

Connector Tensor Networks: A Renormalization-Type Approach to Quantum Certification

Miguel Navascués,¹ Sukhbinder Singh,² and Antonio Acín^{3,4}

¹*Institute for Quantum Optics and Quantum Information (IQOQI) Vienna Austrian Academy of Sciences, Vienna 1090., Austria*

²*Max-Planck Institute for Gravitational Physics (Albert Einstein Institute), Potsdam 14476, Germany*

³*ICFO-Institut de Ciències Fotoniques, The Barcelona Institute of Science and Technology, 08860 Castelldefels (Barcelona), Spain*

⁴*ICREA-Institució Catalana de Recerca i Estudis Avançats, Lluís Companys 23, 08010 Barcelona, Spain*



(Received 5 August 2019; revised manuscript received 6 March 2020; accepted 15 April 2020; published 19 June 2020)

As quantum technologies develop, we acquire control of an ever-growing number of quantum systems. Unfortunately, current tools to detect relevant quantum properties of quantum states, such as entanglement and Bell nonlocality, suffer from severe scalability issues and can be computed only for systems of a very modest size of around six sites. In order to address large many-body systems, we propose a renormalization-type approach based on a class of local linear transformations called *connectors*, which can be used to *coarse grain* the system in a way that preserves the property under investigation. Repeated coarse graining produces a system of manageable size, whose properties can then be explored by means of the usual techniques for small systems. In the case of a successful detection of the desired property, the method outputs a linear witness which admits an exact *tensor network* representation composed of connectors. We demonstrate the power of our method by certifying entanglement, Bell nonlocality, and supraquantum Bell nonlocality in systems with hundreds of sites using a normal desktop computer.

DOI: [10.1103/PhysRevX.10.021064](https://doi.org/10.1103/PhysRevX.10.021064)

Subject Areas: Quantum Information

I. INTRODUCTION

A central goal in quantum-information theory is to detect interesting global properties of few- or many-body systems. For example, traditionally, one may be interested in detecting whether a given quantum-many-body state is entangled [1], or whether a given conditional probability distribution contains nonclassical correlations, in the sense of violating a Bell inequality [2]. More recent ventures include detecting quantum causality properties [3] or the minimal local Hilbert space dimension of each party in a Bell test necessary for a given violation [4]. The basic underlying approach to all these tasks is the same: We consider a property (entanglement, nonlocality, dimensionality, etc.) of the system that we wish to falsify, derive an operational limitation on the set of all systems satisfying this property, and then show that this limitation is violated in the experiment.

Effective numerical tools to derive the operational limitations of small systems [5–11] are available. These allow detection of global properties such as entanglement and nonlocality in three- or even four-partite systems in a few

minutes using a regular desktop computer. Unfortunately, the analysis of large systems presents two problems.

The first is that the number of parameters required to fully specify the operational behavior of a many-body system increases exponentially with its size, so do the resources (e.g., number of experiments) needed to estimate all such parameters. That is, even prior to detection, one cannot efficiently specify the state of the system in general. However, many natural quantum states admit an efficient *tensor network* representation (see, e.g., Refs. [12–17]), which has been exploited to develop tomographic protocols to characterize such states with a number of experiments that scale only polynomially with the system size [18,19]. Assuming that the quantum states underlying our experiments are somehow typical and can be represented by a tensor network, it is therefore possible to circumvent this problem.

The second problem is that, even when the system can be efficiently represented, the computational resources required to detect the relevant global properties of the system also scale exponentially with the system size. Consequently, as experiments with quantum simulators and condensed-matter systems progress, we get access to larger and larger systems whose nonclassical properties cannot be detected with current theoretical tools.

In this work, we propose a general approach to solve this second problem. Our approach relies heavily on the framework of tensor networks, and also, somewhat

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

surprisingly, on a central concept from quantum foundations: the framework of generalized probabilistic theories (GPTs) [20–24]. It also requires techniques from convex optimization theory [25,26]. The result, *connector theory*, allows us to detect global properties in many-body quantum systems of unprecedented size.

A. The main idea

The key insight underlying connector theory is that one can construct local transformations to *coarse grain* the many-body system to an effective small system, say, with two or three sites, while preserving the global property of interest. Subsequently, if one can detect that the resulting small system does not have the desired property, neither does the original system.

Our inspiration comes from renormalization approaches and, in particular, coarse-graining techniques that have proved very effective in, e.g., diagonalizing large quantum-many-body Hamiltonians in condensed-matter physics where one is often interested only in the low-energy subspace. The many-body Hamiltonian can be coarse grained to a few-site Hamiltonian—which can be exactly diagonalized—while preserving its low-energy subspace. This strategy has led to the invention of ground-breaking simulation algorithms for large condensed-matter systems.

Before presenting all the details of the method, it is useful to illustrate the main idea with an example. Suppose you are given a quantum state ρ of m particles and your task is to certify that the state is entangled. An m -body quantum state is separable if it can be expressed as

$$\rho = \sum_i p_i \rho_i^1 \otimes \cdots \otimes \rho_i^m, \quad (1)$$

where $p_i \geq 0$, $\sum_i p_i = 1$, and ρ_i^j are normalized quantum states. The state ρ is entangled if it does not admit such a decomposition. Now, assume you have a linear transformation mapping two systems into one,

$$T: (\mathbb{C} \otimes \mathbb{C}) \rightarrow \mathbb{C}, \quad (2)$$

and such that product states are transformed into valid quantum states, that is, $T(\rho \otimes \sigma) \geq 0$ for all ρ and σ . Note that we do not require the map to be physical; that is, it may produce nonpositive states when applied to an initial (entangled) state. Clearly, the application of this map to a separable m -body state results in a separable $(m-1)$ -body state. By repeatedly applying maps of this form to the initial state, it is possible to reach a size in which standard entanglement-detection methods, including state positivity, can be tested. If any of these methods fails, we can certify that the initial m -body state was entangled. However, to apply this idea in practice, many aspects need to be sorted out. For instance, one needs to find a way of applying the maps T without having to deal with the complete m -body

quantum state and provide the tools to construct them. These and other issues are presented in what follows and constitute the main technical results of this work.

B. Consequences of our results

There are two scenarios which might greatly benefit from our connector-based approach to entanglement or nonlocality detection.

(a) Given a general density matrix or probability distribution of the system, we wish to determine whether the latter is entangled or Bell nonlocal.

General methods for entanglement detection, such as the Doherty-Parrilo-Spedalieri (DPS) hierarchy [11], require vast memory resources to operate. More specifically, an implementation of the second level of the DPS hierarchy (i.e., beyond the positivity under partial transposition criterion [27]) in a normal desktop would experience difficulties accepting six-qubit or four-qudit states as input. Similarly, the best linear programming solvers just allow assessing Bell nonlocality in systems with fewer than eight sites (assuming two inputs and two outputs per party). Deciding whether a multipartite distribution or box is compatible with the laws of quantum mechanics is even more complicated: In the simplest Bell scenario of two inputs and two outputs per party, even a characterization of the related almost quantum set [28] is unrealistic for six-partite boxes.

In contrast, the applicability of connector theory for entanglement, Bell nonlocality, or supraquantum nonlocality detection is essentially limited by the computer memory required to store the whole description of the quantum state or box. In a computer with 8 GB of RAM (and MATLAB), this means operating with 13-qubit quantum states and 18-partite boxes.

(b) Given a tensor network description of the density matrix or probability distribution of the system, we wish to determine whether the latter is entangled or Bell nonlocal.

This problem is not hypothetical. Such a compressed description can arise, for instance, as the result of a numerical optimization [13–17]. Alternatively, one could assume that the experimental quantum state prepared in the lab is close to the ground or thermal state of a 1D gapped system with local interactions. In this predicament, the state (and thus, any n -partite distribution it might produce) is known to admit a matrix product operator decomposition with low bond dimension [29]: The parameters specifying such states or distributions can then be determined, as we mention above, through a number of tomographic experiments which grow polynomially on the system size [18,19].

There are currently no known methods for entanglement, Bell nonlocality, or supraquantum nonlocality detection which exploit the promise of a tensor network structure on the input of the problem. Hence, once more we would be limited to study systems of the order of six sites. In fact, all the methods listed above would require, first and foremost,

decompressing the state or multipartite box before inputting it to the corresponding algorithm to detect the property of interest. More worryingly, even if we manage to run the methods above, storing the generated entanglement witnesses, Bell inequalities, or supraquantum Bell inequalities, or proving their soundness, would require, respectively, unreasonable memory and time resources.

Connector theory changes this state of affairs: In the particular case that the state or multipartite box admits an efficient matrix product operator (MPO) decomposition, the time and memory complexity of our connector-based methods scales *linearly* with the system size. This linear complexity lets us access relevant systems made of hundreds of sites with a regular desktop, as we see in the following lines. Furthermore, the entanglement witnesses, Bell inequalities, or supraquantum Bell inequalities output by our method admit an efficient tensor network decomposition, and their soundness can be established with a number of operations linear on the system size.

C. Structure of the paper

The article is organized as follows. In Sec. II, we begin by reviewing the three essential ingredients of connector theory: tensor networks (as efficient representation of quantum-many-body states), convex optimization, and the formalism of GPTs. This section also introduces the graphical notation for tensor networks that is convenient to explain the basics of connector theory, and is used in the rest of the paper. In Sec. III, we explain the use of connector theory for detecting Bell nonlocality as a case study. In Secs. IV and V, we describe how to apply the formalism of connector theory to detect supraquantum nonlocality and entanglement in quantum-many-body systems. Finally, in Sec. VI, we present our conclusions.

II. BACKGROUND

The main objective of this first section is to review the main ingredients used in our construction: tensor networks, techniques from convex optimization theory, and GPTs.

A. Tensor networks

We start by providing a broad introduction to the formalism of tensor networks as it appears in quantum-information theory and condensed-matter physics for efficiently representing quantum-many-body states. For a review, see Ref. [17].

1. Notation and definitions

Broadly speaking, a tensor network is a set of *tensors* that are interconnected or contracted according to a given network. By a tensor, we simply mean a multidimensional array of complex coefficients—an object that generalizes the notion of vectors and matrices. More precisely, a tensor $T_{j_1 j_2 \dots j_n}^{i_1 i_2 \dots i_m}$ is a linear map from the tensor product of a set of

input vector spaces to the tensor product of some output vector spaces,

$$T: (\mathbb{V}_1 \otimes \mathbb{V}_2 \otimes \dots \mathbb{V}_m) \rightarrow (\mathbb{W}_1 \otimes \mathbb{W}_2 \otimes \dots \mathbb{W}_n). \quad (3)$$

The indices i_1, i_2, \dots, i_m label an orthonormal basis in the input spaces $\mathbb{V}_1, \mathbb{V}_2, \dots, \mathbb{V}_m$, respectively, whereas the indices j_1, j_2, \dots, j_n label an orthonormal basis in the output spaces $\mathbb{W}_1, \mathbb{W}_2, \dots, \mathbb{W}_n$, respectively. For convenience, we represent tensors graphically as illustrated in Fig. 1. A tensor is depicted by a shape and its indices are depicted by directed lines emanating from the shape. Input and output indices, which we later need to distinguish, are indicated by attaching incoming and outgoing arrows to the corresponding lines.

One can obtain a new tensor by *contracting* or multiplying together a set of tensors. Tensor contraction generalizes the notion of matrix multiplication. Two matrices M and N can be multiplied to obtain a new matrix $R \equiv MN$. In tensor notation, we write

$$R_j^i = \sum_k M_k^i N_j^k. \quad (4)$$

We graphically depict this by connecting an output index of matrix M with an input index of matrix N , as shown in Fig. 1(d). Of course, the dimension of the output index of M must equal the dimension of the input index of N . The *dimension* of an index is the number of values the index runs over. For example, if R in Eq. (4) is a 2×3 matrix, then the dimension of indices i and j is equal to 2 and 3, respectively.

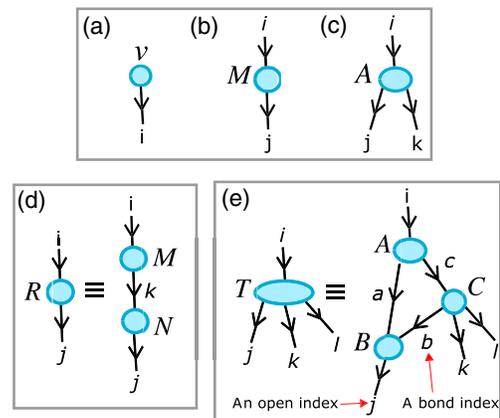


FIG. 1. Graphical representation of tensors and tensor operations. (a) A vector $|v\rangle \equiv \sum_i v_i |i\rangle$. (b) A matrix $M \equiv \sum_{ij} M_{ij} |j\rangle \langle i|$. (c) A three-index tensor $A = \sum_{ijk} A_{ijk} |j\rangle \langle k| \langle i|$. (d) Matrix multiplication, Eq. (4). (e) Example of a more general tensor contraction, Eq. (5). Two types of indices are distinguished in a contraction: *open indices* that emanate only from one tensor and are left uncontracted, and *bond indices* that connect two tensors and get summed over.

A more general tensor contraction is illustrated in Fig. 1(e), where three tensors A , B , and C are contracted to obtain a four-index tensor T ,

$$T_{jkl}^i = \sum_{abc} A_{ac}^i B_j^{ab} C_{bkl}^c. \quad (5)$$

In a contraction, the indices that are left uncontracted are called *open indices*, e.g., i , j , k , and l . On the other hand, indices that are summed over are called *bond indices*, e.g., a , b , and c .

The generic pure quantum state $|\Psi\rangle$ of a large many-body system, e.g., a lattice of qubits, requires specifying 2^N probability amplitudes, where N is the number of qubits:

$$|\Psi\rangle \equiv \sum_{i_1, i_2, \dots, i_N} \Psi_{i_1, i_2, \dots, i_N} |i_1\rangle \otimes |i_2\rangle \otimes \dots \otimes |i_N\rangle. \quad (6)$$

Even more parameters, of the order of 4^N , are needed if the state is mixed. However, often interesting states such as ground states or thermal states of local Hamiltonians contain a limited amount of correlations. This can be exploited to efficiently represent them by decomposing the large quantum-many-body wave function encoded in the exponentially N -index tensor $\Psi_{i_1, i_2, \dots, i_N}$ into a product of small tensors, namely, as a tensor network.

Figure 2 illustrates two popular tensor network decompositions—matrix product states (MPSs) [12,13] and the

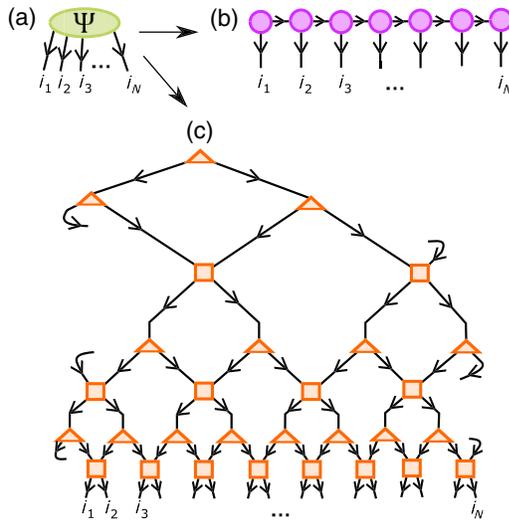


FIG. 2. Examples of popular tensor network decompositions of quantum-many-body states. (a) An N -index tensor Ψ that, e.g., stores the probability amplitudes of a quantum-many-body state in Eq. (6). (b) A MPS decomposition of tensor Ψ . It corresponds to a tensor network composed from tensors with at most three indices. Tensor Ψ is recovered by contracting together the tensor network. (c) A different and more elaborate tensor network decomposition of Ψ , called the MERA, which is composed of three-index tensors (triangles) and four-index tensors (squares). (Once again, tensor Ψ is recovered by contracting together the tensor network.)

multiscale entanglement renormalization *Ansatz* (MERA) [15]—that are used to efficiently represent the ground state of local Hamiltonians acting on a one-dimensional quantum lattice. Later, we propose the use of connector tensor networks that are structurally similar to these decompositions.

Here, the sites of the lattice correspond to the open indices of tensor networks, while the bond indices carry the entanglement and correlations in the state. The dimensions of the bond indices generally indicate the amount of entanglement and correlations in the state: A larger bond dimension generally corresponds to larger entanglement and correlations. By contracting together all the tensors in a tensor network, one recovers the probability amplitudes in Eq. (6). The choice of the network pattern of the tensor network decomposition of a given state is dictated by the specific structure of entanglement in the state.

In a condensed-matter context, one typically uses these tensor networks as an *Ansatz* for the unknown ground state (or a low-energy subspace) of a given Hamiltonian and determines the tensors numerically by means of, say, a variational energy minimization. The maximum bond dimension in a tensor network *Ansatz* determines both the cost of the numerical optimization and the accuracy of the approximation. Here, we propose a novel application of tensor networks in the context of certification of relevant quantum properties, for instance, as witnesses for entanglement and nonlocality. Moreover, and as describe below, these tensor networks can be understood as measurements in a general probabilistic theory, thus, extending the formalism of tensor networks beyond quantum theory.

2. Implementing tensor networks in MATLAB

In this section, we briefly describe how to implement basic tensor network operations—such as the ones that are needed to implement the numerical optimizations described in this paper—in MATLAB. (The numerical plots presented in this paper are obtained by using a MATLAB code.) This section may be skipped on first reading or by the reader familiar with implementing tensor network code.

In MATLAB, tensors simply correspond to multidimensional numeric arrays. For instance, the command

```
A=rand(2,3,5)
```

creates a three-index tensor A initialized with random components and with index dimensions two, three, and five, respectively. The three indices are ordered—the index in the first position has dimension two, index in the second position has dimension three, and the final index has dimension five. A new tensor Z may be obtained from A by changing the order of the indices by calling the function `permute()`. For example,

```
Z=permute(A,[1 3 2]).
```

Here, tensor Z is the same as A but with the position of the second and third index swapped.

Another way to obtain a new tensor from A is by *reshaping* its dimensions. For example,

$$M = \text{reshape}(A, [6 \ 5])$$

produces a 6×5 matrix M by reorganizing the components of tensor A . Reshaping generalizes the more familiar operation of *vectorizing* a matrix. For example, one can vectorize the matrix M ,

$$M = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

by concatenating the columns of M in a single row to obtain the vector $v = [1 \ 3 \ 2 \ 4]$. In MATLAB, this is implemented by

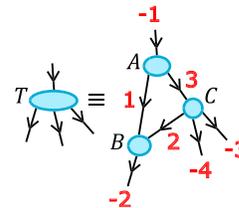
$$v = \text{reshape}(M, [1 \ 4]).$$

The most common tensor network operation is contracting a given tensor network to obtain a single tensor, e.g., Eq. (5) [Fig. 1(e)]. A straightforward, but possibly slower way, to contract a tensor network is to simply loop and sum over all the bond indices and multiply the tensor components.

However, a more efficient way to contract a pair of tensors is by first reshaping them appropriately to obtain matrices, multiplying the matrices, and then reshaping the resulting matrix to obtain the output tensor. (This approach is particularly fast when using fast matrix multiplication libraries such as LAPACK, which is also used by MATLAB). A tensor network consisting of more than two tensors may then be contracted by performing a sequence of pairwise tensor contractions. But contracting a large tensor network may require some care since different sequences of pairwise contractions will generally incur different computational cost. In particular, an optimal sequence of contractions can incur a substantially lower computational cost. A useful heuristic estimate of the computational cost of carrying out a tensor network contraction is simply given by the product of the dimension of all the indices in the contraction (each bond index is counted only once). See Ref. [30] for a useful algorithm and accompanying MATLAB code to determine optimal contraction sequences for tensor networks.

We use an open source function called `NCON()` for all tensor network contractions [31]. Here we briefly illustrate how this function can be used to perform the contraction shown in Fig. 1(e). For more details, examples, and the source code, see Ref. [31]. We begin by labeling each index in the contraction with an integer. Open indices are labeled with negative values, while bond indices are labeled with positive values. The absolute value of an open index corresponds to the location of the index in the output

tensor. On the other hand, the values of the bond indices specify the sequence of pairwise tensor contractions. Namely, the bond indices are contracted in ascending order. For example, one way to number the indices in Fig. 1(e) is shown below (in red):



In this case, the tensors A and B are first contracted to obtain an intermediate tensor X , which is then contracted with C to obtain the output tensor T . This tensor network contraction can be implemented with `ncon()` as `T=ncon({A,B,C}, {[−1 1 3],[1 2 −2],[2 3 −4 −3]})`. Here, the first input argument is a list (a “cell array” in MATLAB) of the tensors, and the second argument is a list of vectors, one for each tensor. Each vector in this list specifies the indices of the corresponding tensor in the order that they appear in the tensor.

B. Convex optimization theory

Let X be a vector space, and let $\mathcal{X} \subset X$ be a convex subset thereof. The goal of convex optimization is to solve problems of the form

$$\begin{aligned} \min f(\bar{x}), \\ \text{such that } \bar{x} \in \mathcal{X}, \end{aligned} \quad (7)$$

where f is a convex function, i.e., $f[p\bar{x}_1 + (1-p)\bar{x}_2] \leq pf(\bar{x}_1) + (1-p)f(\bar{x}_2)$, for $\bar{x}_1, \bar{x}_2 \in \mathcal{X}$, $0 \leq p \leq 1$. Any vector of variables $\bar{x} \in X$ satisfying $\bar{x} \in \mathcal{X}$ is said to be a *feasible point*.

Linear programming (LP) [25] is a branch of convex optimization where \mathcal{X} is a polytope (a convex set defined by a finite number of linear inequalities), and f is a linear function of the variables $\bar{x} \in \mathbb{R}^n$ of the optimization problem. Linear programming is thus concerned with optimization problems of the sort

$$\begin{aligned} \min \bar{c} \cdot \bar{x}, \\ \text{such that } A\bar{x} \geq \bar{b}. \end{aligned} \quad (8)$$

Here, the $m \times n$ matrix A , $\bar{b} \in \mathbb{R}^m$, and $\bar{c} \in \mathbb{R}^n$ are the inputs of the problem. For any pair of vectors \bar{y}, \bar{z} of identical size, the notation $\bar{y} \geq \bar{z}$ indicates that $y_i - z_i \geq 0$ for all i . As we see, linear programming is an instrumental tool for nonlocality detection.

In order to deal with entanglement and quantum nonlocality, we use a more sophisticated tool, namely, *semidefinite programming* (SDP) [26]. A semidefinite program is an optimization problem of the form

$$\begin{aligned} & \min \bar{c} \cdot \bar{x}, \\ & \text{such that } F_0 + \sum_{i=1}^n x_i F_i \geq 0. \end{aligned} \quad (9)$$

This time, the $m \times m$ matrices $F_0, \{F_i\}$ and the vector \bar{c} constitute the problem input. Beware of the change in notation: If A is a square matrix, then $A \geq 0$ is used to denote that A is positive semidefinite; i.e., it is self-adjoint and all its eigenvalues are non-negative.

There exist free solvers available to solve both linear and semidefinite programs. These solvers exploit convex optimization theory to provide, not only an approximate solution of the problem, but also rigorous bounds on how this figure differs from the exact value. For linear programs of any size, we recommend the MATLAB solver GUROBI [32]; the packages SEDUMI [33] and MOSEK [34] are appropriate, respectively, to solve small and large instances of semidefinite programs. We recommend not to work with these solvers directly, but through general optimization MATLAB packages, such as YALMIP [35] or CVX [36,37]. The advantage of using either of these packages is that the user does not need to write the programs in the standard form (8), (9): It is enough to indicate what linear or semidefinite constraints the variables \bar{x} of the problem must be subjected to.

Unless otherwise specified, in all our numerical computations we make use of YALMIP [35] in combination with GUROBI [32] (for LPs) or MOSEK [34] (for SDPs).

C. Generalized probabilistic theories

The formalism of GPTs was conceived to reason about physical theories beyond quantum physics. In a sense, it conveys an operational description of what one can do within a physical theory, but without a correspondence principle to relate the mathematical formalism of the theory to the instruments of an experimental workshop. Viewed as a GPT, quantum physics is a theory where each system type is labeled by a natural number D (the Hilbert space dimension). Normalized (subnormalized) states of a system of dimension D are described by $D \times D$ complex positive semidefinite matrices with trace (smaller than or equal to) 1; measurements are defined by positive-operator-valued measures (POVMs), and transformations, by completely positive trace-preserving maps. Also, states of a bipartite system of dimensions D, D' are in one-to-one correspondence with the states of a system of dimension DD' .

Analogously, a general GPT is specified by a list of possible system types, together with composition rules specifying which system type describes the combination of several other types. In a GPT, the state of a given system of type S is identified with a vector [38] \bar{s} living in a space \mathcal{H}_S . The set of possible states of S corresponds to a convex set $\mathcal{C}_S \subset \mathcal{H}_S$. For every system S , we assume the existence of a vector $\bar{e}_S \in \mathcal{H}_S$, the *unit effect*, whose scalar product with

any state returns the *norm* of the state, or the probability that the state was successfully prepared. It follows that, for all $\bar{s} \in \mathcal{C}_S$, $0 \leq \bar{e}_S^T \cdot \bar{s} \leq 1$. Moreover, $\bar{s} \in \mathcal{C}_S$ is a deterministic preparation if and only if $\bar{e}_S^T \cdot \bar{s} = 1$. Sometimes, for simplicity, we use the notation $E(\bar{s}) = \bar{e}_S^T \cdot \bar{s}$.

In the following we consider only GPTs which satisfy local tomography [20]. In our language, this implies that $\mathcal{H}_{S \otimes S'} = \mathcal{H}_S \otimes \mathcal{H}_{S'}$, where $S \otimes S'$ denotes the composition of systems S, S' . To recover the marginal state \bar{v}_S of system S from the joint state $\bar{v}_{SS'}$ of systems S, S' , we apply the unit effect over the space $\mathcal{H}_{S'}$, i.e., $\bar{v}_S = (\mathbb{I}_S \otimes \bar{e}_{S'}^T) \cdot \bar{v}_{SS'}$.

Any (nondeterministic) transformation of system S into another system of type S' corresponds to a linear map $W: \mathcal{H}_S \rightarrow \mathcal{H}_{S'}$ with the property that, for any system type R ,

$$\begin{aligned} \bar{v}_{S \otimes R} \in \mathcal{C}_{S \otimes R} & \Rightarrow (W \otimes \mathbb{I}_R) \cdot \bar{v}_{S \otimes R} \in \mathcal{C}_{S' \otimes R}, \\ E(W\bar{v}_S) & \leq E(\bar{v}_S). \end{aligned} \quad (10)$$

The last line demands the linear transformation to be norm decreasing. That way, $E(W\bar{s})$ can be interpreted as the probability that the physical process represented by W is successful when acting over the state \bar{s} .

When the output system S' has dimension one, the transformation W corresponds to a vector $\bar{w} \in \mathcal{H}_S$. Then, \bar{w} represents an *effect*, or the result of a dichotomic measurement. For any state $\bar{s} \in \mathcal{C}_S$, the probability that the event signified by the effect \bar{w} occurs is then given by $\bar{w}^T \cdot \bar{s}$. More generally, a d -outcome measurement M corresponds to a set of effects $\{\bar{w}_a\}_{a=1}^d$ with the property $\sum_{a=1}^d \bar{w}_a = \bar{e}_S$. The probability $p(a|\bar{s}, M)$ of observing outcome a when measurement M is conducted on state $\bar{s} \in \mathcal{C}_S$ is $p(a|\bar{s}, M) = \bar{w}_a^T \cdot \bar{s}$.

Effects must not be confused with *witnesses*. A normalized witness is a vector $\bar{w} \in \mathcal{H}_S$ with the property $0 \leq \bar{w} \cdot \bar{v} \leq 1$ for all $\bar{v} \in \mathcal{C}_S$. An effect has, in addition, the property that $(\bar{w}^T \otimes \mathbb{I}_T) \cdot \bar{v}_{ST}$ is a state in \mathcal{C}_T if $\bar{v}_{ST} \in \mathcal{C}_{ST}$. While all effects are normalized witnesses (in general, GPTs), not all normalized witnesses are effects.

To fix ideas, consider the following example of a nonquantum GPT. This GPT, that hereafter we call Newton's world, is the result of modeling discrete classical physics in the GPT framework. Intuitively, each physical system in Newton's world represents a random variable a . The type of system corresponds to the number A of values that the corresponding random variable a can take. The composition of two systems of types A, A' is a system of type AA' : This follows from the fact that the cardinality of a joint variable (a, a') is the product of the cardinalities of variables a, a' . The state space of a system of type A is \mathbb{R}^A , and the corresponding set of states \mathcal{C}_A is the set of subnormalized probability distributions for a random variable of cardinality A . Namely, $\mathcal{C}_A = \{\vec{\lambda} \in \mathbb{R}^A : \lambda_a \geq 0, \sum_a \lambda_a \leq 1\}$. The unit effect is the "all 1's" vector

$\bar{e} = (1, 1, \dots)^T$. Note that, for $\bar{\lambda} \in \mathcal{C}_A$, $\bar{e}^T \cdot \bar{\lambda} = \sum_a \lambda_a$, and so $0 \leq E(\bar{\lambda}) \leq 1$.

In Newton's world, the set of transformations mapping a system of type A to a system of type B is in one-to-one correspondence with the set of $B \times A$ substochastic matrices (i.e., matrices W such that $W_{ba} \geq 0$, $\sum_b W_{ba} \leq 1$). Let us next prove why any such matrix satisfies criterion (10). Let $\bar{\lambda} \in \mathcal{C}_{A \otimes C}$, and define $\bar{\mu} = (M \otimes \mathbb{I}_C) \bar{\lambda}$. Since $\bar{\lambda}$, $\bar{\mu}$ are vectors living in a tensor product of two vector spaces, we label their components with a double index. That is, $\bar{\lambda} = \sum_{a,c} \lambda_{ac} |a\rangle_A \otimes |c\rangle_C$, $\bar{\mu} = \sum_{b,c} \lambda_{bc} |b\rangle_B \otimes |c\rangle_C$. Now, from the entrywise positivity of W , it follows that all the components of $\bar{\mu}$ are non-negative. Furthermore,

$$\begin{aligned} E(\bar{\mu}) &= \sum_{b,c} \mu_{bc} = \sum_{a,c} \left(\sum_b W_{ba} \right) \lambda_{ac} \\ &\leq \sum_{a,s} \lambda_{ac} = E(\bar{\lambda}). \end{aligned} \quad (11)$$

We thus have that $\bar{\mu}$ is a state of the composite system $B - C$ with norm smaller than or equal to $\bar{\lambda}$. The set of measurements in Newton's world can be similarly shown to correspond to collections of vectors $\{\bar{w}_a\}_a$ with non-negative components, and adding up to the unit effect \bar{e}_S .

If the GPT formalism could describe only classical and quantum physics, then it would not be of much use for quantum foundations. However, as shown in Refs. [39,40], there exist GPTs which cannot be embedded in any quantum theory (and, by inclusion, in any classical theory). In the following sections, we exploit this feature of the GPT framework to obtain a physical intuition for otherwise abstract optimization problems.

III. A CASE STUDY: BELL NONLOCALITY

Once the main building blocks of the construction have been presented, in what follows we illustrate how to use connector theory to detect relevant properties of large systems. We do so by means of different relevant examples, starting with the detection of Bell nonlocality. We explain the connector formalism in this scenario, characterize connectors for detecting Bell nonlocality, and the techniques for optimizing them. After nonlocality in the following sections, we also show how to use connectors in the context of supraquantum nonlocality and entanglement detection.

Consider an m -partite Bell scenario, where each party interacts with a black box, to which it can input a symbol x and then obtain an output a ; see Fig. 3. The operational description of this box is given by the probabilities $P(a_1, \dots, a_m | x_1, \dots, x_m)$. We represent these probabilities also as a *tensor* with m incoming and m outgoing indices, namely,

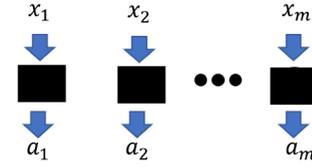


FIG. 3. Bell test. A Bell test consists of m -distant parties performing different measurements on their systems. The choice of measurements by party i is labeled by x_i and the corresponding output by a_i . The different systems are seen as back boxes, each producing the classical output a_i after receiving the classical input x_i . The whole scenario is described by the conditional probability distribution $P(a_1, \dots, a_m | x_1, \dots, x_m)$.

$$P_{a_1, \dots, a_m}^{x_1, \dots, x_m} \equiv P(a_1, \dots, a_m | x_1, \dots, x_m), \quad (12)$$

and graphically represent this tensor as shown in Fig. 4(a). We assume that this box is *nonsignaling*; i.e., for any $k \in \{1, \dots, m\}$, the marginal probability distribution

$$\sum_{a_k} P_{a_1, \dots, a_m}^{x_1, \dots, x_m} \quad (13)$$

does not depend on x_k .

We are asked whether box P is *Bell local*, i.e., whether the correlations $P_{a_1, \dots, a_m}^{x_1, \dots, x_m}$ can be expressed as

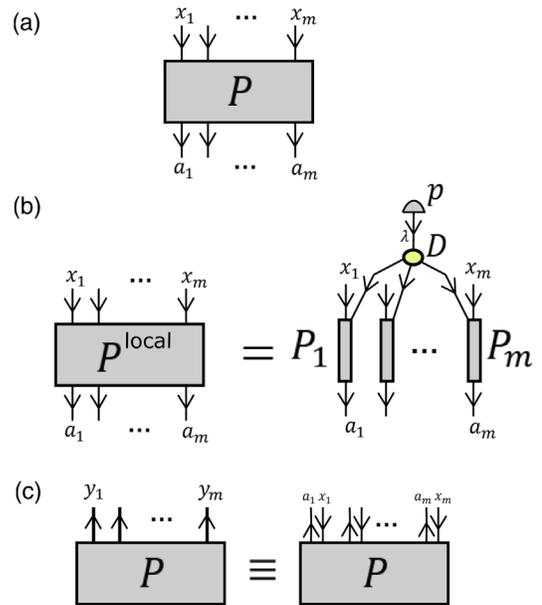


FIG. 4. Boxes as tensors. (a) A nonsignaling box as a tensor $P_{a_1 a_2, \dots, a_m}^{x_1 x_2, \dots, x_m}$. (b) A Bell-local box P^{local} decomposes as shown in terms of three-index tensors P_1, P_2, \dots, P_m , each of which is a conditional probability distribution. Here, D is an m -index identity tensor (also called a *copy tensor*), namely, a tensor whose only nonzero components are $D_{ii, \dots, i} = 1$, and p_λ is a vector of probabilities $p(\lambda)$. (d) For convenience, we combine each pair (a_k, x_k) of input and output indices of a box into a single outgoing index $y_k \equiv (a_k, x_k)$.

$$P_{a_1, \dots, a_m}^{x_1, \dots, x_m} = \sum_{\lambda} p_{\lambda}(P_1)_{a_1}^{x_1, \lambda}, \dots, (P_m)_{a_m}^{x_m, \lambda}, \quad (14)$$

where $p(\lambda)$, $(P_j)_{a_j}^{x_j, \lambda} \equiv P_j(a_j|x_j, \lambda)$ are arbitrary probability distributions; see Fig. 4(c). We also refer to Bell-local boxes as *classical boxes*.

To help us answer this question, we introduce a GPT, that we call LOC world. Intuitively, each system in LOC world corresponds to a multipartite box, with a number of possible inputs and outputs. A general system is thus labeled by a vector of natural numbers of the form $[O_1, \dots, O_m, I_1, \dots, I_m]$, which denotes that the k th party's box has I_k inputs and O_k possible outputs. In LOC world, the set of states of any system of type $[O_1, \dots, O_m, I_1, \dots, I_m]$ corresponds to the set of un-normalized probabilities $P_{a_1, \dots, a_m}^{x_1, \dots, x_m}$ of the form Eq. (14). We define the norm of a state P in LOC world as $E(P)$,

$$E(P) = \sum_{a_1, \dots, a_m} P_{a_1, \dots, a_m}^{0, \dots, 0}. \quad (15)$$

Valid transformations W in LOC world correspond to linear maps which, acting on part of a classical box P , return a classical box $P' = (W \otimes \mathbb{I})P$. To be interpreted as non-deterministic transformations, such maps must satisfy the condition $E(P') \leq E(P)$. We call such transformations *connectors* (in LOC world).

A general $m \rightarrow q$ connector takes as input a system of type $[O_1, \dots, O_m, I_1, \dots, I_m]$ and returns as output a system of type $[O'_1, \dots, O'_q, I'_1, \dots, I'_q]$. Any $m \rightarrow q$ connector can be interpreted as a tensor of the form $W_{y_1, \dots, y_m}^{y'_1, \dots, y'_q}$ (notice that all indices are doubled indices). This tensor is graphically represented as shown in Fig. 5(a). Acting with W on an $(m+r)$ -partite box P results in a $(q+r)$ -partite box P' . The action of W over the first m systems of P can be expressed as the contraction of the first m outgoing legs of P with the incoming legs of C :

$$P'_{y'_1, \dots, y'_q, y_{m+1}, \dots, y_{m+r}} \equiv \sum_{y_1, \dots, y_m} W_{y_1, \dots, y_m}^{y'_1, \dots, y'_q} P_{y_1, \dots, y_{m+r}} \quad (16)$$

depicted in Fig. 5(b).

Now, suppose that, given a nonsignaling box P for nine parties, we apply to it connectors W^1, W^2, \dots, W^8 as depicted in Fig. 5(c). Consider the case in which the output system of the topmost connector W^8 is of type $[1, 1]$, resulting on a joint probability distribution for a one-input and one-output situation, that is, a probability $0 \leq p \leq 1$. Then, the action of all the connectors W^1, W^2, \dots, W^8 can be interpreted as a measurement W in LOC world. Clearly, if $W(P) < 0$ or $W(P) > 1$, it follows that P did not belong to the class of states of LOC world. In other words, P is nonlocal.

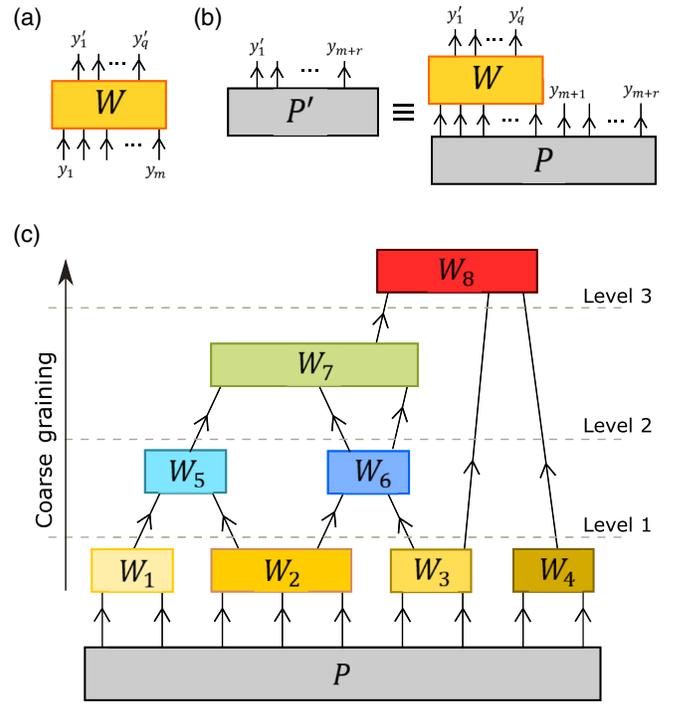


FIG. 5. Connectors. (a) Graphical representation of an $m \rightarrow q$ connector C . (b) Contraction of C with an $(m+r)$ -partite nonsignaling box P results in a $(q+r)$ -partite nonsignaling box P' . (c) Connectors can be used to coarse grain a box to a manageable size. The coarse-grained box at any “level” (indicated by a dashed line) is obtained by contracting the original box P with all the connectors up to that level. Here, at level 3 we obtain a three-party box, whose nonlocality can be probed exactly with a known three-party witness W_8 (red).

This observation is the basis of connector theory. Namely, any tensor network of connectors that has

- (1) no outgoing arrows and
- (2) no cycles

defines a normalized Bell inequality [41], i.e., a linear functional with the property that $0 \leq W(P) \leq N(P)$ for all classical boxes P . It is easy to see that any attempt at enlarging the set of connectors with extra tensors will result in the loss of this property.

Note that, in principle, there is a more general strategy to detect that our original tensor does not belong to the set of states of the theory. Namely, use connectors to coarse grain the system and then apply a witness to prove the non-physicality of the resulting coarse-grained network. For instance, in Fig. 5(c) we would replace the connector W_8 by an arbitrary normalized witness. Since the set of normalized witnesses contains the set of effects in a GPT, this method should allow us to detect more instances of Bell nonlocality. However, in LOC world, as well as in the other two GPTs we define in the following for supraquantum and entanglement detection, normalized witnesses happen to be connectors as well, so it is enough when we consider connector tensor networks.

A. Characterizing connectors in LOC world

Because of the structure (14) of the set of classical boxes, LOC world has the convenient property that any linear map fulfilling condition (10) with $T = \emptyset$ constitutes a valid transformation. That is, if a given map in LOC world is valid when acting on a system, it is also a valid map when acting on parts of a larger system. Let us see why.

Any box of the form P_a^x can be expressed as

$$P_a^x = \sum_{\bar{a}} p_{\bar{a}} P_a^{x,\bar{a}}, \quad (17)$$

where $p_{\bar{a}}$ is a probability distribution over $\{1, \dots, d\}^n$ and $P_a^{x,\bar{a}} = \delta_a^{x,\bar{a}}$ are deterministic boxes. Absorbing $p_{\bar{a}}$ in the definition of the hidden variable λ in Eq. (14), we have that an m -partite box is classical if and only if it can be expressed as a convex combination of m -partite deterministic boxes of the form $P_{a_1, \dots, a_m}^{x_1, \dots, x_m} = \prod_{k=1}^m P_{a_k}^{x_k, \bar{a}^k}$. Each of these deterministic boxes is an *extreme point*, namely, a point that cannot be decomposed as a convex decomposition of other points within the set of classical boxes.

Given an $m \rightarrow q$ connector Ω , deciding whether it satisfies Eq. (10) amounts to verifying that $\Omega \otimes \mathbb{I}_T$ maps deterministic boxes to classical boxes. The general result then follows by applying convexity. Now, suppose that Ω satisfies Eq. (10) with $T = \emptyset$, consider an arbitrary $(m+r)$ -partite deterministic box $P \equiv P^1 P^2, \dots, P^{m+r}$ and let Ω act over the first m systems (equivalently, take $T = \{m+1, \dots, m+r\}$). The result will be the (in general, un-normalized) box $P' = Q P^{m+1}, \dots, P^{m+r}$, with $Q = \Omega(P^1, \dots, P^m)$. By hypothesis, Q is classical, and therefore, so is P' . We thus conclude that $\Omega \otimes \mathbb{I}_T$ will map classical boxes to classical boxes for all T .

Note that this property does not hold in quantum mechanics. Indeed, there exist positive linear maps Ω , like the transposition map $\Omega(\rho) = \rho^T$, which satisfy $\Omega(\rho) \geq 0$ for all $\rho \geq 0$ (i.e., they are positive), despite the fact that $\Omega \otimes \mathbb{I}$ is not positive [42,43]. In other words, in the case of local correlations, it is impossible to find maps that are analog to the positive but not completely positive maps for quantum states.

We thus reduce the problem of characterizing connectors in LOC world to the problem of identifying those transformations which map extreme classical boxes to classical boxes. Since classical boxes form a convex set with a finite number of extreme points, it follows that characterizing or conducting linear optimizations over connectors in LOC world can be cast as a LP. See the Appendix A for a detailed description of the LPs, together with some tips to reduce their time and memory complexity.

What is the form of connectors in LOC world? Some of them correspond to *wirings* [44], namely, transformations that correspond to feeding the outputs of some parties to the inputs of some other parties. To fix ideas, consider connectors from $(2,2,2,2)$ systems to $(2,2)$ systems, with

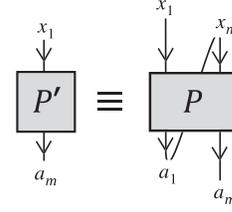


FIG. 6. Example of a wiring. A two-party box P is mapped into a single-party box P' by using as input for the second system the output generated by the first.

input indices $(a_1, x_1), (a_2, x_2)$ and output indices (b, y) . If we denote by P, P' the input and output boxes, then a possible wiring would be given by

$$P_b^{y'} = \sum_a P_{a,b}^{y,a} = \sum_{a,a',x,x'} W_{a,a',x,x'}^{b,y} P_{a,a'}^{x,x'}, \quad (18)$$

with $W_{a,a',x,x'}^{b,y} = \delta_{y,x} \delta_{b,a'} \delta_{a,x'}$. This wiring is just the result of inputting y on the first part of P , reading the result a , and using it as an input in the second part of P . The final outcome of such an effective box is the output b produced by the second part of P ; see Fig. 6. Although wirings map nonsignaling boxes to nonsignaling boxes—and therefore, they are examples of connectors—the contraction of any network of wirings with a nonsignaling box always results in a non-negative number. This means that wirings, by themselves, cannot be used to detect Bell nonlocality. Fortunately, there exist more general connectors in LOC world, as we show next.

B. Connectors built from Bell inequalities

Consider the following two-to-one connector defined by the relations

$$\begin{aligned} C_{axd'x'}^{0,0} &= C_{axd'x'}^{1,0} = (E - C)_{axd'x'}, \\ C_{a,x,d',x'}^{0,1} &= C'_{axd'x'}, \quad C_{a,x,d',x'}^{1,1} = (E - C')_{axd'x'}, \end{aligned} \quad (19)$$

where C, C' correspond to normalized forms of the Clauser-Horne-Shimony-Holt (CHSH) Bell inequality, i.e.,

$$\begin{aligned} C_{axd'x'} &\equiv -\delta_a^0 \delta_{a'}^0 (\delta_x^0 \delta_{x'}^0 + \delta_x^1 \delta_{x'}^0 + \delta_x^0 \delta_{x'}^1 - \delta_x^1 \delta_{x'}^1) \\ &\quad + \delta_x^0 \delta_{x'}^0 (\delta_a^0 + \delta_{a'}^0), \end{aligned} \quad (20)$$

$$\begin{aligned} C'_{axd'x'} &\equiv -\delta_a^0 \delta_{a'}^0 (-\delta_x^0 \delta_{x'}^0 + \delta_x^1 \delta_{x'}^0 + \delta_x^0 \delta_{x'}^1 + \delta_x^1 \delta_{x'}^1) \\ &\quad + \delta_x^1 \delta_{x'}^1 (\delta_a^0 + \delta_{a'}^0). \end{aligned} \quad (21)$$

It can be verified that $0 \leq C(P), C'(P) \leq 1$ for any bipartite classical probability distribution P with two inputs and two outputs. It follows that, for any classical box P , the new box $P' = C(P)$ will be such that $P_b^{y'} \geq 0$, for $b, y = 0, 1$ and

$\sum_b P_b^{xy} = E(P)$. This connector corresponds to a deterministic transformation in LOC world.

How would we implement transformation C in practice? C is neither a wiring nor a convex combination thereof: This follows from the fact that, applied over any box violating the CHSH inequality, it will return a “box” with negative probabilities. Now, suppose that the input box P is indeed classical. This implies that, hidden within P , there exist variables $(a_0, a_1, a'_0, a'_1) \in \{0, 1\}^4$ which will determine the outcomes of the box: If we input $x, y \in \{0, 1\}$, we will obtain the outputs a_x, a'_y . The values of (a_0, a_1, a'_0, a'_1) can change every time we initialize the box; we assume that they are distributed according to a measure $\mu(a_0, a_1, a'_0, a'_1)$.

Now, consider the functions

$$\begin{aligned} f(a_0, a_1, a'_0, a'_1) &= -\bar{a}_0 \bar{a}'_0 - \bar{a}_1 \bar{a}'_0 - \bar{a}_0 \bar{a}'_1 + \bar{a}_1 \bar{a}'_1 + \bar{a}_0 + \bar{a}'_0, \\ g(a_0, a_1, a'_0, a'_1) &= \bar{a}_0 \bar{a}'_0 - \bar{a}_1 \bar{a}'_0 - \bar{a}_0 \bar{a}'_1 - \bar{a}_1 \bar{a}'_1 + \bar{a}_1 + \bar{a}'_1, \end{aligned} \quad (22)$$

where $\bar{c} \equiv 1 - c$. It is easy to check that $f(a_0, a_1, a'_0, a'_1), g(a_0, a_1, a'_0, a'_1) \in \{0, 1\}$ for all $a_0, a_1, a'_0, a'_1 \in \{0, 1\}$. Moreover,

$$\begin{aligned} \sum_{a_0, a_1, a'_0, a'_1} \mu(a_0, a_1, a'_0, a'_1) f(a_0, a_1, a'_0, a'_1) &= \mathcal{C}(P), \\ \sum_{a_0, a_1, a'_0, a'_1} \mu(a_0, a_1, a'_0, a'_1) g(a_0, a_1, a'_0, a'_1) &= \mathcal{C}'(P). \end{aligned} \quad (23)$$

To implement C in the lab over a local distribution P , it suffices to set up a device inside the box that can read the values a_0, a_1, a'_0, a'_1 . On input $y = 0$ ($y = 1$), the device would return $b = 0$, if $f(a_0, a_1, a'_0, a'_1) = 1$ ($g(a_0, a_1, a'_0, a'_1) = 1$), and $b = 1$, otherwise. Obviously, such an operation is just possible if P is a classical box to begin with. Inside a quantum box, for instance, a_0, a_1 could correspond to the outcomes of noncommuting measurements. Therefore, we could not have simultaneous access to them and the above scheme would be unrealizable.

The fact that we use a Bell inequality to devise a $2 \rightarrow 1$ connector is not coincidental. Actually, all $m \rightarrow 1$ connectors can be related to Bell inequalities. Let \mathcal{B} be the set of m -partite classical boxes, and let \mathcal{B}' be its dual, i.e., the set of linear functionals U which map any box inside \mathcal{B} to a non-negative number. Then, all nondeterministic $m \rightarrow 1$ connectors W in LOC world are of the form

$$W_{c|z} \in \mathcal{B}', \quad \sum_c W_{c|z} = W_0, \quad E - W_0 \in \mathcal{B}'. \quad (24)$$

Now, it turns out that \mathcal{B}' is in one-to-one correspondence with the set of Bell inequalities. Indeed, let $B(a_1, \dots, a_m, x_1, \dots, x_m), K \in \mathbb{R}$ be such that

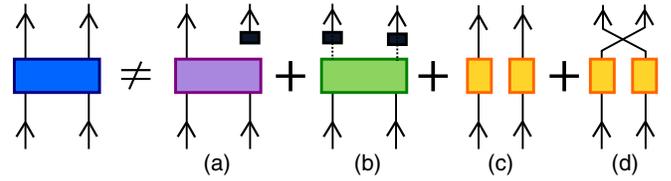


FIG. 7. Example of a nontrivial $2 \rightarrow 2$ connector. The $2 \rightarrow 2$ connector identified in Appendix E cannot be decomposed as a combination of nondeterministic transformations of the following forms: (a) a $2 \rightarrow 1$ connector followed by a preparation, (b) a (global) transformation of the bipartite box into a probability distribution, which we use as a local hidden variable model to build a new box, (c) local mappings on both boxes, and (d) local mappings followed by swapping the two parties.

$$\sum_{\bar{a}, \bar{x}} B_{\bar{x}}^{\bar{a}} P_{\bar{a}}^{\bar{x}} \geq K \quad (25)$$

for all classical boxes $P_{a_1, \dots, a_m}^{x_1, \dots, x_m}$. Then, the linear functional U given by $U(P) \equiv \sum_{\bar{a}, \bar{x}} B_{\bar{x}}^{\bar{a}} P_{\bar{a}}^{\bar{x}} - KE(P)$ satisfies $U(P) \geq 0$ for all classical boxes.

It is worth remarking that the notion of composing $m \rightarrow 1$ connectors to form new Bell inequalities is implicit in the work of Wu *et al.* [45]. There, the authors propose a scheme to generate a new $(m + 1)$ -partite Bell inequality in the two-input and two-output Bell scenario, given two m -partite Bell inequalities. In our language, their scheme can be interpreted as a contraction between an $m \rightarrow 1$ connector and the CHSH inequality [46].

To our knowledge, though, $m \rightarrow m'$ connectors have never been considered in Bell nonlocality, so we cannot relate them to past literature on the subject. A preliminary exploration of this class of transformations revealed rather intriguing objects. In this regard, in Appendix E we present a $2 \rightarrow 2$ connector that does not admit a decomposition in terms of nondeterministic $1 \rightarrow 1$ and $2 \rightarrow 1$ connectors; see Fig. 7.

C. Applications

Now that we have nontrivial connectors, the next step is to contract them to generate new Bell inequalities. One possibility is to take C in Eq. (19) and contract multiple copies thereof according to a tree network; see Fig. 8.

How useful are these new inequalities? As it turns out, computing the minimum value of an arbitrary Bell inequality under nonsignaling distributions can be cast as a linear program. This allows us to calculate, numerically, the corresponding maximal violations of each “CHSH tree,” which seem to increase with the number of parties. Note that this value is a meaningful quantity, since all CHSH trees are normalized by construction.

However, the ultimate goal of Bell nonlocality detection is not to devise arbitrary Bell inequalities, but to detect the nonclassicality of specific experimental systems. In this context, we now discuss two applications. First, detection

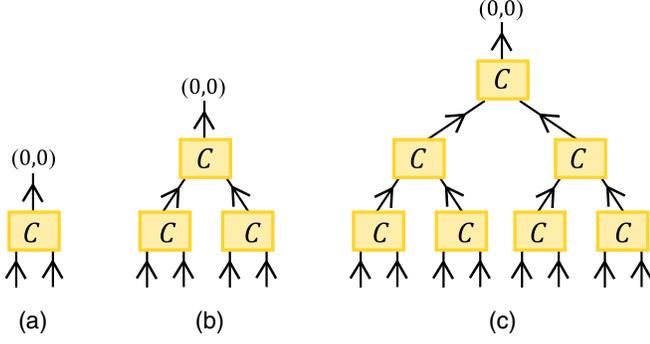


FIG. 8. CHSH trees. Contracting many copies of connector (19) according to a tree network and inputting (0,0) on the top connector, we obtain Bell inequalities for $n = 2, 4, 8$ parties. The minimum nonsignaling value of inequalities (a), (b), (c) are, respectively, $-0.5, -1.5, -7.5$.

of nonlocality in a experimental setup where the actual preparation of the underlying quantum state is known. And second, detection of nonlocality for more general boxes.

1. Nonlocality detection in many-body states

We find that connector theory provides a simple heuristic that relates the detection of nonlocality in a particular quantum experimental setup with the actual preparation of the underlying quantum state. Consider a scenario where several copies of the maximally entangled state $|\phi\rangle \equiv (e^{i\pi\sigma_y/8} \otimes \mathbb{I}_2)(1/\sqrt{2})(|0,0\rangle + |1,1\rangle)$ are acted with random two-qubit unitaries U^1, \dots, U^m ; see Fig. 9 (left). The resulting many-body state is distributed among $2n$ parties, who probe it with Pauli measurements, thus, obtaining a $2m$ -partite nonsignaling box P with three inputs and two outputs at each site. Our goal is to certify that the P is Bell nonlocal.

Denote by P' the three-input or two-output box that results when we distribute $|\phi\rangle$ among two parties and allow

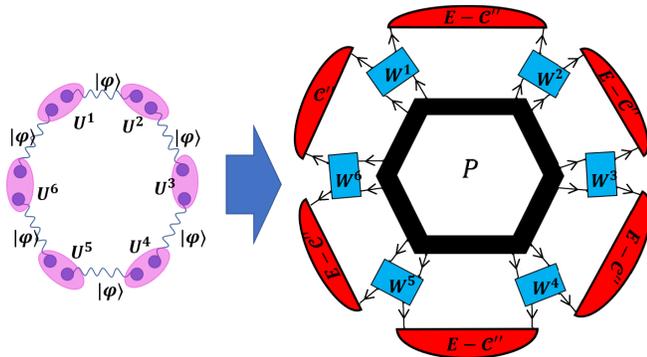


FIG. 9. Left: A $2m$ -partite quantum state built by acting with random two-qubit unitaries U_1, \dots, U_m over many copies of the bipartite state $|\phi\rangle$. Right: Contraction of $2 \rightarrow 2$ and $2 \rightarrow 1$ connectors used to detect the Bell nonlocality of the associated $2m$ -partite box (in black).

each to measure its qubit with Pauli operators. Since Pauli measurements form a complete operator basis, we can identify any two-qubit operator U with the way it transforms linear combinations of $\sigma_i \otimes \sigma_j$. It follows that there exist matrices V^1, \dots, V^m such that $P = (V_{2,3}^1 \otimes V_{4,5}^2, \dots, \otimes V_{2m,1}^m)(P')^{\otimes m}$. On the other hand, P' violates one of the forms C'' of the normalized CHSH Bell inequality when each party measures with σ_x, σ_z ; more specifically, $C''(P') = \frac{1}{2} - (1/\sqrt{2}) < 0$. It follows that

$$\begin{aligned} & \{C'' \otimes (E - C'') \otimes \dots \otimes (E - C'')\} \\ & ([V_{2,3}^1]^{-1} \otimes \dots \otimes [V_{2m,1}^m]^{-1})P \\ & = \left(\frac{1}{2} - \frac{1}{\sqrt{2}}\right) \left(\frac{1}{2} + \frac{1}{\sqrt{2}}\right)^{m-1}. \end{aligned} \quad (26)$$

Unfortunately, that does not prove that P is Bell nonlocal, since $(V^1)^{-1}, (V^2)^{-1}, \dots$ are not connectors (that is, each one of them does not necessarily map classical boxes to classical boxes). Suppose, though, that we identify $2 \rightarrow 2$ connectors W^1, W^2, \dots whose action on P was analogous to that of $(V^1)^{-1}, (V^2)^{-1}, \dots$; see Fig. 9 (right). Then, there would be a fair chance that the newly devised Bell inequality

$$B \equiv \{C'' \otimes (E - C'') \otimes \dots \otimes (E - C'')\} \circ (W_{2,3}^1 \otimes \dots \otimes W_{2n,1}^n) \quad (27)$$

were such that $B(P) < 0$.

To find a guess for W^1 , we consider the box $Q = (\mathbb{I}_1 \otimes V_{23}^1 \otimes \mathbb{I}_4)P' \otimes P'$. It is easy to see that identifying the connector W^1 that minimizes $\{C'' \otimes (E - C'')\}(\mathbb{I}_1 \otimes W_{23}^1 \otimes \mathbb{I}_4)Q$ can be cast as a linear program. Heuristically, W^1 is approximately inverting the action of U^1 . Next, we consider the problem of identifying the connector W^2 such that $\{C'' \otimes (E - C'') \otimes (E - C'')\}(\mathbb{I}_1 \otimes W^1 V^1 \otimes W^2 V^2 \otimes \mathbb{I}_6)(P'^{\otimes 3})$ is minimized; again a LP. We iterate this procedure until we obtain suitable guesses for W^1, \dots, W^{m-1} . The last step is to identify the connector W^n that minimizes the contraction shown on the right side of Fig. 9. If the result is negative, we have detected the Bell nonlocality of P .

To assess how well this method works, we generate 300 m -tuples of random unitaries (U^1, \dots, U^m) for different values of m and apply the procedure to the resulting $2m$ -partite box. The results are shown in Table I, method I. Note that, for $m \geq 5$, the algorithm detects nonlocality *always*.

Notice that, if the intuition behind the heuristic is taken literally, the detection of Bell nonlocality should depend only on the values of just U^1, U^2, U^n . Indeed, intuitively, the C'' red connector in Fig. 9 is associated with the first negative term on the right-hand side of Eq. (26): The remaining red connectors $E - C''$ give the positive contribution on the right of the equation and are just meant to enhance the magnitude of the Bell violation. This intuition

TABLE I. Cases (out of 300) where the nonlocality of the chain state in Fig. 9 (left) is detected via the connector contraction in Fig. 9 (right) for two different methods to choose the connectors. The numbers in the first row are the same because both methods are equivalent for $m = 2$.

m	Method I	Method II
2	120	120
3	201	52
4	282	59
5	300	56
6	300	65

leads one to anticipate that, for $m \geq 3$, the probability of a Bell violation should not depend on the system size. In fact, we observe just the opposite: As the system grows in size, the probability of detecting nonlocality with the contraction in Fig. 9 increases with m ; see Table I.

A possible explanation is that, as the index k runs from 1 to n , the action of W^k becomes less and less the inversion of V^k . On the contrary, the heuristic seems to be exploiting the structure of the correlations between the remaining parties after the application of $\{C'' \otimes (E - C'')^{\otimes k}\}$ in order to boost the overall Bell violation even further. Actually, if we modify the heuristic and, for $k \geq 2$, we derive each W^k by maximizing the contraction

$$[E \otimes (E - C) \otimes (E - C) \otimes E],$$

$$(\mathbb{I}_1 \otimes W^{k-1} \otimes W^k \otimes \mathbb{I}_6)(P')^{\otimes 3}, \quad (28)$$

then the dependence in m disappears; see Table I, method II.

2. Nonlocality detection in more general boxes

In some situations, we want to decide the Bell nonlocality of a multipartite box for which one just has a theoretical description. (That is, no preparation information is available as in the previous discussion.) To address nonlocality detection in more general boxes, we introduce the matrix product connector tensor network (MPCTN): a witness composed of $2 \rightarrow 1$ connectors that are contracted similar to tensors in a matrix product state; see Fig. 2(b). Figure 10 illustrates a MPCTN for a nine-party box.

Denote by n_I and n_O the number of inputs and the number of outputs that appear on the bond indices of the MPCTN, respectively. We call the pair (n_I, n_O) the *bond dimension* of the MPCTN. The bond dimension determines the size of the output index of each connector. In principle, one can choose a different bond dimension for each connector. However, in the numerical simulations that we present here, we fix the same bond dimension for all the connectors in the MPCTN. We remark that even though in this work all numerical results are obtained by using a MPCTN witness, we can at least construct a MERA-like

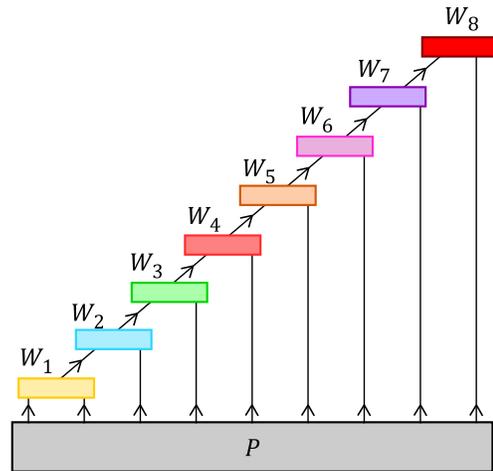


FIG. 10. A matrix product connector tensor network (MPCTN). A generic witness composed from $2 \rightarrow 1$ connectors that are contracted similar to tensors in a matrix product state [see Fig. 2(b)]. We find the resulting witness—the MPCTN—useful for detecting nonlocality and entanglement in several systems. The sequence of $2 \rightarrow 1$ connectors applied from left to right coarse grain the box to an effective two-site box, whose global properties can be explored by means of a two-site witness W_8 . When P also has an efficient tensor network representation, then a MPCTN provides a witness that can be scaled to hundreds of sites.

witness, since nontrivial $2 \rightarrow 2$ connectors exist (e.g., the one depicted in Fig. 7).

Having thus fixed MPCTN as the *Ansatz* for nonlocality witness, the next task is to (numerically) determine the connectors $\{W_i\}$ which minimize the tensor network contraction illustrated in Fig. 10 for a given box P . In numerical optimizations, the bond dimension controls both the computational cost and the value of possible violations. We use two types of optimization techniques.

Seesaw optimization.—We initialize all the connectors to random values (within the space of $2 \rightarrow 1$ connectors of the given bond dimension). Denote by $W(P)$ the value of the contraction illustrated in Fig. 10. We single out one connector, say, at location j , and rewrite $W(P)$ as the contraction of W_j with its “environment” E_j . Here E_j is the tensor that results from contracting the remaining tensors in the network (see Fig. 11), and it can be computed by the MATLAB function `ncon()` as we describe in Sec. II A 2.

Since the contraction of W_j with E_j can be interpreted as a linear function on the entries of W_j , minimizing $W(P)$ over W_j while keeping the remaining tensors fixed amounts to a linear optimization over the set of $2 \rightarrow 1$ connectors. This optimization, in turn, can be cast as a linear program. We solve this LP, identify the optimal connector at position j , and update the value of W_j accordingly. Next, we choose another connector W_k and repeat the operation. Iterating, we obtain a sequence of decreasing values for $W(P)$. We can stop the protocol as soon as $W(P)$ becomes

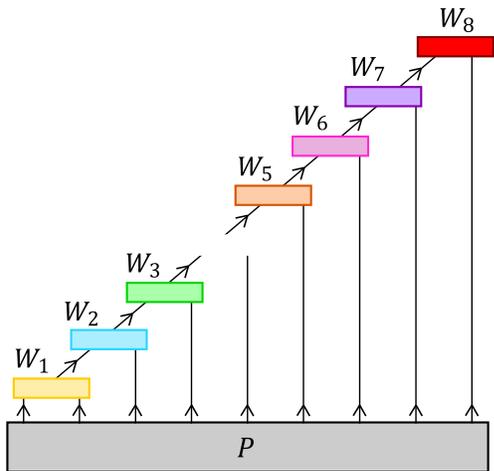


FIG. 11. Example of an environment tensor. Shown here is a possible environment tensor obtained from Fig. 10 by removing tensor W_4 and contracting all remaining tensors, for example, using the MATLAB function `ncon()` (see Sec. II A 2).

negative, although it is advisable to keep it running for a while in order to identify a stronger Bell violation.

These sorts of optimization procedures are called *seesaw methods* [47,48], and they have proven very helpful in condensed-matter physics and quantum nonlocality. In our numerical simulations, though, we find that, unless $W(P)$ is negative from the very beginning, very often one of the optimal connectors becomes 0. In such cases, a projected gradient method [49] seems to be a better choice to minimize $W(P)$.

Gradient descent optimization.—Choose $\epsilon > 0$, $\epsilon \ll 1$, and let E_j denote the environment of connector W_j . Adapted to this problem, the subgradient method consists of updating the connectors via the iterative equation:

$$\Omega_j^{k+1} = \pi_C(\Omega_j^k - \epsilon E_j^k). \quad (29)$$

Here, for any tensor A , $\pi_C(A)$ denotes the projection onto the set of valid connectors. That is, $\pi_C(A)$ is the connector that best approximates A in 2-norm (when viewed as a multipartite vector). Computing projections can be formulated as a SDP, and hence, it can be solved efficiently, as long as the cardinality of the indices of the connector is kept at a reasonable value.

Guess for initial connectors.—While sometimes using random connectors as the initial guess for the optimization (both the seesaw and the gradient descent) work well, we observe that in some cases making an educated guess for the initial connectors produces a violation when starting with random initial connectors fails to do so. (An educated guess for the initial connectors also sometimes enhances the violation in cases where there is a violation with random

initial connectors.) A guess that often works when exploring the same box for larger and larger number of parties is to use the *optimized* connectors for the smaller box as initial connectors for the larger number of parties. Sometimes, the reverse works—optimized connectors for a larger number of parties provides a good guess for a smaller number of parties. In practice, we try such different schemes to determine a method that works well for a given box.

Scalability.—There exist relevant scenarios in which the nonsignaling box P can also be efficiently represented as a tensor network, e.g., a matrix product state of a low bond dimension. This is possible when the box is the result of measuring a quantum state with limited correlations, for example, the thermal state of a 1D local Hamiltonian. In this case, the contraction illustrated in Fig. 10 to compute $W(P)$ can be carried out with a cost that scales only *linearly* with the number of parties [13]. This means that we can apply the method above to assess the Bell nonlocality of boxes shared by hundreds or even thousands of parties (see Fig. 12). However, note that increasing the bond dimension—either of the box P or of the MPCTN—increases the prefactor in the scaling of the computational cost.

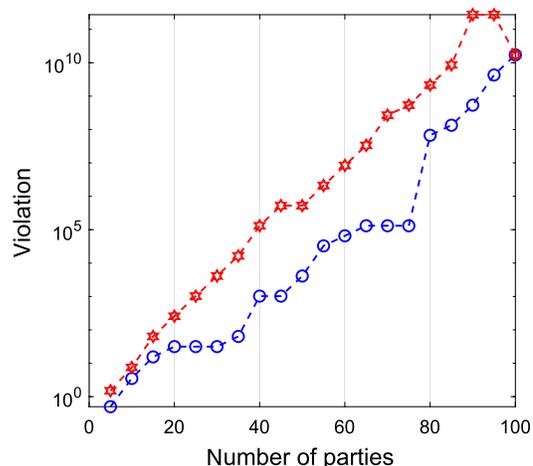


FIG. 12. Bell violations for the GHZ state with fixed measurements for number of parties = {5, 10, 15, ..., 100}. Bond dimension = (2 2). The blue points (circles) are the violations obtained using the seesaw optimization scheme with randomly chosen initial connectors until 75 parties. The use of random initial connectors fails to produce violations for larger number of parties. The violations for larger sizes (between 75 and 100) are obtained by using the *optimized* connectors for L parties as the initial guess for the optimization of $L + 5$ parties. The red points (stars) are substantially enhanced violations obtained by using the *optimized* connectors for 100 parties as the initial guess for smaller number of parties, suggesting that a simple seesaw optimization scheme may not be optimal. (The 100-party solution stops working as the initial guess for less than 35 parties. We still manage to obtain enhanced violations in this domain by feeding the optimized connectors for L parties as the initial guess for $L - 5$ parties.)

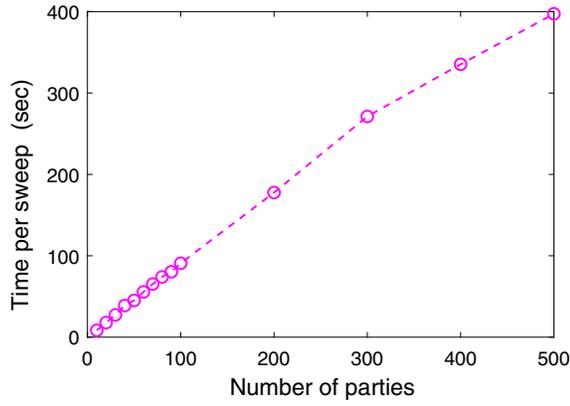


FIG. 13. Scalability. Computational time (in seconds) for a single seesaw sweep for optimization of Bell nonlocality connectors where bond dimension $= (2, 2)$ grows approximately linearly with number of parties. Simulations are run on a regular laptop with 2.5-GHz CPU and 8-GB RAM.

Benchmarking result: GHZ state.—To test how these methods work, we consider the quantum box that results when m parties share the GHZ state [50] $|\psi\rangle = (1/\sqrt{2})(|0\rangle^{\otimes m} + |1\rangle^{\otimes m})$ and are allowed to measure it with the settings $\{[(\mathbb{I} \pm \sigma_x)/2], [(\mathbb{I} \pm \sigma_z)/2]\}$. The result is a box which can be expressed as a MPS of bond dimension four.

Figure 13 shows the violations that we obtain for $m = \{5, 10, 15, \dots, 100\}$ using two seesaw optimization schemes. Note that the magnitude of the violation seems to increase approximately exponentially with the number of parties. (In fact, we are unable to proceed for number of parties much larger than 100 because of instabilities in optimizing the connectors owing to the presence of very large coefficients in connector tensors.) This exponential violation rules out the possibility that the optimization algorithm is simply determining wirings to project the last three parties into a GHZ state and then probing the resulting tripartite Greenberger–Horne–Zeilinger (GHZ) state with, e.g., the Mermin inequality [51]. The algorithm is finding a cleverer solution. Furthermore, we find violations only after a couple of seesaw sweeps, even when starting from random connectors. Note that detecting the nonlocality of a system made of hundreds parties, as done here, is completely out of reach with the existing techniques.

IV. SUPRAQUANTUM NONLOCALITY DETECTION

In this section, we sketch how to use connector theory to determine whether a given nonsignaling box $P_{a_1, \dots, a_m}^{x_1, \dots, x_m}$ admits a quantum realization. The object of interest is the same as above, a probability distribution for the measurement outputs conditioned on the inputs, but now the goal is to understand whether these given correlations can be

reproduced within the quantum formalism. That is, whether there exist Hilbert spaces $\mathcal{H}_1, \dots, \mathcal{H}_m$ and operators $\{M_{a|x}^k : \mathcal{H}_k \rightarrow \mathcal{H}_k : k = 1, \dots, m, a = 1, \dots, d, x = 1, \dots, n\}$, $\rho : \mathcal{H}_1 \otimes \dots \otimes \mathcal{H}_m \rightarrow \mathcal{H}_1 \otimes \dots \otimes \mathcal{H}_m$ such that

$$P_{a_1, \dots, a_m}^{x_1, \dots, x_m} = \text{tr}\{\rho_{1, \dots, n}(M_{a_1|x_1}^1 \otimes \dots \otimes M_{a_m|x_m}^m)\}, \quad (30)$$

where ρ is a non-normalized quantum state and $M_{a|x}^k$ is to be understood as the POVM element corresponding to party k inputting x in its box and obtaining the result a . That is,

$$M_{a|x}^k \geq 0, \quad \sum_a M_{a|x}^k = \mathbb{I}_{\mathcal{H}_k}. \quad (31)$$

The natural GPT to consider here would be QUANT world, whose states are quantum boxes of arbitrarily many inputs and outputs. In this case, the dual set \mathcal{Q}' of the set of quantum boxes corresponds to the set of coefficients $W_{x_1, \dots, x_n}^{a_1, \dots, a_n}$ such that, given any set of operators $\{M_{a|x}^k\}$ satisfying Eq. (31), the operator

$$\sum_{a_1, \dots, a_n, x_1, \dots, x_n} W_{x_1, \dots, x_n}^{a_1, \dots, a_n} M_{a_1|x_1}^1 \otimes \dots \otimes M_{a_n|x_n}^n \quad (32)$$

is positive semidefinite. As we explain in Appendix B, nontrivial SDP *Ansätze* on \mathcal{Q}' can be derived from the Navascués–Pironio–Acín (NPA) hierarchy [5,52,53] and variants [54].

We claim that, replacing \mathcal{B}' in Eq. (24) by \mathcal{Q}' , we arrive at a characterization of the set of $m \rightarrow 1$ transformations between quantum boxes. Indeed, consider a general transformation W such that $\{W_{c|z} : c, z\}$, acting on the POVM elements $\{M_{a|x}^k\}$ via Eq. (32), generate positive semidefinite operators $N_{c|z}$ with the property that $\sum_c N_{c|z} = N \leq \mathbb{I}$. When applied over the box $P_{a_1, \dots, a_{m+r}}^{x_1, \dots, x_{m+r}}$, the resulting $(r+1)$ -partite box $P_{c, a_{m+1}, \dots, a_{m+r}}^{z, x_{m+1}, \dots, x_{m+r}}$ admits the decomposition

$$P_{c, a_{m+1}, \dots, a_{m+r}}^{z, x_{m+1}, \dots, x_{m+r}} = \text{tr}\{\tilde{\rho}(\tilde{N}_{c|z} \otimes M_{a_{m+1}|x_{m+1}}^{m+1} \otimes \dots \otimes M_{a_{m+r}|x_{m+r}}^{m+r})\}, \quad (33)$$

where $\tilde{N}_{a|x} \equiv N^{-1/2} N_{a|x} N^{-1/2}$ and $\tilde{\rho} = (N^{1/2} \otimes \mathbb{I}_{m+1, \dots, m+r}) \rho_{1, \dots, m} (N^{1/2} \otimes \mathbb{I}_{m+1, \dots, m+r})$. Note that $\{\tilde{N}_{c|z}\}_c$ can be interpreted as the POVM elements of an m -partite global measurement z with outcome c on the postselected state $\tilde{\rho}$. This state, in turn, is the result of affecting the trace-decreasing map $N^{1/2}$ on part of ρ . Hence, $P_{c, a_{m+1}, \dots, a_{m+r}}^{z, x_{m+1}, \dots, x_{m+r}}$ defines an $(r+1)$ -partite quantum box.

To test the method, next we introduce a number of nonsignaling supraquantum boxes which admit a MPS decomposition of bond dimension linear on the system size, or even bounded. A limited bond dimension allows us to test their nonquantumness for high system sizes.

A. Generalized Svetlichny box

Consider a scenario where m parties have two measurements, each with two outcomes, i.e., $x_1, \dots, x_m, a_1, \dots, a_m \in \{0, 1\}$, and let $f(x_1, \dots, x_m)$ be any Boolean function. It can be verified that the box with statistics

$$P_{a_1, \dots, a_m}^{x_1, \dots, x_m} = \frac{1}{2^{n-1}} \delta_{\bigoplus_k a_k, f(x_1, \dots, x_m)} \quad (34)$$

is normalized and nonsignaling. The proof is simple: If we trace out any party, the probability of any sequence of outputs equals $1/2^{m-1}$, independent of the sequence of inputs. Almost all such boxes allow us, by wirings, to produce a perfect Popescu-Rohrlich box, and hence, they cannot be realized within quantum theory. Moreover, there exist important supraquantum boxes within this family, such as the Svetlichny box [55]. Next, we generalize the Svetlichny box to arbitrarily many parties and then show that it admits a MPS representation with bond dimension 16.

The original Svetlichny box is tripartite, with statistics given by

$$P_{a_1, a_2, a_3}^{x_1, x_2, x_3} = \frac{1}{2^2} \delta_{a_1 \oplus a_2 \oplus a_3, x_1 x_2 \oplus x_2 x_3 \oplus x_3 x_1}. \quad (35)$$

It is thus of the form Eq. (34), with $f(x_1, x_2, x_3) = x_1 x_2 \oplus x_2 x_3 \oplus x_3 x_1$. We generalize it to a box of the form Eq. (34), with $f(x_1, \dots, x_m) = \bigoplus_{k=1}^{m-1} x_k x_{k+1} \oplus x_m x_1$. It can be verified that all such ‘‘generalized Svetlichny boxes’’ can be simulated by distributing a PR box to each party and its near neighbor and let each party wire its two boxes together. Again, reaching similar results for systems of hundreds of particles is impossible with existing methods.

In turn, Svetlichny boxes can be seen to admit a MPS decomposition

$$P_{a_1, \dots, a_m}^{x_1, \dots, x_m} = \Lambda_{a_1, x_1}^{[1]}, \dots, \Lambda_{a_{m-1}, x_{m-1}}^{[m-1]} \Lambda_{a_m, x_m}^{[m]} \quad (36)$$

involving matrices $\Lambda_{a, x}^{[k]}$ of size at most 16×16 ; see Appendix D for their exact expression.

Using the above MPS representation of the Svetlichny boxes and by using a MPCTN witness composed of $2 \rightarrow 1$ connectors from QUANT world, we are able to detect quantum nonlocality violation in these boxes for a large number of parties; see Fig. 14. We find that the violation increases exponentially with the number of parties.

B. Other boxes

Another family of boxes is obtained by defining

$$f_r(x_1, \dots, x_m) = 1, \quad \text{if } x_1, \dots, x_m \text{ contains } r \text{ consecutive } 1\text{s}, \\ 0, \text{ otherwise.} \quad (37)$$

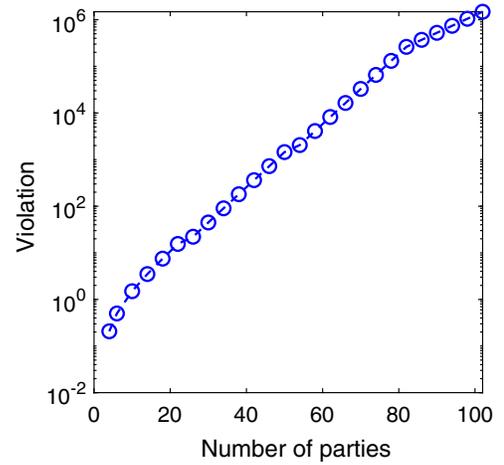


FIG. 14. Quantum nonlocality violation of the Svetlichny box for number of parties = 4, 8, 12, ..., 102. Maximum bond dimension = (44). The violations are obtained using the seesaw optimization scheme. We use the *optimized* connectors for $L - 4$ parties as the initial guess for connectors for L parties.

This box also admits a MPS representation (36) for bond dimension $2(r + 1)$; see Appendix D.

A fully symmetric ‘‘majority voting’’ box is given by the function

$$\text{maj}(x_1, \dots, x_m) = 1, \quad \text{if half or more of the inputs are } 1\text{s}, \\ 0, \text{ otherwise.} \quad (38)$$

This time, the bond dimension of the box scales linearly with the system size m .

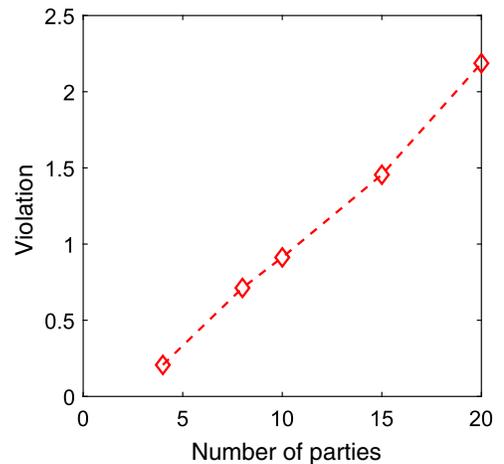


FIG. 15. Quantum nonlocality violation of the box defined in Eq. (37) with $r = 2$. Maximum bond dimension = (44). With the optimizations techniques described in this paper, we manage to find violations only up to 20 parties. It is possible that violations will also be found for more parties by using more sophisticated seesaw optimization techniques, better guess for initial connectors, and/or using larger bond dimensions. With our current implementation, we manage to run only simulations with a maximum bond dimension = (44) in a reasonable time.

Figure 15 shows the violations we find for the box defined in Eq. (37) with $m = 2$, for up to 20 parties. We also try the same box with $m = 3$, but manage to find a small violation of approximately -0.1 for four parties. Even for this case, we use a bond dimension $= (4, 4)$. Optimizations over connectors with a larger bond dimension are very slow.

V. ENTANGLEMENT DETECTION

As a final application of connector theory, we come back to the problem of entanglement detection, which we use in the Introduction to illustrate the main intuition of our construction. To apply connector theory to detect entanglement, we define a GPT whose states coincide with the fully separable quantum states, call it SEP world. In this theory, the norm of a state ρ is defined as $E(\rho) \equiv \text{tr}(\rho)$. As in LOC world, due to the structure (1) of the set of separable states, SEP world has the property that any linear-map-fulfilling condition (10) with $T = \emptyset$ constitutes a valid transformation. The set of connectors in SEP world thus corresponds to the set of linear maps which transform separable states into separable states. In general, these operations cannot be implemented in quantum theory.

When studying LOC world, we note that the structure of $m \rightarrow 1$ connectors is closely linked to that of Bell inequalities. Similarly, in SEP world there is a one-to-one correspondence between scaled connectors and $(m + 1)$ -partite entanglement witnesses [56]. We remind the reader that a k -partite entanglement witness W is an operator acting in $\otimes_{i=1}^k \mathcal{H}_i$ such that

$$\text{tr}(W\rho) \geq 0, \quad (39)$$

for all fully separable states ρ .

We claim that a linear map $\Omega: B(\otimes_{i=1}^m \mathcal{H}_i) \rightarrow B(\mathcal{H}_{m+1})$ corresponds to an $m \rightarrow 1$ connector if and only if

- (1) W_Ω defined via the relation $\text{tr}\{W_\Omega(\sigma \otimes \beta)\} = \text{tr}\{\Omega(\sigma)\beta\}$ is an $(m + 1)$ -partite entanglement witness,
- (2) $\mathbb{I}_{1,\dots,m} - \text{tr}_{m+1}(W_\Omega)$ is an m -partite entanglement witness.

Let us see why this correspondence holds. Let W be an $(m + 1)$ -partite entanglement witness and ρ any separable state of the form $\rho = \sigma_{1,\dots,m} \otimes \beta$. Consider the map $\Omega_W(\sigma) = \text{tr}_{1,\dots,m}\{(\sigma \otimes \mathbb{I}_{m+1})W\}$. By definition, we have that $\text{tr}[\Omega(\sigma)\beta] = \text{tr}(W\rho) \geq 0$. Since $\beta \geq 0$ is arbitrary, this means that $\Omega_W(\sigma) \geq 0$, for all fully separable states σ . That is, Ω_W is a scaled connector. Conversely, any linear map Ω with the property that $\Omega(\sigma) \geq 0$ for all fully separable states σ can be mapped to an entanglement witness W_Ω as defined in condition 1. This establishes that Ω is a scaled connector if and only if condition 1 holds. Condition 2 is easily seen to ensure that the norm is nonincreasing, i.e., $\text{tr}\{\Omega(\sigma)\} \leq \text{tr}\{\sigma\}$.

The above observation allows us to link the connector theory of SEP world with the existing literature on entanglement detection. In principle, we can promote any k -partite witness to a $k - 1 \rightarrow 1$ connector and contract several copies thereof, as we do with the CHSH Bell inequality in Fig. 8. The result would be a novel entanglement witness for m -partite entangled states.

Take, for instance, the family of m -qubit entanglement witnesses derived by Tóth *et al.* in Ref. [57]:

$$\sum_{i=x,y,z} \langle (J_i - \langle J_i \rangle)^2 \rangle \geq \frac{m}{2}, \quad (40)$$

with $J_i = \frac{1}{2} \sum_{j=1}^m \sigma_i^{(j)}$. Equation (40) does not correspond to a linear entanglement witness, but it can be turned into one just by replacing $\langle J_i \rangle$ by arbitrary real numbers:

$$\sum_{i=x,y,z} \langle (J_i - \lambda_i)^2 \rangle \geq \frac{m}{2}. \quad (41)$$

Taking $\lambda_i = 0$, one can contract the connectors associated with the four- and two-qubit entanglement witnesses as shown in Fig. 16 to produce the witness W_6 . Numerically, we find that there exist six-qubit states which, while satisfying all forms of Eq. (41), can be detected by W_6 . This shows that new detection properties can arise from composition alone. We come back to this in Sec. V C.

Constructing witnesses which detect the entanglement of a given quantum state is a more complicated task, due to the difficulty of certifying that $W_\Omega, \mathbb{I}_{1,\dots,m} - \text{tr}_{m+1}(W_\Omega)$ are indeed entanglement witnesses. A possible approach to this problem is to prove instead that the average values of those two operators are non-negative when evaluated over a *relaxation* (a superset) of the set of separable states. The family of relaxations which we consider in our numerical examples is called the Doherty-Parrilo-Spedaliery (DPS) hierarchy [9–11]. Combining this idea with the observation in Ref. [58] that a small perturbation of the DPS sets projects them to the interior of the set of separable states, in Appendix C we present a family of SDP *Ansätze* on the set of $m \rightarrow m'$ connectors. Throughout the rest of this section, we use those *Ansätze* whenever a linear optimization over feasible connectors is required.

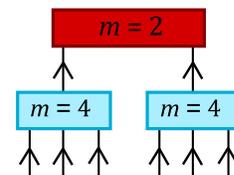


FIG. 16. Connectors for entanglement detection. Six-qubit entanglement witness W_6 resulting from the composition of witness (41) with $m = 2$ and $m = 4$.

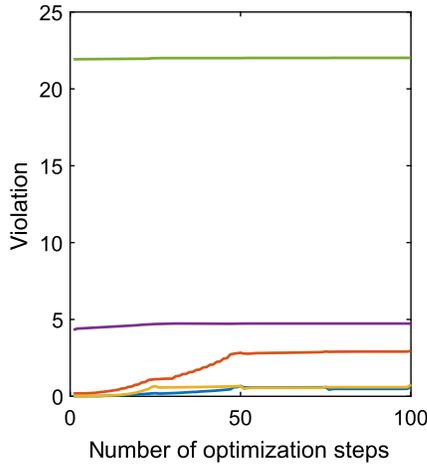


FIG. 17. Entanglement detection in 50-qubit mixed states using $2 \rightarrow 2$ connectors. We randomly sample five states and find a violation in each case. Each optimization step consists of optimizing one connector.

A. PPT states

To test how useful connectors are for entanglement detection, we first consider a famous class of multipartite entangled states which are positive under partial transposition (PPT) [27]. An unextendible product basis (UPB) is a collection of m -partite orthogonal product states $\{|\psi_i\rangle\}_{i=1}^K$ with the property that no other product vector is orthogonal to their span. Given any UPB $\{|\psi_i\rangle\}_{i=1}^K$, the m -partite state

$$\rho \propto \mathbb{I}_{1,\dots,m} - \sum_i |\psi_i\rangle\langle\psi_i| \quad (42)$$

can be shown entangled and PPT [59]. In Ref. [60], a family of six-qubit UPBs parametrized by three-qubit unitaries is presented. We sample ten unitary triples randomly according to the Haar measure, built the corresponding six-qubit quantum states (42), and use a MPCTN to detect their entanglement. The output system of each connector is a qubit. In all cases, a seesaw algorithm finds a normalized entanglement witness whose average value of the state is $\simeq -0.5$.

B. Many-body mixed states

We generate mixed states following a preparation similar to the one described in Sec. III C 1. We consider distributing m singlets $|\Psi^-\rangle \equiv \frac{1}{\sqrt{2}}(|0\rangle|1\rangle - |1\rangle|0\rangle)$ among $2m$ parties as illustrated in Fig. 9 (where state $|\phi\rangle$ is replaced with $|\Psi^-\rangle$). We also replace the action of unitaries U_1, U_2, \dots, U_m shown in the figure by conjugation with a convex combination of two unitaries (drawn randomly for each pair of sites). The resulting state is mixed, has an efficient MPO representation, and may be separable. We want to certify whether such states are entangled or not using connectors.

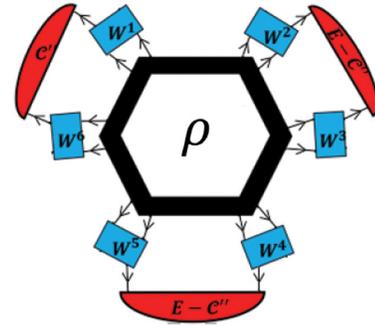


FIG. 18. Contraction of $2 \rightarrow 1$ connectors used to detect entanglement of a many-body mixed state ρ . Here, $C' = \text{SWAP}$ is the witness for certifying entanglement in the singlet state $|\Psi^-\rangle$.

We use a witness to the one illustrated in Fig. 9 (right-hand side). The intuition is the same. We want to find $2 \rightarrow 2$ connectors in SEP world that approximately invert the randomizing quantum channel, thus, exposing the initial singlets. The singlets can be certified to be entangled by a two-party witness which is simply the SWAP gate, which evaluates to -1 for the singlet $|\Psi^-\rangle$. (That is, we replace $C' = \text{SWAP}$ in Fig. 9; $E - \text{SWAP}$ connectors are once again used to amplify the violation.)

We generate such finitely correlated mixed states from randomly chosen unitaries (using the Haar measure) for a system of 50 qubits, and find a violation (certifying the presence of entanglement) almost each time. Figure 17 shows the violations obtained for five such randomly drawn states.

We also use another witness, one composed of only $2 \rightarrow 1$ connectors as illustrated in Fig. 18. Again, we easily find violations for randomly chosen states; see Fig. 19.

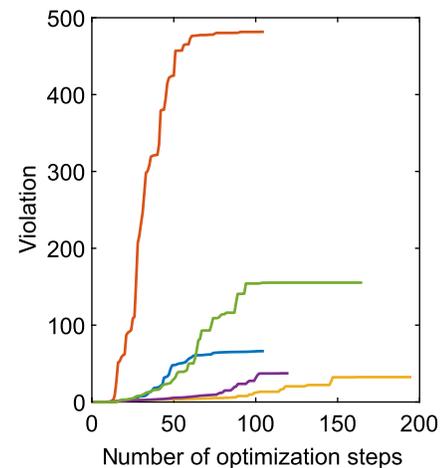


FIG. 19. Entanglement detection in FCS mixed states of 60 qubits using $2 \rightarrow 1$ connectors. We randomly sample five states and find a violation in each case. Each optimization step consists of optimizing one connector. We find much larger violations than when using $2 \rightarrow 2$ connectors.

We also find that the violations are larger than those obtained by using $2 \rightarrow 2$ connectors.

C. Entanglement detection through hybrid GPTs

In the previous section, we link the entanglement problem to connector theory by defining a GPT, SEP LOC, where the set of physical states coincides with the set of fully separable states. The purpose of this section is to demonstrate that it is even possible to follow a hybrid approach in which different GPTs are connected.

Consider, for example, a theory where there are two types of basic systems: quantum systems and boxes. The state of a composite system of, say, two boxes and two quantum systems, would be a steering ensemble of the form $\{\rho_{CD}^{a|x,b|y}\}$, with the property that there exists a fully separable quantum state σ_{ABCD} and measurement operators $M_{a|x}^A, M_{b|y}^B$, such that

$$\rho_{CD}^{a|x,b|y} = \text{tr}_{AB} \{ \sigma M_{a|x}^A \otimes M_{b|y}^B \otimes \mathbb{I}_{CD} \}. \quad (43)$$

We dub this theory STEER world.

Now, suppose that we wish to assess the entanglement of a three-qubit state. One possibility would be to regard it as a possible state of STEER world and then apply the connectors depicted in Fig. 20. There, the three-partite quantum state is transformed into a bipartite box, which we then evaluate with the normalized CHSH inequality \mathcal{C} (20). The diagram depicted in Fig. 20 looks like device-independent certification of entanglement. Actually, it would be equivalent to this primitive, if the transformations U, V acted on single systems. Indeed, in that case, connectors from quantum systems to boxes correspond to conducting quantum measurements on the quantum systems, and entanglement is detected if and only if the corresponding box violates a Bell inequality.

As we see, the $2 \rightarrow 1$ connector U mapping bipartite quantum systems to a single box changes things completely. Let U, V be defined via

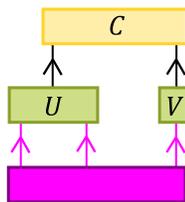


FIG. 20. Entanglement detection through hybrid GPTs. Purple lines indicate quantum systems; black lines, boxes. Starting from a quantum state, we effect two transformations U, V to map it to a bipartite box, which we then probe with the normalized CHSH inequality.

$$\begin{aligned} U^{0|0}(\rho) &= \text{tr}(\text{SWAP}\rho), \\ U^{0|1}(\rho) &= \text{tr} \left\{ \rho \frac{1}{2} (\mathbb{I}_4 + \sigma_y \otimes \sigma_y) \right\}, \\ V^{0|0}(\beta) &= \text{tr} \left\{ \beta \frac{\mathbb{I}_2 + \sigma_x}{2} \right\}, \\ V^{0|1}(\beta) &= \text{tr} \left\{ \beta \frac{\mathbb{I}_2 + \sigma_z}{2} \right\}, \end{aligned} \quad (44)$$

where the superindex in each tensor corresponds to the index of its upper, black leg. Here, SWAP denotes the permutation operator $\sum_{i,j=0,1} |i\rangle\langle j| \otimes |j\rangle\langle i|$, and $\{\sigma_i\}_{i=x,y,z}$, the three Pauli matrices. We assume that U, V are deterministic transformations, i.e., $U^{1|x}(\rho) = E(\rho) - U^{0|x}(\rho)$, $V^{1|y}(\beta) = E(\beta) - V^{0|y}(\beta)$, for $x, y = 0, 1$.

Note that $U^{0|1}, V^{0|0}, V^{0|1}$ are projectors, and hence, for normalized quantum states ρ, β , $P_U(a|1) = U^{a|1}(\rho)$, $P_V(b|y) = V^{b|y}(\beta)$ satisfy $0 \leq P_U(a|1), P_V(b|y) \leq 1$. SWAP , despite not being positive semidefinite, is a normalized entanglement witness; hence, $0 \leq U^{0|0}(\rho) \leq 1$ for all separable states ρ . Both U and V thus represent valid deterministic transformations from separable states to classical boxes.

Contracting U, V, \mathcal{C} , we obtain a three-qubit entanglement witness X , that we can express as an operator acting on $\mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2$. Now, consider an optimization over PPT three-qubit states ρ_{ABC} , i.e., consider the problem

$$\begin{aligned} \min \text{tr}(X\rho_{ABC}), \\ \text{such that } \rho_{ABC}, \rho_{ABC}^{T_A}, \rho_{ABC}^{T_B}, \rho_{ABC}^{T_C} \geq 0, \\ \text{tr}(\rho_{ABC}) = 1. \end{aligned} \quad (45)$$

This problem can be cast as a SDP; hence, we can solve it. The solution is -0.0721 . This is surprising because neither the CHSH inequality nor the SWAP operator can, by themselves, detect PPT entanglement. Their composition, however, does. So, even if we hadn't been aware of the existence of nondecomposable entanglement witnesses (those which can detect PPT states), we could have derived them from compositional arguments alone.

VI. CONCLUSION

We present a general method to analyze complex networks, be they classical, quantum, or supraquantum. In essence, our method consists of acting on the many-body system in question with a number of linear transformations—the connectors—which iteratively coarse grain the system to one that is small enough to analyze with the existing mathematical tools. While we can relate $m \rightarrow 1$ connectors to past literature on Bell nonlocality and entanglement theory, $m \rightarrow m'$ connectors seem to be a completely different beast. We show that connector theory is powerful enough to detect Bell nonlocality (quantum and supraquantum) and

entanglement in networks composed of hundreds of sites. Even though we focus on these three areas, we suspect that connector theory will soon find application in other scenarios, for example, to build new dimension witnesses.

Connectors are a natural tool to analyze large, complex many-body systems, and we feel that future research should focus on understanding their mathematical properties. In this regard, our work leaves open important theoretical questions.

One of them is to understand the limitations of the new formalism. Could there be, e.g., entangled tripartite quantum states undetectable by the composition of a $2 \rightarrow 1$ connector and a bipartite witness? If not, one wonders how difficult it is in general to find the “right” connectors to detect a particular state or box. The performance of our current numerical methods oscillates between disappointing (it sometimes takes ages to identify the appropriate connectors, even for $m = 3$) and excellent (100 sites in less than 2 min). Actually, in some scenarios, like QUANT world, we altogether avoid discussing how to optimize over general $m \rightarrow m'$ connectors.

Another question pertains the practical use in experiments of connector-generated witnesses. Estimating the average value of a witness on a many-body state or box generally requires a number of experiments that scale exponentially with the system size. Is there any way to exploit the tensor network structure of a witness in order to estimate its value with a polynomial number of experiments?

Finally, it is an intriguing idea whether more complicated connectors could be devised by working on a GPT where states are identified with the connectors themselves (CONNECTOR world), or even with connectors of connectors.

ACKNOWLEDGMENTS

We thank S. Pironio and D. Roset for interesting discussions. This work is funded by the Austrian Science Fund stand-alone Project No. P 30947, the ERC AdG CERQUITE and CoG QITBOX, the AXA Chair in Quantum Information Science, the Spanish MINECO (Grant No. QIBEQI FIS2016-80773-P and Severo Ochoa Grant No. SEV-2015-0522), Fundació Cellex and Mir-Puig, and Generalitat de Catalunya (SGR 1381, QuantumCAT and CERCA Programme). As part of the Gravity, Quantum Fields and Information Group at the Albert Einstein Institute, S.S. acknowledges the support of the provided by the Alexander von Humboldt Foundation and the Federal Ministry for Education and Research through the Sofja Kovalevskaja grant.

APPENDIX A: CHARACTERIZATION OF THE CONNECTORS IN LOC WORLD

Consider a system in LOC world of the type $(O_1, \dots, O_m, I_1, \dots, I_m)$. The state of any such system can be

expressed as a convex combination of $\prod_{k=1}^m O_k^{I_k}$ extreme classical boxes. If we order them as $\{P^i(a_1, \dots, a_m | x_1, \dots, x_m)\}_i$, then verifying whether a (normalized) box is classical can thus be cast as the LP:

$$\begin{aligned} & \min 0, \\ & \text{such that } P(a_1, \dots, a_m | x_1, \dots, x_m) \\ & \quad = \sum_i p_i P^i(a_1, \dots, a_m | x_1, \dots, x_m), \\ & p_i \geq 0, \quad \sum_i p_i = 1. \end{aligned} \quad (\text{A1})$$

The distribution $\{p_i\}_i$ is sometimes called the local hidden variable model. For general nonsignaling boxes, it does not exist and, when it does, in general it is not unique.

Also, due to this finiteness of extreme points, the problem of optimizing over $m \rightarrow q$ connectors can be cast as a linear program too. Indeed, let $\{P^i\}_i$ ($\{\tilde{P}^j\}_j$) be the set of extreme points of the output (input) q -partite (m -partite) LOC-world system, and suppose that, for some fixed tensor C , we wish to minimize $W(C)$ over all nondeterministic connectors W . The corresponding program would be

$$\begin{aligned} & \min W(C), \\ & \text{such that } W(\tilde{P}^i) = \sum_j p_j^i \tilde{P}^j, \\ & p_j^i \geq 0, \quad \sum_j p_j^i \leq 1, \end{aligned} \quad (\text{A2})$$

where the last condition enforces that the norm of the box does not increase after we apply the connector. To see that any linear functional W satisfying the feasibility conditions transforms local boxes into subnormalized local boxes, note that any initial classical box P admits a decomposition $\sum_i p_i P^i$, with $p_i \geq 0$, $\sum_i p_i = 1$. The result of applying W over such a box is thus the box $\tilde{P} = \sum_{i,j} p_i P_j^i \tilde{P}^j$. Identifying $\mu_j \equiv \sum_i p_i P_j^i$ with our local hidden variable model, we find that \tilde{P} is also Bell local.

Program (A2), although correct, can be greatly improved. In the following, we show how to do so by exploiting the lessons learned from the monogamy of nonlocal correlations. We do so in three stages. First, we introduce a convenient notation to deal with nonsignaling boxes that is also useful to minimize the complexity of LPs like Eq. (A2). A characterization of the dual of the set of nonsignaling boxes follows. Finally, building on the above two results, we present our proposal for linear optimizations over $2 \rightarrow 1$ connectors.

1. Notation of nonsignaling boxes

Because of the nonsignaling conditions, any non-normalized nonsignaling box of the form $P(a_1, \dots, a_k | x_1, \dots, x_k)$ with $a_i \in \{1, \dots, d_i\}$, $x_i \in \{1, \dots, m_i\}$ can

be expressed in terms of just $\prod_{i=1}^k [m_i(d_i - 1) + 1]$ parameters. One way to do so is to adopt what we call from now on the abbreviated form $P(A_1, A_2, \dots, A_n)$, where $A_i \in \{\emptyset\} \cup \{1, \dots, d_i - 1\} \times \{1, \dots, m_i\}$. Here, \emptyset denotes that the random variable was not measured. That way, e.g., for $a_2 \in \{1, \dots, d_2 - 1\}$, $P(\emptyset, [a_2, x_2])$ represents the probability that the second-party-conducted measurement x_2 obtains the result a_2 , i.e., $P(\emptyset, [a_2, x_2]) = P_2(a_2|x_2)$.

If we represent the probabilities $P(a_1, \dots, a_k|x_1, \dots, x_k)$ $[P(A_1, \dots, A_k)]$ as a vector \bar{P} , \bar{Q} , there exists a matrix S such that $\bar{P} = S\bar{Q}$.

2. The dual of nonsignaling boxes

Consider the k -partite nonlocality scenario $\{(d_i, m_i)\}_{i=1}^k$. In abbreviated form, the corresponding set of (non-normalized) nonsignaling distributions is $\mathcal{B} = \{\bar{q} : S\bar{q} \geq 0\}$, where $\bar{q} \in \mathbb{R}^D$, with $D = \prod_{i=1}^k [m_i(d_i - 1) + 1]$, and S is the matrix that transforms a box from its abbreviated representation $P(A_1, \dots, A_k)$ to its standard representation $P(a_1, \dots, a_k|x_1, \dots, x_k)$; see Sec. A 1. The condition $S\bar{q} \geq 0$ enforces that all the probabilities of the box are non-negative.

The set of positive linear functionals in abbreviated representation is given by the set $\mathcal{B}' = \{S^T \bar{c} : \bar{c} \geq 0\}$. Indeed, by definition, any $\bar{v} \in \mathcal{B}'$ satisfies $\bar{v} \cdot \bar{q} \geq 0$ for all $\bar{q} \in \mathcal{B}$, and so the dual set of \mathcal{B} contains \mathcal{B}' . It rests to show that any vector outside \mathcal{B}' cannot belong to the dual of \mathcal{B} . First note that, for any $\bar{q} \notin \mathcal{B}$, there exists $\bar{v} \in \mathcal{B}'$ such that $\bar{v} \cdot \bar{q} < 0$ [take, e.g., $\bar{v} = S^T \bar{c}$, with $c_j = \Theta(-\langle S^T \bar{q}, j \rangle)$]. Now, let $\bar{w} \notin \mathcal{B}'$. By the separation theorem, there exists \bar{q} such that $\bar{v} \cdot \bar{q} \geq 0$ for all $\bar{v} \in \mathcal{B}'$, and $\bar{w} \cdot \bar{q} < 0$. The first condition implies that $\bar{q} \in \mathcal{B}$, and so the second condition implies that \bar{w} is not in the dual of \mathcal{B} .

With the formulation above, it is clear that linear optimizations over the set of positive functionals of nonsignaling boxes can be carried out via linear programming [25].

3. Faster codes for optimization over $2 \rightarrow 1$ connectors

First, we define a (non-normalized) local box in a nonstandard way.

Definition 1. The probabilities $\{P(a, b|x, y) : x = 1, \dots, n_A; y = 1, \dots, n_B; a = 1, \dots, d_A; b = 1, \dots, d_B\}$ define a local box if and only if there exist $\{P(a, b_1, \dots, b_{n_B}|x)\}$, such that

$$\begin{aligned} P(a, b_y|x, y) &= \sum_{b_z : z \neq y} P(a, b_1, \dots, b_{n_B}|x), \\ \sum_a P(a, b_1, \dots, b_{n_B}|x) &= P(b_1, \dots, b_{n_B}), \\ P(a, b_1, \dots, b_{n_B}|x) &\geq 0. \end{aligned} \quad (\text{A3})$$

That this definition implies bipartite locality can be seen by noting that the variables b_1, \dots, b_{n_B} play the role of local hidden variables in the decomposition above.

Conversely, let $P(a, b|x, y) = \sum_\lambda P(\lambda) P_A(a|x, \lambda) P_B(b|y, \lambda)$. Then, one can verify that $P(a, b_1, \dots, b_{n_B}|x) \equiv \sum_\lambda P(\lambda) P_A(a|x, \lambda) P_B(b_1|1, \lambda), \dots, P_B(b_{n_B}|n_B, \lambda)$ satisfies the conditions in Eq. (A3). From now on, we refer to the object $P(a, b_1, \dots, b_{n_B}|x)$ as an *extended box*, and represent it by the vector of probabilities \bar{P} .

Note that we can regard an extended box as a non-signaling box where all the parties but the first have just one input. Therefore, we can represent extended boxes in abbreviated form, as a vector of probabilities $\bar{Q} = P(A, B_1, \dots, B_{n_B})$, with $A \in \{\emptyset\} \cup \{1, \dots, d_A - 1\} \times \{1, \dots, n_A\}$, $B_i \in \{\emptyset\} \cup \{1, \dots, d_B - 1\}$. Let S be the matrix that effects the transformation $S\bar{Q} = \bar{P}$. In abbreviated form, the set of non-normalized local boxes is thus described by $L \equiv \{\bar{q} : S\bar{q} \geq 0\}$. As we prove in Sec. A 2, the dual of this set, i.e., the set of vectors \bar{v} such that $\bar{v} \cdot \bar{q} \geq 0$ for all $\bar{q} \in L$, corresponds to the set of vectors $L' \equiv \{S^T \bar{c} : \bar{c} \geq 0\}$.

Any positive linear functional over local boxes must remain positive if we embed it into the space of extended boxes. This provides us with a computationally efficient characterization of the set of positive Bell functionals.

Proposition 2. $\{U(A, B)\}$ is a positive functional over the set of local boxes if and only if there exists a vector $\bar{c} \geq 0$, such that

$$\begin{aligned} S^T \bar{c}(A, B_1, \dots, B_{n_B}) &= 0, \quad \text{if } \exists i, j : i \neq j, B_i, B_j \neq \emptyset, \\ S^T \bar{c}(A, B_1, \dots, B_{n_B}) &= U(A, B), \quad \text{if } B_y = B, \\ B_z &= \emptyset, \quad \forall z \neq y. \end{aligned} \quad (\text{A4})$$

With this formulation, one can carry out linear optimizations over the set of positive functionals of local boxes via linear programming [25]. The computational cost is bearable provided that n_B is not very large. n_A can take high values, though.

APPENDIX B: $m \rightarrow 1$ CONNECTORS FOR QUANTUM BOXES

The set of quantum boxes can be formulated as the set of all boxes of the form

$$P(a_1, \dots, a_k|x_1, \dots, x_k) = \text{tr}(\rho E_{a_1, x_1}^1 \otimes \dots \otimes E_{a_k, x_k}^k), \quad (\text{B1})$$

where ρ is a positive semidefinite matrix with $\text{tr}(\rho) \leq 1$ and $\{E_k^{a, x}\}$ satisfy

$$\begin{aligned} (E_j^{a, x})^\dagger &= (E_j^{a, x})^2 = E_j^{a, x}, \\ \sum_a E_j^{a, x} &= \mathbb{I}_j. \end{aligned} \quad (\text{B2})$$

In order to derive positive functionals for quantum boxes, we rely on noncommutative polynomial optimization theory [5, 52, 61]. Let X_0, X_1, \dots, X_n be a number of Hermitian operators acting on the same Hilbert space, with $X_0 = \mathbb{I}$, and let ρ be a normalized quantum state. The s th-order moment matrix Γ of this system is the matrix whose

rows are columns are labeled by words of the alphabet $\{0, \dots, n\}$ of length s or smaller, and whose entries are given by

$$\Gamma_{\vec{i}, \vec{j}} = \text{tr}(\rho X_{\vec{i}}^\dagger X_{\vec{j}}), \quad (\text{B3})$$

where, for any word \vec{i} of length t , $X_{\vec{i}} \equiv X_{i_1}, \dots, X_{i_t}$. It can be verified that all moment matrices are positive semi-definite [61].

Now, consider the moment matrices defined by the quantum state ρ and operators $\{\mathbb{I}_{1, \dots, j-1} \otimes E_j^{a,x} \otimes \mathbb{I}_{j+1, \dots, k}\}$ generating our quantum box (B1). It is immediate to see that, for moment matrices Γ of high enough order, there exist matrices F_{A_1, \dots, A_k} such that $P(A_1, \dots, A_k) = \text{tr}(\Gamma F_{A_1, \dots, A_k})$. It is also easy to see that the moment matrices of un-normalized quantum boxes are subject to nontrivial linear constraints [5, 52]. That is, there exists a set of matrices $\{G_j\}_j$ such that any moment matrix Γ can be expressed as $\Gamma = \sum_j c_j G_j$, for some choice of coefficients $\{c_j\}_j$.

From the above, it follows that a sufficient condition for $w(A_1, \dots, A_k)$ to be a positive linear functional over the set of quantum boxes is that there exists a matrix $Z \geq 0$, such that

$$\text{tr}\left(\sum_{A_1, \dots, A_k} w(A_1, \dots, A_k) F_{A_1, \dots, A_k} G_j\right) = \text{tr}(Z G_j), \quad (\text{B4})$$

for all j .

Indeed, note that

$$\begin{aligned} & \sum_{A_1, \dots, A_k} w(A_1, \dots, A_k) P(A_1, \dots, A_k) \\ &= \text{tr}\left(\sum_{A_1, \dots, A_k} w(A_1, \dots, A_k) F_{A_1, \dots, A_k} \Gamma\right) \\ &= \text{tr}(Z \Gamma) \geq 0. \end{aligned} \quad (\text{B5})$$

Here, the second equality stems from the fact that $\Gamma = \sum_j c_j G_j$. The inequality holds because both Γ and Z are positive semidefinite matrices.

APPENDIX C: $m \rightarrow m'$ CONNECTORS FOR SEP WORLD

First, we explain how to optimize over $m \rightarrow 1$ connectors which transform multipartite separable states with Hilbert space dimension $d_1 \times d_2 \times \dots \times d_m$ into a quantum state in dimension d_{m+1} . We use the Choi-Jamiołkowski notation to represent connectors; i.e., each connector Ω is identified with an $(m+1)$ -partite operator $W_{1, \dots, m+1}$ such that $\Omega(\rho_{1, \dots, m}) = \text{tr}_{1, \dots, m}\{W_{1, \dots, m+1}(\rho_{1, \dots, m} \otimes \mathbb{I}_{m+1})\}$.

Call $\mathcal{H}_{\text{sym}}^{k,d}$ the symmetric subspace of k identical particles of dimension d . Then, S^k is the set of $(m+1)$ -partite states σ , such that there exists a $(km+1)$ -partite

state $\beta_{1, \dots, 1, 2, \dots, 2, \dots, m, \dots, m, m+1} \in B(\mathcal{H}_{\text{sym}}^{k,d_1}) \otimes \dots \otimes B(\mathcal{H}_{\text{sym}}^{k,d_m}) \otimes B(\mathbb{C}^{d_{m+1}})$, satisfying

- (1) $\beta \geq 0$,
- (2) $\beta^{T_A} \geq 0$ has a PPT [27] for all bipartitions A of the systems $\{1, \dots, 1, 2, \dots, 2, m, \dots, m, m+1\}$,
- (3) $\text{tr}_\Lambda(\beta) = \sigma$, where Λ denotes any set of indices with $k-1$ 1's, 2's, ..., m 's,
- (4) $\beta \Pi_{\text{sym}}^k = \beta$.

Here, Π_{sym}^k denotes the tensor product of the symmetric projectors on the spaces $(\mathbb{C}^{d_j})^{\otimes k}$, for $j = 1, \dots, m$, times the identity on $\mathbb{C}^{d_{m+1}}$.

Intuitively, β represents the PPT state of an ensemble of m groups of k identical bosons, plus a third particle labeled $m+1$. Any $\beta \in B(\mathcal{H}_{\text{sym}}^{k,d_1}) \otimes \dots \otimes B(\mathcal{H}_{\text{sym}}^{k,d_m}) \otimes B(\mathbb{C}^{d_{m+1}})$ satisfying the properties above is called a Bose-symmetric PPT k extension of σ .

As proven in Ref. [11], any separable state admits a Bose-symmetric PPT k extension for all k . Indeed, let $\sigma = \sum_i p_i \otimes_{j=1}^{m+1} |u_i^j\rangle\langle u_i^j|$. Then it can be verified that the state $\beta = \sum_i p_i \otimes_{j=1}^m |u_i^j\rangle\langle u_i^j|^{\otimes k} \otimes |u_i^{m+1}\rangle\langle u_i^{m+1}|$ satisfies the above constraints. Most importantly, the limiting set $\lim_{k \rightarrow \infty} S^k$ is the set of fully separable states [11].

As we explain in the main text, rather than over general entanglement witnesses, we conduct optimizations over a subset thereof. More precisely, we consider a set \mathcal{W}_m^k of multipartite operators $W_{1, \dots, m+1}$ such that $\text{tr}\{W\sigma\} \geq 0$ for all states $\sigma \in S^k$.

This set is composed by operators $W_{1, \dots, m+1}$, such that

$$\Pi_{\text{sym}}^k(W_{1, \dots, m+1} \otimes \mathbb{I}_{d_1}^{\otimes k} \otimes \dots \otimes \mathbb{I}_{d_m}^{\otimes k}) \Pi_{\text{sym}}^k = \sum_A V_A^{T_A}, \quad (\text{C1})$$

with the sum on the left running over all bipartitions A of the $km+1$ parties and $V_A \geq 0$, for all partitions A .

We next prove that any such an operator satisfies $\text{tr}(W_{1, \dots, m+1}\sigma) \geq 0$ for all states σ admitting a Bose-symmetric PPT k extension on systems $1, 2, \dots, m$. This implies, in particular, that W is an entanglement witness.

Let then σ admit a Bose-symmetric PPT k extension $\beta_{1, \dots, 1, 2, \dots, 2, \dots, m, \dots, m, m+1}$. Then, we have that

$$\begin{aligned} \text{tr}(W\sigma) &= \text{tr}\{(W \otimes \mathbb{I}_{d_1}^{\otimes k} \otimes \dots \otimes \mathbb{I}_{d_m}^{\otimes k})\beta\} \\ &= \text{tr}\left\{\Pi_{\text{sym}}^k(W \otimes \mathbb{I}_{d_1}^{\otimes k} \otimes \dots \otimes \mathbb{I}_{d_m}^{\otimes k}) \Pi_{\text{sym}}^k \beta\right\} \\ &= \text{tr}\left\{\sum_A V_A^{T_A} \beta\right\} = \sum_A \text{tr}\{V_A \beta^{T_A}\} \geq 0. \end{aligned} \quad (\text{C2})$$

Here, the first equality follows from the fact that β is an extension, the second, from it living in the symmetric subspace, and the third, from Eq. (C1). The last inequality follows from the fact that β is PPT and that $V_A \geq 0$ for all bipartitions A .

It hence follows that any map Ω satisfying the SDP conditions

- (1) $W_\Omega \in \mathcal{W}_m^k$,
- (2) $\mathbb{I}_{1,\dots,m} - \text{tr}_{m+1}(W_\Omega) \in \mathcal{W}_{m-1}^k$

is an $m \rightarrow 1$ connector in SEP world. Clearly, linear optimizations over this set can be cast as a SDP.

The DPS hierarchy also provides us with tools to define SDP *Ansätze* of $m \rightarrow m'$ connectors. In Ref. [58], it is shown that, for any state $\rho_{1,\dots,m'}$ admitting a Bose-symmetric extension (note that the PPT condition is not necessary), the state $\tilde{\Omega}(\rho) \equiv \otimes_{j=1}^{m'-1} \Omega_j^{d_j,k}(\rho)$ is separable. Here, $\Omega^{d,k}$ denotes the partially depolarizing channel

$$\Omega^{d,k}(\sigma) = \frac{k}{k+d}\sigma + \frac{d}{k+d}\frac{\mathbb{I}}{d}. \quad (\text{C3})$$

It follows that any $m \rightarrow 1$ connector with output in $\mathcal{B}(\mathcal{H}_{\text{sym}}^{k,d'_1} \otimes \dots \otimes \mathcal{H}_{\text{sym}}^{k,d'_{m'+1}} \otimes \mathbb{C}^{d_{m'}})$ can be transformed into a $m \rightarrow m'$ connector just by tracing out the extra systems and applying $\tilde{\Omega}$ at the output.

Let us finish with a trick to optimize over $m \rightarrow 2$ connectors when the output is a $\mathbb{C}^2 \otimes \mathbb{C}^2$ or a $\mathbb{C}^2 \otimes \mathbb{C}^3$ system. We again start from an $m \rightarrow 1$ connector W , with output spaces A_{m+1}, B_{m+1} . The key idea is to enforce that both W_Ω and $W_\Omega^{T_{A_{m+1}}}$ are entanglement witnesses. If that is the case, then the output of the map will be a PPT state, and so, by Ref. [62], a separable state. Imposing that $W_\Omega, W_\Omega^{T_{A_{m+1}}} \in \mathcal{W}_m^k$ is again a SDP.

APPENDIX D: NONSIGNALING BOXES ADMITTING A MPS DECOMPOSITION

The tensors $\Lambda_{a_k, x_k}^{[k]}$ of the general Svetlichny box are given by

$$\begin{aligned} \Lambda_{a,x}^{[1]} &= \frac{1}{2} \langle x | \langle x | \langle 0 | \langle a |, \\ \Lambda_{a,x}^{[k]} &= \frac{1}{2} \mathbb{I}_2 \otimes \sum_{y=0,1} |y\rangle \langle x| \otimes X^{yx} \otimes X^a, \quad \text{for } 1 < k < m, \\ \Lambda_{a,x}^{[m]} &= \sum_{y,z,s=0,1} |z\rangle |y\rangle |s\rangle \langle X^{x(y+z)+a} | s \rangle. \end{aligned} \quad (\text{D1})$$

Here, X denotes the Pauli matrix $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

The result can be proved by induction, but, to get an intuition on the construction, consider the vector $\langle \psi_k | \equiv \Lambda_{a_1, x_1}^{[1]} \Lambda_{a_2, x_2}^{[2]} \dots \Lambda_{a_k, x_k}^{[k]}$. It can be verified that $\langle \psi_k | = (1/2^k) \langle x_1 | \langle x_k | \langle x_1 x_2 \oplus \dots \oplus x_{k-1} x_k | \langle a_1 \oplus \dots \oplus a_k |$. That is, the first qubit register contains a copy of the value of x_1 , the second, the value of x_k , the third, the part of $f(x_1, \dots, x_m)$ computed so far, and the last one, the part of $a_1 \oplus \dots \oplus a_m$ computed so far.

The tensors defining box (37) are given by

$$\begin{aligned} \Lambda_{a,x}^{[1]} &= \frac{1}{2} \langle x | \langle a |, \\ \Lambda_{a,x}^{[k]} &= \frac{1}{2} M_{x,r} \otimes X^a, \quad \text{for } 1 < k < m, \\ \Lambda_{a,x}^{[m]} &= (M_{x,r} \otimes X^a) \left(|r\rangle |1\rangle + \sum_{j=0}^{r-1} |j\rangle |0\rangle \right), \end{aligned} \quad (\text{D2})$$

where $M_{x,r} = |r\rangle \langle r| + \sum_{j=0}^{r-1} |j\rangle \langle x(j+1)|$. In this case, the first register has $r+1$ levels ($|0\rangle, \dots, |r\rangle$), and it represents a counter. The second register is a qubit carrying the sum modulo 2 of the outputs.

Finally, it can be verified that a MPS representation for box (38) is given by the matrices

$$\begin{aligned} \Lambda_{a,x}^{[1]} &= \frac{1}{2} \langle x | \langle a |, \\ \Lambda_{a,x}^{[k]} &= \frac{1}{2} \tilde{M}_{x,m} \otimes X^a, \quad \text{for } 1 < k < m, \\ \Lambda_{a,x}^{[m]} &= (\tilde{M}_{x,m} \otimes X^a) \left(\left| \left\lfloor \frac{m}{2} \right\rfloor \right\rangle |1\rangle + \sum_{j=0}^{\lfloor m/2 \rfloor} |j\rangle |0\rangle \right), \end{aligned} \quad (\text{D3})$$

where $\tilde{M}_{x,m} = |\lceil (m/2) \rceil\rangle \langle \lceil (m/2) \rceil| + \sum_{j=0}^{\lceil (m/2) \rceil - 1} |j\rangle \langle j+x|$. This time, the first register has $\lceil (m/2) \rceil + 1$ levels ($|0\rangle, \dots, |\lceil (m/2) \rceil\rangle$).

APPENDIX E: NONTRIVIAL $2 \rightarrow 2$ CONNECTOR

Consider the $[2, 2] \rightarrow [2, 2]$ connector W given by the matrix

$$W = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & 0 & 0 & \frac{1}{4} & \frac{1}{2} & -\frac{1}{4} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & -\frac{1}{4} & \frac{1}{2} & \frac{1}{2} & -\frac{1}{4} & -\frac{1}{2} & 0 & \frac{1}{4} & -\frac{3}{4} \\ \frac{3}{4} & -\frac{1}{2} & 0 & \frac{1}{4} & \frac{1}{2} & -\frac{1}{4} & -\frac{1}{2} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{4} & 0 & -\frac{1}{4} & \frac{1}{4} & \frac{1}{2} & -\frac{1}{4} & 0 & -\frac{1}{4} & \frac{3}{4} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{4} & \frac{1}{2} & 0 & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{4} \\ \frac{3}{4} & 0 & 0 & -\frac{1}{4} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & 0 & -\frac{1}{4} & -\frac{1}{4} & 0 & \frac{1}{2} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{2} \\ \frac{1}{4} & 0 & \frac{1}{2} & \frac{1}{4} & -\frac{1}{2} & 0 & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{2} \end{pmatrix}. \quad (\text{E1})$$

We are using the abbreviated notation; i.e., both the input and output distributions $P(A, B), Q(C, D)$ are represented as nine-entry vectors, with $Q = W\dot{P}$. It can be verified that W acting over any deterministic point generates a local bipartite box. However, $W \cdot \bar{P}$ generates a box with negative probabilities when \bar{P} corresponds to the one of

the variants of the PR box, $P(a, b|x, y) = \frac{1}{2}\delta(\bar{a} \oplus \bar{b}, \bar{x} \bar{y})$ [63], where $\delta(s, t)$ denotes the Kronecker delta. W is thus not a wiring. Furthermore, it can be verified, using semi-definite programming, that

$$\min_{W'} \|W - W'\|_{\infty} \approx 0.078, \quad (\text{E2})$$

where the minimum is taken over all nondeterministic transformations W' (in matrix form) which can be factored out as shown in Fig. 7.

-
- [1] O. Gühne and G. Tóth, *Entanglement Detection*, *Phys. Rep.* **474**, 1 (2009).
- [2] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, *Bell Nonlocality*, *Rev. Mod. Phys.* **86**, 419 (2014).
- [3] J.-M. A. Allen, J. Barrett, D. C. Horsman, C. M. Lee, and R. W. Spekkens, *Quantum Common Causes and Quantum Causal Models*, *Phys. Rev. X* **7**, 031021 (2017).
- [4] N. Brunner, M. Navascués, and T. Vértesi, *Dimension Witnesses and Quantum State Discrimination*, *Phys. Rev. Lett.* **110**, 150501 (2013).
- [5] M. Navascués, S. Pironio, and A. Acín, *Bounding the Set of Quantum Correlations*, *Phys. Rev. Lett.* **98**, 010401 (2007).
- [6] M. Navascués and T. Vértesi, *Bounding the Set of Finite Dimensional Quantum Correlations*, *Phys. Rev. Lett.* **115**, 020501 (2015).
- [7] M. Navascués, A. Feix, M. Araújo, and T. Vértesi, *Characterizing Finite-Dimensional Quantum Behavior*, *Phys. Rev. A* **92**, 042117 (2015).
- [8] E. Wolfe, R. W. Spekkens, and T. Fritz, *The Inflation Technique for Causal Inference with Latent Variables*, *J. Causal Infer.* **7**, 20170020 (2019).
- [9] A. C. Doherty, P. A. Parrilo, and F. M. Spedalieri, *Distinguishing Separable and Entangled States*, *Phys. Rev. Lett.* **88**, 187904 (2002).
- [10] A. C. Doherty, P. A. Parrilo, and F. M. Spedalieri, *Complete Family of Separability Criteria*, *Phys. Rev. A* **69**, 022308 (2004).
- [11] A. C. Doherty, P. A. Parrilo, and F. M. Spedalieri, *Detecting Multipartite Entanglement*, *Phys. Rev. A* **71**, 032333 (2005).
- [12] M. Fannes, B. Nachtergaele, and R. Werner, *Finitely Correlated States on Quantum Spin Chains*, *Commun. Math. Phys.* **144** (1992).
- [13] F. Verstraete and J. I. Cirac, *Matrix Product States Represent Ground States Faithfully*, *Phys. Rev. B* **73**, 094423 (2006).
- [14] M. B. Hastings, *Locality in Quantum and Markov Dynamics on Lattices and Networks*, *Phys. Rev. Lett.* **93**, 140402 (2004).
- [15] G. Vidal, *Class of Quantum Many-Body States That Can Be Efficiently Simulated*, *Phys. Rev. Lett.* **101**, 110501 (2008).
- [16] N. Schuch, I. Cirac, and D. Perez-García, *PEPS as Ground States: Degeneracy and Topology*, *Ann. Phys. (Amsterdam)* **325**, 2153 (2010).
- [17] R. Orus, *A Practical Introduction to Tensor Networks*, *Ann. Phys. (Amsterdam)* **349**, 117 (2014).
- [18] M. Cramer, M. B. Plenio, S. T. Flammia, D. Gross, S. D. Bartlett, R. Somma, O. Landon-Cardinal, Y.-K. Liu, and D. Poulin, *Efficient Quantum State Tomography*, *Nat. Commun.* **1**, 149 (2010).
- [19] T. Baumgratz, D. Gross, M. Cramer, and M. B. Plenio, *Scalable Reconstruction of Density Matrices*, *Phys. Rev. Lett.* **111**, 020401 (2013).
- [20] L. Hardy, *Quantum Theory from Five Reasonable Axioms*, [arXiv:quant-ph/0101012](https://arxiv.org/abs/quant-ph/0101012).
- [21] B. Daki and A. Brukner, in *Deep Beauty: Understanding the Quantum World through Mathematical Innovation*, edited by H. Halvorson (Cambridge University Press, Cambridge, England, 2011), pp. 365–392.
- [22] G. Chiribella, G. M. D’Ariano, and P. Perinotti, *Probabilistic Theories with Purification*, *Phys. Rev. A* **81**, 062348 (2010).
- [23] H. Barnum, M. P. Mueller, and C. Ududec, *Higher-Order Interference and Single-System Postulates Characterizing Quantum Theory*, *New J. Phys.* **16**, 123029 (2014).
- [24] A. J. Short and J. Barrett, *Strong Nonlocality: A Trade-Off between States and Measurements*, *New J. Phys.* **12**, 033034 (2010).
- [25] E. D. Nering and A. W. Tucker, *Linear Programs and Related Problems*, Computer Science and Scientific Computing (Academic Press, New York, 1993).
- [26] L. Vandenberghe and S. Boyd, *Semidefinite Programming*, *SIAM Rev.* **38**, 49 (1996).
- [27] S. Peres, *Separability Criterion for Density Matrices*, *Phys. Rev. Lett.* **77**, 1413 (1996).
- [28] M. Navascués, Y. Guryanova, M. J. Hoban, and A. Acín, *Almost Quantum Correlations*, *Nat. Commun.* **6**, 6288 (2015).
- [29] M. Berta, F. G. S. L. Brandão, J. Haegeman, V. B. Scholz, and F. Verstraete, *Thermal States as Convex Combinations of Matrix Product States*, *Phys. Rev. B* **98**, 235154 (2018).
- [30] R. N. C. Pfeifer, J. Haegeman, and F. Verstraete, *Faster Identification of Optimal Contraction Sequences for Tensor Networks*, *Phys. Rev. E* **90**, 033315 (2014).
- [31] R. N. C. Pfeifer, G. Evenbly, S. Singh, and G. Vidal, *NCOV: A Tensor Network Contractor for MATLAB*, [arXiv:1402.0939](https://arxiv.org/abs/1402.0939).
- [32] Gurobi Optimization, Inc., *GUROBI Optimizer Reference Manual*, 2016, https://www.gurobi.com/wp-content/plugins/hd_documentations/documentation/9.0/refman.pdf.
- [33] J. F. Sturm, *Using SEDUMI.02, a MATLAB Toolbox for Optimization over Symmetric Cones*, *Optim. Methods Software* **11**, 625 (1999).
- [34] L. Vandenberghe and S. Boyd, *The MOSEK Optimization Toolbox for MATLAB Manual*, Version 7.0 (Revision 140) (MOSEK ApS, Copenhagen).
- [35] J. Löfberg, in *Proceedings of the CACSD Conference, Taipei, Taiwan, 2004*, <https://ieeexplore.ieee.org/document/1393890>.
- [36] M. Grant and S. Boyd, *CVX: MATLAB Software for Disciplined Convex Programming*, Version 2.1, <http://cvxr.com/cvx> (2014).
- [37] M. Grant and S. Boyd, in *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, edited by V. Blondel, S. Boyd, and H. Kimura (Springer-Verlag, Berlin, 2008), pp. 95–110.
- [38] In quantum theory, one can construct this vector by writing density matrices in vector form.

- [39] J. Barrett, *Information Processing in Generalized Probabilistic Theories*, *Phys. Rev. A* **75**, 032304 (2007).
- [40] A. J. Short and J. Barrett, *Strong Nonlocality: A Trade-Off between States and Measurements*, *New J. Phys.* **12**, 033034 (2010).
- [41] A. B. Sainz, Y. Guryanova, A. Acín, and M. Navascués, *Almost-Quantum Correlations Violate the No-Restriction Hypothesis*, *Phys. Rev. Lett.* **120**, 200402 (2018).
- [42] A. Jamiolkowski, *An Effective Method of Investigation of Positive Maps on the Set of Positive Definite Operators*, *Rep. Math. Phys.* **3**, 275 (1972).
- [43] M.-D. Choi, *Positive Linear Maps on c^* -Algebras*, *Can. J. Math.* **24**, 520 (1972).
- [44] J. Allcock, N. Brunner, N. Linden, S. Popescu, P. Skrzypczyk, and T. Vértesi, *Closed Sets of Nonlocal Correlations*, *Phys. Rev. A* **80**, 062107 (2009).
- [45] Yu.-C. Wu, P. Badziag, and M. Żukowski, *Clauser-Horne-Shimony-Holt-Type Bell Inequalities Involving a Party with Two or Three Local Binary Settings*, *Phys. Rev. A* **79**, 022110 (2009).
- [46] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, *Proposed Experiment to Test Local Hidden-Variable Theories*, *Phys. Rev. Lett.* **23**, 880 (1969).
- [47] R. F. Werner and M. M. Wolf, *All-Multipartite Bell-Correlation Inequalities for Two Dichotomic Observables per Site*, *Phys. Rev. A* **64**, 032112 (2001).
- [48] K. F. Pál and T. Vértesi, *Maximal Violation of the I3322 Inequality Using Infinite Dimensional Quantum Systems*, *Phys. Rev. A* **82**, 022116 (2010).
- [49] S. Boyd, L. Xiao, and A. Mutapcic, *Subgradient Methods*, Stanford University, 2004.
- [50] W. Dür, G. Vidal, and J. I. Cirac, *Three Qubits Can Be Entangled in Two Inequivalent Ways*, *Phys. Rev. A* **62**, 062314 (2000).
- [51] N. D. Mermin, *Extreme Quantum Entanglement in a Superposition of Macroscopically Distinct States*, *Phys. Rev. Lett.* **65**, 1838 (1990).
- [52] M. Navascués, S. Pironio, and A. Acín, *A Convergent Hierarchy of Semidefinite Programs Characterizing the Set of Quantum Correlations*, *New J. Phys.* **10**, 073013 (2008).
- [53] A. C. Doherty, Y.-C. Liang, B. Toner, and S. Wehner, in *Proceedings of the 2008 IEEE 23rd Annual Conference on Computational Complexity, CCC '08* (IEEE Computer Society, Washington, DC, 2008), pp. 199–210.
- [54] M. Berta, O. Fawzi, and V. B. Scholz, *Quantum Bilinear Optimization*, *SIAM J. Optim.* **26**, 1529 (2016).
- [55] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts, *Nonlocal Correlations as an Information-Theoretic Resource*, *Phys. Rev. A* **71**, 022101 (2005).
- [56] A. C. Doherty, P. A. Parrilo, and F. M. Spedalieri, *Bell Inequalities and the Separability Criterion*, *Phys. Lett. A* **271**, 319 (2000).
- [57] G. Tóth, C. Knapp, O. Gühne, and H. J. Briegel, *Spin Squeezing and Entanglement*, *Phys. Rev. A* **79**, 042334 (2009).
- [58] M. Navascués, M. Owari, and M. B. Plenio, *Power of Symmetric Extensions for Entanglement Detection*, *Phys. Rev. A* **80**, 052306 (2009).
- [59] C. H. Bennett, D. P. DiVincenzo, T. Mor, P. W. Shor, J. A. Smolin, and B. M. Terhal, *Unextendible Product Bases and Bound Entanglement*, *Phys. Rev. Lett.* **82**, 5385 (1999).
- [60] N. Johnston, *The Structure of Qubit Unextendible Product Bases*, *J. Phys. A* **47**, 424034 (2014).
- [61] M. Navascués, S. Pironio, and A. Acín, *Convergent Relaxations of Polynomial Optimization Problems with Non-Commuting Variables*, *SIAM J. Optim.* **20**, 2157 (2010).
- [62] M. Horodecki, P. Horodecki, and R. Horodecki, *Separability of Mixed States: Necessary and Sufficient Conditions*, *Phys. Lett. A* **223**, 1 (1996).
- [63] S. Popescu and D. Rohrlich, *Quantum Nonlocality as an Axiom*, *Found. Phys.* **24**, 379 (1994).