotic communication complexity,  $C_\infty(\mathbf{p})$ :** the limit  $\lim_{n \rightarrow \infty} \bar{C}(\mathbf{p}^{\otimes n})/n$ , where  $\mathbf{p}^{\otimes n}$  is the probability distribution obtained by the straightforward  $n$ -fold parallelization of  $\mathbf{p}$ .

In the case  $\mathbf{p} = \mathbf{p}_{\dim=3}$ , Toner and Bacon showed that the asymptotic communication is strictly less than one, because conditionally on the shared randomness, the bit communicated in their protocol is not uniformly distributed, which allows Shannon compression when the protocol is repeated in parallel. More precisely, they showed that  $C_\infty(\mathbf{p}) \leq \text{Si}(\pi)/(\pi \ln 2) \approx 0.85$  bits, where  $\text{Si}(x)$  is the sine integral function. We argue that the method used to reduce the communication for parallel repetitions of the problem is not optimal by first showing that it is limited by a notion of *entropic* communication complexity, which we introduce in Section 3 and denote  $C_H(P)$ . In Section 4, using a method similar to Pironio [Pir03], we then prove that  $C_H(P) \geq \sqrt{2} - 1 \approx 0.41$  bits. In Section 5 we use an approach presented by Degorre, Laplante and Roland [DLR05], in conjunction with the reverse Shannon theorem [BSST02] to show that a mere  $1 - 1/(2 \ln 2) \approx 0.28$  bits are sufficient in the asymptotic case, significantly improving on the previous best protocol, while also beating the lower bound on the entropic communication complexity. Finally, we introduce in Section 6 a new method, which we call the difference

method, that allows us to prove a lower bound of 0.13 bits on the asymptotic communication complexity of this problem when restricting to one-way communication.

The higher dimensional case  $\mathbf{p} = \mathbf{p}_{\dim=d}$  has been studied by Degorre, Laplante and Roland [DLR06], who proved that the average communication complexity scaled at most as  $\bar{C}(\mathbf{p}) = O(\log d)$ . This was significantly improved by Regev and Toner [RT07], who showed that bounded worst-case communication was sufficient, by providing an explicit 2-bit protocol, so that  $C_w(\mathbf{p}) \leq 2$ . They conjecture that even though 1 bit was sufficient for the 3-dimensional case, 2 bits were necessary for sufficiently large  $d$ . When considering average communication, they could improve their protocol to 1.82 bits. In that case, the best lower bound is also the 1-bit lower bound proved by Barrett, Kent and Pironio [BKP06].

### 3 Entropic and asymptotic complexities

In this section, we review different notions of communication complexity, including a new type which we call entropic communication complexity, and we prove relations between them. These definitions are distributional, that is, we suppose that the input distribution  $D$  is fixed and known by the players, but similar definitions and relations hold when maximizing the costs over the input distribution.

Let  $P$  be a communication protocol simulating  $\mathbf{p}(a, b|x, y)$  using shared randomness  $\Lambda$ . In addition to the worst-case cost  $C_w(P)$  and average cost  $\bar{C}(P)$ , we define the entropic cost of  $P$  as follows

**Entropic cost**  $C_H(P)$ : 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 then define the corresponding entropic communication complexity for a task  $\mathbf{p}$  as  $C_H(\mathbf{p}) = \min_P C_H(P)$ , where the minimum is taken over all protocols  $P$  implementing  $\mathbf{p}$ .

Finally, we define the following cost for communication protocols using *private* randomness only, first introduced by Chakrabarti *et al* [CSWY01]:

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

For private randomness protocols, we define  $C_\infty^\emptyset(\mathbf{p})$  similarly to the shared randomness case, as well as the information complexity  $C_I^\emptyset(\mathbf{p}) = \min_P C_I^\emptyset(P)$ , where the minimum is taken over all private randomness protocols  $P$  implementing  $\mathbf{p}$  (note that for shared randomness protocols, we would have  $C_I(\mathbf{p}) = 0$ , so this quantity is not relevant).

We show the following:

**Proposition 1.** *The communication complexities satisfy the following relations:*

$$C_\infty(\mathbf{p}) \leq C_H(\mathbf{p}) \leq \bar{C}(\mathbf{p}) \leq C_w(\mathbf{p}) \tag{7}$$

$$C_\infty(\mathbf{p}) \leq C_I^\emptyset(\mathbf{p}) \stackrel{(*)}{\leq} C_\infty^\emptyset(\mathbf{p}), \tag{8}$$

where all relations are valid for any input distribution  $D$ , except the relation marked  $(*)$  which is only valid for product input distributions ( $x$  and  $y$  being independent).

*Proof.*  $\bar{C}(\mathbf{p}) \leq C_w(\mathbf{p})$  is immediate. The other relations are based on fundamental propositions in information theory, namely Shannon's source coding theorem [CT91] and the reverse Shannon theorem [BSST02].

$[C_H(\mathbf{p}) \leq \bar{C}(\mathbf{p})]$  Let  $\bar{C}(\mathbf{p})$  be achieved by a protocol  $P$ . By applying Shannon's source coding theorem to each message exchanged during the protocol, we may lower bound the average length of the whole transcript as  $E(|M|) \geq H(M|\Lambda)$ , so that  $C_H(P) \leq \bar{C}(P)$  and, in turn,  $C_H(\mathbf{p}) \leq \bar{C}(\mathbf{p})$ .

$[C_\infty(\mathbf{p}) \leq C_H(\mathbf{p})]$  Let  $C_H(\mathbf{p}) = H(M|\Lambda)$  be achieved by a protocol  $P$ . We build a protocol  $P_n$  for  $\mathbf{p}^{\otimes n}$  by repeating  $n$  times the protocol  $P$  and compressing the concatenation of the messages from the different repetitions. By using an optimal code for each round of communication, the concatenated transcript  $\vec{M} = (M_1, \dots, M_n)$  may be compressed in such a way that  $H(\vec{M}|\vec{\Lambda}) \leq \bar{C}(P_n) \leq H(\vec{M}|\vec{\Lambda}) + 1$ . Since the  $n$  repetitions are independent, we have  $H(\vec{M}|\vec{\Lambda}) = \sum_i H(M_i|\Lambda_i) = nC_H(\mathbf{p})$ , which, together with  $\bar{C}(\mathbf{p}^{\otimes n}) \leq \bar{C}(P_n)$ , implies

$$C_\infty(\mathbf{p}) = \lim_{n \rightarrow \infty} \frac{\bar{C}(\mathbf{p}^{\otimes n})}{n} \leq C_H(\mathbf{p}).$$

$[C_\infty(\mathbf{p}) \leq C_I^\varnothing(\mathbf{p})]$  Let  $C_I^\varnothing(\mathbf{p}) = I(XY : M)$  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 alternatively sent by Alice to Bob and vice-versa. Let us denote  $M_k$  the  $k^{\text{th}}$  message and  $M_{[k]}$  the restriction of the transcript to the first  $k$  messages. We may express the information cost of  $\mathbf{p}$  as:

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

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

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

where 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  $X \rightarrow M_k$ . How much can she compress the message if the protocol is repeated? The reverse Shannon theorem [BSST02] ensures that for each  $n$ , there is a protocol that simulates  $n$  repetitions of the channel  $X \rightarrow M_k$  using shared randomness and communication with average length per repetition that tends to  $I(X : M_k | M_{[k-1]} = m)$  when  $n$  tends to infinity. By compressing similarly each successive message, and averaging over the possible messages, we get that  $C_\infty(\mathbf{p}) \leq I(XY : M) = C_I^\varnothing(\mathbf{p})$ .

$[C_I^\varnothing(\mathbf{p}) \leq C_\infty(\mathbf{p})]$  Suppose  $\bar{C}(\mathbf{p}^{\otimes n}) = E(|M|)$  is achieved by some protocol  $P_n$  (with private but no shared randomness). From Shannon's source coding theorem, we have that  $E(|M|) \geq H(M) \geq I(\vec{X}\vec{Y} : M)$ , where  $\vec{X} = (X_1, \dots, X_n)$  is the vector of Alice's inputs for the different repetitions, and similarly  $\vec{Y}$  for Bob's inputs. Since the inputs for different repetitions are independent, we have

$$I(\vec{X}\vec{Y} : M) \geq \sum_{i=1}^n I(X_i Y_i : M).$$

For each input pair  $(X_i, Y_i)$ , we may build a protocol  $P_i$  for  $\mathbf{p}$ , by using the protocol  $P_n$  for  $\mathbf{p}^{\otimes n}$  and replacing the additional inputs  $X_j, Y_j$  ( $\forall j \neq i$ ) by random values chosen by Alice and Bob using private randomness (this is possible since we have assumed here that  $X_j$  and  $Y_j$  are independent). The information cost of protocol  $P_i$  is then  $C_I^\varnothing(P_i) = I(X_i Y_i : M)$ . Since  $C_I^\varnothing(\mathbf{p}) \leq C_I^\varnothing(P_i)$ , we get by summing over  $i$  and dividing by  $n$  that  $C_I^\varnothing(\mathbf{p}) \leq \frac{\bar{C}(\mathbf{p}^{\otimes n})}{n}$  and, in the limit  $n \rightarrow \infty$ ,  $C_I^\varnothing(\mathbf{p}) \leq C_\infty(\mathbf{p})$ .  $\square$

## 4 Lower bound on the entropic complexity

The previous best upper bound on the asymptotic communication complexity is due to Toner and Bacon, who proved that  $C_\infty(\mathbf{p}_{\text{dim}=3}) \lesssim 0.85$  bits. Indeed, they showed that in their one-bit protocol, the conditional entropy of the messages given the shared randomness 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. As a consequence, the technique they use is limited by what we defined as the entropic communication complexity. In this section we prove that the one-way entropic complexity of  $\mathbf{p}_{\text{dim}=3}$  is at least 0.41 bits, 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 [Pir03]. Let us first recall our notations. Consider a communication protocol  $P$  for a given simulation task  $\mathbf{p}$ , using shared randomness  $\Lambda$ . The protocol  $P$  for  $\mathbf{p}$  consists in running different deterministic protocols  $P_\lambda$ , for other tasks  $\mathbf{p}_\lambda$ , with probability  $p(\lambda)$ , where  $p(\lambda)$  defines a probability distribution over  $\Lambda$  such that  $\mathbf{p} = \sum_\lambda p(\lambda)\mathbf{p}_\lambda$ . We extend  $\Lambda$  so that it is in bijection with the set of all deterministic protocols and consider the subset  $\Lambda_0$  of all deterministic protocols that do not use any communication. By definition, each  $\lambda \in \Lambda_0$  then corresponds to a protocol that simulates a task  $\mathbf{p}_\lambda$  in LHV. Moreover, the set of tasks in LHV is exactly those that can be solved by using a distribution  $p(\lambda)$  over protocols in  $\Lambda_0$  only.

Let  $B$  be a linear functional on a task  $\mathbf{p}$ . We call Bell inequality an inequality satisfied by  $B$  for all  $\mathbf{p}$  in LHV. More specifically, a Bell inequality reads:

$$B(\mathbf{p}) \leq B_0 \quad \forall \mathbf{p} \in \text{LHV},$$

where  $B_0 = \max_{\mathbf{p} \in \text{LHV}} B(\mathbf{p})$ . Note that since a task  $\mathbf{p}$  in LHV may be expressed as a  $\sum_{\lambda \in \Lambda_0} p(\lambda)\mathbf{p}_\lambda$ ,  $B_0$  may equivalently be defined as  $B_0 = \max_{\lambda \in \Lambda_0} B(\mathbf{p}_\lambda)$ .

The idea behind the following theorem is that a task  $\mathbf{p}$  outside LHV may violate a Bell inequality, so that they it will require to use protocols  $P_\lambda$  outside  $\Lambda_0$  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 least 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 2.** *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:*

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

where  $P_{\lambda^*}$  is a deterministic protocol such that

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

The proof is heavily inspired by the proof of Proposition 1 in [Pir03], it is given in Appendix A.

We may now completely determine the entropic communication complexity of  $\mathbf{p}_\mu$ :

**Theorem 3.** *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 showed to be tight by giving an explicit protocol.

$[C_H(\mathbf{p}_\mu) \geq 2\mu - 1]$  We use the CHSH inequality [CHSH69], 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  $B_0 = 2$ , so that the CHSH inequality reads  $B(\mathbf{p}) \leq 2$  for all  $\mathbf{p}$  in LHV. 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,  $\max_{P_\lambda} (B(P_\lambda) - B_0)/C_H(P_\lambda)$  is attained by a protocol  $P_{\lambda^*}$  where one player sends his input to the other, such that  $C_H(P_{\lambda^*}) = 1$  and  $B(\mathbf{p}_{\lambda^*}) = 4$ . We then obtain

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

$[C_H(\mathbf{p}_\mu) \leq 2\mu - 1]$  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  $C_H(P_{1/2}) = 0$ . 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,  $C_H(P_1) = 1$ . 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 suffice 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  $2\mu - 1$ .  $\square$

Using a reduction from  $\mathbf{p}_{1/\sqrt{2}}$  to  $\mathbf{p}_{\text{dim}=3}$ , we may now derive a bound on the (one-way) entropic complexity of  $\mathbf{p}_{\text{dim}=3}$ .

**Theorem 4.**  $C_{H,\text{one-way}}(\mathbf{p}_{\text{dim}=3}) \geq C_H(\mathbf{p}_{1/\sqrt{2}})$ .

*Proof.* The key observation is that the task  $\mathbf{p}_{1/\sqrt{2}}$  for uniformly distributed inputs is equivalent to the task  $\mathbf{p}_{\text{dim}=3}$  for a special distribution, where the inputs are uniform over two vectors  $\{\vec{x}_0, \vec{x}_1\}$  for Alice and two vectors  $\{\vec{y}_0, \vec{y}_1\}$  for Bob, laid out such that  $\vec{x}_i \cdot \vec{y}_j = (-1)^{i \cdot j} / \sqrt{2}$ .

We now show how, from a one-way communication protocol for  $\mathbf{p}_{\text{dim}=3}$  with uniformly distributed inputs, we may build a protocol for any distribution of the inputs. The idea is to apply the same random rotation  $R$  (using shared randomness) to both Alice's input  $\vec{x}$  and Bob's input  $\vec{y}$ , and then to execute the protocol for uniformly distributed inputs using  $R(\vec{x})$  and  $R(\vec{y})$  as inputs. This will keep the scalar product of the inputs unchanged. Moreover, since the protocol is one-way, the communication only depends on the input of one of the players, which is now uniformly distributed (note that the joint distribution of the inputs may not be uniform, but the marginals are).  $\square$

Since  $C_H(\mathbf{p}_\mu) = 2\mu - 1$ , we get that  $C_{H,\text{one-way}}(\mathbf{p}_{\text{dim}=3}) \geq \sqrt{2} - 1 \approx 0.41$  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 0.41 bits. We show in the next section that we can go below this lower bound using another technique, based on the reverse Shannon theorem.

## 5 A New Protocol

In this section, we show how to reduce the communication to 0.28 bits for parallel repetitions of the problem of simulating  $\mathbf{p}_{\text{dim}=3}$ , that is, we show that  $C_\infty(\mathbf{p}_{\text{dim}=3}) \leq 0.28$  bits. We use a result due to Degorre *et al*, which shows that the problem reduces to a distributed sampling task:

**Theorem 5** ([DLR05]). *Let  $\vec{x}$  and  $\vec{y}$  be Alice's and Bob's inputs. If Alice and Bob share a random variable  $\vec{\xi} \in \mathbb{S}_2$  distributed according to a biased distribution with probability density*

$$\rho(\vec{\xi}|\vec{x}) = \frac{|\vec{x} \cdot \vec{\xi}|}{2\pi}, \quad (9)$$

*then they are able to simulate  $\mathbf{p}_{\text{dim}=3}$  without any further resource.*

This observation leads to an apparently very bad communication protocol for  $\mathbf{p}_{\text{dim}=3}$  with private randomness only: using her input and private randomness, Alice locally samples  $\vec{\xi}$  according to the distribution  $\rho(\vec{\xi}|\vec{x})$ , and then communicates  $\vec{\xi}$  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, we may significantly reduce the communication using the reverse Shannon theorem. In particular, we prove the following upper bound on the asymptotic communication complexity:

**Theorem 6.**  $C_\infty(\mathbf{p}_{\text{dim}=3}) \leq 1 - 1/(2 \ln 2) \approx 0.28$  bits.

*Proof.* Let  $P$  be the following private randomness protocol for  $\mathbf{p}_{\text{dim}=3}$ : using her input together with private randomness, Alice samples a random variable  $\Xi$  according to the distribution  $\rho(\vec{\xi}|\vec{x})$  defined above and communicates the obtained sample  $\vec{\xi}$  to Bob. If the players then set their respective outputs as  $a = \text{sgn}(\vec{x} \cdot \vec{\xi})$  and  $b = \text{sgn}(\vec{y} \cdot \vec{\xi})$ , they reproduce  $\mathbf{p}_{\text{dim}=3}$ , as was shown by Degorre *et al* [DLR05]. Let us now compute the information cost of this protocol,  $C_I(P) = I(XY : \Xi) = H(\Xi) - H(\Xi|X)$ . Since  $X$  is uniformly distributed on the sphere, we have  $\rho(\vec{x}) = 1/(4\pi)$ , so that  $\rho(\vec{\xi}) = 1/(4\pi)$  and the entropy of  $\Xi$  is simply:

$$H(\Xi) = - \int_{\mathbb{S}_3} d\vec{\xi} \rho(\vec{\xi}) \log \rho(\vec{\xi}) = \log 4\pi.$$

As for the conditional entropy  $H(\Xi|X = \vec{x})$ , it does not depend on the particular value of  $\vec{x}$  so that

$$H(\Xi|X) = \int_{\mathbb{S}_3} d\vec{x} \rho(\vec{x}) H(\Xi|X = \vec{x}) = H(\Xi|X = \vec{x}_0),$$

where  $\vec{x}_0$  is an arbitrary vector. By taking  $\vec{x}_0$  aligned with the  $z$ -axis of a coordinate system, we have

$$H(\Xi|X) = - \int_{\mathbb{S}_3} d\vec{\xi} \rho(\vec{\xi}|\vec{x}_0) \log \rho(\vec{\xi}|\vec{x}_0) \quad (10)$$

$$= - \frac{1}{2\pi} \int_0^{2\pi} d\varphi \int_0^\pi d\theta \sin \theta |\cos \theta| \log \frac{|\cos \theta|}{2\pi} \quad (11)$$

$$= -2 \int_0^1 du u \log \frac{u}{2\pi} = \log 2\pi - 2 \int_0^1 du u \log u \quad (12)$$

$$= \log 2\pi + \frac{1}{2 \ln 2}. \quad (13)$$

Finally, we find  $C_I(P) = 1 - 1/(2 \ln 2)$ , and we have shown in the previous section that the reverse Shannon theorem implies that  $C_\infty(\mathbf{p}_{\text{dim}=3}) \leq C_I(\mathbf{p}_{\text{dim}=3}) \leq C_I(P)$ .  $\square$

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

**Theorem 7.** *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  $\xi$  be a random bit correlated with  $x$ , such that  $p(x = \xi) = \mu$ . The channel defined by the Markov process  $X \rightarrow \Xi$  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

$$a = (-1)^{(x \oplus \xi \oplus 1) \cdot \lambda_0} (-1)^{(x \oplus \xi) \cdot \lambda_1}, \quad (14)$$

$$b = (-1)^{\lambda_0} (-1)^{y \oplus \xi}, \quad (15)$$

where  $\lambda_0, \lambda_1$  are shared unbiased random bits. The reverse Shannon theorem then ensures that asymptotically, the channel  $X \rightarrow \Xi$  may be simulated using on average  $1 - H[\mu]$  bits per repetition.  $\square$

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

## 6 The Difference-Method

Asymptotic lower bounds are notoriously hard to prove. Examples include the Shannon capacity of graphs and the Parallel Repetition Theorem of Raz. In some lucky cases the situation is better. Quantum values of XOR games and communication complexity of correlations are examples, where mathematics seems to be in our favor. Incidentally, both topics have relevance to lower bounding asymptotic communication complexity. To illustrate our difficulties consider simulation task  $\mathbf{p}_1$ . Recall from Section 1 that  $\mathbf{p}_1$  is specified by  $\mathcal{X} = \mathcal{Y} = \{0, 1\}$ ,  $\mathcal{A} = \mathcal{B} = \{1, -1\}$ ,

$$\mathbf{p}_1(a, b|x, y) = \begin{cases} 1/2 & \text{if } ab = -1^{xy}, \\ 0 & \text{otherwise.} \end{cases}$$

Study  $\mathbf{p}_1^{\otimes n}$ . Alice and Bob, upon getting  $(x_1, \dots, x_n)$  and  $(y_1, \dots, y_n)$  output  $(a_1, \dots, a_n)$  and  $(b_1, \dots, b_n)$  such that  $a_i b_i = -1^{x_i y_i}$  for every  $1 \leq i \leq n$ . This lets Alice and Bob with just one extra bit of communication compute  $-1^{\sum_i x_i y_i} = \prod_i a_i \prod_i b_i$  by Alice sending  $\prod_i a_i$  to Bob. Thus the INNER PRODUCT function is reduced to  $\mathbf{p}_1^{\otimes n}$ . The former has communication complexity at least  $n/2$ , so the latter must have communication complexity at least  $n/2 - 1$ . One can tackle lower bounding  $\overline{C}_\infty(\mathbf{p}_c)$  in general with a slightly more sophisticated version of the method and gets:

**Theorem 8.** *For any  $c > 1/\sqrt{2}$  we have  $\overline{C}_\infty(\mathbf{p}_c) > 0$ .*

This theorem, the proof of which we leave to the full version of this paper, is not hard, and is based on the Lindsey lemma. That the proof breaks down at  $1/\sqrt{2}$  (which is the value we would be most interested in) is not an accident. First, it can be shown that the problem which we get, when we apply the above reduction, has communication complexity two, when  $c < 1/\sqrt{2}$ , by an interesting application of [RT07]. Secondly, we can prove that for all XOR games, when making a communication game out of them, the type of lower bound as in the theorem is typically provable exactly up to the quantum value of the game. These proofs we defer to the full version.

This section is about how to get beyond the notorious  $c = \frac{1}{\sqrt{2}}$  limit, and in fact reach  $c = 1/2 + \epsilon$ , which is the ultimate limit, since  $\mathbf{p}_{1/2}$  is LHV, so requires no communication. 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 9.** *For any  $0.5 \leq c \leq 1$  we have  $\overline{C}_{\infty, \text{one-way}}(\mathbf{p}_c) \geq 1 - H[c]$ .*

Note:  $H(c) = c \log \frac{1}{c} + (1-c) \log \frac{1}{1-c}$ , thus we always get something non-trivial unless  $c = 0.5$ . What we need to show is that  $\mathbf{p}_c^{\otimes n}$  requires  $(1 - H(c))n$  bits of communication.

By  $\circ$  we will denote the point-wise product of two vectors. The idea of the proof is to reduce the problem to communication complexity problem of a correlation *à la* Harsha *et al* [HJMR06], and then use their theorem that the mutual information between the outputs of Alice and Bob is a lower bound on the communication. To apply [HJMR06] 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. But we show that if we take both  $\vec{0}$  and  $\vec{1}$  on Bob's side, run the protocol with a randomly chosen input,  $\vec{x}$ , of Alice, and receive outputs  $\vec{b}$  and  $\vec{b}'$ , respectively from Bob, then  $\vec{b} \circ \vec{b}'$  will contain a lot of information about Alice's input,  $\vec{x}$ . Since the communication is one way, and it depends only on Alice's input, when Bob receives Alice's input, he can just compute the output on any set of input vectors he wants to, but we just use the set  $\{\vec{0}, \vec{1}\}$  for this argument.

Observe that  $(\vec{a} \circ \vec{b}) \circ (\vec{a} \circ \vec{b}') = \vec{b} \circ \vec{b}'$ . Focusing on a particular index  $1 \leq i \leq n$  the specification of  $\mathbf{p}_c$  tells us that  $a_i b_i$  should take 1 with probability  $(1+c)/2$  and 0 with probability  $(1-c)/2$ . Also,  $a_i b'_i$  should take  $(-1)^{x_i}$  with probability  $(1+c)/2$  and  $(-1)^{x_i+1}$  with probability  $(1-c)/2$ . The union bound gives that the probability that  $b_i b'_i = (-1)^{x_i}$  is at least  $1-c$ . This shows that the mutual information  $I(D_i : X_i)$  between  $d_i = b_i b'_i$  and  $x_i$  is at least  $1 - H[c]$ , where we have used the fact that  $H(X_i) = 1$ . Since the  $x_i$ 's are independent, we get for the mutual information between  $\vec{d}$  and  $\vec{x}$

$$I(D : X) \geq \sum_i I(D : X_i) \geq \sum_i I(D_i : X_i) = n(1 - H[c]),$$

which, by Shannon's channel coding theorem, implies that the average communication is at least  $n(1 - H[c])$ . Note that, as expected, we get  $\overline{C}_{\infty, \text{one-way}}(\mathbf{p}_c) \geq 0$  bit for  $c = \frac{1}{2}$ , and  $\overline{C}_{\infty, \text{one-way}}(\mathbf{p}_c) \geq 1$  bit for  $c = 1/2$ . Moreover, we have  $\overline{C}_{\infty, \text{one-way}}(\mathbf{p}_c) \gtrsim 0.13$  bits for  $c = \frac{1}{\sqrt{2}}$ .

## References

- [AGT06] A. Acín, N. Gisin, and B. Toner. Grothendieck's constant and local models for noisy entangled quantum states. *Phys. Rev. A*, 73:062105, 2006.
- [BA57] D. Bohm and Y. Aharonov. Discussion of Experimental Proof for the Paradox of Einstein, Rosen, and Podolsky. *Phys. Rev.*, 108:1070–1076, 1957.
- [BCT99] Gilles Brassard, Richard Cleve, and Alain Tapp. Cost of Exactly Simulating Quantum Entanglement with Classical Communication. *Phys. Rev. Lett.*, 83:1874–1877, 1999. quant-ph/9901035.
- [Bel64] John S. Bell. On the Einstein Podolsky Rosen paradox. *Physics*, 1:195, 1964.
- [BKP06] Jonathan Barrett, Adrian Kent, and Stefano Pironio. Maximally nonlocal and monogamous quantum correlations. *Physical Review Letters*, 97(17):170409, 2006.
- [BSST02] C.H. Bennett, P.W. Shor, J.A. Smolin, and A.V. Thapliyal. Entanglement-assisted capacity of a quantum channel and the reverse shannon theorem. *IEEE Transactions on Information Theory*, 48(10), 2002.
- [CGM00] Nicolas J. Cerf, Nicolas Gisin, and Serge Massar. Classical Teleportation of a Quantum Bit. *Phys. Rev. Lett.*, 84:2521–2524, 2000. quant-ph/9906105.
- [CGMP05] Nicolas J. Cerf, Nicolas Gisin, Serge Massar, and Sandu Popescu. Simulating Maximal Quantum Entanglement without Communication. *Phys. Rev. Lett.*, 94(22):220403, 2005. quant-ph/0410027.

- [CHSH69] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt. Proposed Experiment to Test Local Hidden-Variable Theories. *Phys. Rev. Lett.*, 23:880–884, 1969.
- [CSWY01] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *42nd IEEE Symposium on Foundations of Computer Science*, pages 270–278, 2001.
- [CT91] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley & Sons, New York, 1991.
- [DLR05] Julien Degorre, Sophie Laplante, and Jérémie Roland. Simulating quantum correlations as a distributed sampling problem. *Phys. Rev. A*, 72:062314, 2005. e-print quant-ph/0507120.
- [DLR06] Julien Degorre, Sophie Laplante, and Jérémie Roland. Classical simulation of traceless binary observables on any bipartite quantum state. *Physical Review A*, 75:012309, 2006. e-print quant-ph/0608064.
- [EPR35] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 47:777–780, 1935.
- [GG99] Nicolas Gisin and Bernard Gisin. A local hidden variable model of quantum correlation exploiting the detection loophole. *Phys. Lett. A*, 260(5):323–327, 1999. quant-ph/9905018.
- [HJMR06] Prahladh Harsha, Rahul Jain, David McAllester, and Jaikumar Radhakrishnan. The communication complexity of correlation. *Electronic Colloquium on Computational Complexity*, 13(151), 2006.
- [Mau92] T. Maudlin. Bell’s inequality, information transmission, and prism models. In *Biennial Meeting of the Philosophy of Science Association*, pages 404–417, 1992.
- [Pir03] Stefano Pironio. Violations of bell inequalities as lower bounds on the communication cost of nonlocal correlations. *Phys. Rev. A*, 68(6):062102, 2003. e-print quant-ph/0304176.
- [RT07] Oded Regev and Ben Toner. Simulating quantum correlations with finite communication. In *48th Annual IEEE Symposium on Foundations of Computer Science*, pages 384–394, 2007.
- [Ste00] Michael Steiner. Towards quantifying non-local information transfer: finite-bit non-locality. *Phys. Lett. A*, 270:239–244, 2000. quant-ph/9902014.
- [TB03] Ben F. Toner and Dave Bacon. Communication Cost of Simulating Bell Correlations. *Phys. Rev. Lett.*, 91:187904, 2003. quant-ph/0304076.
- [Yao79] Andrew Yao. Some complexity questions related to distributive computing. In *Proceedings of the 11th annual ACM Symposium on Theory of computing*, pages 209–213, 1979.

## A Proof of Theorem 2

Let us prove the theorem in the restricted case where the input sets  $\mathcal{X}, \mathcal{Y}$  are finite. We will use the two following facts:

**Fact 1:**  $B(\mathbf{p}_{\lambda^*}) > B_0$ .

**Fact 2:** For all  $\lambda \in \Lambda$ , we have

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

We first show that these facts imply the theorem, and then prove the facts themselves. Let  $C_H(\mathbf{p})$  be achieved by a communication protocol  $P$  using shared randomness  $\Lambda$ . Each possible value  $\lambda$  of the shared randomness then defines a deterministic communication protocol  $P_\lambda$ , simulating some probability distribution  $\mathbf{p}_\lambda$ , such that  $\mathbf{p} = \sum_\lambda p(\lambda) \mathbf{p}_\lambda$ . From Fact 2 and the linearity of  $B$ , we find for the entropic communication complexity

$$C_H(P) = \sum_\lambda p(\lambda) C_H(P_\lambda) \tag{16}$$

$$\geq \sum_\lambda p(\lambda) \frac{B(\mathbf{p}_\lambda) - B_0}{B(\mathbf{p}_{\lambda^*}) - B_0} C_H(P_{\lambda^*}) \tag{17}$$

$$= \frac{B(\mathbf{p}) - B_0}{B(\mathbf{p}_{\lambda^*}) - B_0} C_H(P_{\lambda^*}). \tag{18}$$

*Proof of Fact 1.* The key observation is that we may always build a protocol that simulates  $\mathbf{p}$  where Alice sends her input  $x$  to Bob, and Bob sends his input  $y$  to Alice. In that case, both Alice and Bob know the distribution  $\mathbf{p}(x, y)$  they need to simulate, and they may solve the problem using shared randomness only. This protocol has an entropic cost of  $H(X) + H(Y)$  bits. Note that the entropies  $H(X)$  and  $H(Y)$  are well defined due to the assumption that  $\mathcal{X}$  and  $\mathcal{Y}$  are finite (in particular, we have  $H(X) \leq \log |\mathcal{X}|$  and  $H(Y) \leq \log |\mathcal{Y}|$ ). The protocol we obtain will be a distribution over deterministic protocols, each having a cost  $H(X) + H(Y)$ . Since the protocol achieves to simulate  $\mathbf{p}$  with  $B(\mathbf{p}) > B_0$ , it must use with some non-zero probability a deterministic protocol  $P_{\lambda'}$  such that  $B(\mathbf{p}_{\lambda'}) \geq B(\mathbf{p}) > B_0$ . We now have

$$\frac{B(\mathbf{p}_{\lambda^*}) - B_0}{C_H(P_{\lambda^*})} = \max_{\lambda \notin \Lambda_0} \frac{B(\mathbf{p}_\lambda) - B_0}{C_H(P_\lambda)} \geq \frac{B(\mathbf{p}_{\lambda'}) - B_0}{C_H(P_{\lambda'})} \geq \frac{B(\mathbf{p}) - B_0}{H(X) + H(Y)}.$$

This, together with the positivity of the entropies  $H(X), H(Y)$  and  $C_H(P_{\lambda^*})$ , and the fact that  $B(\mathbf{p}) > B_0$ , proves that  $B(\mathbf{p}_{\lambda^*}) > B_0$ .  $\square$

*Proof of Fact 2.* Using Fact 1 and  $C_H(P_{\lambda^*}) > 0$  (since  $\lambda^* \notin \Lambda_0$ ), this is equivalent to

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

For  $\lambda \notin \Lambda_0$ , we have  $C_H(P_\lambda) > 0$  and this follows directly from the definition of  $P_{\lambda^*}$ . For  $\lambda \in \Lambda_0$ , we have  $C_H(P_\lambda) = 0$  and this follows directly from  $B(\mathbf{p}_\lambda) \leq B_0$  (by the definition of  $B_0$ ).  $\square$

Let us note that if we define a new linear functional  $\tilde{B}$  as

$$\tilde{B} = \frac{C_H(P_{\lambda^*})}{B(\mathbf{p}_{\lambda^*}) - B_0} (B - B_0 I),$$

where  $I$  is the linear functional that maps any task  $\mathbf{p}$  to 1, Fact 2 shows that  $\tilde{B}$  corresponds to a Bell inequality  $\tilde{B}(\mathbf{p}) \leq 0$  ( $\forall \mathbf{p} \in \text{LHV}$ ), which is furthermore normalized in such a way that  $\tilde{B}(\mathbf{p}_\lambda) \leq C_H(P_\lambda)$  ( $\forall \lambda \in \Lambda$ ). The theorem then simply follows from that fact that for such an inequality, we have  $\tilde{B}(\mathbf{p}) \leq C_H(P)$  if  $P$  is a protocol for  $\mathbf{p}$ .