Local Hidden Variable Models for Entangled Quantum States Using Finite Shared Randomness

Joseph Bowles, Flavien Hirsch, Marco Túlio Quintino, and Nicolas Brunner
Département de Physique Théorique, Université de Genève, 1211 Genève, Switzerland
(Received 11 December 2014; published 24 March 2015)

The statistics of local measurements performed on certain entangled states can be reproduced using a local hidden variable (LHV) model. While all known models make use of an infinite amount of shared randomness, we show that essentially all entangled states admitting a LHV model can be simulated with finite shared randomness. Our most economical model simulates noisy two-qubit Werner states using only $\log_2(12) \simeq 3.58$ bits of shared randomness. We also discuss the case of positive operator valued measures, and the simulation of nonlocal states with finite shared randomness and finite communication. Our work represents a first step towards quantifying the cost of LHV models for entangled quantum states.

DOI: 10.1103/PhysRevLett.114.120401 PACS numbers: 03.65.Ud

Introduction.—Quantum systems exhibit a wide range of nonclassical and counterintuitive phenomena, such as quantum entanglement [1] and Bell nonlocality [2,3]. In recent years, considerable effort has been devoted to understanding the relation between entanglement and nonlocality; see [3]. While entanglement is necessary to demonstrate nonlocality (i.e., violation of a Bell inequality), it is not yet clear whether all entangled states can lead to nonlocality when considering the most general scenario [4,5]. Nevertheless, entanglement and nonlocality are proven to be different in the simplest scenario in which local (nonsequential) measurements are performed on a single copy of an entangled state. As discovered by Werner [6], there exist entangled states that can provably not violate any Bell inequality, since the state admits a local hidden variable (LHV) model. While Werner focused on projective measurements, Barrett [7] showed that the result holds for the most general nonsequential measurements, so-called positive operator valued measures (POVMs).

Following these early results, plenty of works have investigated these ideas; see [8] for a recent review. LHV models were reported for entangled states with less symmetry than Werner states [9–13]. Multipartite states were discussed as well [14,15]. Interestingly, it was shown that in certain cases, the nonlocality of local entangled states can be activated, e.g., by considering sequential measurements [11,16]. More recently, interest was devoted to a special class of LHV models, referred to as local hidden state (LHS) models, which naturally arise in the context of Einstein-Podolsky-Rosen (EPR) steering [17,18], and essentially require that the local variable represents a quantum state; see [17] for details, and [6,7,9,12,19] for examples of LHS models.

Here we discuss novel types of questions in this context, namely that of quantifying LHV models. Specifically, given a local entangled state, we ask what resources are required to construct a LHV model; i.e., what is the cost of classically simulating the correlations of the state? As a figure of merit, we consider the minimal dimension of the shared local (hidden) variable that is needed; that is, how much classical information (how many bits) is necessary to encode the local variable? Note that all LHV models constructed so far are maximally costly according to our measure as they make use of shared variables which are continuous. Hence, such models would require a communication channel of infinite capacity, the physical relevance of which is questionable. For instance, in Werner's model, the local variables are unit vectors $\vec{\lambda}$ (e.g., vectors on the Bloch sphere). Importantly, although these vectors are of a given dimension, the model requires an infinite number of them, as vectors $\vec{\lambda}$ are taken from the uniform distribution over the sphere.

Hence, a natural question is whether it would be in fact possible to simulate the correlations of an entangled state using shared variables of finite dimension (i.e., a finite number of shared random bits). Here we show that essentially any entangled state admitting a LHV model can be simulated with finite shared randomness, considering arbitrary local projective measurements. We discuss in detail the case of Werner states of two qubits. We also show that the simulation of arbitrary POVMs on certain entangled states is possible using finite shared randomness. Finally, we consider the simulation of nonlocal entangled states (i.e., which can violate a Bell inequality), in which case communication between the parties is necessary. In particular, we show that the simulation of any full rank entangled state can be achieved using only finite communication.

Our work provides a perspective on understanding how the correlations of local entangled states differ from those of fully separable states. On the one hand, it shows that there is no fundamental difference between the two cases, in the sense that finite shared randomness is enough for both (at least for certain entangled states). Recall that the correlations of separable states can always be simulated using $4\log_2(d)$ bits [20], where d denotes the local Hilbert space dimension of the state. On the other hand, our results suggest that the simulation of entangled states is in general more costly compared to that of separable states—despite the fact that both classes of states can never lead to Bell inequality violation.

Preliminaries.—We consider a bipartite Bell scenario. Two distant observers, Alice and Bob, share a quantum state ρ (of Hilbert space dimension $d \times d$) and perform local measurements $A = \{A_a\}$ and $B = \{B_b\}$, respectively. The observed statistics are local (in the sense of Bell), if they can be decomposed as follows [2,3]:

$$\mathrm{Tr}(A_a \otimes B_b \rho) = \int \pi(\lambda) p_A(a|A,\lambda) p_B(b|B,\lambda) d\lambda, \quad (1)$$

where λ represents a shared (hidden) variable, distributed according to density $\pi(\lambda)$. If a decomposition of the form (1) exists for all possible local measurements, we say that the state ρ is local as it will never violate any Bell inequality. The LHV model is then characterized by the distributions $\pi(\lambda)$, and $p_A(a|A,\lambda)$, $p_B(b|B,\lambda)$ which are Alice's and Bob's local response functions.

Trivially, any state ρ that is separable is local. Indeed, one can write $\rho = \sum_{\lambda=1}^{d^4} p_\lambda \rho_A^\lambda \otimes \rho_B^\lambda$ [20,21] with $\sum_\lambda p_\lambda = 1$ and $p_\lambda \geq 0$ (note that for two-qubit states, a more economical decomposition exists, involving only four product states [22]). Here the local variable λ is distributed according to p_λ, and the local response functions are simply $p_A(a|A,\lambda) = \mathrm{Tr}(A_a \rho_A^\lambda)$ for Alice and similarly for Bob. Note that the shared variable takes only d^4 different values here, and can thus be encoded in $4\log_2(d)$ bits (for two-qubit states 2 bits are enough). More interestingly, there exist entangled states ρ which are local. The most famous example is the Werner state, which for the case $d = 2$ takes the form

$$\rho_W(\alpha) = \alpha|\psi^-\rangle\langle\psi^-| + (1-\alpha)\mathbb{1}/4, \quad (2)$$

where $|\psi^-\rangle = (|01\rangle - |10\rangle)/\sqrt{2}$ is the singlet state and $\mathbb{1}/4$ is the maximally mixed two-qubit state. After showing that the state $\rho_W(\alpha)$ is entangled for $\alpha > 1/3$, Werner [6] constructed a local model for arbitrary projective measurements for $\alpha \leq 1/2$; later another local model was constructed for $\alpha \lesssim 0.66$ [10]. Considering the most general nonsequential measurements, i.e., POVMs, a local model was presented for $\alpha \leq 5/12$ [7].

A common feature of these local models (and to the best of our knowledge, of all known LHV models) is the fact that the shared variable λ takes an infinite number of different values; typically, λ denotes a (unit) vector, which is taken randomly from a uniform distribution over the sphere. Hence λ requires an infinite number of bits to be encoded, in stark contrast with the case of separable states, where $4\log_2(d)$ bits are enough. Therefore, it is rather natural to ask if this represents a fundamental difference between local entangled states and separable ones. Below we will show that this is not the case, by exhibiting LHV models for entangled states requiring only finite resources, i.e., where λ can be encoded with a finite number of bits.

Simulating Werner states with finite shared randomness.—We present local models using a finite amount of shared randomness, simulating the correlations of Werner states $\rho_W(\alpha)$ for $\alpha < 0.5$ for all projective measurements; extensions to $\alpha \lesssim 0.66$ are given in the next section. Alice and Bob receive here Bloch vectors $\vec{a}$ and $\vec{b}$ (representing observables $A = \vec{a}\cdot\vec{\sigma}$ and similarly for Bob) and should provide outcomes $a, b = \pm 1$ such that

$$\langle a \rangle = \langle b \rangle = 0, \qquad \langle ab \rangle = -\alpha \vec{a}\cdot\vec{b}. \quad (3)$$

For clarity, we start by presenting a simple model using only $\log_2(12)$ bits of shared randomness, which works for $\alpha \lesssim 0.43$. Our model uses the icosahedron, one of the 5 platonic solids in dimension 3. The icosahedron has 12 vertices represented by the normalized vectors $\vec{v}_\lambda \in V$, which satisfy the following properties:

$$\forall \vec{v}_\lambda \exists\, \vec{v}_j \text{ such that (s.t.) } \vec{v}_\lambda = -\vec{v}_j \quad (4)$$

$$\sum_{j \text{ s.t. } \vec{v}_j\cdot\vec{v}_\lambda \geq 0} \vec{v}_j = \gamma \vec{v}_\lambda \qquad \forall \lambda, \quad (5)$$

with $\gamma = 1 + \sqrt{5}$. Note that the radius of a sphere inscribed inside the icosahedron is given by $\ell = \sqrt{(5+2\sqrt{5})/15}$. In our model the shared variable $\lambda \in \{1, \ldots, 12\}$ is distributed uniformly and represents one of the 12 vertices of the icosahedron. That is, when Alice and Bob receive λ, they will use vector $\vec{v}_\lambda$.

Protocol 1.—Alice and Bob share $\lambda \in \{1, \ldots, 12\}$, uniformly distributed. Upon receiving setting $\vec{a}$, Alice calculates the subnormalized vector $\vec{a}' = \ell\vec{a}$. This ensures that $\vec{a}'$ lies inside the convex hull of V; hence, Alice can find a convex decomposition $\vec{a}' = \sum_i \omega_i \vec{v}_i$ with $\sum_i \omega_i = 1$ and $\omega_i \geq 0$ (note that any convex decomposition can be chosen). Then, with probability ω_i, she outputs $a = \pm 1$ with probability $(1 \pm \mathrm{sgn}[\vec{v}_\lambda \cdot \vec{v}_i])/2$. Bob, upon receiving $\vec{b}$, outputs $b = \pm 1$ with probability $(1 \mp \vec{b}\cdot\vec{v}_\lambda)/2$.

We now show that the protocol reproduces the desired statistics. We start with the correlator:

$$\langle ab \rangle = -\frac{1}{12}\sum_\lambda \sum_i \omega_i \mathrm{sgn}(\vec{v}_i \cdot \vec{v}_\lambda)\vec{v}_\lambda \cdot \vec{b}. \quad (6)$$

Interchanging the sums, we first calculate

$$\sum_\lambda \mathrm{sgn}(\vec{v}_i \cdot \vec{v}_\lambda)\vec{v}_\lambda \cdot \vec{b} = 2\gamma \vec{v}_i \cdot \vec{b}, \quad (7)$$

which follows from (4) and (5); see details in the Supplemental Material [23]. Inserting the last expression in (6), we get

$$\langle ab\rangle = -\frac{\gamma}{6}\sum_i \omega_i \vec{v}_i\cdot\vec{b} = -\frac{\ell\gamma}{6}\vec{a}\cdot\vec{b} \simeq -0.43\ \vec{a}\cdot\vec{b}. \tag{8}$$

Finally, we compute Alice's marginal

$$\langle a\rangle = -\frac{1}{12}\sum_\lambda\sum_i \omega_i\,\mathrm{sgn}(\vec{v}_i\cdot\vec{v}_\lambda) = 0, \tag{9}$$

which can be seen from (4). Similarly, we get that $\langle b\rangle = 0$. Therefore, the model simulates $\rho_W(\alpha)$ for $\alpha \simeq 0.43$. Extension to smaller values of α is straightforward.

The above protocol can be adapted to any polyhedron satisfying conditions (4) and (5). Natural candidates are the Platonic solids, except for the tetrahedron which does not satisfy (4). Among these, the icosahedron turns out to be optimal here; see Supplemental Material [23]. Hence, in order to simulate Werner states which are more entangled, i.e., going beyond $\alpha \simeq 0.43$, we need another method.

We now present a protocol, which will allow us to relax condition (5). Specifically, we consider again a three-dimensional polyhedron V with D vertices $\vec{v}_i$, but only demand that is satisfy condition (4) (which can always be achieved at the expense of doubling the number of vertices of a given polyhedron). As before, the shared variable $\lambda \in \{1,\ldots,D\}$ encodes the choice of vertex, and is uniformly distributed. Having abandoned condition (5), we have for each vertex $\vec{v}_\lambda$:

$$\sum_{j\ \mathrm{s.t.}\ \vec{v}_j\cdot\vec{v}_\lambda\geq 0} \vec{v}_j = \gamma_\lambda \vec{m}_\lambda, \tag{10}$$

where $\vec{m}_\lambda$ is a normalized vector and generally $\vec{m}_\lambda \neq \vec{v}_\lambda$. Let us define $\gamma_{\min} = \min_\lambda(\gamma_\lambda)$. Note that there are now two polyhedra of interest: (i) V, that is defined by the vertices $\vec{v}_\lambda$ and (ii) M, defined by the vertices $\vec{m}_\lambda$, which are in one-to-one correspondence with the $\vec{v}_\lambda$. Consider the following protocol.

Protocol 2.—Alice and Bob share $\lambda \in \{1,\ldots,D\}$ uniformly distributed. Upon receiving setting $\vec{a}$, Alice calculates the subnormalized vector $\vec{a}' = \ell\vec{a}$ where ℓ is the radius of the largest sphere fitting inside M and centered on the origin. This ensures that $\vec{a}'$ lies inside the convex hull of M and Alice can therefore find a convex decomposition $\vec{a}' = \sum_{i=1}^D \omega_i \vec{m}_i$. Then, with probability $p_i = \omega_i\gamma_{\min}/\gamma_i$ she outputs $a = \mathrm{sgn}(\vec{v}_i\cdot\vec{v}_\lambda)$, and with probability $(1-\sum_i p_i)$ she outputs a random bit. Bob, upon receiving $\vec{b}$, outputs $b = \pm 1$ with probability $(1 \mp \vec{b}\cdot\vec{v}_\lambda)/2$.

The resulting correlations are given by

$$\begin{aligned}\langle ab\rangle &= -\frac{1}{D}\sum_\lambda\sum_i \omega_i\frac{\gamma_{\min}}{\gamma_i}\mathrm{sgn}(\vec{v}_i\cdot\vec{v}_\lambda)\vec{b}\cdot\vec{v}_\lambda\\ &= -\frac{2\gamma_{\min}}{D}\sum_i\frac{\omega_i}{\gamma_i}\sum_{\lambda\ \mathrm{s.t.}\ \vec{v}_\lambda\cdot\vec{v}_i\geq 0}\vec{v}_\lambda\cdot\vec{b}\\ &= -\frac{2\ell}{D}\gamma_{\min}\vec{a}\cdot\vec{b},\end{aligned} \tag{11}$$

where we have used Eq. (10) in the last step; see Supplemental Material [23] for details. As for protocol 1, using Eq. (4) we get that the marginals $\langle a\rangle = \langle b\rangle = 0$. Hence, the model reproduces the statistics of $\rho_W(\alpha)$ for $\alpha = (2\ell/D)\gamma_{\min}$.

Starting from a sufficiently regular polyhedron with a large number D of vertices $\vec{v}_\lambda$, we can approximate the unit sphere and the factor ℓ can become arbitrary close to one. In the limit $D\to\infty$ we expect to recover the uniform distribution over the sphere and our model therefore becomes equivalent to Werner's model for $\rho_W(1/2)$ [6]. In Fig. 1 we plot upper bounds on the required shared randomness to simulate $\rho(\alpha)$ as a function of α obtained via protocol 2. We use a family of polyhedra, generated iteratively and starting from the icosahedron. To generate the second polyhedron, we take the union of the icosahedron and its normalized dual (which is the dodecahedron), and so on. One can verify that these polyhedra respect condition (4).

Note that the above protocols are LHS models. Hence the above results can be straightforwardly extended to the simulation of entangled states which are obtained via local filtering on the Werner state, e.g., Ref. [9] (see Supplemental Material [23]). Also, it would be interesting to see if more economical models (i.e., using less shared randomness) exist, and if local entangled states require

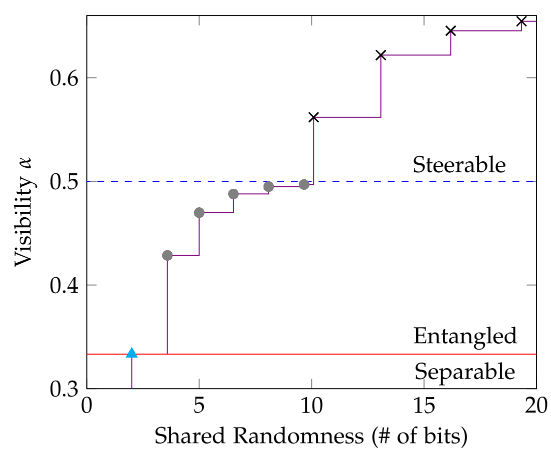

FIG. 1 (color online). Simulation of two-qubit Werner states $\rho_W(\alpha)$ with finite shared randomness. The graph shows the relation between the visibility α (essentially the degree of entanglement) and the amount of shared randomness, quantified in bits. For $\alpha \leq 1/3$ (below the solid line) the state is separable; hence 2 bits of shared randomness suffice (triangle). For $1/3 < \alpha \lesssim 0.43$, the state can be simulated with $\log_2(12)$ bits of shared randomness using protocol 1. For $0.43 \lesssim \alpha \lesssim 0.66$, $\rho_W(\alpha)$ can be simulated with a larger (but nevertheless finite) amount of shared randomness. For $0.43 \lesssim \alpha < 0.5$, we have a LHS model (using protocol 2). For $0.5 < \alpha \lesssim 0.66$ the state becomes steerable but can nevertheless be simulated by a LHV model using finite shared randomness, by applying Result 1 to the model of Ref. [10] (see main text).

more shared randomness compared to separable states. For Werner states, this translates to whether we expect to see a discontinuity at the separable-entangled boundary for $\alpha = 1/3$ (see Fig. 1). We give two partial answers in this direction: (i) for LHS models, the maximum α one can simulate with $D = 4$ is the separable state $\alpha = 1/3$ (see Supplemental Material [23]); (ii) Restricting to equatorial measurements one can achieve $\alpha = 1/2$ with only $D = 4$ (see Supplemental Material [23]).

General results.—In the above, we have focused on a class of highly symmetric states, namely Werner states, and considered only projective measurements. Here we show how local models with finite shared randomness can be constructed for essentially any state that admits a LHV model. We also discuss the case of general measurements, i.e., POVMs.

Result 1: Consider a state ρ (of dimension $d \times d$) admitting a LHV model for all projective measurements. Then, a LHV model using only finite shared randomness can simulate all projective measurements on the state

$$\rho(\eta) = \eta^2 \rho + \eta(1-\eta)\left(\frac{\mathbb{1}}{d} \otimes \rho_B + \rho_A \otimes \frac{\mathbb{1}}{d}\right) + (1-\eta)^2 \frac{\mathbb{1} \otimes \mathbb{1}}{d^2}$$

for any $0 \le \eta < 1$. Here $\rho_{A,B} = \mathrm{Tr}_{B,A}(\rho)$.

Proof.—First, note that it follows from the relation

$$\mathrm{tr}[A_a \otimes B_b \rho(\eta)] = \mathrm{tr}[A_a(\eta) \otimes B_b(\eta)\rho] \qquad (12)$$

that the simulation of projective measurements (given by operators A_a and B_b) on $\rho(\eta)$ is equivalent to the simulation of noisy measurements, given by operators $A_a(\eta) = \eta A_a + (1-\eta)(\mathbb{1}/d)$ and $B_b(\eta) = \eta B_b + (1-\eta)(\mathbb{1}/d)$ on the state ρ. Next, since $A_a(\eta)$ and $B_b(\eta)$ are full rank for any $\eta < 1$, they are not on the border on the set of measurements [24], and can thus be decomposed as convex mixtures over a single set of finitely many projective measurements (more details in the Supplemental Material [23]). Finally, note that the simulation of a finite number of projective measurements on ρ requires only finite shared randomness. This follows from the fact (i) the resulting distribution is local (as ρ admits a LHV model), and (ii) the set of local distributions forms a polytope [3].

Note that the amount of shared randomness needed will depend on the value of η and diverges as $\eta \to 1$.

Result 2: Let us now discuss more general measurements, i.e., POVMs. Consider an entangled state ρ (of dimension $d \times d$) admitting a local model with k bits of shared randomness for projective measurements. We can then construct the state

$$\rho' = \frac{1}{(d+1)^2}[\rho + d(\rho_A \otimes F + F \otimes \rho_B) + d^2 F \otimes F],$$

which admits a local model with k bits of shared randomness for POVMs. Here $F = |d+1\rangle\langle d+1|$ denotes a projector onto a subspace orthogonal to the support of ρ; hence, ρ' is entangled by construction and of local dimension $d+1$. This result follows straightforwardly from Protocol 2 of Ref. [11], since the local model obtained for ρ' makes use of the same shared randomness as the one for ρ.

Finally, we present two examples illustrating the above results. First, applying Result 1 the local model of Ref. [25] allows us to extend our result for two-qubit Werner states. Specifically, we show that $\rho_W(\alpha)$ can be simulated with finite shared randomness for $\alpha \lesssim 0.66$. Upper bounds on the amount of shared randomness are given in Fig. 1 (using again an iterative procedure based on the icosahedron). Notably, this shows that certain states useful for EPR steering can be simulated with finite shared randomness. Second, applying Result 2 to the state [26] $\rho_W(0.43)$, we obtain that the state $\rho = \frac{1}{3}[\rho_W(0.43) + 2|2\rangle\langle 2| \otimes (\mathbb{1}/2)]$ can be simulated for arbitrary POVMs using $\log_2(12)$ bits of shared randomness.

Simulating nonlocal states with finite resources.—Finally, we discuss the simulation of entangled states which are nonlocal. In this case, classical communication from (say) Alice to Bob is required. This communication is sent after Alice has received her input. Two cases can be considered: (i) Alice and Bob are initially uncorrelated (i.e., have no shared randomness), and Alice sends classical information to Bob, (ii) Alice and Bob have access to shared randomness, and Alice sends classical information to Bob. Reference [27] presents a model using no shared randomness and finite expected communication. Other known protocols (see, e.g., [25,28,29]) require, for case (ii), finite communication assisted with infinite shared randomness—hence infinite communication for case (i). Here we present protocols using only finite resources, even in the worst-case scenario.

Considering case (i), we first show that the statistics of any bipartite entangled state ρ of dimension $d \times d$ that is full rank can be simulated with finite communication. Note that a state of full rank does not lie on the border of the set of quantum states [20]. Upon receiving her measurement setting $A = \{A_a\}$, Alice outputs a according to the distribution $p(a) = \mathrm{Tr}(\rho_A A_a)$, where ρ_A is Alice's reduced state. For output a, Bob should hold the (normalized) state $\rho_B^a = \mathrm{Tr}_A(A_a \otimes \mathbb{1}\rho)/p(a)$. Since ρ_B^a is full rank (by construction) there exists a polyhedron V (with D vertices, each representing a pure quantum state of dimension d) such that Alice can decompose ρ_B^a as a convex combination of the vertices of V. With probability ω_i (the coefficient of vertex i in the decomposition) Alice sends label i to Bob, who can then locally reconstruct the corresponding pure state (knowing V). The model thus reproduces the statistics of ρ using $\log_2(D)$ bits of communication.

For case (ii), we show that any state $\rho_W(\alpha)$ [see Eq. (2)], with $\alpha < 1$, can be simulated with finite shared randomness and finite communication (worst case). In particular, for $\alpha \leq 3/4$ a single bit suffices. To construct such a model, we combine the ideas of Protocol 1 and the simulation model (using 1 bit of communication) for the singlet state of Ref. [28]. See Supplemental Material [23] for details.

Conclusion.—We have shown that the correlations of essentially all entangled states that admit a LHV model can be simulated with finite shared randomness for the case of projective measurements. This shows that the requirement of infinite shared randomness (hence channels with infinite capacity) used in previous models can in fact be dispensed with. Whether this result can be extended to the case of POVMs is a relevant issue.

An interesting open question is to find the minimal amount of shared randomness required to simulate a local entangled state. For a state of local dimension d, are more than $4\log_2(d)$ bits of shared randomness always required, that is, is the simulation of local entangled states strictly more costly than that of separable states? We presented a model using only 2 bits for Werner states of two qubits, but our model works only for equatorial measurements; hence, the question remains open.

We thank Antonio Acín and Marcin Pawlowski for discussions. We acknowledge financial support from the Swiss National Science Foundation (Grant No. PP00P2_138917 and QSIT), and SEFRI (COST action MP1006). J. B. and F. H. contributed equally to this work.

[1] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Quantum entanglement, Rev. Mod. Phys. **81**, 865 (2009).

[2] J. S. Bell, On the Einstein-Poldolsky-Rosen paradox, Physics (Long Island City, N.Y.) **1**, 195 (1964).

[3] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Bell nonlocality, Rev. Mod. Phys. **86**, 419 (2014).

[4] L. Masanes, Y.-C. Liang, and A. C. Doherty, All Bipartite Entangled States Display Some Hidden Nonlocality, Phys. Rev. Lett. **100**, 090403 (2008).

[5] T. Vértesi and N. Brunner, Disproving the Peres conjecture: Bell nonlocality from bipartite bound entanglement, Nat. Commun. **5**, 5297 (2014).

[6] R. F. Werner, Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model, Phys. Rev. A **40**, 4277 (1989).

[7] J. Barrett, Nonsequential positive-operator-valued measurements on entangled mixed states do not always violate a Bell inequality, Phys. Rev. A **65**, 042302 (2002).

[8] R. Augusiak, M. Demianowicz, and A. Acín, Local hidden-variable models for entangled quantum states, J. Phys. A **47**, 424002 (2014).

[9] M. L. Almeida, S. Pironio, J. Barrett, G. Tóth, and A. Acín, Noise Robustness of the Nonlocality of Entangled Quantum States, Phys. Rev. Lett. **99**, 040403 (2007).

[10] A. Acín, N. Gisin, and B. Toner, Grothendieck's constant and local models for noisy entangled quantum states, Phys. Rev. A **73**, 062105 (2006).

[11] F. Hirsch, M. T. Quintino, J. Bowles, and N. Brunner, Genuine Hidden Quantum Nonlocality, Phys. Rev. Lett. **111**, 160402 (2013).

[12] J. Bowles, T. Vértesi, M. T. Quintino, and N. Brunner, One-Way Einstein-Podolsky-Rosen Steering, Phys. Rev. Lett. **112**, 200402 (2014).

[13] M. Cong Tran, W. Laskowski, and T. Paterek, The Werner gap in the presence of simple coloured noise, J. Phys. A **47**, 424025 (2014).

[14] G. Tóth and A. Acín, Genuine tripartite entangled states with a local hidden-variable model, Phys. Rev. A **74**, 030306(R) (2006).

[15] R. Augusiak, M. Demianowicz, J. Tura, and A. Acín, Entanglement and nonlocality are inequivalent for any number of particles, arXiv:1407.3114.

[16] S. Popescu, Bell's Inequalities and Density Matrices: Revealing Hidden Nonlocality, Phys. Rev. Lett. **74**, 2619 (1995).

[17] H. M. Wiseman, S. J. Jones, and A. C. Doherty, Steering, Entanglement, Nonlocality, and the Einstein-Podolsky-Rosen Paradox, Phys. Rev. Lett. **98**, 140402 (2007).

[18] M. D. Reid, P. D. Drummond, W. P. Bowen, E. G. Cavalcanti, P. K. Lam, H. A. Bachor, U. L. Andersen, and G. Leuchs, Colloquium: The Einstein-Podolsky-Rosen paradox: From concepts to applications, Rev. Mod. Phys. **81**, 1727 (2009).

[19] S. Jevtic, M. J. W. Hall, M. R. Anderson, M. Zwierz, and H. M. Wiseman, Einstein-Podolsky-Rosen steering and the steering ellipsoid, arXiv:1411.1517.

[20] I. Bengtsson and K. Zyczkowski, *Geometry of Quantum States: An Introduction to Quantum Entanglement* (Cambridge University Press, Cambridge, 2006).

[21] P. Horodecki, Separability criterion and inseparable mixed states with positive partial transposition, Phys. Lett. A **232**, 333 (1997).

[22] W. K. Wootters, Entanglement of Formation of an Arbitrary State of Two Qubits, Phys. Rev. Lett. **80**, 2245 (1998).

[23] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.114.120401 for details of the simulation protocols.

[24] G. Mauro D'Ariano, P. Lo Presti, and P. Perinotti, Classical randomness in quantum measurements, J. Phys. A **38**, 5979 (2005).

[25] O. Regev and B. Toner, Simulating quantum correlations with finite communication, arXiv:0708.0827.

[26] Note that we start here from a LHS, hence the model works directly for POVMs on Bob's side. It is thus sufficient to apply the extension procedure of Result 2 on Alice's side only.

[27] G. Brassard, L. Devroye, and C. Gravel, Exact simulation of the GHZ distribution, arXiv:1303.5942.

[28] B. F. Toner and D. Bacon, Communication Cost of Simulating Bell Correlations, Phys. Rev. Lett. **91**, 187904 (2003).

[29] C. Branciard and N. Gisin, Quantifying the Nonlocality of Greenberger-Horne-Zeilinger Quantum Correlations by a Bounded Communication Simulation Protocol, Phys. Rev. Lett. **107**, 020401 (2011).