Efficient device-independent entanglement detection for multipartite systems

F. Baccari[1],* D. Cavalcanti[1], P. Wittek[1,2], and A. Acín[1,3]
[1]*ICFO - Institut de Ciences Fotoniques, The Barcelona Institute of Science and Technology, 08860 Castelldefels, Spain*
[2] *University of Borås, 50190 Borås, Sweden*
[3]*ICREA - Institució Catalana de Recerca i Estudis Avançats, Lluis Campanys 23, 08010 Barcelona, Spain*

Entanglement is one of the most studied properties of quantum mechanics for its application in quantum information protocols. Nevertheless, detecting the presence of entanglement in large multipartite sates keeps being a great challenge both from the theoretical and the experimental point of view. Most of the known methods either have computational costs that scale inefficiently with the number of parties or require more information on the state than what is attainable in every-day experiments. We introduce a new technique for entanglement detection that provides several important advantages in these respects. First, its scales efficiently with the number of parties, thus allowing for application to systems composed by up to few tens of parties. Second, it needs only the knowledge of a subset of all possible measurements on the state, therefore being apt for experimental implementation. Moreover, since it is based on the detection of nonlocality, our method is device-independent. We report several examples of its implementation for well-known multipartite states, showing that the introduced technique has a promising range of applications.

I. INTRODUCTION

Entanglement is the key ingredient for several protocols in quantum information theory, such as quantum teleportation [1], quantum key distribution [2], measurement based quantum computation [3] and quantum metrology schemes [4]. Therefore, developing techniques to detect the presence of entanglement in quantum states is crucial. In the past years several methods have been introduced. However their application to large multipartite systems turns out to be impractical, both from the theoretical and experimental side.

The most general way to detect entanglement in a given system consists of reconstructing its quantum state using tomography and then applying any entanglement criterion to the resulting state [5]. This, however, is costly both from an experimental and a theoretical perspective. First, determining the state of large quantum systems is impractical in experiments, given that quantum tomography implies measuring a number of observables that increases exponentially with the number of systems, e.g., 3^N observables even in the simplest case of N qubits [6]. Second, determining whether an arbitrary state is entangled is known to be a hard problem – to the best of our knowledge, the computational resources of the most efficient known algorithm scales as $O(N \exp N)$ [7]. Due to these problems, it is desirable to develop entanglement detection techniques for which the experimental and computational requirements scale efficiently with the size and dimensionality of the systems.

One possible approach is to make use of entanglement witnesses. These are criteria for detecting entanglement that require measuring only some expectation values of local observables [8]. In particular, attempts have been made to derive witnesses that adapt to the available set of data [9], especially involving only two-body correlators [10] or a few global measurements [11, 12]. Moreover, a decisive step forward in the direction of efficient multipartite entanglement witnessing has been recently presented in [13]. It consists of a general technique to derive witnesses that require a number of measurement that does not scale with the number of particles in the system. Nonetheless, entanglement witnesses constitute a method that lacks generality, given that the known witnesses are generally tailored to detect very specific states.

A qualitatively different approach is based on Bell nonlocality [14]. The violation of a Bell inequality provides a certificate of the entanglement in the state. Moreover, it has the advantage that it can be assessed in a device-independent manner, i.e. without making any assumption on the actual experimental implementation [15]. In the general case, verifying whether a set of observed correlations violates some Bell inequality can be done via linear programming [16]. However, the number of variables involved grows as 4^N already for the simplest scenarios where only two dichotomic measurements per party are applied [15].

Here we present a novel technique for device-independent detection of entanglement that is efficient both experimentally and computationally. On the one hand, it requires the knowledge of a subset of all possible measurements, most of them consisting of few-body correlation functions, which makes it suitable for practical implementations. On the other hand, it can be applied to any set of observed correlations and can be implemented by semi-definite programming involving a number of variables that grows polynomially with N. Of course, all these nice properties become possible only because our method for entanglement detection is a relaxation of the initial hard problem. However, and despite being a relaxation, we demonstrate the power of our approach by showing how it can be successfully applied to several physically relevant examples.

This article is organized as follows: in Section II and III we introduce the basic idea of the method together with the application to a simple scenario. Section IV is devoted to the

* flavio.baccari@icfo.eu

presentation of its geometrical interpretation together with the comparison to the other techniques. In Section V and VI we list some examples of application of the method to relevant classes of states. Lastly, Section VII contains conclusions and some future perspectives.

II. THE METHOD

We consider an entanglement detection scenario in which N observers, denoted by $A_1, \ldots, A_N$, share an N-partite quantum state ρ_N. Each A_i performs m possible measurements, each having d outcomes. We represent the measurement of party i by $M_{x_i}^{a_i}$, where $x_i \in \{0, \ldots, m-1\}$ denotes the measurement choice and $a_i \in \{0, \ldots, d-1\}$ is the corresponding outcomes.

By repeating the experiment sufficiently many times, the observers can estimate the conditional probabilities

$$p(a_1, \ldots, a_N | x_1, \ldots, x_N) = \mathrm{tr}(M_{x_1}^{a_1} \otimes \ldots \otimes M_{x_N}^{a_N} \rho_N), \quad (1)$$

of getting the different outcomes depending on the measurements they have performed. The conditional probability distribution $p(a_1, \ldots, a_N | x_1, \ldots, x_N)$ describes the correlations observed among the observers when applying the local measurements $M_{x_i}^{a_i}$ on the state ρ_N.

Our method is based on the following chain of implications, proven below:

A quantum state ρ_N is separable

$\Downarrow$

Any local measurements performed on it produce correlations that have a local model

$\Downarrow$

Local correlations can always be realized by performing commuting local measurements on a quantum state

$\Downarrow$

Correlations with commuting measurements define a positive moment matrix with constraints associated to the commutation

Let us now explain all these implications in detail.

First, given a separable quantum state, *i.e.* $\rho_N = \sum_\lambda p_\lambda \bigotimes_i \rho_\lambda^{A_i}$, any set of conditional probability distributions obtained after performing local measurements on it admits a decomposition of the following form

$$\begin{aligned} p(a_1, \ldots, a_N | x_1, \ldots, x_N) &= \sum_\lambda p_\lambda \, \mathrm{tr}(\bigotimes_i M_{x_i}^{a_i} \bigotimes_i \rho_\lambda^{A_i}) \\ &= \sum_\lambda p_\lambda \prod_{i=1}^N p(a_i | x_i, \lambda), \end{aligned} \quad (2)$$

where $p(a_i|x_i, \lambda) = \mathrm{tr}(M_{x_i}^{a_i} \rho_\lambda^{A_i})$. In the context of Bell nonlocality, distributions that can be written in this form are called local [15]. Local correlations do not violate any Bell inequality. If the set of observed distributions (1) is nonlocal, we can conclude that the shared state is entangled.

The second ingredient is that any local set of probability distributions has a quantum realization in terms of local *commuting measurements* applied to a quantum state [17]. In order to see it more explicitly we first realise that any decomposition of the form (2) can be rewritten as

$$p(a_1, \ldots, a_N | x_1, \ldots, x_N) = \sum_\lambda q_\lambda \prod_{i=1}^N D(a_i | x_i, \lambda) \quad (3)$$

where $D(a_i|x_i, \lambda)$ are deterministic functions that give a fixed outcome a for each measurement, *i.e.* $D(a_i|x_i, \lambda) = \delta_{a_i, \lambda(x_i)}$, such that $a_i = \lambda(x_i)$, being $\lambda(\cdot)$ a function from $\{0, \ldots, m-1\}$ to $\{0, \ldots, d-1\}$ [15]. It is easy to see that any such decomposition can be reproduced by choosing the multipartite state $\rho_N = \sum_\lambda q_\lambda |\lambda\rangle\langle\lambda|^{\otimes N}$ and measurement operators of the form $M_{x_i}^{a_i} = \sum_{\lambda'} D(a_i|x_i, \lambda')|\lambda'\rangle\langle\lambda'|$. In particular, $[M_{x_i}^{(i)}, M_{x_i'}^{(i)}] = 0 \; \forall \; i \; x_i$, and x_i'.

The last step consists in using a modified version of the Navascues-Pironio-Acin (NPA) hierarchy [18, 19] that takes into account the commutativity of the local measurements to test if the observed probability distribution is local (a similar idea was introduced in the context of quantum steering [20] – see also [21]). The NPA hierarchy consists of a sequence of tests aimed at certifying if a given set of probability distributions has a quantum realisation (1). In NPA one imposes the commutativity of the measurements between the distant parties. Now, we will impose the extra constraints that the local measurements on each party also commute. The resulting SDP hierarchy is nothing but an application in this context of the more general method for polynomial optimization over non-commuting variables introduced in [22], see also [23]. As noticed there, by imposing commutativity of all the variables this general hierarchy reduces to the well-known Lassere hierarchy, namely the relaxation for polynomial optimization of commuting variables [24]. An application of this relaxation technique to describe local correlations was also proposed in [25]. However, to the best of our knowledge, no systematic analysis of its application to multipartite scenarios has been considered so far.

It is convenient for what follows to recall the main ingredients of the NPA hierarchy [18, 19], which, as said, was designed to characterize probability distributions with a quantum realisation (1). Consider a set $\mathcal{O}$, composed by some products of the measurements operators $\{M_{x_i}^{a_i}\}$ or linear combinations of them. By indexing the elements in the set as $\mathcal{O}_i$ with $i = 1,\ldots k$, we introduce the so-called moment matrix Γ as the $k \times k$ matrix whose entries are defined by $\Gamma_{ij} = tr(\rho_N \mathcal{O}_i^\dagger \mathcal{O}_j)$. For any choice of measurements and state, it can be shown that Γ satisfies the following properties: i) it is positive semidefinite, ii) its entries satisfy a series of linear constraints associated to the commutation relations among measurement operators by different parties and the fact that they correspond to projectors, iii) some of its entries can be computed form the observed probability distribution (1), iv) some of its entries correspond to unobservable numbers (e.g. when O_i and O_j involve non-commuting observables). Based on these facts one can define a hierarchy of tests to check whether a given set of correlations has a quantum realisation. One first defines the sets $\mathcal{O}_\nu$ composed of products of at most ν of the measurement operators, and creates the corresponding Γ matrix using the set of correlations and leaving the unknown ones as variables. Then one seeks for values for these variables that could make the Γ positive. This problem constitutes a semidefinite program (SDP), for which some efficient solving algorithms are known [26]. If no such values are found this means that the set of correlations used does not have a quantum realisation. By increasing the value of ν, one gets a sequence of stricter and stricter ways of testing the belonging of a distribution to the quantum set.

We can now use the same idea to define a hierarchy of conditions to test whether a given set of correlations has a quantum realisations with commuting measurements. To do so we simply impose additional linear constraints on the entries of the moment matrix resulting from assuming that the local measurements also commute (for a more detailed discussion, we refer to Appendix A). Thus, given a set of observed probability distributions one can use them to build an NPA-type matrix with the additional linear constraints associated to the local commutation relations, and run an SDP to check its positivity to certify if the considered set of correlations can not be obtained by measuring a separable state. The convergence of the method follows from the results in [19, 22] and, therefore, any nonlocal correlation would fail the SDP test at a finite step of the sequence.

Depending on the level of the hierarchy, one might not need the knowledge of the full probability distribution. To make this clearer, let us define the marginal distributions

$$p(a_{i_1},\ldots,a_{i_k}|x_{i_1},\ldots x_{i_k}) = \mathrm{tr}(M_{x_{i_1}}^{a_{i_1}} \otimes \ldots \otimes M_{x_{i_k}}^{a_{i_k}} \rho_{i_1,\ldots,i_k}) \quad (4)$$

where $0 \leq i_1 < \ldots < i_k < N$, $1 \leq k \leq N$ and $\rho_{i_1,\ldots,i_k}$ is the reduced state of ρ_N corresponding to the considered subset of parties. Marginals can equivalently be obtained from the full distribution (1) by summing over the remaining outcomes. By looking at (4), it is evident that to define a marginal distribution involving k parties one requires the product of k measurements $M_{x_i}^{a_i}$. Now, given that the operators of the set $\mathcal{O}_\nu$ contain products of at most ν measurement operators, the terms in the moment matrix at level ν can only coincide with the marginals of the observed distribution of up to $k = 2\nu$ parties. Therefore, in the multipartite setting, fixing the level of the hierarchy is also a way to limit the order of the marginals that can be assigned in the moment matrix.

We conclude this section by introducing a bit more of notation, motivated by the fact that, while our method is general and applies to any Bell scenario, in what follows we are mostly going to consider scenarios involving two-output measurements (resulting from projective measurements performed on qubits). Then, all the information in the observed distribution (1) can be expressed in terms of the correlation functions, also known as correlators,

$$\langle M_{j_1}^{(i_1)} \ldots M_{j_k}^{(i_k)} \rangle = \sum_{a_{i_1},\ldots,a_{i_k}} (-1)^{\sum_{l=1}^k a_{i_l}} p(a_{i_1} \ldots, a_{i_k}|j_1,\ldots,j_k) \quad (5)$$

where $0 \leq i_1 < \ldots < i_k < N$, $j_l \in \{0, m-1\}$ and $1 \leq k \leq N$. The value of k represents the order of the correlators: for instance, expectation values $\langle M_{j_1}^{(i_1)} M_{j_2}^{(i_2)} \rangle$ are of order two. Correlators of order N are often referred to as full-body correlators.

III. A SIMPLE EXAMPLE

After presenting the general idea of the method, it is convenient to illustrate it with a simple example consisting of two parties before applying it to relevant multipartite scenarios. In what follows, we present the explicit form of the moment matrix for the case of $N = 2$, two dicothomic measurements per party and level $\nu = 2$ of the hierarchy. For the sake of simplicity, we rename the measurement operators for the two parties as $\mathcal{A}_x$ and $\mathcal{B}_y$, with $x, y = 0, 1$. In this scenario, the set of operators reads as $\mathcal{O}_2 = \{\mathbb{1}, \mathcal{A}_0, \mathcal{A}_1, \mathcal{B}_0, \mathcal{B}_1, \mathcal{A}_0\mathcal{A}_1, \mathcal{A}_0\mathcal{B}_0, \mathcal{A}_0\mathcal{B}_1, \mathcal{A}_1\mathcal{B}_0, \mathcal{A}_1\mathcal{B}_1, \mathcal{B}_0\mathcal{B}_1\}$. The corresponding moment matrix is

$$
\Gamma = \begin{pmatrix}
1 & \langle\mathcal{A}_0\rangle & \langle\mathcal{A}_1\rangle & \langle\mathcal{B}_0\rangle & \langle\mathcal{B}_1\rangle & v_1 & \langle\mathcal{A}_0\mathcal{B}_0\rangle & \langle\mathcal{A}_0\mathcal{B}_1\rangle & \langle\mathcal{A}_1\mathcal{B}_0\rangle & \langle\mathcal{A}_1\mathcal{B}_1\rangle & v_2 \\
\langle\mathcal{A}_0\rangle & 1 & v_1 & \langle\mathcal{A}_0\mathcal{B}_0\rangle & \langle\mathcal{A}_0\mathcal{B}_1\rangle & \langle\mathcal{A}_1\rangle & \langle\mathcal{B}_0\rangle & \langle\mathcal{B}_1\rangle & v_3 & v_4 & v_5 \\
\langle\mathcal{A}_1\rangle & v_1^* & 1 & \langle\mathcal{A}_1\mathcal{B}_0\rangle & \langle\mathcal{A}_1\mathcal{B}_1\rangle & v_6 & v_3^* & v_4^* & \langle\mathcal{B}_0\rangle & \langle\mathcal{B}_1\rangle & v_7 \\
\langle\mathcal{B}_0\rangle & \langle\mathcal{A}_0\mathcal{B}_0\rangle & \langle\mathcal{A}_1\mathcal{B}_0\rangle & 1 & v_2 & v_3 & \langle\mathcal{A}_0\rangle & v_5 & \langle\mathcal{A}_1\rangle & v_7 & \langle\mathcal{B}_1\rangle \\
\langle\mathcal{B}_1\rangle & \langle\mathcal{A}_0\mathcal{B}_1\rangle & \langle\mathcal{A}_1\mathcal{B}_1\rangle & v_2^* & 1 & v_4 & v_5^* & \langle\mathcal{A}_0\rangle & v_7^* & \langle\mathcal{A}_1\rangle & v_8 \\
v_1^* & \langle\mathcal{A}_1\rangle & v_6^* & v_3^* & v_4^* & 1 & \langle\mathcal{A}_1\mathcal{B}_0\rangle & \langle\mathcal{A}_1\mathcal{B}_1\rangle & v_9 & v_{10} & v_{11} \\
\langle\mathcal{A}_0\mathcal{B}_0\rangle & \langle\mathcal{B}_0\rangle & v_3 & \langle\mathcal{A}_0\rangle & v_5 & \langle\mathcal{A}_1\mathcal{B}_0\rangle & 1 & v_2 & v_1 & v_{12} & \langle\mathcal{A}_0\mathcal{B}_1\rangle \\
\langle\mathcal{A}_0\mathcal{B}_1\rangle & \langle\mathcal{B}_1\rangle & v_4 & v_5^* & \langle\mathcal{A}_0\rangle & \langle\mathcal{A}_1\mathcal{B}_1\rangle & v_2^* & 1 & v_{13} & v_1 & v_{14} \\
\langle\mathcal{A}_1\mathcal{B}_0\rangle & v_3^* & \langle\mathcal{B}_0\rangle & \langle\mathcal{A}_1\rangle & v_7 & v_9^* & v_1^* & v_{13}^* & 1 & v_2 & \langle\mathcal{A}_1\mathcal{B}_1\rangle \\
\langle\mathcal{A}_1\mathcal{B}_1\rangle & v_4^* & \langle\mathcal{B}_1\rangle & v_7^* & \langle\mathcal{A}_1\rangle & v_{10}^* & v_{12}^* & v_1^* & v_2^* & 1 & v_{15} \\
v_2^* & v_5^* & v_7^* & \langle\mathcal{B}_1\rangle & v_8^* & v_{11}^* & \langle\mathcal{A}_0\mathcal{B}_1\rangle & v_{14}^* & \langle\mathcal{A}_1\mathcal{B}_1\rangle & v_{15}^* & 1
\end{pmatrix} \tag{6}
$$

where we have defined the following unassigned variables

$$
\begin{aligned}
v_1 &= \langle\mathcal{A}_0\mathcal{A}_1\rangle\,, & v_2 &= \langle\mathcal{B}_0\mathcal{B}_1\rangle\,, & v_3 &= \langle\mathcal{A}_0\mathcal{A}_1\mathcal{B}_0\rangle\,, & v_4 &= \langle\mathcal{A}_0\mathcal{A}_1\mathcal{B}_1\rangle\,,\\
v_5 &= \langle\mathcal{A}_0\mathcal{B}_0\mathcal{B}_1\rangle\,, & v_6 &= \langle\mathcal{A}_1\mathcal{A}_0\mathcal{A}_1\rangle\,, & v_7 &= \langle\mathcal{A}_1\mathcal{B}_0\mathcal{B}_1\rangle\,, & v_8 &= \langle\mathcal{B}_1\mathcal{B}_0\mathcal{B}_1\rangle\,,\\
v_9 &= \langle\mathcal{A}_1\mathcal{A}_0\mathcal{A}_1\mathcal{B}_0\rangle\,, & v_{10} &= \langle\mathcal{A}_1\mathcal{A}_0\mathcal{A}_1\mathcal{B}_1\rangle\,, & v_{11} &= \langle\mathcal{A}_1\mathcal{A}_0\mathcal{B}_0\mathcal{B}_1\rangle\,, & v_{12} &= \langle\mathcal{A}_0\mathcal{A}_1\mathcal{B}_0\mathcal{B}_1\rangle\,,\\
v_{13} &= \langle\mathcal{A}_0\mathcal{A}_1\mathcal{B}_1\mathcal{B}_0\rangle\,, & v_{14} &= \langle\mathcal{A}_0\mathcal{B}_1\mathcal{B}_0\mathcal{B}_1\rangle\,, & v_{15} &= \langle\mathcal{A}_1\mathcal{B}_1\mathcal{B}_0\mathcal{B}_1\rangle\,.
\end{aligned} \tag{7}
$$

Now, if we further impose commutativity of all the measurements, namely $[\mathcal{A}_0, \mathcal{A}_1] = 0$, $[\mathcal{B}_0, \mathcal{B}_1] = 0$, the corresponding linear constraints reduce the number of variables. Explicitly, one gets $v_i^* = v_i$ for any $i = 1, \ldots, 15$, and also

$$
\begin{aligned}
v_6 &= \langle\mathcal{A}_0\rangle\,, & v_8 &= \langle\mathcal{B}_0\rangle\,, & v_9 = v_{14} &= \langle\mathcal{A}_0\mathcal{B}_0\rangle\,,\\
v_{10} &= \langle\mathcal{A}_0\mathcal{B}_1\rangle\,, & v_{15} &= \langle\mathcal{A}_1\mathcal{B}_0\rangle\,, & v_{11} = v_{12} &= v_{13}\,.
\end{aligned} \tag{8}
$$

For a visual representation, the variables that become identical because of the commutativity constraints are represented by the same color in (6). For any set of observed correlations $\{\langle\mathcal{A}_x\rangle, \langle\mathcal{B}_y\rangle, \langle\mathcal{A}_x\mathcal{B}_y\rangle\}$, testing whether it is local can be done in the following steps: assigning the values to the entries of Γ that can be derived from the observed correlations and leaving the remaining terms as variables; then checking whether there is a assignment for such variables such that the matrix is positive semidefinite.

For instance, it is possible to check that any set of correlations that violates the well-known CHSH inequality [27]

$$
\mathcal{I}_{CHSH} = \langle\mathcal{A}_0\mathcal{B}_0\rangle + \langle\mathcal{A}_0\mathcal{B}_1\rangle + \langle\mathcal{A}_1\mathcal{B}_0\rangle - \langle\mathcal{A}_1\mathcal{B}_1\rangle \leq 2 \tag{9}
$$

is incompatible with a positive semidefinite matrix (6). Moreover, we notice that, in this particular scenario, any set of nonlocal correlations has to violate CHSH (or symmetrical equivalent of it) [15]. Therefore, it turns out that in this case the second level of the hierarchy is already capable to detect any nonlocal correlation. That is, even if any finite step in the hierarchy in principle represents a relaxation of the general problem, the second level happens to be tight in this particularly simple case.

IV. GEOMETRICAL CHARACTERIZATION OF CORRELATIONS

Before presenting the applications of our method, we review a geometrical perspective, schematically represented in Figure 1 [15], that is useful when studying correlations among many different parties. It is known that the set of local correlations (2) defines a polytope, i.e. a convex set with a finite number of extremal points. Such points coincide with the deterministic strategies $D(a_i|x_i, \lambda)$ introduced in (3) and can be easily defined for any multipartite scenario. As represented in Figure 1, the set of quantum correlations (1) is strictly bigger than the local set. All the points lying outside the set $\mathcal{L}$ represent nonlocal correlations.

Determining whether some observed correlations are nonlocal corresponds to checking whether they are associated to a point inside the local set. A very simple way to detect nonlocality is by means of Bell inequalities. They are inequalities that are satisfied by any local distribution and geometrically they constitute hyperplanes separating the $\mathcal{L}$ set from the rest of the correlations. Violating a Bell inequality directly implies that the corresponding distribution is nonlocal. However, there can be nonlocal correlations that are not detected by a given inequality, meaning that the fall on the same side of the hyperplane as the local correlations.

On the other hand, a very general technique to check if a point belongs to the local set consists in determining if it can be decomposed as a convex combination of its vertices [16]. Such question is a typical instance of a linear programming problem, for which there exist algorithms that run in a time that is polynomial in the number of variables [28]. Nevertheless, finding a convex decomposition in the multipartite scenario is generally an intractable problem because the number of deterministic strategies grows as d^{mN}. Already in the simplest cases in which each party measures only $m = 2, 3$ dicothomic measurements, the best approach currently known stops at $N = 11$ and $N = 7$ respectively [29].

Coming back to the SDP method presented in the previous section, we can now show how the technique can help in overcoming the limitations imposed to the linear program. Let us define the family of sets $\mathcal{L}_\nu$ as the ones composed by the correlations that are compatible with the moment matrix Γ defined by the observables $\mathcal{O}_\nu$ and the additional constraints of

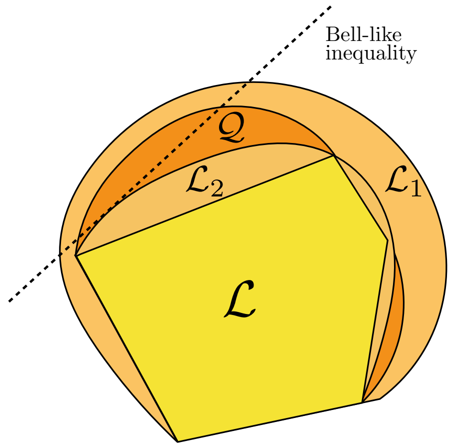

FIG. 1. Pictorial representation of the sets of correlations, together with our approach to detection of multipartite nonlocality. The $\mathcal{L}$ and $\mathcal{Q}$ sets delimit the local and quantum correlations respectively. As it is shown in the picture, the first forms a polytope, namely a convex set delimited by a finite amount of extremal points, while the second, despite being still convex, is not a polytope. The light orange sets are the first representatives of the hierarchy $\mathcal{L}_1 \supseteq \mathcal{L}_2 \supseteq \ldots \supseteq \mathcal{L}$ approximating the local set from outside. It can be seen that some of the quantum correlations lie outside the $\mathcal{L}_2$, meaning that they are detected as nonlocal from the SDP relaxation at the second level. The dotted line shows a Bell-like inequality that can be obtained by the corresponding dual problem.

commuting measurements. Given that any local distribution has a quantum representation with commuting measurements, the series $\mathcal{L}_1 \supseteq \mathcal{L}_2 \supseteq \ldots \supseteq \mathcal{L}$ defines a hierarchy of sets approximating better and better the local set from outside. In Figure 1 we show a schematic representation of the first levels of approximations.

We have been focusing on the first non-trivial level that allows for nonlocality detection, namely the $\mathcal{L}_2$. We notice that, at this level of the hierarchy, specifying the entries $\Gamma_{ij} = \mathrm{tr}(\mathcal{O}_i^\dagger \mathcal{O}_j \rho)$ requires the knowledge of up-to-four-body correlators. Moreover, the amount of terms in the set $\mathcal{O}_2$ scales as the number of possible pairs of measurements $M_{x_i}^{a_i}$, that is, as $N^2 m^2 d^2$. This implies that the size of the moment matrix scales only quadratically with the number of parties and measurements, which is much more efficient compared to the exponential dependence d^{mN} of the linear program.

As mentioned before, checking whether a set of observed correlations belongs to $\mathcal{L}_2$ constitute a SDP feasibility problem. Since we are addressing approximations of the local set, there will be nonlocal correlations that will fall inside $\mathcal{L}_2$ and that will not be distinguishable from the local ones. Therefore, our technique can provide only a necessary conditions for nonlocality. Nonetheless, we were able to find several examples in which this method is able to successfully detect nonlocal correlations arising from various relevant states, proving that it has a practical implementation.

V. APPLICATIONS

The goal of this section is to show that the SDP relaxation can be successfully employed for detection of nonlocality in a broad range of cases. We focused particularly in exploring the efficient scaling of the method in terms of number of parties. To generate the SDP relaxations, we used the software Ncpol2sdpa [30], and we solved the SDPs with Mosek [31].

We collected evidence that, from a computational point of view, the main limiting factor of the technique is not time but the amount of memory required to store the moment matrix. Indeed, the longest time that was taken to run one of the codes amounted to approximately 9 hours [32]. Despite the memory limitation, the SDP technique allowed us to consider multipartite scenarios that cannot be dealt with the standard linear program approach to check locality. Indeed, for the scenarios with $m = 2, 3$, we were able to detect nonlocality for systems of up to $N = 29$ and $N = 15$ respectively, thus overcoming the current limits of [29].

In the following subsections we list the examples of states we considered. Given that we have studied cases with dichotomic measurements only, we present them in the expectation value form, namely by using the notation $M_{x_i}^{(i)} = M_{x_i}^1 - M_{x_i}^0$.

W state

As a first case, we analysed the Dicke state with a single excitation, also known as W state, namely

$$|W_N\rangle = \frac{1}{\sqrt{N}}(|0\ldots01\rangle + |0\ldots10\rangle \ldots + |10\ldots0\rangle). \quad (10)$$

Let us consider the simplest scenario of $m = 2$ dichotomic measurements per party, were each observer performs the same two measurements, that is, $M_0^{(i)} = \sigma_x$ and $M_1^{(i)} = \sigma_z$ for all $i = 1, \ldots, N$. We were able to show that the obtained probability distribution is detected as nonlocal at level $\mathcal{L}_2$ for $N \leq 29$. It is worth estimating the complexity of this test, both from a computational and experimental perspective. For this level, the rows and columns of the moment matrix correspond to the product of at most two observables. Therefore, the elements of the moment matrix involve at most four operators, which implies that the number of measurements to be estimated in the experiment scale as

$$\binom{N}{4} = O(N^4). \quad (11)$$

At the same time, the size of the matrix scales as $O(N^2)$. As announced, our relaxation has a scaling polynomial with the number of parties.

We have also studied the robustness of our technique to white noise,

$$\rho_N(p) = (1-p)|W_N\rangle\langle W_N| + p\frac{\mathbb{1}_N}{2^N} \quad (12)$$

where $0 \leq p \leq 1$ and $\mathbb{1}_N$ represents the identity operator acting on the space of N qubits. We estimated numerically the maximal value of p, referred to as p_{max}, for which the given correlations are still nonlocal according to the SDP criterion. Table I reports the resulting values as a function of the number of parties.

N	p_{max}	N	p_{max}	N	p_{max}
5	0.295	14	0.141	23	0.083
6	0.296	15	0.131	24	0.079
7	0.277	16	0.122	25	0.076
8	0.251	17	0.114	26	0.073
9	0.225	18	0.107	27	0.070
10	0.202	19	0.101	28	0.068
11	0.183	20	0.096	29	0.065
12	0.167	21	0.091		
13	0.153	22	0.087		

TABLE I. Robustness of nonlocality to white noise in the case of the W state, reported as a function of N.

While the robustness to noise decreases with the number of parties, the methods tolerates realistic amounts of noise, always larger than 6%, for all the tested configurations.

Finally, in order to study the robustness of the proposed test with respect to the choice of measurements we also considered a situation where the parties are not able to fully align their measurements and choose randomly two orthogonal measurements (ROM)[33]. More precisely, we assumed that $M_0^{(i)} = \vec{x}_0^{(i)} \cdot \vec{\sigma}$ and $M_1^{(i)} = \vec{x}_1^{(i)} \cdot \vec{\sigma}$, where $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ and $\vec{x}_0^{(i)}, \vec{x}_1^{(i)}$ are vectors chosen uniformly at random, with the only constraint of being orthogonal, namely $\vec{x}_0^{(i)} \cdot \vec{x}_1^{(i)} = 0$ for all $i = 1, \ldots, N$. We calculated numerically the probability p_{NL} for the corresponding correlations to be detected as nonlocal at the second level of the relaxation. To estimate p_{NL}, we computed the fraction N_{NL}/N_r of N_{NL} nonlocal distributions obtained over a total of $N_r = 1000$ rounds. The corresponding results are reported in the following table as a function of N.

N	p_{NL}	N	p_{NL}
3	50.2 %	7	21.0 %
4	44.4 %	8	12.8 %
5	38.4 %	9	6.3 %
6	28.8 %	10	2.7 %

The results for random measurements also exemplify one of the advantages of our approach with respect to previous entanglement detection schemes. Given some observed correlations, our test can be run and sometimes detects whether the correlations are nonlocal and therefore come from an entangled state. To our understanding, reaching similar conclusions using entanglement witnesses or other entanglement criteria is much harder, as they require solving optimisation problems involving N-qubit mixed states.

GHZ state

Another well studied multipartite state is the Greenberger-Horne-Zeilinger (GHZ) state, given by

$$|GHZ_N\rangle = \frac{1}{\sqrt{2}} \left(|0\rangle^{\otimes N} + |1\rangle^{\otimes N}\right). \tag{13}$$

Contrarily to the W, such state is not suited for detection of nonlocality with few-body correlations because all the k-body distributions arising from measurements on (13) are the same as those obtained by measuring the separable mixed state $\frac{1}{2}(|0\rangle\langle 0|^{\otimes k} + |1\rangle\langle 1|^{\otimes k})$. Therefore, in order to apply our nonlocality detection method to the GHZ we need to involve at least one full-body term.

The solutions we are going to present are inspired by the self-testing scheme for graph states introduced in [34]: the first scenario involves $m = 3$ dichotomic measurements per party, namely $M_0^{(i)} = \sigma_x$, $M_1^{(i)} = \sigma_d = \frac{1}{\sqrt{2}}(\sigma_x + \sigma_z)$ and $M_2^{(i)} = \sigma_z$ for all $i = 1, \ldots, N$. To introduce full-body correlators in the SDP we define the set $\mathcal{O}_{mix} = \{\mathcal{O}_2, \langle M_0^{(1)} M_0^{(2)} \ldots M_0^{(N)}\rangle, \langle M_1^{(1)} M_0^{(2)} \ldots M_0^{(N)}\rangle\}$. The moment matrix corresponding to such set will represent a mixed level of the relaxation, containing also two full-body correlators in the entries. However, since the number of added columns and rows is fixed to 2 for any N, this level is basically equivalent to level $\mathcal{L}_2$. Therefore, we preserve the efficient $O(N^4)$ scaling with the number of parties of elements in the moment of matrix and measurements to implement.

By numerically solving the SDP associated to this mixed level of the hierarchy we were able to confirm nonlocality of the correlations arising from the GHZ state and the given measurement for up to $N \leq 15$ parties. Moreover, we checked that the number of full-body values that is necessary to assign is constant for any of the considered N, coinciding with the two correlators $\langle M_0^{(1)} M_0^{(2)} \ldots M_0^{(N)}\rangle$ and $\langle M_1^{(1)} M_0^{(2)} \ldots M_0^{(N)}\rangle$. Lastly, we estimated that the robustness to noise in this case does not depend on N and it amounts to $p_{max} \approx 0.17$.

As a second scenario, we also noticed that one can produce nonlocal correlations from the GHZ at the level $\mathcal{O}_{mix}$ by considering $m = 2$ measurement choices only. Indeed, if ones considers $M_0^{(i)} = \sigma_x$, $M_1^{(i)} = \sigma_d$, the resulting correlations are detected as nonlocal for any $N \leq 28$ (The fact that we are not able to reach $N = 29$ is due to the mixed level of the relaxations, that results in a bigger matrix compared the scenario for W). Table (II) shows the corresponding robustness to noise, computed in the same way as for the W state. For both configurations, the noise robustness of our scheme in detecting GHZ states seems to saturate for large N even if the computational (and experimental) effort scales polynomially.

Graph states

Graph states [35] constitute another important family of multipartite entangled states. Such states are defined as follows: consider a graph G, i.e. a set of N vertices labeled by i

N	p_{max}	N	p_{max}	N	p_{max}
5	0.107	14	0.135	23	0.145
6	0.112	15	0.137	24	0.146
7	0.116	16	0.138	25	0.147
8	0.120	17	0.140	26	0.147
9	0.123	18	0.141	27	0.148
10	0.127	19	0.142	28	0.148
11	0.129	20	0.143		
12	0.132	21	0.144		
13	0.134	22	0.145		

TABLE II. Robustness of nonlocality to white noise in the case of the GHZ state and 2 dicothomic measurements per party, reported as a function of N.

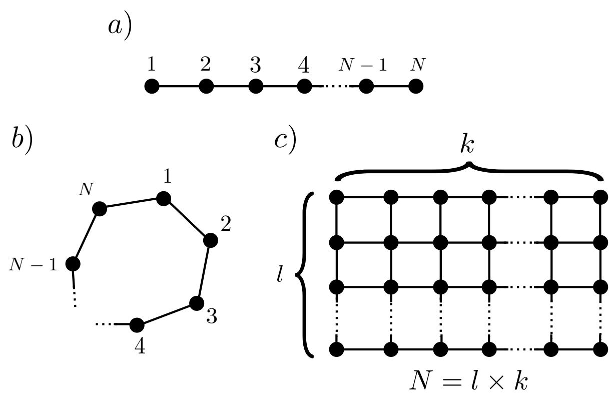

FIG. 2. Representatives of the graphs associated to the classes of states that have been studied with the SDP method: a) Linear graph states. b) Loop graph states. c) 2D cluster states.

connected by some edges $\mathcal{E}_{ij}$ connecting the vertices i and j. We associate a qubit system in the state $|+\rangle_i$ for each edge i and apply a control-Z grate $CZ_{ij} = \text{diag}(1, 1, 1, -1)$ to every pair of qubits i and j that are linked according to the graph G.

We considered some exemplary graph states such as the 1D and 2D cluster states and the loop graph state illustrated in Figure 2. Inspired by the self-testing scheme in [34], we considered that each party applies three measurements given by σ_x, σ_z and σ_d. We were able to detect nonlocality in the obtained correlations at level $\mathcal{L}_2$ for states involving up to $N = 15$ qubits. Again, the method at this level scales as N^4.

Interestingly, our approach for the detection of nonlocality in graph states shows to be qualitatively different from McKauge's scheme in [34]. While the latter requires correlators of an order than depends on the connectivity of the graph (namely, equal to 1 plus the maximal number of neighbours that each vertex has), our method seems to be independent of it. Indeed, we were able to detect nonlocality with four-body correlators in 2D cluster states, whose connectivity would imply five-body correlators for the self-testing scheme.

VI. EXPLICIT BELL INEQUALITIES

Another nice property of our nonlocality criterion comes from the fact that, as it can be put in an SDP form, it immediately provides a method to find experimentally friendly Bell inequalities involving a subset of all possible measurements. In fact, it turns out that the the SDP proposed in Sec. II has a dual formulation that can be interpreted as the optimisation of a linear function of the correlations that can be seen as a Bell-like functional, *i.e.* a functional that has a nontrivial bound for all correlations in $\mathcal{L}_k$ [19] (see Appendix A for details). Thus, if a set of correlations are found to be nonlocal, then the solution of the SDP provides a Bell inequality that is satisfied by correlations in $\mathcal{L}_\nu$ and that is violated by the tested correlations. Importantly, this Bell inequality can further be used to test other sets of correlations.

By using the two sets of correlations obtained by measuring 3 dicothomic observables per party in the GHZ state we were able to find the following Bell inequality:

$$\begin{aligned}\mathcal{I}^3_{mix} = \sum_{i=2}^{N}\langle M_1^{(1)} M_2^{(i)}\rangle - \sum_{i=2}^{N}\langle M_0^{(1)} M_2^{(i)}\rangle + \\ +(N-1)\langle M_0^{(1)} M_0^{(2)} \ldots M_0^{(N)}\rangle + \\ +(N-1)\langle M_1^{(1)} M_0^{(2)} \ldots M_0^{(N)}\rangle \leq 2(N-1)\end{aligned} \tag{14}$$

Numerically, we could only certify the validity of this inequality for up to $N \leq 15$. Moreover, in principle the bound of $\beta_C = 2(N-1)$ is only guaranteed to be satisfied by correlations in $\mathcal{L}_{mix}$. However, motivated by the obtained numerical insight, we could prove that this bound actually coincides with the true local bound and therefore (14) is a valid Bell inequality for all N (for all the analytical proofs regarding this subsection we refer to Appendix B). This shows that, at least in this instance, the $\mathcal{L}_{mix}$ defined by the SDP relaxation associated to $\mathcal{O}_{mix}$ is tight to the local set.

It is also easy to show that (14) is violated by the GHZ state and the previously introduced choice of measurements. In particular, the value reached is $\mathcal{I}^3_{GHZ} = (1+\sqrt{2})(N-1)$ for any N. Given that both the local bound and the violation scale linearly with N, the robustness of nonlocality to white noise is constant and amounts to $p_{max} = \frac{\sqrt{2}-1}{\sqrt{2}+1} \approx 0.174$. We notice that this results is in agreement with what achieved numerically with the SDP for up to $N = 15$.

Similarly, we also found the following Bell inequality by using the set of correlations involving only two measurement per party described for the GHZ state:

$$\begin{aligned}\mathcal{I}^2_{mix} = \sum_{i=2}^{N}\langle M_1^{(1)} M_1^{(i)}\rangle - \sum_{i=2}^{N}\langle M_0^{(1)} M_1^{(i)}\rangle + \\ +(N-1)\langle M_0^{(1)} M_0^{(2)} \ldots M_0^{(N)}\rangle + \\ +(N-1)\langle M_1^{(1)} M_0^{(2)} \ldots M_0^{(N)}\rangle \leq 2(N-1)\end{aligned} \tag{15}$$

Once more, although this inequality was found numerically for up to $N \leq 28$ we could prove that it is valid for any N.

Moreover the bound $\beta_C = 2(N-1)$ is not only valid for correlation in $\mathcal{L}_{mix}$ but for any local set of correlations. The GHZ state and the given measurements result in a violation of $\mathcal{I}^2_{GHZ} = \frac{3+\sqrt{2}}{2}$. Given that in this case the relative violation is lower, we also have a lower robustness to noise, coinciding with $p_{max} = \frac{\sqrt{2}-1}{\sqrt{2}+3} \approx 0.09$ for any N. We notice that this value is different than the ones reported in Table (II). The reason is that, to derive inequality (15) from the dual, we restricted to assign only the values of the two-body correlations and the two full-body ones. On the other hand, the results in Table (II) take also into account the assignment of the three- and four-body correlators, showing that this additional knowledge helps in improving the robustness to noise.

As a final remark, we stress that the measurement settings considered to derive an inequality from the dual might not be the optimal ones. For instance, we were able to identify different measurement choices for the case of (15) which lead to a higher violation of such inequality, hence resulting also in a better robustness to noise (see Appendix B for details).

VII. CONCLUSIONS

We introduced a technique for efficient device-independent entanglement detection for multipartite quantum systems. It relies on a hierarchy of necessary conditions for nonlocality in the observed correlations. By focusing on the second level of the hierarchy, we considered a test that requires the knowledge of up to four-body correlators only. We showed that it can be successfully applied to detect entanglement of many physically relevant states, such as the W, the GHZ and the graph states. Besides being suitable for experimental implementation, our technique has also an efficient scaling in terms of computational requirements, given that the number of variables involved grows polynomially with N. This allowed us to overcome the limitation of the currently known methods and to detect entanglement for states of up to few tens of particles. Moreover, the proposed technique has a completely general approach and it can be applied to any set of observed correlations. This makes it particularly relevant for the detection of new classes of multipartite entangled states.

On a more fundamental side, it would be interesting to study how accurate is the approximation of the local set of correlations provided by the second level of the hierarchy. In some of the scenario that we considered the approximation was actually tight to the local set, but this is not generally the case. A possible approach could be to compare the local bound of some known Bell inequalities with that resulting from the hierarchy.

Lastly, we notice that the second level of the hierarchy has also an efficient scaling with the number of measurements performed by the parties. This would allow to inquire whether an increasing number of measurement choices can provide an advantage for entanglement detection in multipartite systems.

VIII. ACKNOWLEDGEMENTS

This work was supported by the ERC CoG grant QITBOX, Spanish MINECO (Severo Ochoa grant SEV-2015-0522, a Ramón y Cajal fellowship and FOQUS FIS2013-46768-P), the AXA Chair in Quantum Information Science, the Generalitat de Catalunya (SGR875) and Fundacion Cellex. We also acknowledge the computational resources granted by the High Performance Computing Center North (SNIC 2016/1-320).

[1] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, Physical Review Letters **70**, 1895 (1993).
[2] A. K. Ekert, Physical Review Letters **67**, 661 (1991).
[3] H. J. Briegel, D. E. Browne, W. Dür, R. Raussendorf, and M. Van den Nest, Nature Physics **5**, 19 (2009).
[4] V. Giovannetti, S. Lloyd, and L. Maccone, Science **306**, 1330 (2004).
[5] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Review of Modern Physics **81**, 865 (2009).
[6] H. Häffner, W. Hänsel, C. Roos, J. Benhelm, M. Chwalla, T. Körber, U. Rapol, M. Riebe, P. Schmidt, C. Becher, *et al.*, Nature **438**, 643 (2005).
[7] F. G. S. L. Brandão and M. Christandl, Physical Review Letters **109**, 160502 (2012).
[8] M. Bourennane, M. Eibl, C. Kurtsiefer, S. Gaertner, H. Weinfurter, O. Gühne, P. Hyllus, D. Bruß, M. Lewenstein, and A. Sanpera, Physical Review Letters **92**, 087902 (2004).
[9] B. Jungnitsch, T. Moroder, and O. Gühne, Physical Review Letters **106**, 190502 (2011).
[10] G. Tóth and O. Gühne, Applied Physics B: Lasers and Optics **82**, 237 (2006).
[11] B. Lücke, J. Peise, G. Vitagliano, J. Arlt, L. Santos, G. Tóth, and C. Klempt, Physical Review Letters **112**, 155304 (2014).
[12] G. Vitagliano, I. Apellaniz, I. L. Egusquiza, and G. Tóth, Physical Review A **89**, 032307 (2014).
[13] L. Knips, C. Schwemmer, N. Klein, M. Wieśniak, and H. Weinfurter, arXiv:1412.5881 (2014).
[14] J. S. Bell, Physics **1**, 195 (1964).
[15] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Reviews of Modern Physics **86**, 419 (2014).
[16] M. Zukowski, D. Kaszlikowski, A. Baturo, and J.-A. Larsson, arXiv quant-ph/9910058 (1999).
[17] A. Fine, Physical Review Letters **48**, 291 (1982).
[18] M. Navascués, S. Pironio, and A. Acín, Physical Review Letters **98**, 010401 (2007).
[19] M. Navascués, S. Pironio, and A. Acín, New Journal of Physics **10**, 073013 (2008).
[20] I. Kogias, P. Skrzypczyk, D. Cavalcanti, A. Acin, and G. Adesso, Physical Review Letters **115**, 210401 (2015).
[21] D. Cavalcanti and P. Skrzypczyk, arXiv:quant-ph/1604.00501 (2016).
[22] S. Pironio, M. Navascués, and A. Acín, SIAM J. Optim. **20**, 2157 (2010).
[23] A. C. Doherty, Y.-C. Liang, B. Toner, and S. Wehner, Proc. IEEE 23rd Annual Conf. Comp. Compl. , 199 (2008).
[24] J. B. Lasserre, SIAM Journal on Optimization **11**, 796 (2001).
[25] G. V. Steeg and A. Galstyan, arXiv:1106.1636 (2011).

[26] R. Cleve, P. Høyer, B. Toner, and J. Watrous, in *Computational Complexity, 2004. Proceedings. 19th IEEE Annual Conference on* (IEEE, 2004) pp. 236–249.
[27] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Physical Review Letters **23**, 880 (1969).
[28] S. Boyd and L. Vandenberghe, *Convex optimization* (Cambridge university press, 2004).
[29] J. Gondzio, J. A. Gruca, J. J. Hall, W. Laskowski, and M. Żukowski, Journal of Computational and Applied Mathematics **263**, 392 (2014).
[30] P. Wittek, ACM Trans. Math. Softw. **41**, 21:1 (2015).
[31] Available at `http://www.mosek.com/`.
[32] The code used is available under an open source license at `https://github.com/FlavioBaccari/Hierarchy-for-nonlocality-detection`.
[33] J. J. Wallman, Y.-C. Liang, and S. D. Bartlett, Physical Review A **83**, 022110 (2011).
[34] M. McKague, in *Conference on Quantum Computation, Communication, and Cryptography* (Springer, 2011) pp. 104–120.
[35] M. Hein, W. Dür, J. Eisert, R. Raussendorf, M. Van den Nest, and H. . Briegel, arXiv:quant-ph/0602096 (2006).

Appendix A: Details of the method

Here we present in more detail the SDP relaxation associated to quantum realization with commuting measurements. In order to be consistent with the examples we presented in the main text, we will introduce it in the correlators approach, but we stress that a formulation in terms of projector and probabilities for higher numbers of outcomes is also possible.

Let us consider that the N observers A_i are allowed to perform m dicothomic measurements each. We can therefore define the operators $M_{x_i}^{(i)} = M_{x_i}^1 - M_{x_i}^0$ in terms of measurements $M_{x_i}^{a_i}$. It can be easily seen that expectation values of the $M_{x_i}^{(i)}$ corresponds to the correlators (5).

For any quantum realization of such operators, it is possible to show that they satisfy the following properties:

i) $(M_{x_i}^{(i)})^\dagger = M_{x_i}^{(i)}$ for any $i = 1, \ldots N$ and $x_i = 1, \ldots, m$,

ii) $(M_{x_i}^{(i)})^2 = \mathbb{1}$ for any $i = 1, \ldots N$ and $x_i = 1, \ldots, m$,

iii) $[M_{x_i}^{(i)}, M_{x_j}^{(j)}] = 0$ for any $i \neq j$ and $x_i, x_j = 1, \ldots, m$.

Now, let us consider that the sets $\mathcal{O}_\nu$ introduced in section II consist exactly of all the products of the $\{M_{x_i}^{(i)}\}$ up to order ν. Then, by indexing the operators in the sets as $\mathcal{O}_i$ for $i = 1, \ldots, k$, we define the $k \times k$ moment matrix as follows

$$\Gamma_{ij} = tr(\rho_N \mathcal{O}_i^\dagger \mathcal{O}_j)$$

where ρ_N is a generic N-partite quantum state. As it was shown in [18, 19], for any set of quantum correlations P, the properties i)-iii) and the fact that the associated ρ_N is a proper quantum state reflect into the following properties of the moment matrix:

- $\Gamma^\dagger = \Gamma$,
- $\Gamma \succeq 0$,
- the entries of the matrix are constrained by some linear equations of the form

$$\sum_{i,j} (F_m)_{ij} \Gamma_{ij} = g_m(P) \quad m = 1, \ldots, l$$

where $(F_m)_{ij}$ are some coefficients and the $g_m(P)$ can depend on the values of the correlators composing the P vector, as such

$$g_m(P) = (g_m)_0 + \sum_{k=1}^{N} \sum_{\substack{i_1 < \ldots < i_k \\ i_1, \ldots, i_j}} (g_m)_{j_1, \ldots j_k}^{i_1, \ldots, i_k} \langle M_{j_1}^{(i_1)} \ldots M_{j_k}^{(i_k)} \rangle$$

Up to this point, the method we described coincides with the NPA hierarchy [18, 19], which is used to check whether a set of observed correlations is compatible with a quantum realization. In order to define a hierarchy to test for local hidden variables realization, we introduce the additional condition that all the measurements for the same party have also to be commuting, namely

iv) $[M_{x_i}^{(i)}, M_{y_i}^{(i)}] = 0$ for any $i = 1, \ldots, N$ and $x_i \neq y_i = 1, \ldots, m$.

It can be seen that property iv) implies a second set of linear constraints on the Γ matrix, that we identify as

$$\sum_{i,j} (F'_m)_{ij} \Gamma_{ij} = g'_m(P) \quad m = 1, \ldots, l'$$

To make it clearer, we show an example of linear constraint that can come only if we impose condition iv). Let us consider the following four operators: $\mathcal{O}_k = M_{x_i}^{(i)} M_{y_i}^{(i)}$, $\mathcal{O}_l = M_{x_i}^{(i)} M_{x_j}^{(j)}$, $\mathcal{O}_n = M_{y_i}^{(i)}$ and $\mathcal{O}_m = M_{x_j}^{(j)}$. It is easy to see that, by exploiting i)-iii) plus iv), $\Gamma_{kl} = \Gamma_{nm}$ for any choice of $x_i, y_i, x_j = 1, \ldots, m$ and $i, j = 1, \ldots, N$.

Now, for any chosen $\mathcal{O}_\nu$, we can test whether an observed distribution P is compatible with a local model via the following SDP

$$\begin{aligned} \text{maximize} \quad & \lambda, \\ \text{subject to} \quad & \Gamma - \lambda \mathbb{1} \succeq 0, \\ & \sum_{i,j} (F_m)_{ij} \Gamma_{ij} = g_m(P) \quad m = 1, \ldots, l \;, \\ & \sum_{i,j} (F'_m)_{ij} \Gamma_{ij} = g'_m(P) \quad m = 1, \ldots, l' \;, \end{aligned} \tag{A1}$$

which is the *primal form* of the problem. A solution $\lambda_{min}^* < 0$ implies that it is not possible to find a semidefinite positive moment matrix satisfying the given linear constraint. Therefore P has no quantum realization with commuting measurements and we conclude it is nonlocal. We notice that by increasing the value of ν we get a sequence of more and more stringent tests for nonlocality. Indeed, the linear constraints for the level ν are always a subset of the ones coming from $\nu+1$. Moreover, in analogy with the NPA hierarchy, the series of tests is convergent, hence any nonlocal correlation will give a negative solution λ^*_{min} at a finite step of the sequence.

Interestingly, we can also study the *dual form* of the SDP problem, which reads as follows

$$\begin{aligned} \text{minimize} \quad & G(P) = \sum_k y_k g_k(P) + \sum_k {y'}_k {g'}_k(P), \\ \text{subject to} \quad & \sum_k y_k F_k^T + \sum_k {y'}_k {F'}_k^T \succeq 0, \\ & \sum_k y_k \text{tr}(F_k^T) + \sum_k {y'}_k \text{tr}({F'}_k^T) = 1. \end{aligned} \tag{A2}$$

Thanks to the strong duality of the problem, a negative solution for the primal implies also $G(P) = \lambda^*_{min} < 0$. Since any point in $\mathcal{L}_\nu$ satisfies the SDP condition at level ν with $G(P) \geq 0$, we can interpret $G(P)$ as a Bell-like inequality separating the $\mathcal{L}_\nu$ from the rest of the correlations. Indeed, since the $g_k(P)$ and ${g'}_k(P)$ are linear in the terms of the probability distribution, we derive that $G(P) \geq 0$ defines also a linear inequality for P. Violation of such inequality directly implies nonlocality.

Appendix B: Proof of local bound and quantum violation for the inequalities

We start by proving the local bounds for the inequalities introduced in the main text. To do so, we remind that to derive the maximal valued attained by local correlations it is enough to maximize over the vertices of the local set. In the correlator space, the deterministic local strategies (DLS) take the form

$$\langle M_{j_1}^{(i_1)} \dots M_{j_k}^{(i_k)} \rangle = \langle M_{j_1}^{(i_1)} \rangle \dots \langle M_{j_k}^{(i_k)} \rangle \tag{B1}$$

where each $M_{x_i}^{(i)}$ term can take only 1 and -1 values. By using this property, inequality (14) becomes

$$\begin{aligned} \mathcal{I}^3_{mix}(DLS) = (N-1) & \left[\langle M_1^{(1)} \rangle + \langle M_0^{(1)} \rangle \right] \mathcal{T}_0 \\ & + \left[\langle M_1^{(1)} \rangle - \langle M_0^{(1)} \rangle \right] \mathcal{T}_2 \end{aligned} \tag{B2}$$

where $\mathcal{T}_0 = \langle M_0^{(2)} \rangle \dots \langle M_0^{(N)} \rangle$ and $\mathcal{T}_2 = \sum_{i=2}^N \langle M_2^{(i)} \rangle$. For any number of parties N, we have that $\mathcal{T}_0 \leq 1$ and $\mathcal{T}_2 \leq N-1$, therefore

$$\mathcal{I}^3_{mix}(DLS) \leq 2(N-1)\langle M_1^{(1)} \rangle \leq 2(N-1) \tag{B3}$$

Similarly, for any deterministic strategy, inequality (15) takes the from

$$\begin{aligned} \mathcal{I}^2_{mix}(DLS) = (N-1) & \left[\langle M_1^{(1)} \rangle + \langle M_0^{(1)} \rangle \right] \mathcal{T}_0 \\ & + \left[\langle M_1^{(1)} \rangle - \langle M_0^{(1)} \rangle \right] \mathcal{T}_1 \end{aligned} \tag{B4}$$

where $\mathcal{T}_1 = \sum_{i=2}^N \langle M_1^{(i)} \rangle$. As for before, we can use the argument that $\mathcal{T}_0 \leq 1$ and $\mathcal{T}_1 \leq N-1$ to conclude

$$\mathcal{I}^2_{mix}(DLS) \leq 2(N-1)\langle M_1^{(1)} \rangle \leq 2(N-1) \tag{B5}$$

Regarding the quantum violation, we recall that the scenario considered was $|\psi\rangle = |GHZ_N\rangle = \frac{1}{\sqrt{2}}(|0\rangle^{\otimes N} + |1\rangle^{\otimes N})$ with measurements choices $M_0^{(i)} = \sigma_x$, $M_1^{(i)} = \sigma_d = \frac{1}{\sqrt{2}}(\sigma_x + \sigma_z)$ and $M_2^{(i)} = \sigma_z$ for all $i = 1, \dots, N$. It is easy to check that, for a GHZ state of any number of parties, the following is true

- $\langle \sigma_x^{(i)} \sigma_x^{(j)} \rangle = \langle \sigma_x^{(i)} \sigma_z^{(j)} \rangle = 0$ for any $i \neq j = 1, \dots, N$.
- $\langle \sigma_z^{(i)} \sigma_z^{(j)} \rangle = 1$ and therefore $\langle \sigma_d^{(i)} \sigma_z^{(j)} \rangle = \frac{1}{\sqrt{2}}$ and $\langle \sigma_d^{(i)} \sigma_z^{(j)} \rangle = \frac{1}{2}$for any $i \neq j = 1, \dots, N$.
- $\langle \sigma_x^{(1)} \sigma_x^{(2)} \dots \sigma_x^{(N)} \rangle = 1$ and $\langle \sigma_z^{(1)} \sigma_x^{(2)} \dots \sigma_x^{(N)} \rangle = 0$, therefore $\langle \sigma_d^{(1)} \sigma_x^{(2)} \dots \sigma_x^{(N)} \rangle = \frac{1}{\sqrt{2}}$ for any N.

By using the properties listed above, one can check that

$$\langle \mathcal{I}^3_{mix} \rangle_{GHZ_N} = (1+\sqrt{2})(N-1) \approx 2.41(N-1) \tag{B6}$$

and, similarly, that

$$\langle \mathcal{I}^2_{mix} \rangle_{GHZ_N} = \frac{3+\sqrt{2}}{2}(N-1) \approx 2.21(N-1) \tag{B7}$$

Moreover, we notice that by changing the measurement setting, one can achieve a higher violation of $\mathcal{I}^2_{mix}$. Indeed, it is easy to see that by choosing $M_0^{(1)} = \frac{1}{\sqrt{2}}(\sigma_x + \sigma_z)$, $M_1^{(1)} = \frac{1}{\sqrt{2}}(\sigma_x - \sigma_z)$ and $M_0^{(i)} = \sigma_x$,$M_1^{(i)} = \sigma_z$ for $i = 2, \dots, N$, the resulting violation is

$$\langle \mathcal{I}^2_{mix} \rangle_{GHZ_N} = 2\sqrt{2}(N-1) \approx 2.83(N-1) \tag{B8}$$

To conclude, we proceed with the analysis of the robustness to noise. We recall that this implies considering the noisy version of the GHZ state, namely

$$\rho_N(p) = (1-p)\rho_{GHZ_N} + p\frac{\mathbb{1}_N}{2^N} \tag{B9}$$

where $0 \leq p \leq 1$ represent the amount of white noise added to the state. It can be easily seen that the noise affects the values of the correlators for the GHZ in the following way

$$\langle\sigma_{j_1}^{(i_1)}\ldots\sigma_{j_k}^{(i_k)}\rangle_{\rho_N} = (1-p)\langle\sigma_{j_1}^{(i_1)}\ldots\sigma_{j_k}^{(i_k)}\rangle_{GHZ_N} \tag{B10}$$

for any $j_l \in \{x, y, z\}$ and $1 \leq k \leq N$. Therefore we can consider the noise as a simple damping factor in the violation of the inequalities. By using this fact, we get that $\mathcal{I}^3_{mix}$ will be violated as long as $(1-p)(1+\sqrt{2})(N-1) > 2(N-1)$, hence

$$p_{max}(\mathcal{I}^3_{mix}) = \frac{\sqrt{2}-1}{\sqrt{2}+1} \approx 0.17 \tag{B11}$$

By the same argument, we analyse $\mathcal{I}^2_{mix}$ for the two measurements settings that have been introduced. For the first one, the inequality will be violated as long as $(1-p)\frac{3+\sqrt{2}}{2}(N-1) > 2(N-1)$ and therefore

$$p_{max}(\mathcal{I}^2_{mix}) = \frac{\sqrt{2}-1}{\sqrt{2}+3} \approx 0.09 \tag{B12}$$

while for the second one, the violation is preserved for $(1-p)2\sqrt{2}(N-1) > 2(N-1)$, hence

$$p'_{max}(\mathcal{I}^2_{mix}) = 1 - \frac{\sqrt{2}}{2} \approx 0.29 \tag{B13}$$

Clearly, we see that for the second setting a higher violation results also in a significantly higher robustness to noise.