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: December 24, 2013)

Quantum nonlocality concerns correlations among spatially separated systems that cannot be classically explained without post-measurement 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, 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.

The outcomes of measurements performed on spatially separate entangled systems can display nonlocal correlations that cannot be classically explained without some post-measurement communication [1]. 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]. Some stronger-than-quantum nonsignaling correlations can even collapse the communication complexity in a two-party scenario. Namely, 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 [4].

As the violation of a Bell inequality is the signature of nonlocal correlations, a possible measure of nonlocality is the strength of this violation. However, this quantity does not necessarily provide a reliable measure of nonlocality as an information-theoretic resource. A more natural measure relies on the very definition of nonlocality: nonlocality requires some communication to be 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 nonlocal capacity, turns out to provide an ultimate limit to the power of classical communication in a two-party scenario. Indeed, nonlocal resources cannot replace an amount of classical communication bigger than the associated nonlocal capacity. Alternative measures of nonlocality could use different resources as unit of nonlocality, such as nonlocal boxes [3, 5, 6]. 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 [5].

In this paper, we will introduce a method for computing the nonlocal capacity of correlations. Namely, we show that the asymptotic nonlocal capacity is the minimum of a convex functional over a suitable space of probability distributions. Then, we discuss the relation with a previous work on the communication complexity of channels in general probabilistic theories [5]. Finally, we illustrate the method with a numerical example.

![NS-box](image)

**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 $k$.

The object that we will consider is a nonsignaling box, which is an abstract generalization of the following physical scenario. Two parties, say Alice and Bob, simultaneously perform a measurement on two spatially separate entangled systems. In general, Alice and Bob are allowed to choose among two sets of possible measurements. We assume that Bob’s set of measurements 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 $\{1, \ldots, M\}$, where $M$ is the number of measurements that Bob can perform. After the measurements, Alice gets the outcome $r$, Bob gets the outcome $s$. The overall scenario is described by the joint conditional probability $P(r; s|a; b)$. This distribution satisfies the nonsignaling conditions

$$
\sum_a P(r; s|a; b) = \sum_a P(r; s|a; \bar{b}) \equiv P(r|a) \forall a, b, \bar{b}, r,
$$

$$
\sum_r P(r; s|a; b) = \sum_r P(r; s|\bar{a}; b) \equiv P(s|b) \forall a, \bar{a}, b, r.
$$

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 post-measurement commu-
FIG. 2: (a) $N$ identical nonsignaling boxes. On one side, Alice chooses the inputs $(a_1, \ldots, a_N)$ and gets the outcomes $(r_1, \ldots, r_N)$. On the other side, Bob chooses the inputs $(b_1, \ldots, b_N)$ and gets the outcomes $(s_1, \ldots, s_N)$. (b) Simulation of the $N$ nonsignaling boxes through a shared stochastic variable $y$ and the communication of the variable $k$. We assume that only a one-way communication from Alice to Bob is allowed. The classical protocol is as follows (see Fig. 1b for reference). Alice generates an outcome $r$ with probability $P(r|a)$ depending on the variable $a$ and some stochastic variable $y$ shared with Bob and generated with probability $\rho(y)$. Then, she sends Bob a variable, say $k$, generated with probability $P(k|y)\,b\,k$ depending on $y$, $r$, and $k$. Protocol simulates the NS-box $P(r; s|a; b)$ if

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

As done in Refs. [9, 10] in a different context, we define the communication cost, say $\mathcal{C}$, of the simulation as the maximum, over the space of distributions $P(a)$, of the conditional Shannon entropy $H(K|Y) \equiv -\int dy \rho(y) \sum_k P(k|y) \log_2 P(k|y)$. That is,

$$\mathcal{C} \equiv \max_{P(a)} H(K|Y).$$

We define the nonlocal capacity (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.

More generally, we can perform a parallel simulation of $N$ identical NS-boxes, schematically represented in Fig. 2a. The $i$th 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 $(a_1, \ldots, a^N) = \vec{a}$ and gets the outcome $(r_1, \ldots, r^N) = \vec{r}$. Similarly, Bob chooses the input $(b_1, \ldots, b^N) = \vec{b}$ and gets the outcome $(s_1, \ldots, s^N) = \vec{s}$. The simulation of the NS-boxes through a shared variable $y$ and the communication of a variable $k$ is represented in Fig. 2b. The simulation scheme is the same as for a single box, with $a$, $b$, $r$ and $s$ replaced by $\vec{a}$, $\vec{b}$, $\vec{r}$ and $\vec{s}$. The protocol exactly simulates the $N$ boxes if

$$\sum_k \int dy P(s|y) \vec{b}(k) P(k|y) \vec{r}(k) a(r) \rho(y) = \prod_j P(r^j; s^j|a^j; b^j).$$

Each parallelized protocol has $N$ as a free parameter. The asymptotic communication cost of the protocol is equal to $\lim_{N \to \infty} \mathcal{C}^{par}/N \equiv \mathcal{C}^{asym}$, 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 $P(a_1 \ldots a_N)$. We define the asymptotic nonlocal capacity of the NS-box as the minimum of $\mathcal{C}^{asym}$ among the parallelized protocols. The asymptotic nonlocal capacity is denoted by $\mathcal{C}^{asym}_{nl}$.

Our task is to reduce the computation of $\mathcal{C}^{asym}_{nl}$ 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 $\rho(r|s|a)$ over $r$ and the sequence $s = (s_1, \ldots, s_M)$ 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 $\rho(r|s|a)$ satisfying the constraints

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

where

$$\sum_{s, s_m = s} \to \sum_{s_1, \ldots, s_M = s_1, \ldots, s_M}$$

is the summation over every index in $s$ but the $m$th one, which is set equal to $s$. The set $\mathcal{V}$ is surely non-empty. A function in $\mathcal{V}$ is $\rho(r|s|a) = P(s_1|r|a = 1) \times \cdots \times P(s_M|r|a = M) P(r|a)$. Note that $\rho(r|s|a)$ can be defined only if the first nonsignaling condition (1) is satisfied. The conditional probability $\rho(r|s|a)$ 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. [9] in a different context, we introduce the following protocol that simulates a NS-box through one of its HV-boxes.

**Master protocol.** Alice generates the outcome $r$ and the array $s$ according to a conditional probability $\rho(r|s|a) \in \mathcal{V}$. Then, she sends $s$ 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. [3] and used in Ref. [9] for quantum channels, it is possible to turn the master protocol into a child protocol for parallel simulations whose asymptotic communication cost is the capacity of the channel $a \rightarrow s$ described by the conditional probability $\rho(s|a) \equiv \sum_r \rho(r|s|a)$. Let us recall that a channel $x_1 \rightarrow x_2$ is identified by a conditional probability distribution $\rho(x_2|x_1)$ and its capacity is the mutual information between $x_1$ and $x_2$ over the space of probability distributions $\rho(x_1)$. 

**Lemma 1.** Alice and Bob share $N$ identical HV-boxes. The $i$th box has input $a_i$ and outcome $r_i$ on Alice’s side, and outcome $s_i^r$ on Bob’s side. The outcomes are generated with conditional probability $\rho(r_i,s_i^r|a_i)$. Alice chooses the $N$ inputs $a_1, \ldots, a_N$ and gets the $N$ outcomes $r_1, \ldots, r_N$. Bob gets the outcomes $s_1, \ldots, s_N$. The boxes can be locally simulated in parallel with an additional asymptotic communication cost from Alice to Bob equal to the capacity of the channel $\rho(s|a) \equiv \sum_r \rho(r|s|a)$.

**Proof.** The simulation is as follows. Alice chooses the inputs $a_1, \ldots, a_N$. Then she sends Bob an amount of information that allows Bob to generate the variables $s_1, \ldots, s_N$ according to the conditional probability $\rho(s_i|a_i)$. The reverse Shannon theorem [12] states that there is a protocol for this task with asymptotic communication cost equal to the capacity of the channel $\rho(s|a)$, provided that Alice and Bob share some stochastic variable, say $\chi$. It is always possible to have a deterministic protocol, so that the outcomes $s_1, \ldots, s_N$ are uniquely determined by $\chi$ and the communicated information. Since $\chi$ is shared with Alice, Alice knows Bob’s outcomes. Thus, Alice generates her outcomes $r_1, \ldots, r_N$ according to the conditional probability $\rho(r_i|a_i,s_i) \equiv (\rho(r_i|a_i)\rho(s_i|a_i))/\rho(s_i|a_i)$. The overall set of outcomes is generated according to the joint distribution $\rho(r_i,s_i|a_i)$. □

**Lemma 2.** Identical NS-boxes can be simulated in parallel with an asymptotic communication cost equal to the capacity of the channel $\rho(s|a) \equiv \sum_r \rho(r|s|a)$, where $\rho(r|s|a) \in \mathcal{V}$ is an associated HV-box.

**Proof.** This lemma 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 $\rho(r,s|a)$, which can be simulated in parallel with an asymptotic communication cost equal to the capacity of the channel $\rho(s|a)$. □

Now, we have enough tools to prove the main result.

**Theorem.** The asymptotic nonlocal capacity $C_{nl}^{asym}$ of a NS-box $P(r;s|a;b)$ is the minimum of the capacity of the channel $\rho(s|a)$ over the space $\mathcal{V}$ of associated HV-boxes. In other words,

$$C_{nl}^{asym} = D \equiv \min_{\rho(r,s|a) \in \mathcal{V}} \max_{P(a)} I(S;A),$$

where $I(S;A)$ is the mutual information between the stochastic variables $s$ and $a$.

**Proof.** The inequality $C_{nl}^{asym} \leq D$ is a consequence of Lemma 2. Indeed, Lemma 2 implies that there is a parallel simulation of the NS-box with asymptotic communication cost equal to $D$. Let us prove that no protocol can have a communication cost smaller than $D$. Let $C$ be the asymptotic communication cost of a parallel protocol that simulates the $N$ NS-boxes. The protocol is defined by the conditional probabilities $P(r^j|y^j), P(k^j|y^j)$ and $P(s^j|b^j)$ satisfying constraint (1). Through a procedure used in the supplemental material of Ref. [3], it is possible to build a multivariate HV-box with conditional probability

$$\rho(r^1 \ldots r^N s^1 \ldots s^N|a^1 \ldots a^N) = \rho(r^j s^j a^j)$$

over $r^j$ and the sequences $s^j = (s^j_1 \ldots s^j_M)$ so that the following properties are satisfied,

1. the capacity of the channel $\rho(s|a)$ is smaller than or equal to $NC + o(N)$.
2. The marginal distribution of the variables $r^j$ and $s^j_1$ is equal to $P(r^j,s^j_1|a^j,b^j = i)$, that is,

$$\sum_{s_1 r_1} \rho(r^j s^j a^j) = P(r^j s^j|a^j,b^j = i).$$

The first property comes from the data-processing inequality [11] since the marginal distribution of $r^j = r$ and $s^j_1 = s$ in constraint (10) depends only on $a^j$ and $b^j$, it is easy to show through the Karush-Kuhn-Tucker conditions [13] that the channel $\rho(s^j|a^j)$ with minimal capacity satisfying the constraints (1) takes the form $\rho(s^j|a^j) \times \cdots \times \rho(s^j|a^N)$, where $\rho(s^j|a^j)$ is a channel with capacity $D$. This and property 1 imply that $ND \leq NC + o(N)$ for every $N$, that is, $D \leq C$. □

This theorem reduces the evaluation of the nonlocal capacity of a NS-box to a convex optimization problem, which can be solved with standard methods [13]. A specialized numerical method that is particularly efficient for this problem will be discussed elsewhere.

The reverse Shannon theorem [12] and the data-processing inequality [11] are the key ingredients for proving that $C_{nl}^{asym} = D$. As there is a single-shot version of the reverse Shannon theorem [14] giving an upper bound to the communication cost required for the simulation of a single noisy channel, it is easy to adapt the proof and derive the inequalities

$$D \leq C_{nl} \leq D + 2 \log_2(D + 1) + 2 \log_2 e$$

for the single-shot nonlocal capacity $C_{nl}$.

There is a relationship between the nonlocal capacity of a NS-box and the communication complexity of a channel in a general probabilistic theory. This quantity has been defined in Ref. [3]. 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. 3, a C-box is called game 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. 9 for details). Let us denote by \( C_{ch}^{asym} \) this quantity (denoted by \( C_{min}^{asym} \) in Ref. 3).

**Corollary 1.** The nonlocal capacity \( C_{nl}^{asym} \) of an NS-box \( P(r; s|a;b) \) satisfies the inequalities

\[
C_{ch}^{asym} + \min_a \log_2 P(r|a) \leq C_{nl}^{asym} \leq C_{ch}^{asym} - \min_a \max_b I(R; S|a;b),
\]

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

**Proof.** The first inequality can be proved by using a procedure described in Sec. IIIB of Ref. 13, 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 \( -\min_a \log_2 P(r|a) \). Given the NS-box \( P(r; s|a;b) \), let \( V \) and \( V' \) be the space of associated HV-boxes \( \rho(rs|a) \) and distributions \( \rho(s|ra) \equiv \rho(rs|a)/\rho(r|a) \), respectively. The chain rule [11]

\[
I(S; A) = I(S; R, A) - I(S; R|A)
\]

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

\[
I(S; A) \leq I(S; R, A) - \min_a \max_b I(R; S|a;b).
\]

As the marginal distribution of \( r \) given \( a \) is the same for every distribution \( \rho(rs|a) \) in \( V \), the maximization in Eq. (7) can be performed over the space \( \rho(s|ra) \in V' \). Thus, we have that

\[
C_{nl}^{asym} \leq \min_{\rho(s|ra) \in V'} \max_a I(S; R, A) - \min_b \max_a I(R; S|a;b).
\]

We also have that

\[
\min_{\rho(s|ra) \in V'} \max_a I(S; R, A) \leq \min_{\rho(s|ra) \in V'} \max_{\rho(s|ra) \in V'} P(r,a).
\]

In Ref. 9, we showed that the right-hand side is equal to \( C_{ch}^{asym} \). Thus, the inequalities [11] [15] imply the second inequality [11]. □

**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,

\[
C_{nl}^{asym} = C_{ch}^{asym} - n,
\]

where \( C_{ch}^{asym} \) 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 eigenstates of the allowed measurements.

**Proof.** The corollary comes directly from Corollary 1. Indeed, the lower and upper bounds in Ineqs. (11) collapse to the same value, as \( P(r|a) = 2^{-n} \) and \( I(R; S|a;b = a) = n \).

To illustrate the presented method, we have numerically evaluated the nonlocal capacity for the Werner state \( \rho_{\gamma} = \gamma/2(|00\rangle + |11\rangle)(|00\rangle + |11\rangle + (1 - \gamma)1/4 \). 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 \( C_{nl}^{asym} \) is equal to zero for \( \gamma \leq 1/\sqrt{2} \) and equal to about \( 9/4(\gamma - 1/\sqrt{2})^2 \) for \( \gamma \in [1/\sqrt{2}, 1] \), with a maximum error lower than 3%.

In conclusion, we have presented a method for evaluating the asymptotic nonlocal capacity of correlations, and provided a tight lower and upper bound for 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 an open problem concerning the range of \( \gamma \) where a Werner state is nonlocal.

**Acknowledgments.** This work is supported by the Swiss National Science Foundation, the NCCR QSIT, and the COST action on Fundamental Problems in Quantum Physics.

[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] W. van Dam. Nonlocality and Communication Complexity. PhD thesis, University of Oxford, Department of Physics (2000); W. van Dam, arXiv:quant-ph/0501159.
[5] J. Barrett, S. Pironio, Phys. Rev. Lett. 95, 140401 (2005).
[6] N. Brunner, N. Gisin, V. Scarani, New J. Phys. 7, 88 (2005).
[7] F. Dupuis, N. Gisin, A. Hasidim, A. Méthot, H. Pilpel, J. Math. Phys. 48, 082107 (2007).
[8] A. Montina, Phys. Rev. Lett. 109, 110501 (2012).
[9] A. Montina, M. Pfaffhauser, S. Wolf, Phys. Rev. Lett. 111, 160502 (2013).
[10] A. Montina, Phys. Rev. A 87, 042331 (2013).
[11] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).

[12] C. H. Bennett, P. Shor, J. Smolin, and A. V. Thapliyal, IEEE Trans. Inf. Theory 48, 2637 (2002).

[13] S. Boyd, L. Vandenberghe *Convex Optimization* (Cambridge University Press, Cambridge, 2004).

[14] P. Harsha, R. Jain, D. McAllester, J. Radhakrishnan, IEEE Trans. Inf. Theory 56, 438 (2010).

[15] A. Montina, Phys. Rev. A 84, 042307 (2011).