Information-based measure of nonlocality

Alberto Montina and Stefan Wolf
Facoltà di Informatica, Università della Svizzera Italiana, Via G. Buffi 13, 6900 Lugano, Switzerland
(Dated: February 26, 2016)

Quantum nonlocality concerns correlations among spatially separated systems that cannot be explained classically without communication among the parties. Thus, a natural measure of nonlocal correlations is provided by the minimal amount of communication required for classically simulating them. In this paper, we present a method to compute the minimal communication cost of parallel simulations, which we call nonlocal capacity, for any general nonsignaling correlations. This measure turns out to have an important role in communication complexity and can be used to discriminate between local and nonlocal correlations, as an alternative to the violation of Bell's inequalities.

I. INTRODUCTION

The outcomes of measurements performed on spatially separate entangled systems can display nonlocal correlations that cannot be explained classically without some communication [1]. In particular, one of the parties needs some information on the measurement choice of the other party. These *nonlocal correlations* can be used as an information-theoretic resource. For example, they can exponentially reduce the amount of communication required to solve some distributed computational problems [2, 3]. Furthermore, for some tasks, the use of nonlocal correlations can make communication unnecessary, such as in pseudo-telepathy games [4]. Some stronger-than-quantum nonsignaling correlations can even collapse the communication complexity in any two-party scenario. Indeed, the access to an unlimited number of Popescu-Rohrlich (PR) nonlocal boxes allows two parties to solve any communication complexity problem with the aid of a constant amount of classical communication [5]. Nonlocal correlations have also a fundamental role in device-independent applications, such as key agreement in cryptography [6–12] and randomness amplification [13, 14].

As the violation of a given Bell inequality is the signature of nonlocal correlations, a possible measure of nonlocality is the strength of this violation. However, since this quantity has no obvious relation with information, it does not necessarily provide a reliable measure as an information-theoretic resource. A more natural measure has been employed in Refs. [4, 15–18] and relies on the very definition of nonlocality; nonlocal correlations require some communication to be classically simulated, thus the minimal amount of required classical communication can be used as a measure of the strength of nonlocality. This measure, which we call *communication complexity* of the nonlocal resource, provides an ultimate limit to the power of nonlocal correlations in terms of classical communication in a two-party scenario. Indeed, nonlocal resources cannot replace an amount of classical communication bigger than the associated communication complexity. As shown in Ref. [19], the strength of the Bell inequality violation and the communication complexity of nonlocal resources turn out to be identical if the average amount of communication is employed as measure of the communication cost and the optimal inequality is taken for the given nonlocal correlations. In this paper, we mainly focus on the minimal asymptotic communication cost of parallel simulations in the asymptotic limit of infinite instances. This quantity, which we call *nonlocal capacity*, turns out to be much easier to be computed than its single-shot counterpart. Furthermore, tight lower and upper bounds on the minimal average communication cost are given in terms of the nonlocal capacity, as discussed later. Thus, the nonlocal capacity also gives tight bounds on the maximal violation of the Bell inequalities. Alternative measures of nonlocality could use different resources as unit of nonlocality, such as nonlocal boxes [20, 21]. For example, the strength of nonlocality could be defined as the number of PR-boxes necessary to simulate the correlations. However, no finite set of PR-boxes can simulate all bipartite nonlocal correlations [22, 23].

By definition, the computation of the nonlocal capacity is an optimization problem, but it is not convex in its original form. This makes it very hard to find the global minimum, expecially when the set of allowed measurements is large. In this paper, we show that the problem can be reduced to a convex minimization problem, which can be numerically solved with very efficient algorithms [24]. Then, we discuss the relation with a previous work on the communication complexity of channels in general probabilistic theories [25]. Finally, we illustrate the method with a numerical example.

II. COMMUNICATION COST OF NONLOCAL CORRELATIONS

In this paper, we will discuss the general case of nonsignaling correlations, which satisfy the minimal requirements of relativity and causality. Namely, the object that we will consider is a nonsignaling box, which is an abstract generalization of the following quantum scenario. Two parties, say Alice and Bob, simultaneously perform a measurement on two spatially separate parts of an entangled system. In general, Alice and Bob are allowed to choose among their respective sets of possible measurements. We assume that Bob's set of mea-

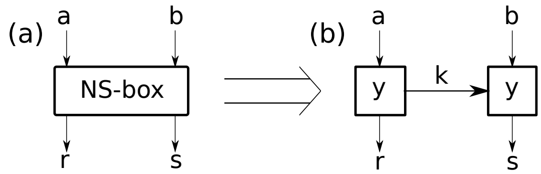

FIG. 1: (a) Nonsignaling box with inputs a and b and outcomes r and s. (b) Simulation of the nonsignaling box through shared stochastic variable y and communication of the variable k.

surements is finite, but arbitrarily large. For the sake of simplicity, we also assume that Alice's set is discrete, although this is not strictly necessary. Let us denote by the indices a and b the measurements performed by Alice and Bob, respectively. The index b takes a value in $\{1, \dots, M\}$, where M is the number of measurements that Bob can perform. After the measurements, Alice gets an outcome r and Bob an outcome s. The overall scenario is described by the joint conditional probability $P(r, s|a, b)$. This distribution satisfies the nonsignaling conditions

$$\begin{array}{c} \sum_s P(r,s|a,b) = \sum_s P(r,s|a,\bar{b}) \equiv P(r|a)\ \forall a,b,\bar{b},r, \\ \sum_r P(r,s|a,b) = \sum_r P(r,s|\bar{a},b) \equiv P(s|b)\ \forall a,\bar{a},b,s. \end{array} \quad (1)$$

These conditions are implied by causality and relativity. In the following discussion, we consider a more general scenario including non-quantum correlations and we just assume that the joint conditional probability satisfies the nonsignaling condition. The abstract machine producing the correlated variables r and s from the inputs a and b will be called *nonsignaling box* (briefly, NS-box). The NS-box, schematically represented in Fig. 1a, is identified with the conditional probability $P(r, s|a, b)$.

In general, a classical simulation of the joint distribution $P(r, s|a, b)$ requires some communication between the parties. We assume that only a one-way communication from Alice to Bob is allowed. The classical protocol is as follows (as illustrated in Fig. 1b). Alice generates an outcome r and a variable k with probability $P(k, r|y\,a)$ depending on the variable a and some stochastic variable y shared with Bob and generated with probability $\rho(y)$. The variable k is sent to Bob. Finally, Bob generates an outcome s with probability $P(s|y\,b\,k)$ depending on y, b and k. The protocol simulates the NS-box $P(r, s|a, b)$ if

$$\sum_k \int dy P(s|y\,b\,k)P(k,r|y\,a)\rho(y) = P(r,s|a,b). \quad (2)$$

We define the *communication complexity* (denoted by $\mathcal{C}_{nl}$) of the NS-box as the minimal amount of communication $\mathcal{C}$ required for an exact simulation of the NS-box.

There are different measures of *amount of communication*. Here we employ the entropic definition, although the presented results apply also to the case of average communication. Let us introduce the conditional probability

$$P(k|y) \equiv \sum_{r,a} P(k,r|y,a)P(a)$$

and the corresponding conditional Shannon entropy of the variable K given Y

$$H_{P(a)}(K|Y) \equiv -\int dy \rho(y) \sum_k P(k|y) \log_2 P(k|y),$$

which depends on $P(a)$. We define the communication cost $\mathcal{C}$ of the simulation as the maximum, over the space of distributions $P(a)$, of $H_{P(a)}(K|Y)$, that is,

$$\mathcal{C} \equiv \max_{P(a)} H_{P(a)}(K|Y) \quad (3)$$

(see also Refs. [25, 27] and later discussion for the operational interpretation). Note the abuse of notation in Eq. (3). The maximization is performed with the respect to $P(a)$ as a function of a. The argument of the function $P(a)$ is used to distinguish it from the other distributions, such as $P(s|y\,b\,k)$ and $P(k, r|y\,a)$. The same representation is used for the label of $H_{P(a)}(K|Y)$. For the sake of simplicity, we will use this notation whenever the meaning is clear from the context.

The operational interpretation of $\mathcal{C}$ is provided by Shannon's source coding theorem and the *wrong code* theorem (Theorem 5.4.3 in Ref. [28]). Given a compression code for k, let us denote by $L(a)$ the expected length of the codeword of k for a given input a and by $\bar{L}[P(a)] \equiv \sum_a P(a)L(a)$ the expected length averaged over a with the distribution $P(a)$. The interpretation of $\mathcal{C}$ is given by the following properties. There is an optimal coding such that the minimal worst-case *expected* length $\max_a L(a)$ is equal to $\mathcal{C}$ up to one additional bit, that is,

$$\mathcal{C} \le \max_a L(a) \le \mathcal{C} + 1. \quad (4)$$

In other words, for the optimal code, Alice needs to send not more than $\mathcal{C} + 1$ on average for every choice of the input a and this bound is strict for some input a up to one bit. Furthermore, the optimal code minimizing $\max_a L(a)$ also minimizes $\bar{L}[P(a)]$ for the worst-case distribution $P(a)$ and the minimum is equal to $\mathcal{C}$ up to one bit. It is worth to stress that the upper bound collapses to $\mathcal{C}$ if block-coding of k is employed, as discussed later. Let us prove Ineqs. (4). Suppose that Alice and Bob employ the optimal code minimizing $\max_a L(a)$ and Alice chooses the input a according to the distribution $P(a)$ maximizing $H_{P(a)}(K|Y)$, denoted by $P_M(a)$. From Shannon's source coding theorem, we have that the expected length of the codeword of k, $\sum_a P_M(a)L(a)$, is not smaller than $\mathcal{C} = H_{P_M(a)}(K|Y)$. Thus,

$$\mathcal{C} \le \max_a L(a). \quad (5)$$

Let us define the distribution

$$P_M(k|y) \equiv \sum_{r,a} P(k,r|y,a)P_M(a). \quad (6)$$

From Shannon's theorem, it is clear that the optimal code minimizing $\bar{L}[P(a)]$ for the worst-case distribution $P(a)$ is the code that minimizes $\bar{L}[P_M(a)]$ and the minimum is equal to $\mathcal{C}$ up to one bit. Now, we show that the optimal code minimizing $\bar{L}[P_M(a)]$ also minimizes $\max_a L(a)$ up to one additional bit. Namely, employing the optimal code for $P_M(a)$, the *wrong code* theorem implies that $\max_a L(a)$ is not bigger than $\mathcal{C}+1$. Indeed, if Alice generates a according to a different distribution $P(a)$ and uses the code that is optimal for $P_M(a)$, the expected codeword length $\bar{L}[P(a)]$ is equal to the Shannon entropy $H_{P(a)}(K|Y)$ plus the relative entropy $D[P(k|y)||P_M(k|y)]$ up to an additional bit [28], where the relative entropy is [28]

$$D[P(k|y)||P_M(k|y)] \equiv -\int dy\rho(y)\sum_k P(k|y)\log\frac{P(k|y)}{P_M(k|y)}. \tag{7}$$

Thus, defining the quantity

$$\begin{aligned}\mathcal{C}[P(a)] \equiv H_{P(a)}(K|Y) + D[P(k|y)||P_M(k|y)] = \\ -\int dy\rho(y)P(k|y)\log P_M(k|y),\end{aligned} \tag{8}$$

we have that

$$\bar{L}[P(a)] \le \mathcal{C}[P(a)] + 1. \tag{9}$$

Let us prove that

$$\mathcal{C}[P(a)] \le \mathcal{C}. \tag{10}$$

Given the distribution $P_\alpha(a) \equiv \alpha P(a) + (1-\alpha)P_M(a)$ with $\alpha \in [0,1]$, we have that

$$\left.\frac{dH_{P_\alpha(a)}(K|Y)}{d\alpha}\right|_{\alpha=0} \le 0, \tag{11}$$

as $P_M(a)$ maximizes the conditional entropy. This equation implies Eq. (10), which can be seen by explicitly performing the derivation of the conditional entropy. Thus, from Eqs. (9,10), we have that $\bar{L}[P(a)] \le \mathcal{C}+1$. Since this inequality holds for every $P(a)$, we have that $L(a) \le \mathcal{C}+1$.

If block-coding of many parallel instances of k is employed, the minimal expected length $L(a)$ per instance turns out to be equal to $\mathcal{C}$ in the asymptotic limit of infinite instances. However, block-coding of k is not the most general compression method in the case of a parallel simulation of NS-boxes. A more general protocol simulating N NS-boxes, which is schematically represented in Fig. 2a, is as follows. The ith box has input a^i and outcome r^i on one side (Alice side), and input b^i and outcome s^i on the other side (Bob side). Alice chooses the input $(a^1,\dots,a^N) \equiv \vec{a}$ and gets the outcome $(r^1,\dots,r^N) \equiv \vec{r}$. Similarly, Bob chooses the input $(b^1,\dots,b^N) \equiv \vec{b}$ and gets the outcome $(s^1,\dots,s^N) \equiv \vec{s}$. Hereafter, we always use a superscript as a label of the NS-boxes. The parallel simulation of N NS-boxes is the same as for a single box, with a, b, r and s replaced by $\vec{a}$, $\vec{b}$, $\vec{r}$ and $\vec{s}$. This more general scheme produces a global variable k with a probability depending on the overall input $\vec{a}$. The protocol exactly simulates the N boxes if

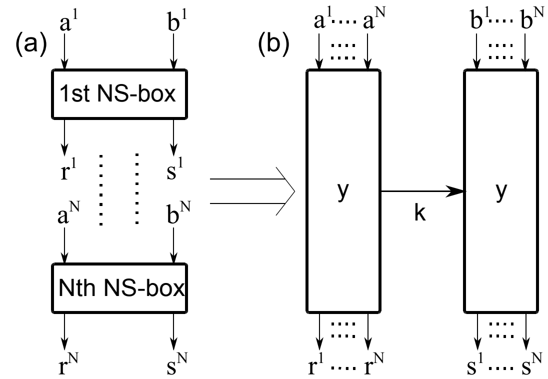

FIG. 2: (a) N identical nonsignaling boxes. On one side, Alices chooses the inputs $(a^1,\dots,a^N)$ and gets the outcomes $(r^1,\dots,r^N)$. On the other side, Bob chooses the inputs $(b^1,\dots,b^N)$ and gets the outcomes $(s^1,\dots,s^N)$. (b) Simulation of the N nonsignaling boxes through a shared stochastic variable y and the communication of the variable k.

$$\sum_k\int dyP(\vec{s}|y\,\vec{b}\,k)P(k,\vec{r}|y\,\vec{a})\rho(y) = \prod_i P(r^i,s^i|a^i,b^i). \tag{12}$$

Each parallelized protocol has N as a free parameter. The asymptotic communication cost of the protocol is defined as $\lim_{N\to\infty}\mathcal{C}^{par}/N \equiv \mathcal{C}^{asym}$, where $\mathcal{C}^{par}$ is the communication cost of the parallelized simulation. In this case, the maximization in Eq. (3) is performed over the space of joint input distributions $P(a_1\dots a_N)$. We define the *nonlocal capacity* of the NS-box as the minimum of $\mathcal{C}^{asym}$ among the parallelized protocols. The nonlocal capacity is denoted by $\mathcal{C}_{nl}^{asym}$.

III. COMPUTATION OF NONLOCAL CAPACITY AS A CONVEX OPTIMIZATION PROBLEM

Our task is to reduce the computation of $\mathcal{C}_{nl}^{asym}$ to the minimization of a functional over a suitable space of distributions. Let us define this space.

Definition 1. Given a nonsignaling box with conditional probability $P(r,s|a,b)$, the set $\mathcal{V}$ contains any conditional probability $\rho(r\,\mathbf{s}|a)$ over r and the sequence $\mathbf{s}=(s_1,\dots,s_M)$ whose marginal distribution of r and the m-th variable is the distribution $P(r,s|a,b=m)$. In other words, the set $\mathcal{V}$ contains any $\rho(r\,\mathbf{s}|a)$ satisfying the constraints

$$\sum_{\mathbf{s},s_m=s}\rho(r\,\mathbf{s}|a) = P(r,s|a,b=m), \tag{13}$$

where the summation is over every index in $\mathbf{s}$ except the m-th one, which is set equal to s. The subscript "$\mathbf{s}, s_m =$

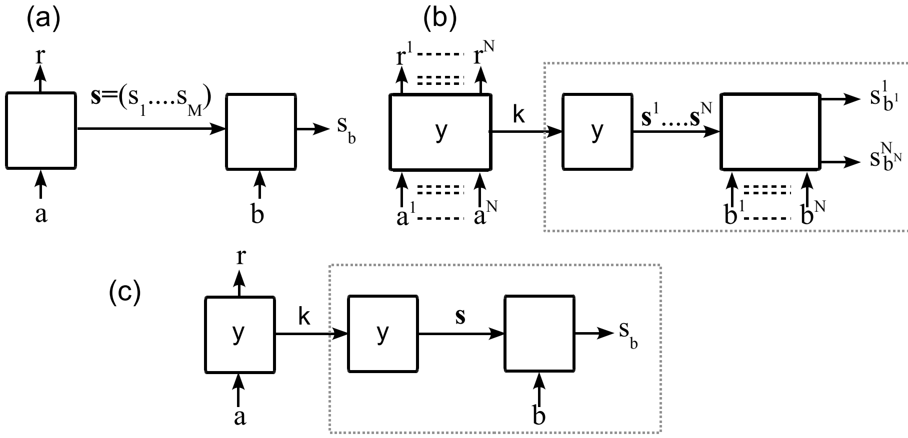

FIG. 3: (a) Master protocol using a HV-box $\rho(r,\mathbf{s}|a)\in\mathcal{V}$ for simulating a NS-box. Alice sends $\mathbf{s}$, generated according to $\rho(r,\mathbf{s}|a)\in\mathcal{V}$. Bob generates the outcome s_b. (b) Child protocol using the simulation of N HV-boxes by employing shared randomness and communication. First, N identical HV-boxes $\rho(r^1,\mathbf{s}^1|a^1),\dots,\rho(r^N,\mathbf{s}^N|a^N)\in\mathcal{V}$ are simulated as follows. Alice, who chooses $a^1,\dots,a^N$ and generates $r^1,\dots,r^N$, sends k, enabling Bob to generate the variables $\mathbf{s}^1,\dots,\mathbf{s}^N$ of N instances according to the distributions $\rho(r^i,\mathbf{s}^i|a^i)\in\mathcal{V}$. The two parties share the random variable y. Finally, Bob generates the outcomes $s^1_{b^1},\dots,s^N_{b^N}$, as done in the master protocol for a single instance. (c) One-shot child protocol. A single HV-box is simulated.

s" means that the summation is done over $\vec{s}$ with the constraint that the component s_m is taken equal to s, that is,

$$\sum_{\mathbf{s},s_m=s}=\sum_{\mathbf{s}}\delta_{s_m,s}, \tag{14}$$

$\delta_{a,b}$ being the Kronecker delta.

The set $\mathcal{V}$ is surely non-empty. Indeed, the distribution $\rho(r\,\mathbf{s}|a)=P(s_1|r\,a\,b=1)\times\dots\times P(s_M|r\,a\,b=M)P(r|a)$ is an element of $\mathcal{V}$. Note that $\rho(r\,\mathbf{s}|a)$ can be defined only if the first nonsignaling condition (1) is satisfied. The conditional probability $\rho(r\,\mathbf{s}|a)$ defines a new box with a single input, a. We call this box 'HV-box', where HV stands for 'hidden variable'. Indeed, this box gives simultaneously the outcomes for every query b of Bob, whereas this information is partially hidden in a query of the original NS-box.

There is a trivial protocol that simulates a NS-box through its HV-box. Using the same terminology introduced in Ref. [25] in the context of channels, we introduce the following protocol (Fig. 3a) that simulates a NS-box through one of its HV-boxes.
Master protocol. Alice generates the outcome r and the array $\mathbf{s}$ according to a conditional probability $\rho(r\,\mathbf{s}|a)\in\mathcal{V}$. Then, she sends $\mathbf{s}$ to Bob. Finally, Bob chooses the input b and gives the outcome $s=s_b$.

It is obvious that r and s are generated according to the conditional probability $P(r,s|a,b)$.

Through the procedure discussed in Ref. [26] and used in Ref. [25] for quantum channels, we show that it is possible to turn the master protocol into a child protocol (Fig. 3b) for parallel simulations whose asymptotic communication cost is the capacity of the channel $a\to\mathbf{s}$ associated to the conditional probability $\rho(\mathbf{s}|a)\equiv\sum_r\rho(r\,\mathbf{s}|a)$. Let us recall that a channel $x_1\to x_2$ is identified by a conditional probability distribution $\rho(x_2|x_1)$ and its capacity is the maximum of the mutual information between x_1 and x_2 over the space of probability distributions $\rho(x_1)$ [28]. Let us denote by $C(x_1\to x_2)$ the capacity of the channel $x_1\to x_2$. The procedure in Ref. [26] is based on the Reverse Shannon theorem [31]. Using the single-shot version of the reverse Shannon theorem [29], we also show that there is a single-shot simulation of a NS-box (Fig. 3c), with associated HV-box $\rho(r,\mathbf{s}|a)$, whose communication cost is equal to $C(a\to\mathbf{s})$ plus an additional term scaling as $\log C(a\to\mathbf{s})$.

Let us first prove the following.
Lemma 1. Multiple instances of identical HV-boxes $\rho(r,\mathbf{s}|a)$ can be simulated in parallel through local randomness and communication with asymptotic communication cost $\mathcal{C}^{asym}$ equal to the capacity of the channel $\rho(\mathbf{s}|a)\equiv\sum_r\rho(r\,\mathbf{s}|a)$. That is,

$$\mathcal{C}^{asym}=C(a\to\mathbf{s}) \tag{15}$$

The communication is established from Alice to Bob. Alice chooses an input a^i and gets an outcome r^i in each instance i, whereas Bob gets the outcomes $\mathbf{s}^i$. The two parties can use shared random variables. Furthermore, there is a single-shot simulation of a HV-box with communication cost $\mathcal{C}$ such that

$$\mathcal{C}^{asym} \le \mathcal{C} \le \mathcal{C}^{asym} + 2\log_2[\mathcal{C}^{asym}+1] + 2\log_2 e. \quad (16)$$

Proof. The simulation is as follows. The ith instance has input a^i and outcome r^i on Alice's side, and outcome $\mathbf{s}^i$ on Bob's side. The outcomes are generated with conditional probability $\rho(r^i, \mathbf{s}^i|a^i)$. Alice chooses the inputs $a^1,\dots,a^N$. Then she sends Bob an amount of information that allows Bob to generate the variables $\mathbf{s}^1,\dots,\mathbf{s}^N$ according to the conditional probability $\rho(\mathbf{s}^i|a^i)$. The reverse Shannon theorem [31] states that there is a protocol for this task with asymptotic communication cost equal to the capacity of the channel $\rho(\mathbf{s}|a)$, provided that Alice and Bob share some stochastic variable, say χ. It is always possible to have a deterministic protocol, so that the outcomes $\mathbf{s}^1,\dots,\mathbf{s}^N$ are uniquely determined by χ and the communicated information. Since χ is shared with Alice, Alice knows Bob's outcomes. Thus, Alice generates her outcomes $r^1,\dots,r^N$ according to the conditional probability $\rho(r^i|a^i\,\mathbf{s}^i) \equiv \rho(r^i\,\mathbf{s}^i|a^i)/\rho(\mathbf{s}^i|a^i)$. The overall set of outcomes is generated according to the joint distribution $\rho(r^i, \mathbf{s}^i|a^i)$. The last statement of the lemma has a similar proof and uses the result in Ref. [29]. The single-shot version of the reverse Shannon theorem proved in Ref. [29] states that a single-shot simulation of the channel $a \to \mathbf{s}$ can be performed with a communication cost $\mathcal{C}$ satisfying the inequalities (16). □

Lemma 1 directly implies the following.

Lemma 2. Identical NS-boxes can be simulated in parallel with an asymptotic communication cost $\mathcal{C}^{asym}$ equal to the capacity of the channel $\rho(\mathbf{s}|a) \equiv \sum_r \rho(r\,\mathbf{s}|a)$, where $\rho(r\,\mathbf{s}|a) \in \mathcal{V}$ is an associated HV-box. The parallel protocol is obtained by using a parallel simulation of HV-boxes that employs shared randomness and communication. The overall protocol, called *child protocol*, is represented in Fig. 3b. Furthermore, there is a single-shot simulation of a NS-box with communication cost $\mathcal{C}$ satisfying the inequalities (16) (Fig. 3c).

Proof. This is a trivial consequence of Lemma 1. Indeed, a NS-box $P(r,s|a,b)$ can be simulated by a master protocol through an associated HV-box $\rho(r,\mathbf{s}|a)$, which can be simulated with asymptotic communication cost equal to the capacity of the channel $\rho(\mathbf{s}|a)$ and with single-shot communication cost $\mathcal{C}$ satisfying constraints (16). □

Let us define the quantity

$$\mathcal{D} \equiv \min_{\rho(r,\mathbf{s}|a)\in\mathcal{V}} \max_{P(a)} I(\mathbf{S};A) \equiv \min_{\rho(r,\mathbf{s}|a)\in\mathcal{V}} C(a\to\mathbf{s}) \quad (17)$$

associated to a NS-box $P(r,s|a,b)$, where $I(\mathbf{S};A)$ is the mutual information between the stochastic variables $\mathbf{s}$ and a and $C(a\to\mathbf{s})$ the capacity of the channel $\rho(\mathbf{s}|a) \equiv \sum_r \rho(r,\mathbf{s}|a)$. Clearly, $\mathcal{D}$ is the minimum of the capacity of the channels $a\to\mathbf{s}$ that are the marginals of the channels $a\to r,\mathbf{s}$ in the set $\mathcal{V}$. Lemma 2 implies that the nonlocal capacity of the NS-box $P(r,s|a,b)$ satisfies the inequality

$$\mathcal{C}_{nl}^{asym} \le \mathcal{D}. \quad (18)$$

The next main task is to prove that the optimal parallel protocol simulating identical NS-boxes is given by a child protocol (schematized in Fig. 3b) employing a simulation of parallel HV-boxes. The proof is a consequence of the data-processing inequality [28], which implies that $\mathcal{C}_{nl}^{asym} \ge \mathcal{D}$, and therefore

$$\mathcal{C}_{nl}^{asym} = \mathcal{D}. \quad (19)$$

Let us first consider a similar proof for the single-shot case, which is less intricate and elucidates the key ideas used in the proof of the main theorem. Namely, we show that $\mathcal{C}_{nl} \ge \mathcal{D}$.

Theorem 1. The communication complexity $\mathcal{C}_{nl}$ of a NS-box $P(r,s|a,b)$ satisfies the bounds

$$\mathcal{D} \le \mathcal{C}_{nl} \le \mathcal{D} + 2\log_2(\mathcal{D}+1) + 2\log_2 e. \quad (20)$$

In few words, the procedure used in the proof of the Theorem 1 is as follows. Given an optimal protocol with communication cost $\mathcal{C} = \mathcal{C}_{nl}$, we build a HV-box such that the associated capacity $C(a\to\mathbf{s})$ is not greater than $\mathcal{C}$. This and the definition of $\mathcal{D}$ imply that $\mathcal{C}_{nl} \ge \mathcal{D}$.

Proof. The second inequality is consequence of Lemma 2. Let us prove the first inequality. Let $P(s|y\,b\,k)$, $P(k,r|y\,a)$ and $\rho(y)$ be the probability distributions defining the optimal single-shot protocol. Thus, the associated communication cost $\mathcal{C}$ is equal to the communication complexity $\mathcal{C}_{nl}$. Now, let us show that there is a channel $\rho(r\,\mathbf{s}|a) \in \mathcal{V}$ such that the capacity of the reduced channel $\rho(\mathbf{s}|a)$ is not greater than $\mathcal{C}_{nl}$. The channel is

$$\rho(r\,\mathbf{s}|a) \equiv \int dy \sum_k \left[\prod_b P(s_b|y\,b\,k)\right] P(k,r|y\,a)\rho(y)$$

By construction, the distribution $\rho(r\,\mathbf{s}|a)$ is in the set $\mathcal{V}$. Indeed,

$$\sum_{\mathbf{s},s_b=s} \rho(r\,\mathbf{s}|a) = \int dy \sum_{k,y} P(s|y\,b\,k)P(k,r|y\,a)\rho(y), \quad (21)$$

which is reduced to Eq. (13) in view of Eq. (2). The stochastic variables a, $\mathbf{s}$ and k satisfy the Markov chain $a \xrightarrow{y} k \xrightarrow{y} \mathbf{s}$, where the label above that arrows denotes the shared variable y. From the data-processing inequality [28], we have that

$$\max_{P(a)} I(\mathbf{S};A) \le \max_{P(a)} I(K;A|Y),$$

where $I(K;A|Y)$ denotes the conditional mutual information between k and a given y. The mutual information between two variables is always less than or equal to the entropy of each variable. Thus, $\max_{P(a)} I(K;A|Y) \le \max_{P(a)} H_{P(a)}(K|Y)$ and, from the definition of communication cost, we have that

$$\max_{P(a)} I(\mathbf{S};A) \le \max_{P(a)} I(K;A|Y) \le \max_{P(a)} H_{P(a)}(K|Y) = \mathcal{C} = \mathcal{C}_{nl}. \quad (22)$$

Finally, from the definition of $\mathcal{D}$ and the fact that $\rho(r\,\mathbf{s}|a) \in \mathcal{V}$, we have that

$$\mathcal{D} \le \mathcal{C} = \mathcal{C}_{nl}.$$

□

The inequalities (20) provide tight lower and upper bounds on $\mathcal{C}_{nl}$ and establish that the single-shot communication complexity is equal to $\mathcal{D}$ up to an additional term scaling as the logarithm of $\mathcal{D}$. As stated in the next theorem, the additional cost disappears in the case of parallel simulations and the strict Eq. (19) holds. The proof is more intricate and needs some further final efforts. As we said, a parallel protocol simulating N NS-boxes is described by the conditional probabilities $P(\vec{s}|y\,\vec{b}\,k)$, $P(k,\vec{r}|y\,\vec{a})$ and $\rho(y)$. The components of the sequences $\vec{r} = (r^1, \dots, r^N)$, $\vec{s} = (s^1, \dots, s^N)$, $\vec{a} = (a^1, \dots, a^N)$ and $\vec{b} = (b^1, \dots, b^N)$ are the inputs and outcomes of each NS-box. Adapting the construction used in the proof of theorem 1 to the parallel case, we build a multivariate HV-box $\rho(\vec{r}\,\vec{\mathbf{s}}|\vec{a}) = \rho(r^1 \dots r^N\,\mathbf{s}^1 \dots \mathbf{s}^N|a^1 \dots a^N)$ associated to the overall collection of NS-boxes, where $\mathbf{s}^i = (s^i_1, \dots, s^i_M)$. The multivariate distribution is built so that the marginal distribution of r^i and s^i_b given a^i is equal to $P(r^i, s^i_b|a^i, b)$, that is,

$$\sum_{\vec{\mathbf{s}},\vec{r},r^i=r,s^i_b=s} \rho(\vec{r}\,\vec{\mathbf{s}}|\vec{a}) \equiv \rho(r^i = r, s^i_b = s|\vec{a}) = P(r, s|a^i, b), \tag{23}$$

where $\rho(r^i, s^i_b|\vec{a})$ is the conditional marginal distribution of the variables r^i and s^i_b given $\vec{a}$. Let us denote by $\mathcal{V}^{par}$ the set of multivariate HV-boxes satisfying this property on the marginals. The main key ingredient used in the proof of the next theorem is the equality

$$N \times \min_{\rho(r,\mathbf{s}|a)\in\mathcal{V}} C(a \to \mathbf{s}) = \min_{\rho(\vec{r},\vec{\mathbf{s}}|\vec{a})\in\mathcal{V}^{par}} C(\vec{a} \to \vec{\mathbf{s}}). \tag{24}$$

In particular, if $\rho(r, \mathbf{s}|a)$ minimizes $C(a \to \mathbf{s})$ in the set $\mathcal{V}$, then the factorized distribution

$$\rho(\vec{r}, \vec{\mathbf{s}}|\vec{a}) = \prod_i \rho(r^i, \mathbf{s}^i|a^i), \tag{25}$$

minimizes $C(\vec{a} \to \vec{\mathbf{s}})$ in the set $\mathcal{V}^{par}$. It is clear that

$$N \times \min_{\rho(r,\mathbf{s}|a)\in\mathcal{V}} C(a \to \mathbf{s}) \ge \min_{\rho(\vec{r},\vec{\mathbf{s}}|\vec{a})\in\mathcal{V}^{par}} C(\vec{a} \to \vec{\mathbf{s}}), \tag{26}$$

as the capacity of a factorized channel $\rho(\vec{\mathbf{s}}|\vec{a}) = \prod_{i=1}^N \rho(r^i, \mathbf{s}^i|a^i) \in \mathcal{V}^{par}$ is equal to the sum of the capacity of the channels $\rho(r^i, \mathbf{s}^i|a^i) \in \mathcal{V}$. To prove Eq. (24), it is sufficient to show that the minimum at the right-hand side is attained by a factorized function. The proof is quite technical and is provided in Appendix A.

Theorem 2. The nonlocal capacity $\mathcal{C}^{asym}_{nl}$ of an NS-box $P(r, s|a, b)$ is the minimum of the capacity of the channel $\rho(\mathbf{s}|a)$ over the space $\mathcal{V}$ of associated HV-boxes. That is,

$$\mathcal{C}^{asym}_{nl} = \mathcal{D}. \tag{27}$$

Proof. The inequality $\mathcal{C}^{asym}_{nl} \le \mathcal{D}$ is a consequence of Lemma 2. Let us prove that $\mathcal{C}^{asym}_{nl} \ge \mathcal{D}$. Let $\mathcal{C}^{par}$ be the communication cost of the optimal parallel protocol simulating N NS-boxes. Thus, from the definition of nonlocal capacity, we have that

$$\lim_{N\to\infty} \mathcal{C}^{par}/N = \mathcal{C}^{asym}_{nl}.$$

The protocol is defined by the conditional probabilities $P(k, \vec{r}|y\,\vec{a})$ and $P(\vec{s}|y\,\vec{b}\,k)$ satisfying constraint (12). Through a procedure described in Appendix B, it is possible to build a multivariate HV-box with conditional probability

$$\rho(r^1 \dots r^N\,\mathbf{s}^1 \dots \mathbf{s}^N|a^1 \dots a^N) = \rho(\vec{r}\,\vec{\mathbf{s}}|\vec{a}) \tag{28}$$

over r^i and the sequences $\mathbf{s}^i = (s^i_1 \dots s^i_M)$ so that the following properties are satisfied.

1. The capacity of the channel $\rho(\vec{\mathbf{s}}|\vec{a})$ is smaller than or equal to $\mathcal{C}^{par}$. That is,
$$C(\vec{a} \to \vec{\mathbf{s}}) \le \mathcal{C}^{par}. \tag{29}$$
2. The marginal distribution of the variables r^i and s^i_b is equal to $P(r^i, s^i_b|a^i, b)$, that is, constraints (23) are satisfied.

Ineq. (29) is similar to Ineq. (22). The proof is identical and uses the data-processing inequality [28] (see Appendix B). From Eq. (24), we have that

$$N \times \min_{\rho(r,\mathbf{s}|a)\in\mathcal{V}} C(a \to \mathbf{s}) \le C(\vec{a} \to \vec{\mathbf{s}}) \le \mathcal{C}^{par}. \tag{30}$$

In the limit $N \to \infty$, we obtain

$$\mathcal{D} = \min_{\rho(r,\mathbf{s}|a)\in\mathcal{V}} C(a \to \mathbf{s}) \le \lim_{N\to\infty} \frac{\mathcal{C}^{par}}{N} = \mathcal{C}^{asym}_{nl}. \tag{31}$$

□

This theorem reduces the evaluation of the nonlocal capacity of a NS-box to the computation of the quantity $\mathcal{D}$, defined by Eq. (17) as the minimum of the capacity $C(a \to \mathbf{s})$ over the convex set $\mathcal{V}$. As the mutual information $I(\mathbf{S}; A)$ is convex in $\rho(\mathbf{s}|a)$ and the maximum over a set of convex functions is a convex function [24], the capacity $C(a \to \mathbf{s})$ is convex in $\rho(\mathbf{s}|a)$. Thus, the computation of $\mathcal{D}$ is a convex optimization problem, which is the main advantage of the presented method. Indeed, convexity implies that every local minimum is a global minimum. A different formulation of the problem has been introduced in Ref. [30], but it does not have the form of a convex minimization problem. This makes it harder to find the global minimum.

It is worth to note that the capacity $C(a \to \mathbf{s})$ does not have generally an explicit analytic form and is not necessarily differentiable everywhere. This makes it harder to use standard methods of convex optimization [24]. However, provided that the optimal distribution $P(a)$

maximizing the mutual information $I(\mathbf{S};A)$ is known, the dual form of our optimization problem is a geometric programming problem. This has been shown for the related problem of computing the communication complexity of quantum channels [32, 33]. Geometric programming is an extensively studied class of nonlinear optimization problems and can be solved by robust and very efficient algorithms [34, 35]. A commercial implementation is provided by the MOSEK package (see http://www.mosek.com). Even if $P(a)$ is not known, the minimization

$$\min_{\rho(r,\mathbf{s}|a)} I(\mathbf{S};A) \equiv C_- \tag{32}$$

with any fixed $P(a)$ provides a lower bound on the nonlocal capacity, as a consequence of the minimax theorem [32, 33]. Furthermore, if $P(a)\ne 0$ for every input a, then $\min_{\rho(r,\mathbf{s}|a)} I(\mathbf{S};A)$ is different from zero only and only if the correlations are nonlocal. Indeed, if the correlations are local, C_- is obviously equal to zero for every $P(a)$. Conversely, if there is a $\rho(r,\mathbf{s}|a)$ and a $P(a)\ne 0$ for every a such that $I(\mathbf{S};A)=0$, then $I(\mathbf{S};A)=0$ for every $P(a)$ and $\mathcal{D}$ is equal to zero. Thus, if we are interested to discriminate between local and nonlocal correlations, we can fix $P(a)$, for example by taking a uniform distribution, and solve the minimization (32). A specialized numerical method that is particularly efficient for this problem and computes also the optimal $P(a)$ will be discussed elsewhere. A similar method for computing the communication complexity of quantum channels is discussed in Ref. [36].

As the number of variables defining the probability distribution $\rho(r,\mathbf{s}|a)$ scales exponentially with respect to the number of Bob's measurements, our method displays an exponential computational cost. Thus, it does not provide a better scaling complexity than the computation of the distance from the nonlocal polytope. However, numerical experiments show a speed difference of many decades between the two methods. Our method can solve a problem with 20 measurements in few minutes or even few seconds, whereas the computation of the distance turns out to be very time-demanding even with 6 measurements. This difference is relevant if one needs to test many different experimental configurations even if the number of measurements is relatively small. Furthermore, the dual form of our optimization problem displays very interesting properties, as stressed in Refs. [32, 33]. First, the number of dual variables scales linearly with the size of the input of the problem, that is, with the number of variables defining the conditional probability $P(r,s|a,b)$. Second, although the number of constraints grows exponentially, they are completely independent of $P(r,s|a,b)$. Thus, given a feasible point of the dual constraints, the computation of a lower bound for *every* nonlocal correlation has a linear computational cost. This feature is employed in Ref. [37] to efficiently compute the setting of measurements maximizing the nonlocal capacity and, thus, providing the highest degree of nonlocality.

Finally, it is worth to note that the distribution $\rho(r,\mathbf{s}|a)$ solving the minimization problem is equal to zero for most of the values of r and $\mathbf{s}$. Indeed, the numerical simulations and theoretical arguments show that the support of the distribution grows linearly with the size of the problem. Thus, most of the computational effort is to determine where the distribution is equal to zero. It is an open question if the support can be determined efficiently or even analytically in some relevant cases. This problem is related to the determination of feasible points of the dual constraints. In Ref. [32], we showed an example involving infinite measurements, for which we found analytically a nontrivial feasible point, from which we determined nontrivial lower bounds for the communication complexity.

IV. RELATIONSHIP WITH COMMUNICATION COMPLEXITY OF CHANNELS

There is a relationship between the nonlocal capacity of a NS-box and the communication complexity of a channel in a general probabilistic theory and, under some conditions, the computation of the former can be reduced to the computation of the latter, which requires less sophisticated algorithms [36]. Furthermore, the relationship allows us to directly infer results on the nonlocal capacity from known results on the communication complexity of channels. For example, the analytical solution found in Ref. [25] provides an analitical solution for maximally entangled qubits and measurements associated to planar Bloch vectors. The communication complexity of a channel has been defined in Ref. [25]. The central scenario studied there is the process of state preparation, communication through a channel, and subsequent measurement. This process is described by a conditional probability $P(s|a;b)$, where a and b are inputs chosen by the sender (Alice) and the receiver (Bob), respectively, and s is an outcome obtained by Bob. In a general abstract setting, we will just assume that $P(s|a;b)$ is any conditional probability depending on two spatially separated inputs. We call this object C-box, where C stands for channel. In Ref. [25], a C-box is called game $\mathbf{G}$. The asymptotic communication complexity of a C-box is the minimal asymptotic communication cost of a parallel simulation of many copies of the C-box (See Ref. [25] for details). Let us denote this quantity by $\mathcal{C}_{ch}^{asym}$ (denoted by $\mathcal{C}_{min}^{asym}$ in Ref. [25]).

Corollary 1. The nonlocal capacity $\mathcal{C}_{nl}^{asym}$ of an NS-box $P(r,s|a,b)$ satisfies the inequalities

$$\begin{array}{c}\mathcal{C}_{ch}^{asym}+\min_{a\,r}\log_2 P(r|a)\le \mathcal{C}_{nl}^{asym}\le \mathcal{C}_{ch}^{asym} \\ -\min_a \max_b I(R;S|a\,b),\end{array} \tag{33}$$

where $\mathcal{C}_{ch}^{asym}$ is the asymptotic communication complexity of the C-box $P(s|r\,a;b)\equiv P(r,s|a,b)/P(r|a)$ with Alice's inputs r and a, and Bob's input b.

The first inequality can be proved by using a procedure described in Sec. IIIB of Ref. [38], the second inequality follows from Theorem 2 and the theorem proved in Ref. [25]. The proof of the Corollary is given in Appendix C.

Corollary 2. Let $P(r,s|a,b)$ be a NS-box implemented through a maximally entangled state of two pairs of n qubits (n ebits of entanglement). The two parties perform projective measurements and they share the same set of allowed measurements. Then,

$$\mathcal{C}_{nl}^{asym} = \mathcal{C}_{ch}^{asym} - n, \tag{34}$$

where $\mathcal{C}_{ch}^{asym}$ is the capacity of the associated C-box $P(s|ra;b)$. The C-box can be implemented through a quantum channel with capacity n qubits; the receiver can perform any measurement allowed in the NS-box case, whereas the sender can prepare any state corresponding to the eigestates of the allowed measurements.

Proof. The corollary follows directly from Corollary 1. Indeed, the lower and upper bounds in Ineqs. (33) collapse to the same value, as $P(r|a) = 2^{-n}$ and $I(R;S|a,b=a) = n$.

V. NUMERICAL EXAMPLE

To illustrate the presented method with an example, let us consider the case of two-qubits in the Werner state $\rho_\gamma = \gamma/2(|00\rangle+|11\rangle)(\langle 00|+\langle 11|)+(1-\gamma)\mathbf{1}/4$. What is the critical value of γ, denoted by γ_0, below which the Werner state becomes local? This question is particularly interesting from an experimental point of view, as γ_0 provides the amount of noise that makes a maximally entangled state local. The Werner state admits a local model for $\gamma < 0.659$ [39] and it is nonlocal for $1/\sqrt{2} \le \gamma \le 1$, as the CHSH inequalities are violated. In Ref. [40], Vértesi derived a family of Bell inequalities that are violated for $\gamma > 0.7056$, which is slightly below the bound $1/\sqrt{2}$. This family requires 465 measurement settings on each side. Thus, γ_0 is between about 0.659 and 0.7056. Is it possible to derive a better upper bound on γ_0 with a much smaller set of measurements? To answer this question, we have computed the nonlocal capacity for a number of measurements up to 20 by trying a high number of different settings, such as highly symmetric settings and random configurations. Each computation of the nonlocal capacity requires just few seconds on a standard laptop for the considered maximal set of measurements. We always found local correlations for $\gamma \le 1/\sqrt{2}$. For example, given a set of 13 measurements corresponding to Bloch vectors pointing to the faces, edges and vertices of a cube (2 opposite vectors for each measurement), we find that $\mathcal{C}_{nl}^{asym}$ is equal to zero for $\gamma \le 1/\sqrt{2}$ and is well-approximated by the analytic expression $9/4(\gamma - 1/\sqrt{2})^2$ for $\gamma \in [1/\sqrt{2}, 1]$, with a maximum error lower than 3% for $\gamma = 1$. Our calculations suggest that, for a reasonable number of measurements, the CHSH inequalities are the best solution for testing nonlocality of a singlet in presence of noise. In a recent paper [32], we derived the dual optimization problem for the computation of the communication complexity of quantum channels and we used it to derive an analytical lower bound on the communication complexity in the case of infinite measurements. A similar dual problem can be used to derive a lower bound on the nonlocal capacity of Werner states and, thus, an upper bound on γ_0. We are currently studying the possibility of analytically deriving the exact value of the nonlocal capacity in the case of infinite measurements by using the dual formulation.

VI. CONCLUSIONS

In conclusion, we have presented a method for evaluating the nonlocal capacity of correlations, and provided a tight lower and upper bound on the communication complexity in the single-shot case. The introduced measure of nonlocality can be used as an alternative to Bell's inequalities for testing if some given theoretical or experimental data display nonlocal correlations. Furthermore, this measure provides an upper bound to the power of nonlocal correlations, as an information-theoretic resource, in terms of classical communication. In a subsequent work, we will present an efficient numerical method for evaluating the nonlocal capacity and will derive a dual optimization problem which can help to solve the open problem concerning the range of γ where a Werner state is nonlocal. A similar dual problem was recently derived in Ref. [32] for the optimization problem introduced in Ref. [25].

Acknowledgments.

This work is supported by the Swiss National Science Foundation, the NCCR QSIT, the COST action on Fundamental Problems in Quantum Physics and Hasler Foundation under the project number 14030 "Information-Theoretic Analysis of Experimental Qudit Correlations". The authors wish to thank Arne Hansen for useful comments and the careful reading of the manuscript.

[1] J. Bell, Physics **1**, 195 (1964).
[2] R. Cleve and H. Buhrman, Phys. Rev. A **56**, 1201 (1997).
[3] H. Buhrman, R. Cleve, S. Massar, and R. de Wolf, Rev. Mod. Phys. **82**, 665 (2010).
[4] G. Brassard, R. Cleve, and A. Tapp, Phys. Rev. Lett. **83**, 1874 (1999).

[5] W. van Dam. *Nonlocality and Communication Complexity. PhD thesis*, University of Oxford, Department of Physics (2000); W. van Dam, arXiv:quant-ph/0501159.
[6] J. Barrett, L. Hardy, and A. Kent, Phys. Rev. Lett. **95**, 010503 (2005).
[7] A. Ac n, N. Gisin, and L. Masanes, Phys. Rev. Lett. **97**, 120405 (2006)
[8] V. Scarani et al., Phys. Rev. A **74**, 042339 (2006).
[9] A. Ac n, S. Massar, and S. Pironio, New J. Phys. **8**, 126 (2006).
[10] A. Aci n at al., Phys. Rev. Lett. **98**, 230501 (2007).
[11] Ll. Masanes, R. Renner, M. Christandl, A. Winter, J. Barrett, IEEE Trans. Inf. Theory, **60**, 4973 (2014).
[12] E. Hänggi, R. Renner, S. Wolf, Theor. Comp. Sci. **486**, 27 (2013).
[13] R. Colbeck and R. Renner, Nature Physics **8**, 450 (2012).
[14] R. Gallego, Ll. Masanes, G. de la Torre, C. Dhara, L. Aolita, A. Acin, Nature Communications **4**, 2654 (2013).
[15] T. Maudlin, *Proceedings of the Biennial Meeting of the Philosophy of Science Association* (D. Hull, M. Forbes, and K. Okruhlik. Philosophy of Science Association, East Lansing, MI, 1992), vol. 1, pp. 404-417.
[16] M. Steiner, Phys. Lett. A **270**, 239 (2000).
[17] N. Gisin and B. Gisin, Phys. Lett. A **260**, 323 (1999).
[18] C. Branciard, N. Gisin, Phys. Rev. Lett. **107**, 020401 (2011).
[19] S. Pironio, Phys. Rev. A **68**, 062102 (2003).
[20] J. Barrett, S. Pironio, Phys. Rev. Lett. **95**, 140401 (2005).
[21] N. Brunner, N. Gisin, V. Scarani, New J. Phys. **7**, 88 (2005).
[22] F. Dupuis, N. Gisin, A. Hasidim, A. Méthot, H. Pilpel, J. Math. Phys. **48**, 082107 (2007).
[23] M. Forster, S. Wolf, Phys. Rev. A **84**, 042112 (2011).
[24] S. Boyd, L. Vandenberghe *Convex Optimization* (Cambridge University Press, Cambridge, 2004).
[25] A. Montina, M. Pfaffhauser, S. Wolf, Phys. Rev. Lett. **111**, 160502 (2013).
[26] A. Montina, Phys. Rev. Lett. **109**, 110501 (2012).
[27] A. Montina, Phys. Rev. A **87**, 042331 (2013).
[28] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 2006).
[29] P. Harsha, R. Jain, D. McAllester, J. Radhakrishnan, IEEE Trans. Inf. Theory **56**, 438 (2010).
[30] M. H. Yassaee, A. Gohari, M. R. Aref, arXiv:1203.3217.
[31] C. H. Bennett, P. Shor, J. Smolin, and A. V. Thapliyal, IEEE Trans. Inf. Theory **48**, 2637 (2002).
[32] A. Montina, S. Wolf, IEEE Int. Symp. Inform. Theory (ISIT), 1484 (2014).
[33] A. Montina, S. Wolf, Phys. Rev. A, **90**, 012309 (2014).
[34] S. Boyd, S.-J. Kim, L. Vandenberghe, A. Hassibi, Optim. Eng. **8**, 67 (2007).
[35] M. Chiang, Found. Trends Commun. Inf. Theory **2**, 1 (2005).
[36] A. Hansen, A. Montina, S. Wolf, to be published.
[37] A. Montina, S. Schwarz, A. Stefanov, S. Wolf, to be published.
[38] A. Montina, Phys. Rev. A **84**, 042307 (2011).
[39] A. Acin, N. Gisin, B. Toner, Phys. Rev. A **73**, 062105 (2006).
[40] T. Vértesi, Phys. Rev. A **78**, 032112 (2008).

Appendix A: Minimizing the capacity with constraints on the marginals

Let us consider the set $\mathcal{V}^{par}$ of channels $a_1,\dots,a_N \to (x_1,y_1,\dots,x_N,y_N)$ with conditional probability $\rho(\vec x,\vec y|\vec a)$ satisfying the constraints

$$\sum_{\vec x,\vec y} B_{k,i}(x_i,y_i)\rho(\vec x,\vec y|\vec a)=A_{k,i}(a_i)\ \ \forall k=1,\dots,K,\ \ i=1,\dots,N, \tag{A1}$$

where $B_{k,i}$ and $A_{k,i}$ are real numbers. The constraints (23) take this form, once we identify $\mathbf{\vec r}$, $\mathbf{\vec s}$ and k with $\vec x$, $\vec y$ and (r,s,b), respectively.

Note that the constraints can be rewritten in the form

$$\sum_{x_i,y_i} B_{k,i}(x_i,y_i)\rho(x_i,y_i|\vec a)=A_{k,i}(a_i)\ \ \forall k=1,\dots,K,\ \ i=1,\dots,N, \tag{A2}$$

where $\rho(x_i,y_i|\vec a)$ are the marginal conditional distributions of the variables x_i and y_i. Furthermore, whereas $\rho(x_i,y_i|\vec a)$ can depend on every element of the sequence $a_i,\dots,a_N$, the sum at the left-hand side of Eq. (A2) depends only on a_i, as the right-hand side depends only on a_i.

Theorem. There is a factorized distribution $\rho_f(\vec x,\vec y|\vec a)=\prod_i \rho(x_i,y_i|a_i)$ such that the capacity of the reduced channel $\vec a\to\vec y$ is minimal in $\mathcal{V}^{par}$.

Proof. Let $\rho_m(\vec x,\vec y|\vec a)$ be a channel that minimizes the capacity of the reduced channel $\rho_m(\vec y|\vec a)$ under the constraints (A2). We denote the channel $\rho_m(\vec y|\vec a)$ by M. Now, we build another distribution that is factorized and minimal in the set $\mathcal{V}^{par}$ of constrained channels. Let us take a factorized distribution $\rho(\vec a)=\prod_i \rho(a_i)$ for the variable $\vec a$ and introduce the probability distribution $\rho(\vec x,\vec y,\vec a)$. Note the abuse of notation. We should introduce some index, such as $\rho_i(a_i)$, for distinguishing different distributions. For the sake of simplicity, we will distinguish two distributions from their argument, so that $\rho(a_1)$ and $\rho(a_2)$ are not meant to be the same function. Let $\rho(x_i,y_i,a_i)$ be the marginal distributions of the variables x_i, y_i and a_i. The conditional probability distributions $\rho(x_i,y_i|a_i)$ define the channels $a_i\to x_i,y_i$. By construction, the multivariate channel

$$\rho_f(\vec x,\vec y|\vec a)\equiv\prod_i \rho(x_i,y_i|a_i) \tag{A3}$$

satisfies the constraints (A2), as the right-hand side of the constraints only depend on one component a_i. Let the reduced factorized channel

$$\rho_f(\vec{y}|\vec{a}) \equiv \prod_i \rho(y_i|a_i) \tag{A4}$$

be denoted by F.

Let us choose the distributions $\rho(a_1),\dots,\rho(a_N)$ so that the mutual information $I(y_i;a_i)$ is maximal in $\rho(a_i)$ for every i. Note that each conditional probability $\rho(x_i|a_i)$ depends on the full set of distributions $\rho(a_1),\dots,\rho(a_N)$, apart from $\rho(a_i)$. Thus, the maximizations of the functions $I(y_i;a_i)$ are not independent optimization problems. Anyway, the overall problem has a solution and this choice of $\rho(a_1),\dots,\rho(a_N)$ is always possible.

Thus, by definition of channel capacity [28], $I(y_i;a_i)$ is the capacity of the reduced channel $a_i \rightarrow y_i$, say $C(a_i \rightarrow y_i)$. Furthermore, the capacity of the channel F, say $C(F)$, is

$$C(F) = \sum_i C(a_i \rightarrow y_i). \tag{A5}$$

We denote the mutual information between the stochastic variables $\vec{a}$ and $\vec{y}$ with conditional probability $\rho_m(\vec{y}|\vec{a})$ and marginal distribution $\rho(\vec{a}) = \prod_{i=1}^N \rho(a_i)$ by $I_m(\vec{y};\vec{a})$. Using the chain rule for the mutual information [28]

$$I(x,y;z) = I(x;z) + I(y;z|x) \tag{A6}$$

and the fact that $I(a_i;a_{i'}) = 0$ for $i \ne i'$, let us prove that

$$I_m(\vec{y};\vec{a}) \ge \sum_i I(y_i;a_i) = \sum_i C(a_i \rightarrow y_i) = C(F). \tag{A7}$$

Intuitively, this inequality says that the variable $\vec{y}$ contains less information about $\vec{a}$ if the conditional probability $\rho_m(\vec{y}|\vec{a})$ is replaced by the factorized form (A4), provided that the marginal distribution $\rho(\vec{a})$ is factorized. Indeed, the factorized form $\rho_f(\vec{y}|\vec{a})$ looses the correlations among the components of $\vec{y}$, and these correlations can contain extra-information about $\vec{a}$. Conversely, if $\rho(\vec{a})$ is not factorized, the factorized form $\rho_f(\vec{y}|\vec{a})$ can increase the information about $\vec{a}$ for majority vote reasons. Let us prove this inequality for $N = 2$. The general case can be proved recursively.

$$\begin{gathered} I_m(y_1\,y_2;a_1\,a_2) = I_m(y_1\,y_2;a_1) + I_m(y_1\,y_2;a_2|a_1) = I_m(y_1\,y_2;a_1) + I_m(y_1\,y_2,a_1;a_2) - I_m(a_1;a_2) = \\ I_m(y_1\,y_2;a_1) + I_m(y_1\,y_2,a_1;a_2) = I_m(y_1;a_1) + I_m(y_2;a_1|y_1) + I_m(y_2;a_2) + I_m(y_1\,a_1;a_2|y_2) \ge \\ I(y_1;a_1) + I(y_2;a_2). \end{gathered} \tag{A8}$$

As the capacity $C(M)$ of the channel M is the maximum of the mutual information $I(\vec{y};\vec{a})$ with respect to the whole space of distributions $\rho(\vec{a})$, Ineq. (A7) implies the inequality

$$C(M) \ge C(F). \tag{A9}$$

Since $C(M)$ is minimal under the constraints (A1) and $\rho_f(\vec{x},\vec{y}|\vec{a})$ satisfies the constraints, the factorized channel $\rho_f(\vec{x},\vec{y}|\vec{a})$ also minimizes the capacity of the reduced channel $\vec{a} \rightarrow \vec{y}$ in the set $\mathcal{V}^{par}$. $\square$

Appendix B: Proof of Properties 1 and 2 in Theorem 2

The protocol is defined by the conditional probabilities $P(k,\vec{r}|y,\vec{a})$, $P(\vec{s}|y\,\vec{b}\,k)$ and $\rho(y)$ satisfying the constraints (12).

First, we note that Bob does not need to generate the outcome of the i-th NS-box with a probability depending on the inputs of the other boxes. If this independence property is not satisfied by $P(\vec{s}|y\,\vec{b}\,k)$, it is always possible to replace Bob's protocol with one satisfying the property, without affecting Alice's protocol and, thus, the communication cost. Thus, we can safely assume that $P(\vec{s}|y\,\vec{b}\,k)$ is factorized as follows

$$P(s^1\dots s^N|y,b^1\dots b^N,k) = \prod_i P^i(s^i|y,b^i,k). \tag{B1}$$

Let us introduce the conditional probability

$$P^i(s_1,s_2,\dots,s_M|y,k) \equiv \prod_{b=1}^M P^i(s_b|y,b,k), \tag{B2}$$

We will concisely denote $P^i(s_1, s_2, \dots, s_M|y,k)$ by $P^i(\mathbf{s}|y,k)$. We use the probabilities $P^i(\mathbf{s}|y,k)$ to build the conditional probabilities

$$P(\mathbf{s}^1, \dots, \mathbf{s}^N|y,k) = \prod_i P^i(\mathbf{s}^i|y,k). \tag{B3}$$

Finally, from this distribution and $P(k,\vec{r}|y,\vec{a})$, we build the conditional probability

$$\rho(\vec{r}, \mathbf{s}^1, \dots, \mathbf{s}^N|\vec{a}) = \sum_k \int dy \rho(y) P(\mathbf{s}^1, \dots, \mathbf{s}^N|y,k) P(k,\vec{r}|y,\vec{a}). \tag{B4}$$

As seen in the proof of Theorem 1, we obtain from the data processing inequality and Eq. (B4) that the capacity $C(\vec{a} \to \vec{\mathbf{s}})$ of the channel $\vec{a} \to (\mathbf{s}^1, \dots, \mathbf{s}^N)$ satisfies the inequality

$$C(\vec{a} \to \vec{\mathbf{s}}) \le \max_{P(\vec{a})} I(K;A|Y) \le \max_{P(\vec{a})} H_{P(a)}(K|Y) = \mathcal{C}^{par}, \tag{B5}$$

which is Property 1.

By construction and Eq. (12), we also have that the constraints (23) are satisfied, i.e. Property 2.

Appendix C: Proof of Corollary 1.

The first inequality can be proved by using a procedure described in Sec. IIIB of Ref. [38], where we showed that a protocol for simulating a maximally entangled state of n qubits can be used to simulate the communication of n qubits with an additional cost of n classical bits. More generally, the additional cost is not more than $-\min_{r\,a} \log_2 P(r|a)$. Let us prove it. Any protocol simulating N NS-boxes is deterministic or can be made deterministic by introducing some additional random variables shared by Alice and Bob. This means that the two parties have a common list noise realizations, say $y_1, y_2, \dots$. In the NS-box simulation, Alice chooses a, but she cannot choose her outcome r, which is determined by a and the noise. Conversely, in the C-box simulation, Alice chooses both a and r such that the conditional probability $P(s|r\,a;b)$ of Bob's outcome in the C-box and NS-box simulations are identical. The C-box can be simulated as follows. Alice starts reading the noise list from the first element and stops at the value y_i that generates the outcome r chosen by her. Then, she uses the communication procedure of the NS-box protocol and, furthermore, she sends the index i. Finally, Bob uses the noise realization y_i in the NS-box protocol. The additional cost is not bigger than $-\min_{a,r} \log_2 P(r|a)$.

Now, let us prove the second inequality. Given the NS-box $P(r,s|a,b)$, let $\mathcal{V}$ and $\mathcal{V}'$ be the space of associated HV-boxes $\rho(r\mathbf{s}|a)$ and distributions $\rho(\mathbf{s}|ra) \equiv \rho(r\mathbf{s}|a)/\rho(r|a)$, respectively. The chain rule [28]

$$I(\mathbf{S};A) = I(\mathbf{S};R,A) - I(\mathbf{S};R|A) \tag{C1}$$

for the mutual information, the definition of a HV-box and the data-processing inequality imply that

$$I(\mathbf{S};A) \le I(\mathbf{S};R,A) - \min_a \max_b I(R;S|ab). \tag{C2}$$

As the marginal distribution of r given a is the same for every distribution $\rho(r\mathbf{s}|a)$ in $\mathcal{V}$, the maximization in Eq. (27) giving the nonlocal capacity can be performed over the space $\rho(\mathbf{s}|ra) \in \mathcal{V}'$. Thus, we have that

$$\mathcal{C}_{nl}^{asym} \le \min_{\rho(\mathbf{s}|ra)\in\mathcal{V}'} \max_{P(a)} I(\mathbf{S};R,A) - \min_a \max_b I(R;S|a\,b). \tag{C3}$$

We also have that

$$\min_{\rho(\mathbf{s}|ra)\in\mathcal{V}'} \max_{P(a)} I(\mathbf{S};R,A) \le \min_{\rho(\mathbf{s}|ra)\in\mathcal{V}'} \max_{P(r,a)} I(\mathbf{S};R,A). \tag{C4}$$

In Ref. [25], we showed that the right-hand side is equal to $\mathcal{C}_{ch}^{asym}$. Thus, the inequalities (C3, C4) imply the second inequality.