

Entanglement Certification Using Noncontextuality Inequalities

Yujie Zhang^{1,2,*} , Jonah Spodek¹ , David Schmid² , Carter Reid¹ , Liam J. Morrison^{1,3} ,
Thomas Jennewein,^{1,4} Kevin J. Resch,¹ and Robert W. Spekkens² 

¹*Institute for Quantum Computing and Department of Physics and Astronomy, University of Waterloo,
200 University Avenue West, Waterloo, Ontario, N2L 3G1, Canada*

²*Perimeter Institute for Theoretical Physics, 31 Caroline Street North, Waterloo,
Ontario, N2L 2Y5, Canada*

³*Department of Physics and Atmospheric Science, Dalhousie University,
Halifax, Nova Scotia, B3H 3J5, Canada*

⁴*Department of Physics, Simon Fraser University, 8888 University Drive West,
Burnaby, British Columbia, V5A 1S6, Canada*

 (Received 24 July 2025; revised 19 March 2026; accepted 21 July 2026; published 1 September 2026)

By combining the assumptions of Bell locality with those of generalized noncontextuality, we define classes of noncontextuality inequalities for correlations arising in a bipartite Bell circuit. These classes are distinguished by which subsets of the full set of operational identities are taken as input to the principle of noncontextuality; certain natural subsets form a hierarchy that provides a way of understanding and classifying different forms of quantum correlations, including entanglement, steering, and nonlocality. Each level of this hierarchy gives rise to a corresponding class of noncontextuality inequalities whose violation witnesses one of these forms of bipartite quantum resourcefulness, thereby yielding different sufficient conditions for entanglement. The resulting entanglement certification paradigm requires no prior characterization of the measurements, delivers a verdict independent of tomographic gauge freedom, and can certify any entangled state without auxiliary entangled sources. To illustrate the power of this paradigm, we show that noncontextuality inequalities can certify entanglement for families of two-qubit isotropic states for which certification by Bell or steering inequalities is known to be impossible. We also show that, compared with the Bell test, this approach certifies a much larger fraction of entangled states, while the associated membership problem is computationally more tractable. On the experimental side, we describe techniques to ensure nontrivial operational identities in the presence of noisy and imperfect implementations. We also identify the sufficient condition under which these techniques are valid, namely, a particular notion of *tomographic completeness* of the implemented Bell circuit, which ensures that the operational identities are independent of tomographic gauge. Finally, we provide an experimental demonstration of the superior performance of this entanglement certification technique using polarization-entangled photons.

DOI: [10.1103/dxpr-wp6l](https://doi.org/10.1103/dxpr-wp6l)

Subject Areas: Quantum Physics, Quantum Information

I. INTRODUCTION

Distinguishing whether a given operational behavior is classically explainable or genuinely quantum is a central problem in the foundations of quantum mechanics. As is well known, certain quantum experiments exhibit correlations that cannot be explained within a classical model that respects the natural causal structure of the experiment. For example, there is no local (i.e., common-cause) explanation for quantum correlations in a Bell circuit [1,2]. In addition

to Bell nonlocality [3], other types of correlations that arise naturally in this scenario include entanglement [4] and steerable correlations [5].

These concepts play a central role in quantum information science and have been extensively studied [6–8], yet there is no consensus on how to best understand the relationship between them [9,10]. While all pure entangled states exhibit nonlocality [11,12], for mixed states the connection between entanglement and nonlocality is less clear [6,13]. As a consequence, Bell inequalities [14], which serve as device-independent tests of nonlocality, cannot detect all entangled states in a standard Bell circuit, and alternative approaches such as quantum state tomography, entanglement witnesses, and measurement device-independent tests have been developed to verify entanglement in practical settings [15–17].

*Contact author: yujie4physics@gmail.com

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

We suggest, alternatively, that classical nonexplainability should be understood as the failure of operational statistics to be realizable by a noncontextual ontological model (NCOM) [18], a property that we shall refer to as *NCOM-nonrealizability* [19]. A noncontextual ontological model imposes constraints on the representation of each component of a quantum circuit based on the specific operational identities that they satisfy (where an operational identity is simply any equality holding among the different possibilities for a component in the circuit [22]). This approach generalizes the notion of Kochen-Specker noncontextuality from projective measurements to arbitrary quantum processes and generalizes the notion of negativity of the Wigner function to arbitrary quasiprobabilistic representations [23,24]. NCOM-nonrealizability also coincides with the failure of the natural notion of classical explainability in generalized probabilistic theories [24–26]. Recent efforts have made some connections between NCOM-nonrealizability and various notions of resourcefulness of bipartite quantum states [27–29]. However, these studies have focused on comparing bipartite quantum resources to NCOM-nonrealizability in certain prepare-measure scenarios where the set of preparations is induced by the bipartite state through steering. As we show, examining NCOM-nonrealizability within a bipartite scenario reveals more direct connections to the different classes of bipartite quantum resources.

The first part of this work develops this theoretical connection between generalized noncontextuality and bipartite quantum resources. We introduce the notion of NCOM-realizability for operational statistics within a bipartite Bell circuit, where we denote the two parties as Alice and Bob, who can perform local measurements. We demonstrate that different sets of operational identities yield, via the principle of noncontextuality, different sufficient conditions for establishing NCOM-nonrealizability. More precisely, we explore four distinct subsets of the full set of operational identities: (i) the union of those that hold among Alice’s measurements and those that hold among Bob’s measurements, (ii) only those that hold among Bob’s measurements, (iii) only those that hold among Alice’s measurements, and (iv) the trivial set. We prove that the quantum states that define Bell circuits that are NCOM-nonrealizable relative to these classes correspond, respectively, to the entangled states, the Alice-to-Bob steerable states, the Bob-to-Alice steerable states, and the nonlocal states.

This unified framework provides tools to witness membership in these classes via distinct sets of noncontextuality (NC) inequalities. The inequalities derived from considering class (i), i.e., the union of the set of all operational identities associated to each party, are particularly noteworthy, as they are capable of certifying the entanglement of *any* entangled state, including those that are local and unsteerable. This significantly extends the reach of

entanglement certification beyond the limitations of standard Bell and steering inequalities. As a concrete illustration, we derive several such inequalities and experimentally demonstrate the violation of one of these, thereby certifying the entanglement of a class of two-qubit isotropic states that are both local and unsteerable and which, consequently, cannot be certified using conventional Bell or steering inequalities.

As with Bell- and steering-inequality-based approaches, constructing a noncontextuality inequality tailored to a given target state is generally nontrivial: It typically requires optimizing over the set of measurements, a task that quickly becomes technically and computationally demanding as the number of measurements grows. Despite this, we show that, even with a fixed set of local measurements, our framework can certify a significantly broader set of entangled states than standard Bell tests can. For two-qubit entangled states drawn according to the Hilbert-Schmidt measure, we show that a noncontextuality test for a fixed set of measurements can certify entanglement for more than 75% of the sampled entangled states, whereas the Clauser-Horne-Shimony-Holt (CHSH) test certifies only about 1% of these using the same sampling procedure [30].

The second part of this work addresses the practical question of how to turn the theoretical connection between NCOM-nonrealizability and entanglement into a concrete entanglement-certification protocol. The key challenge is to obtain the operational identities which, via the principle of noncontextuality, imply constraints on the ontological model and ultimately the noncontextuality inequalities. This is important, because one can be confident in inferring entanglement from the violation of a noncontextuality inequality only to the extent that one is confident in the operational identity from which the noncontextuality inequality was derived. Unlike a conventional device-dependent entanglement witness, our protocol does not assume a prior (“trusted”) characterization of the measurement devices. Instead, it relies solely on operational identities among the implemented measurement procedures. We describe two general techniques for inferring or establishing the operational identities. In the cut-to-your-cloth approach, one performs self-consistent tomography to characterize the implemented Bell circuit and to infer the linear identities that the realized procedures actually satisfy. In the secondary-procedures approach, one defines new procedures by classical randomization and postprocessing of the realized procedures, so that the desired operational identities hold by construction. Unlike a device-independent Bell test, our noncontextuality-inequality test depends on the realized preparations (the steered states) and measurements being tomographically complete. This is precisely the role played by tomographic completeness in self-consistent tomography [31,32]: It is the condition under which the observed operational

TABLE I. A summary of the advantages of various methods of entanglement certification.

Entanglement certification	Features				
	No prior characterization of measurements	Independent of tomographic gauge	Able to certify any entangled state	No additional entangled sources	No assumption of tomographic completeness ^a
Separability test with quantum state tomography	✗	✓	✓	✓	✗
Entanglement witness [15]	✗	✓	✓	✓	✓
Separability test with GPT or gate-set tomography [31,32]	✓	✗	✓	✓	✗
Bell inequality test in Bell circuit [3,35]	✓	✓	✗	✓	✓
Steering inequality test in Bell circuit [33,36]	✗	✓	✗	✓	✗
Inequality test in non-Bell network scenario [34]	✓	✓	✓	✗	✓
Noncontextuality inequality test (our approach)	✓	✓	✓	✓	✗

^aIn quantum state tomography, tomographic completeness of the set of characterized measurements guarantees a unique reconstructed state. In self-consistent tomography, such as GPT tomography and gate-set tomography, tomographic completeness instead ensures that the operational data determine the states and measurements only up to an invertible linear gauge (i.e., up to a tomographic gauge freedom [31]). Just as tomographic completeness is a sufficient condition for implementing self-consistent tomography, in our entanglement certification technique, it is a sufficient condition for ensuring that the identities respected by the operational statistics suffice to infer the operational identities relevant to the noncontextuality test (see Sec. VI A). In other words, it ensures that the operational identities are independent of the tomographic gauge.

statistics determine the realized procedures up to an invertible tomographic gauge. In our setting, this means that, although the reconstructed states and effects are not unique, the linear identities they satisfy are unique. Equivalently, all tomographic reconstructions compatible with the same data agree on the operational identities relevant to the noncontextuality test. In both the cut-to-your-cloth and secondary-procedures techniques discussed below, the condition of tomographic completeness ensures that the operational identities used in the test are gauge-independent facts about the implemented procedures rather than artifacts of a particular representation. Consequently, the quantities entering the noncontextuality-inequality test, namely, the operational statistics together with these identities, do not rely on a prior characterization of the measurement devices and are independent of the tomographic gauge.

This perspective also situates our method among existing entanglement-certification techniques, such as quantum state tomography, entanglement witnesses [15], self-consistent tomography [i.e., gate-set tomography [31] and generalized probabilistic theory (GPT) tomography [32]], Bell inequality tests [3], steering inequality tests [33], and other causal compatibility tests within more complex causal networks [34]. As summarized in Table I, our approach combines several desirable features that are not achieved simultaneously by these existing methods.

The remainder of the paper is organized as follows. We first separate the ideal certification problem from its

practical implementation and then spell out the conditions under which the certification method is valid. In Sec. II, we define NCOM-realizability for bipartite Bell circuits and introduce the hierarchy of operational identities, together with its relation to bipartite quantum resources. In Sec. III, we study the ideal certification problem: Assuming that the relevant operational identities are exactly specified, we derive the corresponding noncontextuality inequalities and show how they can certify entanglement, including in regimes inaccessible to Bell and steering inequalities. In Sec. IV, we turn to practical implementation and explain how the required operational identities can be inferred or enforced when the implemented measurements are noisy and imperfect. In Sec. V, we apply these techniques in a photonic experiment and demonstrate entanglement certification beyond steering- and Bell-inequality-based methods. In Sec. VI, we discuss the tomographic incompleteness loophole of the protocol and the lack of a loophole associated to detector inefficiency. In Sec. VII, we compare our approach with other entanglement-certification methods.

II. NONCONTEXTUAL ONTOLOGICAL MODEL

A. NCOM-realizability of the statistics of a prepare-measure circuit

The notion of NCOM-realizability provides a rigorous framework for assessing the classical explainability of the statistics obtained in a prepare-measure circuit [18]. A prepare-measure circuit is a circuit that consists of a process

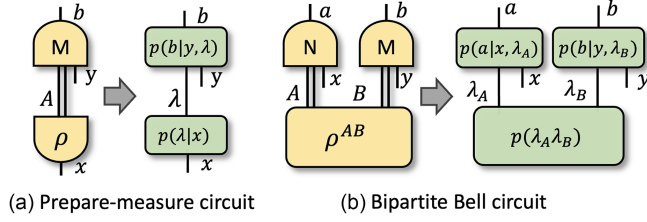


FIG. 1. (a) A unipartite prepare-measure (PM) circuit; (b) a bipartite prepare-measure circuit, i.e., a *bipartite Bell circuit* involving two parties labeled Alice (A) and Bob (B). A classical explanation for a given quantum circuit is given by a linear and diagram-preserving map from the quantum representation to an ontological representation—i.e., to a circuit wherein systems are classical variables and transformations are (sub)stochastic maps on these variables.

of the preparation variety (one having a single quantum output) followed by a process of the measurement variety (one having a single quantum input) [37]; see Fig. 1(a) for a simple example. On the preparation side, the process is one that has a classical input (a setting variable), denoted x . Such a process is represented by a set of quantum states $\mathbf{P} := \{\rho_x\}_x$ and is termed a *multi-state* in Ref. [37]. On the measurement side, the process is one that has both a classical output (the measurement outcome), denoted b , and a classical input (a setting variable), denoted y . Such a process is represented by a set of positive-operator-valued measurements (POVMs) $\mathbf{M} := \{\{M_{b|y}\}_b\}_y$ and is termed a *multi-measurement* in Ref. [37]. The statistics generated by such a prepare-measure circuit have the form of a conditional probability

$$p(b|xy) := \text{Tr}[M_{b|y}\rho_x]. \quad (1)$$

Such a quantum circuit is *classically explainable* if the statistics it predicts in Eq. (1) can be realized by a noncontextual ontological model [38]. This is the case if it is possible to write

$$p(b|xy) = \sum_{\lambda} p(b|y, \lambda) p(\lambda|x), \quad (2)$$

where $p(\lambda|x)$ denotes a probability distribution over ontic states $\lambda \in \Lambda$ given preparation x and $p(b|y, \lambda)$ is the response function specifying the probability of outcome b of measurement y given that the ontic state was λ . The principle of generalized noncontextuality implies that $p(\lambda|x)$ must be a linear function of the quantum state ρ_x and $p(b|y, \lambda)$ must be a linear function of the effect $M_{b|y}$. A given theory or experiment is then said to be NCOM-nonrealizable if and only if no such representation is possible. NCOM-realizability is the notion of classical explainability we use in this work.

Although the question of NCOM-realizability has usually been studied only in the prepare-measure circuit just

discussed, it is equally applicable to circuits that have more complex structures and more quantum processes [24,37]. In this work, we focus on the particular prepare-measure circuit depicted in Fig. 1(b); on the preparation side, it consists of a process having a *bipartite* quantum system AB as outputs and no classical input (i.e., no setting variable). Such a process is associated to a single bipartite state $\mathbf{P} = \rho^{AB}$. On the measurement side, it consists of a multi-measurement on A (Alice), denoted $\mathbf{N} := \{\{N_{a|x}\}_a\}_x$, and a multi-measurement on B (Bob), denoted $\mathbf{M} := \{\{M_{b|y}\}_b\}_y$. We refer to such a triple $\mathcal{B} := (\mathbf{N}, \mathbf{M}, \mathbf{P})$ as a *Bell circuit* (or a bipartite Bell circuit). A multipartite Bell circuit can be defined similarly.

B. NCOM-realizability of the statistics of a bipartite Bell circuit

Quantum theory predicts that, for the Bell circuit $\mathcal{B} := (\mathbf{N}, \mathbf{M}, \mathbf{P})$, the correlations obtained from performing a local multi-measurement on A with setting x and outcome a and a local multi-measurement on B with setting y and outcome b are given by

$$p(ab|xy) = \text{Tr}[(N_{a|x}^A \otimes M_{b|y}^B) \rho^{AB}]. \quad (3)$$

These are referred to as the *operational statistics*. Similar to the case of a prepare-measure circuit, we take these operational statistics to be *classically explainable* if they are realizable by a noncontextual ontological model [24]. We now review the two key structural features that characterize a noncontextual ontological model when specialized to the Bell circuit.

Diagram preservation (locality)—The ontological representation must preserve the circuit structure. The quantum system AB is represented by an ontic states space Λ_{AB} , which is the Cartesian product of the local ontic state spaces, i.e., $\Lambda_{AB} = \Lambda_A \times \Lambda_B$, the bipartite state is represented by a joint probability distribution $p(\lambda_A \lambda_B)$, and the measurements $\{N_{a|x}^A\}_a$ and $\{M_{b|y}^B\}_b$ are represented by sets of response functions $\{p(a|x, \lambda_A)\}_a$ and $\{p(b|y, \lambda_B)\}_b$, respectively. Accordingly, diagram preservation implies that the operational statistics must be reproduced as

$$p(ab|xy) = \sum_{\lambda_A \lambda_B} p(\lambda_A \lambda_B) p(a|x, \lambda_A) p(b|y, \lambda_B). \quad (4)$$

Note that this is exactly the standard form of a local hidden variable model for a Bell circuit [3].

Linearity (noncontextuality)—An operational identity satisfied by the multi-measurement $\mathbf{N} = \{\{N_{a|x}^A\}_a\}_x$ has the form $\sum_{a,x} \alpha_{a,x} N_{a|x}^A = \mathbb{0}$ and is, therefore, defined by the vector of coefficients $\vec{\alpha} := (\alpha_{a,x})_{a,x}$. Let $\mathcal{O}_{\text{all}}(\mathbf{N})$ represent the set of all vectors associated to the operational identities of $\mathbf{N}$. Let a similar definition hold for multi-measurement $\mathbf{M}$. Explicitly,

$$\mathcal{O}_{\text{all}}(\mathbf{N}) := \left\{ \vec{\alpha} \left| \sum_{a,x} \alpha_{a,x} N_{a|x}^A = \mathbb{0}^A \right. \right\}, \quad (5a)$$

$$\mathcal{O}_{\text{all}}(\mathbf{M}) := \left\{ \vec{\beta} \left| \sum_{b,y} \beta_{b,y} M_{b|y}^B = \mathbb{0}^B \right. \right\}. \quad (5b)$$

For an ontological model to be noncontextual, the set of response functions $\{\{p(a|x, \lambda_A)\}_a\}_x$ representing the multi-measurement $\mathbf{N} = \{\{N_{a|x}^A\}_a\}_x$ must be obtained from the latter by a linear map; similarly for the set of response functions $\{\{p(b|y, \lambda_B)\}_b\}_y$ representing the multi-measurement $\mathbf{M} = \{\{M_{b|y}^B\}_b\}_y$. This implies that these response functions must satisfy the same identities as are satisfied by the effects, that is,

$$\sum_{a,x} \alpha_{a,x} p(a|x, \lambda_A) = 0 \quad \forall \vec{\alpha} \in \mathcal{O}_{\text{all}}(\mathbf{N}), \quad (6a)$$

$$\sum_{b,y} \beta_{b,y} p(b|y, \lambda_B) = 0 \quad \forall \vec{\beta} \in \mathcal{O}_{\text{all}}(\mathbf{M}). \quad (6b)$$

Unlike for a unipartite prepare-measure circuit, there exist nontrivial compositions of processes within a bipartite Bell circuit, namely,

$$\mathbf{N} \circ \mathbf{P} := \{\{\text{Tr}_A[(N_{a|x}^A \otimes \mathbb{1}^B) \rho^{AB}]\}_a\}_x = \{\{\tilde{\rho}_{a|x}^B\}_a\}_x, \quad (7a)$$

$$\mathbf{M} \circ \mathbf{P} := \{\{\text{Tr}_B[(\mathbb{1}^A \otimes M_{b|y}^B) \rho^{AB}]\}_b\}_y = \{\{\tilde{\rho}_{b|y}^A\}_b\}_y. \quad (7b)$$

Here, $\{\{\tilde{\rho}_{a|x}^B\}_a\}_x$ and $\{\{\tilde{\rho}_{b|y}^A\}_b\}_y$ are steering assemblages [40,41] on parties B and A , respectively. (One can also consider them as *multi-sources* on B and A , respectively [37].)

An operational identity satisfied by a steering assemblage $\{\{\tilde{\rho}_{a|x}^B\}_a\}_x$ has the form $\sum_{a,x} \alpha_{a,x} \tilde{\rho}_{a|x}^B = \mathbb{0}$ and is, therefore, defined by the vector of coefficients $\vec{\alpha} := (\alpha_{a,x})_{a,x}$. Let $\mathcal{O}_{\text{all}}(\mathbf{N} \circ \mathbf{P})$ represent the set of all vectors associated to operational identities of $\mathbf{N} \circ \mathbf{P}$, and similarly for $\mathbf{M} \circ \mathbf{P}$, that is,

$$\mathcal{O}_{\text{all}}(\mathbf{N} \circ \mathbf{P}) := \left\{ \vec{\alpha} \left| \sum_{a,x} \alpha_{a,x} \tilde{\rho}_{a|x}^B = \mathbb{0}^B \right. \right\}, \quad (8a)$$

$$\mathcal{O}_{\text{all}}(\mathbf{M} \circ \mathbf{P}) := \left\{ \vec{\beta} \left| \sum_{b,y} \beta_{b,y} \tilde{\rho}_{b|y}^A = \mathbb{0}^A \right. \right\}. \quad (8b)$$

The steering assemblage on B , $\{\{\tilde{\rho}_{a|x}^B\}_a\}_x$, is represented in the ontological model by a set of subnormalized distributions on Λ_B , $\{\{\tilde{p}(a, \lambda_B|x)\}_a\}_x$. Similarly, the steering assemblage on A is represented by a set of subnormalized distributions on Λ_A , $\{\{\tilde{p}(b, \lambda_A|y)\}_b\}_y$. These are defined as

$$\tilde{p}(a, \lambda_B|x) := \sum_{\lambda_A} p(a|x, \lambda_A) p(\lambda_A \lambda_B), \quad (9a)$$

$$\tilde{p}(b, \lambda_A|y) := \sum_{\lambda_B} p(b|y, \lambda_B) p(\lambda_A \lambda_B). \quad (9b)$$

The principle of noncontextuality implies that each of these sets of distributions must be the image of a linear map on the respective steering assemblage, implying that these distributions must satisfy the same identities as are satisfied by the elements of the steering assemblage, i.e., those in Eq. (8):

$$\sum_{a,x} \alpha_{a,x} \tilde{p}(a, \lambda_B|x) = 0 \quad \forall \vec{\alpha} \in \mathcal{O}_{\text{all}}(\mathbf{N} \circ \mathbf{P}), \quad (10a)$$

$$\sum_{b,y} \beta_{b,y} \tilde{p}(b, \lambda_A|y) = 0 \quad \forall \vec{\beta} \in \mathcal{O}_{\text{all}}(\mathbf{M} \circ \mathbf{P}). \quad (10b)$$

Finally, consider the parallel compositions of the elements of the two steering assemblages $\{\{\tilde{\rho}_{a|x}^B\}_a\}_x$ and $\{\{\tilde{\rho}_{b|y}^A\}_b\}_y$, i.e., the product states $\tilde{\rho}_{b|y}^A \otimes \tilde{\rho}_{a|x}^B$. There exists one final set of operational identities that links these product states to the bipartite state ρ^{AB} , namely,

$$\mathcal{O}_{\text{all}}(\tilde{\mathbf{P}}) := \left\{ \vec{\gamma} \left| \rho^{AB} = \sum_{a,x,b,y} \gamma_{a,x,b,y} \tilde{\rho}_{b|y}^A \otimes \tilde{\rho}_{a|x}^B \right. \right\}. \quad (11)$$

Here, we use $\tilde{\mathbf{P}}$ as a shorthand for the tuple $(\mathbf{P}, \mathbf{N} \circ \mathbf{P}, \mathbf{M} \circ \mathbf{P})$ that is relevant to Eq. (11). The principle of noncontextuality then implies additional constraints on the distributions $p(\lambda_A \lambda_B)$, namely,

$$p(\lambda_A \lambda_B) = \sum_{a,x,b,y} \gamma_{a,x,b,y} \tilde{p}(a, \lambda_B|x) \tilde{p}(b, \lambda_A|y) \quad \forall \vec{\gamma} \in \mathcal{O}_{\text{all}}(\tilde{\mathbf{P}}). \quad (12)$$

For a specific Bell circuit $\mathcal{B} = (\mathbf{N}, \mathbf{M}, \mathbf{P})$, the operational statistics $p(ab|xy)$ are said to be NCOM-realizable if they can be realized by a model that is diagram preserving (and hence local), as in Eq. (4), and that respects the implications of noncontextuality for the full set of operational identities on each of $\mathbf{N}$, $\mathbf{M}$, $\mathbf{N} \circ \mathbf{P}$, $\mathbf{M} \circ \mathbf{P}$, and $\tilde{\mathbf{P}}$, i.e., that respects Eqs. (6), (10), and (12). We denote the set of all such statistics by $\mathcal{NC}(\mathcal{B})$.

Suppose one wishes to find conditions on the operational statistics that are merely *necessary* but not sufficient for NCOM-realizability. This is achieved when one applies the principle of noncontextuality to *subsets* of the full set of operational identities for one or more of $\mathbf{N}$, $\mathbf{M}$, $\mathbf{N} \circ \mathbf{P}$, $\mathbf{M} \circ \mathbf{P}$, and $\tilde{\mathbf{P}}$. In what follows, for each of them, we consider a single alternative to the complete set of operational identities thereon, namely, what we term the *trivial set*. The trivial set of operational identities for $\mathbf{N}$, denoted $\mathcal{O}_{\text{triv}}(\mathbf{N})$, is all and

only those that arise from the completeness condition $\sum_a N_{a|x}^A = \sum_a N_{a|x'}^A = 1$. These identities are deemed to be trivial, because the corresponding ontological identities are automatically satisfied simply by the definition of a set of response functions, i.e., $\sum_a p(a|x, \lambda_A) = \sum_a p(a|x', \lambda_A) = 1$. Similarly, the set $\mathcal{O}_{\text{triv}}(\mathbf{N} \circ \mathbf{P})$ represents those operational identities that arise from the no-signaling condition $\sum_a \tilde{p}_{a|x}^B = \sum_a \tilde{p}_{a|x'}^B$, which are trivial because the corresponding ontological identities $\sum_a \tilde{p}(a\lambda_B|x) = \sum_a \tilde{p}(a\lambda_B|x')$ are automatically satisfied due to the assumption of diagram preservation. For the case of $\tilde{\mathbf{P}}$, the trivial set of operational identities is defined to be the null set.

Hence, in a specific Bell circuit $\mathcal{B} = (\mathbf{N}, \mathbf{M}, \mathbf{P})$, for each of $\mathbf{N}$, $\mathbf{M}$, $\mathbf{N} \circ \mathbf{P}$, $\mathbf{M} \circ \mathbf{P}$, and $\tilde{\mathbf{P}}$, we can associate a binary variable that stipulates whether we use $\mathcal{O}_{\text{all}}$ or $\mathcal{O}_{\text{triv}}$ for it. That is, letting $i, j, u, v, w \in \{a, t\}$, where a stands for “all” and t stands for “trivial,” one can define

$$\begin{aligned} \mathcal{O}_{ijuvw}(\mathcal{B}) \\ := \mathcal{O}_i(\mathbf{N}) \cup \mathcal{O}_j(\mathbf{M}) \cup \mathcal{O}_u(\mathbf{N} \circ \mathbf{P}) \cup \mathcal{O}_v(\mathbf{M} \circ \mathbf{P}) \cup \mathcal{O}_w(\tilde{\mathbf{P}}). \end{aligned} \quad (13)$$

We thereby define 32 different subsets of operational identities: $\{\mathcal{O}_{ijuvw}(\mathcal{B})\}_{ijuvw \in \{a,t\}^5}$. Clearly, we have inclusion relations such as

$$\begin{aligned} \mathcal{O}_{aaaa}(\mathcal{B}) \supseteq \mathcal{O}_{aaat}(\mathcal{B}) \supseteq \mathcal{O}_{aatat}(\mathcal{B}), \mathcal{O}_{aatat}(\mathcal{B}) \\ \supseteq \mathcal{O}_{atatt}(\mathcal{B}) \supseteq \mathcal{O}_{atttt}(\mathcal{B}), \mathcal{O}_{atttt}(\mathcal{B}) \supseteq \mathcal{O}_{ttttt}(\mathcal{B}). \end{aligned} \quad (14)$$

In this paper, we focus on the four cases of $\mathcal{O}_{aatat}(\mathcal{B})$, $\mathcal{O}_{atatt}(\mathcal{B})$, $\mathcal{O}_{atttt}(\mathcal{B})$, and $\mathcal{O}_{ttttt}(\mathcal{B})$, because these define, respectively, the separable-entangled boundary, the unsteerable-steerable boundary (in the $B \rightarrow A$ and $A \rightarrow B$ directions, respectively), and the local-nonlocal boundary. Other subsets are also of interest. $\mathcal{O}_{aaatt}(\mathcal{B})$, for instance, characterizes those bipartite states that can prepare non-classical ensembles of states on B [42], which have been studied in Ref. [29].

Definition 1. Consider a Bell circuit $\mathcal{B} = (\mathbf{N}, \mathbf{M}, \mathbf{P})$ and a subset of its operational identities, denoted $\mathcal{O}(\mathcal{B})$. The operational statistics $\mathbf{p} := \{p(ab|xy)\}_{abxy}$ achieved in this circuit are said to be *NCOM-realizable relative to $\mathcal{O}(\mathcal{B})$* if they can be realized by a model that is diagram preserving (and, hence, local), as in Eq. (4), and that respects the implications of noncontextuality for the operational identities in $\mathcal{O}(\mathcal{B})$, i.e., that respects the subset of ontological identities in Eqs. (6), (10), and (12).

In Ref. [37], the notion of classical explainability of the operational statistics of an experiment—in other words, the notion of NCOM-realizability of these statistics—was leveraged to define a notion of classicality for an *individual*

process within an experiment. In the case of a bipartite state, it is defined to be classical if and only if the operational statistics generated by the set of circuits obtained by composing the state with every possible pair of local measurements are classically explainable.

More formally, let ρ denote the bipartite state in question. We denote the multi-measurements that consist of the *full* set of possible quantum measurements on A by $\mathbf{N}^{\text{full}}$ and the one that consists of the full set on B by $\mathbf{M}^{\text{full}}$ (i.e., these have setting variables that range over all possible measurements). The resulting Bell circuit, denoted $\mathcal{B}^{\text{full}} = (\mathbf{N}^{\text{full}}, \mathbf{M}^{\text{full}}, \mathbf{P})$, is termed a *measurement-full Bell circuit*. The proposal of Ref. [37], then, can be stated as follows.

Definition 2. Consider the measurement-full Bell circuit $\mathcal{B}^{\text{full}} = (\mathbf{N}^{\text{full}}, \mathbf{M}^{\text{full}}, \mathbf{P} = \rho^{AB})$. The bipartite state ρ^{AB} is said to be *classical* [43] if the operational statistics arising in this circuit are NCOM-realizable relative to $\mathcal{O}_{aaaa}(\mathcal{B}^{\text{full}})$ in the sense of Definition 1.

We refer the reader to Ref. [37] for a complete characterization of *classical* bipartite states.

The present work moves beyond the results of Ref. [37] by considering the notion of NCOM-realizability *relative to different subsets* of the set of all operational identities, $\mathcal{O}(\mathcal{B}^{\text{full}}) \subseteq \mathcal{O}_{aaaa}(\mathcal{B}^{\text{full}})$, to define necessary (but, in general, not sufficient) conditions for the classicality of a bipartite state in the Bell circuit $\mathcal{B}^{\text{full}}$.

Definition 3. Consider a measurement-full Bell circuit $\mathcal{B}^{\text{full}} = (\mathbf{N}^{\text{full}}, \mathbf{M}^{\text{full}}, \mathbf{P} = \rho^{AB})$ and a subset of operational identities, denoted $\mathcal{O}(\mathcal{B}^{\text{full}})$. The bipartite state ρ^{AB} is said to be *classical relative to $\mathcal{O}(\mathcal{B}^{\text{full}})$* if the operational statistics arising in this circuit are NCOM-realizable relative to $\mathcal{O}(\mathcal{B}^{\text{full}})$ in the sense of Definition 1.

It was shown in Ref. [37] that the set of bipartite states that are classical (in the sense of Definition 2) is a strict subset of the separable states. Here, we demonstrate that, within this set of bipartite states, considering classicality *relative to various choices of subsets of the full set of operational identities* (in the sense of Definition 3) provides a novel and unified operational definition of the separable-entangled boundary, the steerable-unsteerable boundary, and the local-nonlocal boundary.

Theorem 1. Let $\mathcal{B}^{\text{full}} = (\mathbf{N}^{\text{full}}, \mathbf{M}^{\text{full}}, \mathbf{P} = \rho^{AB})$ denote a measurement-full Bell circuit. For the bipartite state ρ^{AB} , there is an equivalence between the two properties in each of the following pairs:

- (i) classicality relative to $\mathcal{O}_{aatat}(\mathcal{B}^{\text{full}})$ and separability,
- (ii) classicality relative to $\mathcal{O}_{atatt}(\mathcal{B}^{\text{full}})$ and $B \rightarrow A$ unsteerability,
- (iii) classicality relative to $\mathcal{O}_{atttt}(\mathcal{B}^{\text{full}})$ and $A \rightarrow B$ unsteerability, and
- (iv) classicality relative to $\mathcal{O}_{ttttt}(\mathcal{B}^{\text{full}})$ and locality.

We leave the proof of this theorem to Appendix A. The last case follows directly from the fact that if none of the nontrivial operational identities are leveraged, then the only

constraint on the correlations is that of Eq. (4), which is simply realizability by a local hidden variable model [3].

It is worth noting that, in a related work [29], the prepare-measure circuit on system B with preparation that arises from $A \rightarrow B$ steering in a bipartite Bell circuit is considered, and it is shown that a certain witness of NCOM-nonrealizability is a necessary but not sufficient condition for entanglement. The key difference between this and the theorem above lies in the choice of operational identities considered. Specifically, Ref. [29] effectively considers NCOM-realizability relative to $\mathcal{O}_{\text{aatt}}$, whereas case (i) in Theorem 1 in our work focuses on NCOM-realizability relative to $\mathcal{O}_{\text{aattt}}$. By considering the consequences of noncontextuality for a smaller set of operational identities, we are able to exactly single out the separable-entangled divide among bipartite states. Because Ref. [29] considers more operational identities, the constraints from noncontextuality are more stringent and so single out a different divide, between a strict *subset* of the separable states and the complement of this set. This subset of bipartite states can be understood as those that are not capable of steering to a nonclassical set of states on B , for the notion of nonclassicality defined in Refs. [37,42]. More formally, we have the following.

Proposition 1 [29,42]. Consider the measurement-full Bell circuit $\mathcal{B}^{\text{full}} = (\mathbf{N}^{\text{full}}, \mathbf{M}^{\text{full}}, \mathbf{P} = \rho^{AB})$. For the bipartite state ρ^{AB} , there is an equivalence between the two properties in each of the following pairs:

- (v) classicality relative to $\mathcal{O}_{\text{aatt}}(\mathcal{B}^{\text{full}})$ and the impossibility of steering to a nonclassical set of states on B and
- (vi) classicality relative to $\mathcal{O}_{\text{aatat}}(\mathcal{B}^{\text{full}})$ and the impossibility of steering to a nonclassical set of states on A .

The proof of Proposition 1 above proceeds in the same way as the proof of Theorem 1 in Ref. [29]. For completeness, and to highlight its connection to our broader framework, we review the argument in Appendix A using our notation.

Experimentally, one cannot implement *all* local measurements that are deemed possible in quantum theory. Consequently, the multi-measurements $\mathbf{N}$ and $\mathbf{M}$ that arise in an experimental Bell circuit are necessarily distinct from $\mathbf{N}^{\text{full}}$ and $\mathbf{M}^{\text{full}}$. In the following section, we consider different choices of $\mathbf{N}$ and $\mathbf{M}$ and analyze what they imply for witnessing nonclassicality relative to a given set of operational identities and for certifying bipartite resourcefulness.

III. ENTANGLEMENT CERTIFICATION

In a Bell circuit $\mathcal{B} = (\mathbf{N}, \mathbf{M}, \mathbf{P})$, there are different ways of achieving entanglement certification, including those summarized in Table I. They differ in several respects, as we discuss in detail in Sec. VII. Among them, an important distinction concerns the assumptions one has to make about the processes appearing in the Bell circuit. Consider the following alternatives: (i) One has a full characterization of

the processes appearing in the circuit; (ii) one knows only the cardinalities of the setting and outcome variables; and (iii) one has characterized only the operational identities satisfied by the processes in the circuits. We now consider how to achieve entanglement certification relative to each of these alternatives.

- (i) In this case, as long as the sets of measurements in $\mathbf{N}$ and $\mathbf{M}$ are characterized and are informationally complete, then one can tomographically reconstruct $\mathbf{P} = \rho^{AB}$ and certify entanglement in the bipartite state ρ^{AB} using standard separability criteria [15].
- (ii) In this case, entanglement certification can be done in a device-independent manner by looking for a violation of a Bell inequality [3]. Specifically, consider a Bell circuit with outcomes $a, b \in \{0, \dots, o-1\}$ and settings $x \in \{0, \dots, \Delta_N - 1\}$ and $y \in \{0, \dots, \Delta_M - 1\}$. The correlations $\{p(ab|xy)\}_{abxy}$ can be viewed as points in the probability space $\mathbf{p} \in \mathbb{R}^{o^2 \Delta_N \Delta_M}$. Device-independent certification in Bell circuits is typically analyzed relative to the notion of a local hidden variable (LHV) model [3]. The convex set of all LHV-realizable probability distributions $p(ab|xy)$, formalized in Eq. (4), is known as the *local polytope*, denoted $\mathcal{L}(\mathcal{B})$, which depends only on the cardinalities of the setting and outcome variables in $\mathcal{B}$. Bell inequalities are the facet inequalities of $\mathcal{L}(\mathcal{B})$. Because LHV-realizability coincides with NCOM-realizability relative to $\mathcal{O}_{\text{tttt}}(\mathcal{B})$ [i.e., both are defined by Eq. (4)], Bell inequalities can also be seen as the weakest set of noncontextuality inequalities in the bipartite Bell circuit. The simplest and most well-known example of Bell inequalities occurs for $\Delta_N = \Delta_M = o = 2$, leading to the CHSH inequalities [14,44]. For certain Bell circuits, such as the one described in Example 1, Bell inequalities are the *only* noncontextuality inequalities.
- (iii) This is the case of the entanglement certification method proposed here. It differs from (i) in that it does not rely on a full characterization of the measurements in $\mathbf{N}$ and $\mathbf{M}$; it also differs from the Bell test in (ii), since it relies on operational identities in the Bell circuit $\mathcal{B}$. These operational identities can be faithfully inferred from the operational statistics of $\mathcal{B}$ without characterizing the devices, under the sufficient condition of *tomographic completeness*, as we detail in Sec. IV.

Our entanglement certification method relies on characterizing the set of all NCOM-realizable correlations relative to various choices of operational identities. More formally,

Definition 4. The noncontextual polytope $\mathcal{NC}_{\text{ijuvw}}(\mathcal{B})$ is defined to be the set of correlations $\mathbf{p} := \{p(ab|xy)\}_{abxy} \in \mathbb{R}^{o^2 \Delta_N \Delta_M}$ that are NCOM-realizable relative to the set $\mathcal{O}_{\text{ijuvw}}(\mathcal{B})$ of operational identities, i.e.,

$$\mathcal{NC}_{ijuvw}(\mathcal{B})$$

$$:= \{\mathbf{p} \text{ is NCOM-realizable relative to } \mathcal{O}_{ijuvw}(\mathcal{B})\}.$$

Different from the local polytope $\mathcal{L}(\mathcal{B})$, this set depends not only on the cardinalities of the setting variables and outcome variables in $\mathcal{B}$, but also on additional details, namely, those required to determine the operational identities in $\mathcal{B}$. An inequality whose violation witnesses nonmembership in $\mathcal{NC}_{ijuvw}(\mathcal{B})$ will be termed a *noncontextuality inequality relative to $\mathcal{O}_{ijuvw}(\mathcal{B})$* .

Given the subset inclusion relations of Eq. (14), we have that, for any Bell circuit $\mathcal{B}$,

$$\begin{aligned} \mathcal{NC}(\mathcal{B}) &= \mathcal{NC}_{aaaa}(\mathcal{B}) \subseteq \mathcal{NC}_{aattt}(\mathcal{B}) \\ &\subseteq \mathcal{NC}_{atttt}(\mathcal{B}), \mathcal{NC}_{tattt}(\mathcal{B}) \subseteq \mathcal{NC}_{tttt}(\mathcal{B}) = \mathcal{L}(\mathcal{B}), \end{aligned} \quad (15)$$

as depicted schematically in Fig. 2(a). It should be noted that, apart from $\mathcal{NC}(\mathcal{B})$, each of these sets of correlations above is convex, as they are each defined by imposing additional linear constraints [in the form of Eq. (19)] on the convex set $\mathcal{L}(\mathcal{B})$. Therefore, apart from $\mathcal{NC}(\mathcal{B})$, their boundary will be determined by sets of *linear noncontextuality inequalities*.

In the following, we focus mostly on the noncontextual polytope $\mathcal{NC}_{aattt}(\mathcal{B})$ in the context of some particular examples of Bell circuits $\mathcal{B}$. The properties of

$\mathcal{NC}_{aattt}(\mathcal{B})$, namely, its dimension and vertex complexity, are detailed in Appendix B, as are the numerical procedures underpinning its analysis, namely, facet enumeration and polytope membership linear programming.

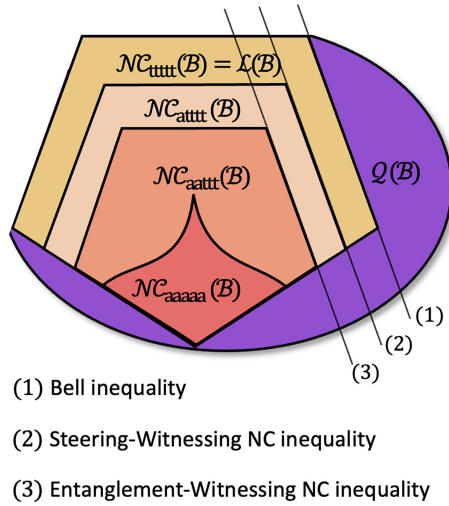
Throughout this section, the relevant operational identities are derived from the theoretical specification of the Bell circuit. The question of how these identities can be inferred from or enforced in experimental data is addressed in Secs. IV and VI A.

A. Certifying two-qubit noisy isotropic states with fixed measurements

Recall that the local polytope $\mathcal{L}(\mathcal{B})$ depends only on the cardinality of setting variables in $\mathbf{N}$ and $\mathbf{M}$, denoted $\Delta_{\mathbf{N}}$ and $\Delta_{\mathbf{M}}$, respectively, and the number of distinct outcomes for each, denoted o . We can, therefore, refer to the class of Bell circuits sharing this same local polytope as the $(\Delta_{\mathbf{N}}, \Delta_{\mathbf{M}}, o, o)$ -Bell circuits, denoted $\text{Bell}_{\Delta_{\mathbf{N}}\Delta_{\mathbf{M}}oo}$, and write $\mathcal{L}(\text{Bell}_{\Delta_{\mathbf{N}}\Delta_{\mathbf{M}}oo})$ for their common local polytope.

By contrast, the noncontextual polytope $\mathcal{NC}_{ijuvw}(\mathcal{B})$ depends also on the set of operational identities $\mathcal{O}_{ijuvw}(\mathcal{B})$ (though it still does not depend on the full characterization of $\mathcal{B}$). In the following, since we are considering only the sets of operational identities $\mathcal{O}_{aattt}(\mathcal{B})$, $\mathcal{O}_{atttt}(\mathcal{B})$, $\mathcal{O}_{tattt}(\mathcal{B})$, and $\mathcal{O}_{tttt}(\mathcal{B})$ and, hence, only the nontrivial operational identities in the multi-measurement $\mathbf{N}$ and $\mathbf{M}$, it is convenient to encode these

(a): Correlation $p(ab|xy)$ in Bell circuit $\mathcal{B}$



(b): Bipartite states ρ^{AB}

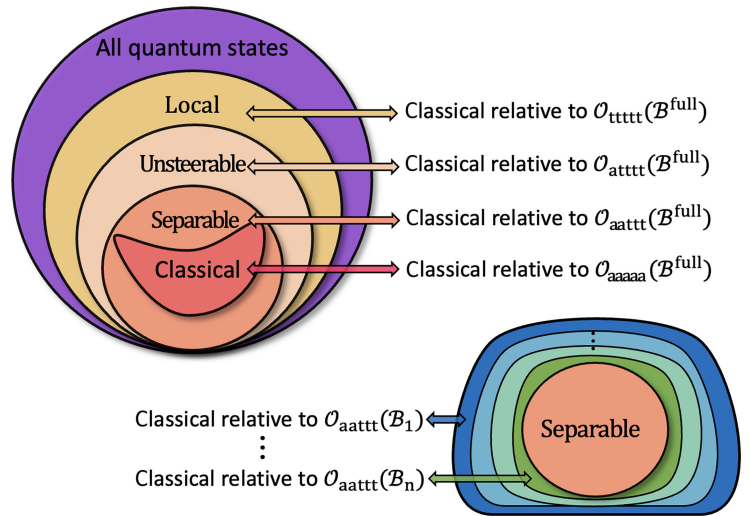


FIG. 2. (a) Given a fixed $(\mathbf{N}, \mathbf{M}, \mathbf{P})$ -Bell circuit $\mathcal{B}$, we sketch different NCOM-realizable polytopes in the space of operational statistics, where $\mathcal{NC}(\mathcal{B}) = \mathcal{NC}_{aaaa}(\mathcal{B}) \subseteq \mathcal{NC}_{aattt}(\mathcal{B}) \subseteq \mathcal{NC}_{atttt}(\mathcal{B}) \subseteq \mathcal{NC}_{tttt}(\mathcal{B}) = \mathcal{L}(\mathcal{B}) \subseteq \mathcal{Q}(\mathcal{B})$. Here, $\mathcal{NC}_{aaaa}(\mathcal{B})$ is defined by considering operational identities; $\mathcal{NC}_{aattt}(\mathcal{B})$ is defined by considering all nontrivial operational identities on Alice's and on Bob's measurements; $\mathcal{NC}_{atttt}(\mathcal{B})$ is defined by considering all nontrivial operational identities only on Alice's measurements; $\mathcal{NC}_{tttt}(\mathcal{B}) = \mathcal{L}(\mathcal{B})$ is defined by not taking into account any nontrivial operational identities, and $\mathcal{Q}(\mathcal{B})$ is the set of all quantum correlations that is realizable by a Bell circuit with the same cardinality of the setting and outcome variables. (b) A bipartite state being classical, separable, unsteerable, or local is equivalent to it being classical relative to $\mathcal{O}_{aaaaa}(\mathcal{B}^{\text{full}})$, $\mathcal{O}_{aattt}(\mathcal{B}^{\text{full}})$, $\mathcal{O}_{atttt}(\mathcal{B}^{\text{full}})$, or $\mathcal{O}_{tttt}(\mathcal{B}^{\text{full}})$, respectively. Inset: using a sequence of Bell circuits $\{\mathcal{B}_i = (\mathbf{N}_i, \mathbf{M}_i, \mathbf{P})\}_i$ that limit to the measurement-full Bell circuit with $\mathbf{N}_1 \subset \cdots \subset \mathbf{N}_n \subset \cdots \subset \mathbf{N}^{\text{full}}$ and $\mathbf{M}_1 \subset \cdots \subset \mathbf{M}_n \subset \cdots \subset \mathbf{M}^{\text{full}}$, the set of states that are classical relative to $\mathcal{O}_{aattt}(\mathcal{B}_i)$ limits to the set of separable states.

identities as the intrinsic geometries of $\mathbf{N}$ and $\mathbf{M}$, that is, as shapes, whose vertices satisfy the linear constraints defined by $\mathcal{O}_{\text{all}}(\mathbf{N})$ and $\mathcal{O}_{\text{all}}(\mathbf{M})$, respectively. Accordingly, we denote this pair of shapes by P and Q , respectively, and refer to the corresponding *class* of Bell circuits as (P, Q) -Bell circuits, denoted Bell_{PQ} . For any $\mathcal{B} \in \text{Bell}_{PQ}$, we can write $\mathcal{N}\mathcal{C}_{\text{ijuvw}}(\text{Bell}_{PQ})$ for the corresponding noncontextual polytope relative to $\mathcal{O}_{\text{ijuvw}}(\mathcal{B})$.

For example, consider the class of the simplest type of Bell circuit with $\Delta_{\mathbf{N}} = \Delta_{\mathbf{M}} = o = 2$. For any pair of binary-outcome measurements $\mathbf{N} = \{\{N_{a|x}^A\}_a\}_x$, the full set of operational identities satisfied by this set of effects is the singleton set consisting of the trivial identity: $N_{0|0} + N_{1|0} = N_{0|1} + N_{1|1}$. The intrinsic geometry of this set of effects is that of a square, since the vertices of the square, denoted $\{\vec{t}_{a,x}\}_{a,x}$, satisfy exactly one linear constraint: $\vec{t}_{0|0} + \vec{t}_{1|0} = \vec{t}_{0|1} + \vec{t}_{1|1}$. Therefore, this class can be denoted either as the class of (2, 2, 2, 2)-Bell circuits, Bell_{2222} , or under our convention as the class of (square, square)-Bell circuits, denoted Bell_{ss} .

Example 1. Consider the class of (square, square)-Bell circuits, i.e., Bell_{ss} . We have $\mathcal{O}_{\text{aattt}}(\text{Bell}_{ss}) = \mathcal{O}_{\text{atttt}}(\text{Bell}_{ss}) = \mathcal{O}_{\text{ttttt}}(\text{Bell}_{ss})$. It follows that we have a collapse of the hierarchy of polytopes in Eq. (15), that is,

$$\mathcal{N}\mathcal{C}_{\text{aattt}}(\text{Bell}_{ss}) = \mathcal{N}\mathcal{C}_{\text{atttt}}(\text{Bell}_{ss}) = \mathcal{N}\mathcal{C}_{\text{ttttt}}(\text{Bell}_{ss}).$$

Because $\mathcal{N}\mathcal{C}_{\text{ttttt}}(\text{Bell}_{ss}) = \mathcal{L}(\text{Bell}_{2222})$, the set of noncontextuality inequalities for this class is equivalent to the set of CHSH inequalities.

Similarly, one can observe that for the class of (octahedron, octahedron)-Bell circuits, denoted Bell_{oo} , the noncontextual polytope associated to $\mathcal{O}_{\text{aattt}}(\text{Bell}_{oo})$ also coincides with the local polytope for (3, 3, 2, 2)-Bell circuits, i.e., $\mathcal{N}\mathcal{C}_{\text{aattt}}(\text{Bell}_{oo}) = \mathcal{L}(\text{Bell}_{3322})$. Consequently, to obtain nontrivial noncontextuality inequalities that offer new avenues for entanglement certification [i.e., beyond the avenues already offered by Bell tests, thereby showing that the final inclusion relation in the hierarchy of Eq. (15) can be made strict], it is necessary to consider a Bell circuit with more structure in the local intrinsic geometries of the measurements.

In the next three examples, we take the shape describing the intrinsic geometry of $\mathbf{N}$ to be the geometric dual of the shape describing the intrinsic geometry of $\mathbf{M}$. This strategy is motivated by the fact that such a geometric duality relation achieves the maximal quantum violation of the CHSH inequalities and the fact that a similar duality is conjectured to also be optimal for quantum violations of noncontextuality inequalities in prepare-measure circuits [42]. In particular, we will use certain pairs of regular polyhedra because they may enable a more noise-robust entanglement certification protocol [33] and because their symmetries simplify the analytical results.

To demonstrate their usefulness, the resulting noncontextuality inequalities are compared with all previously known Bell- and steering-based tests for certifying entanglement in the one-parameter family of two-qubit noisy isotropic states:

$$\rho_{\text{iso}}^v = v|\Phi^+\rangle\langle\Phi^+| + (1-v)\frac{\mathbb{1}}{4}, \quad (16)$$

where $|\Phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ is a maximally entangled state shared between Alice and Bob.

We now turn to the case where P and Q are an octahedron and a cube, respectively, as depicted in Fig. 3(a).

Example 2. Consider the class of (octahedron, cube)-Bell circuits, denoted Bell_{oc} . We have $\mathcal{O}_{\text{aattt}}(\text{Bell}_{oc}) = \mathcal{O}_{\text{tattt}}(\text{Bell}_{oc}) \subset \mathcal{O}_{\text{atttt}}(\text{Bell}_{oc}) = \mathcal{O}_{\text{ttttt}}(\text{Bell}_{oc})$. (The equalities are due to the fact that the octahedron, like the square, describes only trivial operational identities, while the strict inclusion is due to the fact that the cube includes some nontrivial operational identities.) It follows that the inclusion relations in the hierarchy of correlations described in Eq. (15) become

$$\begin{aligned} \mathcal{N}\mathcal{C}_{\text{aattt}}(\text{Bell}_{oc}) &= \mathcal{N}\mathcal{C}_{\text{tattt}}(\text{Bell}_{oc}) \subset \mathcal{N}\mathcal{C}_{\text{atttt}}(\text{Bell}_{oc}) \\ &= \mathcal{N}\mathcal{C}_{\text{ttttt}}(\text{Bell}_{oc}). \end{aligned}$$

The characterization of the polytope $\mathcal{N}\mathcal{C}_{\text{aattt}}(\text{Bell}_{oc})$ is discussed in Appendix B. It is found to have 336 nontrivial facet inequalities (excluding the trivial non-negativity constraints), which can be classified into three distinct

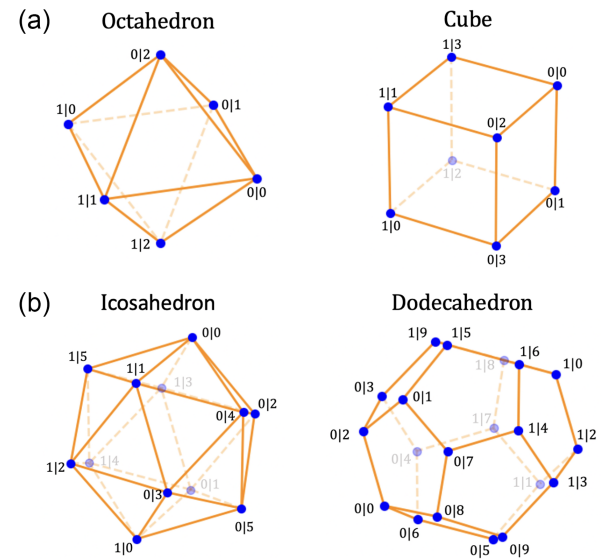


FIG. 3. Geometric configuration of $\mathbf{N} = \{\{N_{a|x}^A\}_a\}_x$ (left) and $\mathbf{M} = \{\{M_{b|y}^B\}_b\}_y$ (right). (a) Example 2: with cubic and octahedral symmetry; (b) Example 3: with icosahedral and dodecahedral symmetry.

TABLE II. Facet-defining inequalities for the polytope $\mathcal{NC}_{\text{aattt}}(\text{Bell}_{\text{oc}})$.

Inequality	Orbit size
$p(01 22) + p(10 03) - p(11 00) + p(11 21) \geq 0$	144
$p(01 22) + p(10 03) - p(11 00) + p(11 20) \geq 0$	144
$p(01 22) + p(10 03) - p(11 00) + p(11 10) - p(11 12) - p(11 20) + p(11 21) + p(11 22) \geq 0$	48

classes, where the elements within each class form an orbit under permutations of the values of the outcome variables and of the values of the setting variables. See Table II.

In particular, the third class of inequalities can be used to certify entangled two-qubit isotropic states for $v > (1/\sqrt{3}) \simeq 0.577$, thereby outperforming all certification methods for such states that rely on Bell inequalities based on projective measurements, which certify entanglement only for $v > \{1/[K_G(3)]\} \in (0.6875, 0.6961)$ [45], where $K_G(3)$ is the Grothendieck constant [46]. (The CHSH inequalities can be used to certify entangled two-qubit isotropic states only if $v > (1/\sqrt{2}) \simeq 0.717$.)

Note that the range of v for which entanglement is witnessable in Example 2 *does not* exceed the range that is witnessable, in principle, by steering inequalities, namely, $v \geq 1/2$ [47,48]. Consequently, to obtain nontrivial non-contextuality inequalities that offer avenues for entanglement certification beyond those offered by steering inequalities, it is necessary to consider a Bell circuit with even more structure in the local intrinsic geometries of the measurements. The next example does so. It considers multi-measurements having the intrinsic geometries of an icosahedron on one side and a dodecahedron on the other, as depicted in Fig. 3(b).

Example 3. Consider the class of (icosahedron, dodecahedron)-Bell circuits, denoted Bell_{id} . We have $\mathcal{O}_{\text{aattt}}(\text{Bell}_{\text{id}}) \subset \mathcal{O}_{\text{atttt}}(\text{Bell}_{\text{id}})$, $\mathcal{O}_{\text{tattt}}(\text{Bell}_{\text{id}}) \subset \mathcal{O}_{\text{ttttt}}(\text{Bell}_{\text{id}})$. It follows that the inclusion relations in the hierarchy of correlations described in Eq. (15) become

$$\begin{aligned}\mathcal{NC}_{\text{aattt}}(\text{Bell}_{\text{id}}) &\subset \mathcal{NC}_{\text{atttt}}(\text{Bell}_{\text{id}}), \\ \mathcal{NC}_{\text{tattt}}(\text{Bell}_{\text{id}}) &\subset \mathcal{NC}_{\text{ttttt}}(\text{Bell}_{\text{id}}).\end{aligned}$$

Importantly, there exists the following facet-defining inequality of the polytope $\mathcal{NC}_{\text{aattt}}(\text{Bell}_{\text{id}})$:

$$\begin{aligned}\mathcal{I} = & p(00|00) + p(00|01) - (\phi + 1)p(00|02) + \frac{1}{\phi}p(00|03) \\ & + \frac{1}{\phi}p(00|10) + \frac{1}{\phi}p(00|11) - 2p(00|12) + p(00|13) \\ & - p(00|20) - \frac{1}{\phi}p(00|21) + (\phi + 1)p(00|22) - p(00|23) \\ & - \frac{1}{\phi^2}p(10|11) + \frac{1}{\phi}p(10|12) \geq 0,\end{aligned}\quad (17)$$

where $\phi = [(\sqrt{5} + 1)/2]$. This inequality can be violated by two-qubit isotropic states with $v > \sqrt{[(1 + \phi^2)/3\phi^4]} \approx 0.4195$ and, consequently, outperforms all entanglement certification methods that rely on steering inequalities.

In the class of Bell circuits Bell_{id} , there are 12 effects on one side and 20 effects on the other, hence, $12 \times 20 = 240$ product effects. The operational identities, however, imply that every effect on one side can be expressed as a linear combination of four effects on that side, so that the 240 product effects can be expressed as linear combinations of just 16 product effects. It follows that the 240 distinct conditional probability distributions $\{p(ab|xy)\}_{abxy}$ can be expressed as a linear combination of 16 of these. For this reason, any noncontextuality inequality can be expressed in terms of at most 16 conditional probabilities. [The inequality of Eq. (17), for instance, has only 14 terms, and some setting variables do not appear.]

Example 4. We now consider a family of dual polyhedra whose vertices become dense on the unit sphere. For each positive integer μ , let P_μ denote the icosahedral geodesic polyhedron $\{3, 5^+\}_{\mu,0}$ and let Q_μ denote its dual Goldberg polyhedron $\{5^+, 3\}_{\mu,0}$, obtained via regular μ -fold subdivisions of the icosahedron and dodecahedron [49]. (The case $\mu = 1$ recovers the icosahedron-dodecahedron pair used in Example 3.) The polyhedra P_μ and Q_μ have $10\mu^2 + 2$ and $20\mu^2$ vertices, respectively. For each μ , this defines the class of (P_μ, Q_μ) -Bell circuits, denoted Bell_μ .

As summarized in Table III, this measurement strategy is highly effective for certifying entanglement of the two-qubit isotropic states of Eq. (16): By using standard qubit projective measurements with $5\mu^2 + 1$ settings for Alice and $10\mu^2$ settings for Bob with the intrinsic geometry

TABLE III. Critical visibility v for entanglement certification of the two-qubit noisy isotropic state ρ_{iso}^v using projective multi-measurements with the intrinsic geometries of the geodesic polyhedron P_μ for Alice and the dual Goldberg polyhedron Q_μ for Bob.

μ	No. of vertices in P_μ	No. of vertices in Q_μ	Visibility v
1	12	20	≈ 0.4195
2	42	80	≈ 0.3568
3	92	180	≈ 0.3434
4	162	320	≈ 0.3394
5	252	500	≈ 0.3372
6	362	720	≈ 0.3360
7	492	980	≈ 0.3353

(P_μ, Q_μ) described above, the visibility v for which entanglement certification is still possible decreases rapidly with μ and approaches the value of the visibility at the separable-entangled boundary, namely, $v = \frac{1}{3}$.

B. Certifying two-qubit noisy isotropic states with randomized measurements

In the previous subsection, we restricted our attention to Bell circuits with multi-measurements whose operational identities define intrinsic geometries with nontrivial symmetries. Doing so allowed us to derive closed-form analytical expressions for the noncontextuality inequalities. However, having such symmetry is not necessary for entanglement certification. We now demonstrate that our noncontextuality-based approach remains effective and efficient when Alice and Bob use a strategy in which the measurements implemented at each party are chosen uniformly at random.

In this subsection, we continue to focus on two-qubit isotropic states but randomly sample a set of projective measurements at each party. For each choice of $m \in \{10, 20, 50, 100, 200\}$, we draw m projective measurements for Alice and m for Bob, independently and uniformly at random, and then determine the smallest visibility v of a two-qubit isotropic state that can be certified as entangled by our noncontextuality-based test.

The minimal visibility v of the two-qubit isotropic state for which entanglement can be witnessed for a given m depends on the specific sets of measurements sampled. For instance, samplings that distribute more uniformly over the sphere are expected to yield smaller values of the minimal visibility v . It is useful, therefore, to report an average value of the minimum visibility v in an ensemble of different samplings. We consider 100 different samplings for each m . The average value of the minimal visibility v for each m is reported in Table IV. As m increases, the average critical visibility decreases steadily and approaches the value at the separable-entangled boundary $v = \frac{1}{3}$.

This randomized measurement strategy offers a useful perspective on the practicality of our entanglement-certification protocol. While employing many measurement

settings may initially seem demanding, randomized measurements can, in fact, simplify the implementation. Randomized measurements underpin protocols such as classical shadows [50] and many other quantum information tasks [51]. A key distinction, however, is that these standard approaches typically rely on precise characterization of the measurement devices, whereas our method requires *no prior characterization*. The fact that our framework accommodates randomized strategies could, therefore, point to a realistic path toward scalable entanglement certification based on generalized noncontextuality.

Beyond their generality, randomized measurement schemes offer a second key advantage: They remove the need to specify *a priori* a particular set of operational identities to be targeted in the experiment. Instead, entanglement certification can proceed as follows: (i) Implement a set of measurements chosen at random, (ii) infer the operational identities that are satisfied by the set using experimental data, and (iii) use these identities to test for entanglement by deriving explicit noncontextuality inequalities tailored to these identities. We return to this strategy in Sec. IV B, where we discuss practical routes to implementing our noncontextuality-based entanglement certification in the presence of imperfect measurements.

C. Certifying arbitrary two-qubit entangled states

We have shown that our noncontextuality-based entanglement certification scheme outperforms Bell-based and steering-based schemes for two-qubit noisy isotropic states. We now demonstrate that it also outperforms Bell-based schemes for *generic* entangled bipartite states.

To do so, we perform numerical simulations on randomly sampled two-qubit entangled states. Following Ref. [30], we draw mixed states from the Hilbert-Schmidt measure by generating 4×4 complex matrices A with independent complex Gaussian entries (zero mean, equal variance) and setting $\rho = [A^\dagger A / \text{Tr}(A^\dagger A)]$, and we postselect those that are entangled. We use fixed projective measurements with Bloch vectors that form dual polyhedra as described in Examples 1–4 with $\mu = 2$ (the absolute orientation of the pair of polyhedra is irrelevant, since the tested states are already uniformly sampled). For each randomly sampled entangled state, we then perform the noncontextuality-based tests introduced above. The results are presented in Table V. One sees that the noncontextuality test significantly outperforms the Bell test.

We note that the measurement bases for the Bell tests in Table V above are not optimized to avoid the additional numerical cost they incur.

However, we can also benchmark our approach against the CHSH test described in Ref. [30], which, for each state, maximizes the CHSH value over all local measurement bases using the Horodecki criterion [52] (equivalently, test in our Example 1 with measurement bases optimized). The question studied in Ref. [30] is as follows: Among

TABLE IV. Average critical visibility v of the two-qubit isotropic state ρ_{iso}^v that can be certified as entangled using m randomly sampled projective measurements. The values are obtained by averaging the critical visibility over 100 independent trials for each m .

No. of bases m	Visibility v
10	0.540 ± 0.027
20	0.445 ± 0.015
50	0.384 ± 0.006
100	0.360 ± 0.003
200	0.348 ± 0.001

TABLE V. Percentage of 10^6 randomly sampled two-qubit entangled states that are certified as entangled by the measurement strategies in Examples 1–4 ($\mu = 2$). Here, we compare the certification methods based on the standard Bell test and those based on our noncontextuality test. We note that computing the percentage for the Bell test becomes numerically infeasible for Example 4, $\mu = 2$, since the corresponding local polytope has 2^{61} vertices; by contrast, the noncontextual polytope has only 8160 vertices, so the noncontextuality test remains feasible.

Bell circuits	(%) certified by	
	Bell test	Noncontextuality test
Example 1	$\approx 0.005\%$	$\approx 0.005\%$
Example 2	$\approx 0.08\%$	$\approx 0.82\%$
Example 3	$\approx 0.24\%$	$\approx 24.40\%$
Example 4, $\mu = 2$	Infeasible	$\approx 76.35\%$

randomly sampled two-qubit entangled states, what fraction are such that their entanglement can be certified by a CHSH inequality violation? When states are sampled according to the same Hilbert-Schmidt measure assumed above, Ref. [30] found that only about 1.09% of entangled states can be certified by a CHSH test alone, even if one optimizes the choice of measurements for each sampled state [specifically, it was shown in Ref. [30] that about 75.8% of all bipartite states are entangled (via the positive partial transpose criterion [53]), while only about 0.82% violate some CHSH inequality (via the Horodecki criterion [52])]. This is an improvement over the best fraction achievable by a Bell test without optimizing measurement bases, namely, 0.24%, but it still falls far short of what can be achieved by noncontextuality tests in Examples 3 and 4 ($\mu = 2$).

Moreover, Bell-test-based certification with many settings quickly becomes computationally prohibitive: The local polytope has $o^{\Delta_N} o^{\Delta_M}$ vertices, i.e., a number of vertices that scales *exponentially* with the cardinality of the setting variables. This is in contrast to the situation with the noncontextual polytope, which has $O(o^2 \Delta_N \Delta_M)$ vertices; i.e., a number of vertices scales only *polynomially* with the cardinality of the setting variables (see also Appendix B 1).

IV. DEALING WITH INEXACT OPERATIONAL IDENTITIES

In Sec. III, we derived noncontextuality inequalities from operational identities that are inferred from a theoretical specification of the Bell circuit. We now explain how these identities can be inferred from or enforced in experimental data.

Real devices never implement the targeted measurements precisely. Moreover, one only ever obtains finite-run statistics, and, consequently, one's estimate of the effects deviates from the true effects because of statistical fluctuations. It follows that the operational identities that hold

among the effects actually realized in the laboratory will never coincide exactly with those that hold for the ideal target effects. It is, therefore, crucial to develop tools that can contend with this problem in order to make our noncontextuality-based entanglement certification protocol practically applicable.

There is a second, more subtle consideration. To make the protocol free from prior characterization assumptions, the operational identities used in our noncontextuality test must be established directly from the experimental data (i.e., the observed statistics) rather than being assumed. The sufficient condition ensuring this is a form of tomographic completeness: Steered states on one side probe enough operational directions to identify the linear identities that hold among the measurement effects on that side. A formal definition of this condition, together with the associated tomographic-incompleteness loophole, is given in Sec. VI A.

In this section, we describe two complementary methods for dealing with this problem. (i) Infer from the observed statistics the operational identities that *happen to hold* for the multi-measurements that were actually implemented. Because the technique fits the aphorism “cut your coat to suit your cloth,” we refer to it as the “cut-to-your-cloth” technique. (ii) Using a generalization of the secondary-procedures technique of Ref. [54], i.e., from the measurement procedures in the convex hull of those realized, find a set, termed the *secondary* measurement procedures, that satisfy the operational identities *exactly* and test the statistics that obtain in the circuit composed of the secondary procedures.

Both methods begin with the standard self-consistent tomography [31,32] to mitigate statistical fluctuations in the raw relative frequencies. Therefore, our approach is well suited to platforms that already use self-consistent tomography [55,56].

A. Self-consistent tomography

This step can be viewed as a regularization step that transforms the raw relative frequencies $f_{ab|xy}$ into a probability distribution that is realizable in a D -dimensional quantum model. Specifically, one varies over choices of bipartite state ρ^{AB} and multi-measurements $\mathbf{N}^{\text{pri}} = \{\{N_{a|x}^{A,\text{pri}}\}_a\}_x$ and $\mathbf{M}^{\text{pri}} = \{\{M_{b|y}^{B,\text{pri}}\}_b\}_y$ to minimize the distance between $f_{ab|xy}$ and the quantum-realizable probability distribution $p^{\text{pri}}(ab|xy)$. The distance can be quantified in various ways—for instance, using a weighted χ^2 and solving

$$\chi^2 := \min_{(\mathbf{N}^{\text{pri}}, \mathbf{M}^{\text{pri}}, \rho^{AB})} \sum_{abxy} \left| \frac{f_{ab|xy} - p^{\text{pri}}(ab|xy)}{\Delta f_{ab|xy}} \right|^2, \quad \text{such that } p^{\text{pri}}(ab|xy) = \text{Tr}[N_{a|x}^{A,\text{pri}} \otimes M_{b|y}^{B,\text{pri}} \rho^{AB}]. \quad (18)$$

We refer to the best-fit $p^{\text{pri}}(ab|xy)$ as the “primary statistics,” the $\mathbf{N}^{\text{pri}}$ and $\mathbf{M}^{\text{pri}}$ as the *primary multi-measurements* (those

actually realized in the experiments), and $\mathbf{P} = \rho^{AB}$ as the primary preparation. The tuple $(\mathbf{N}^{\text{pri}}, \mathbf{M}^{\text{pri}}, \mathbf{P})$ constitutes a nonunique solution of the fitting, and, in general, $\mathbf{N}^{\text{pri}}$ and $\mathbf{M}^{\text{pri}}$ will differ from the $\mathbf{N}$ and $\mathbf{M}$ that are targeted.

Given *tomographic completeness*, the regularization step above is just an instance of “self-consistent tomography” [31,32], where all possible fits for the primary multi-measurements are related by invertible linear maps. This freedom is termed tomographic gauge, a property that will play an important role further on. The particular effect vectors arising from the fit are not gauge invariant, but the linear identities among these are. This is why tomographic completeness is the condition that allows operational identities to be inferred from experimental data using the cut-to-your-cloth technique and to be enforced for the secondary procedures in the secondary-procedures technique.

B. The cut-to-your-cloth technique

We now outline the main steps of our first scheme to address the fact that the experimentally realized measurement processes differ from the targeted ones.

1. Extract the operational identities from the primary statistics

In this first approach, we begin by determining the operational identities satisfied by the primary multi-measurements. Concretely, we define

$$\mathcal{O}_{\text{all}}(\mathbf{N}^{\text{pri}}) := \left\{ \vec{\alpha} \left| \sum_{a,x} \alpha_{a,x} N_{a|x}^{A,\text{pri}} = 0 \right. \right\}, \quad (19a)$$

$$\mathcal{O}_{\text{all}}(\mathbf{M}^{\text{pri}}) := \left\{ \vec{\beta} \left| \sum_{b,y} \beta_{b,y} M_{b|y}^{B,\text{pri}} = 0 \right. \right\}, \quad (19b)$$

capturing the sets of linear relations that hold among the primary effects for Alice and Bob, respectively. Importantly, we note that these operational identities can be determined entirely from the primary statistics $p^{\text{pri}}(ab|xy)$. This follows because different tuples $(\mathbf{N}^{\text{pri}}, \mathbf{M}^{\text{pri}}, \rho^{AB})$ that yield the same $p^{\text{pri}}(ab|xy)$ are related by an invertible linear map on system A and one on system B (i.e., a gauge freedom [31]) under the stated sufficient condition of *tomographic completeness* (see also Appendix C 2), and these invertible linear maps left the operational identities unchanged. Therefore, one can obtain the operational identities $\mathcal{O}_{\text{all}}(\mathbf{N}^{\text{pri}})$ and $\mathcal{O}_{\text{all}}(\mathbf{M}^{\text{pri}})$ by determining the linear constraints directly from $p^{\text{pri}}(ab|xy)$:

$$\mathcal{O}_{\text{all}}(\mathbf{N}^{\text{pri}}) := \left\{ \vec{\alpha} \left| \sum_{a,x} \alpha_{a,x} p^{\text{pri}}(ab|xy) = 0, \quad \forall b, y \right. \right\}, \quad (20a)$$

$$\mathcal{O}_{\text{all}}(\mathbf{M}^{\text{pri}}) := \left\{ \vec{\beta} \left| \sum_{b,y} \beta_{b,y} p^{\text{pri}}(ab|xy) = 0, \quad \forall a, x \right. \right\}. \quad (20b)$$

In other words, $\mathcal{O}_{\text{all}}(\mathbf{N}^{\text{pri}})$ and $\mathcal{O}_{\text{all}}(\mathbf{M}^{\text{pri}})$ defined above are gauge-independent quantities.

2. Derive noncontextuality inequalities

$\mathcal{O}_{\text{all}}(\mathbf{N}^{\text{pri}})$ and $\mathcal{O}_{\text{all}}(\mathbf{M}^{\text{pri}})$ are, in general, more complex than the operational identities that hold for multi-measurements that have the intrinsic geometry of a highly symmetric polytope, as in the examples in Sec. III. Nevertheless, they can still be used to efficiently certify entanglement by simply deriving the explicit noncontextuality inequalities relative to $\mathcal{O}_{\text{all}}(\mathbf{N}^{\text{pri}})$ and $\mathcal{O}_{\text{all}}(\mathbf{M}^{\text{pri}})$. Indeed, the randomized multi-measurement considered in Sec. III B lacked symmetry but nonetheless worked efficiently for certifying entanglement.

3. Test the noncontextuality inequalities on the primary statistics

Finally, we apply the noncontextuality inequalities in step 2 to the primary statistics $p^{\text{pri}}(ab|xy)$ obtained in step 1. If the primary statistics violate any of these inequalities, then the underlying state is entangled.

C. The secondary-procedures technique

The second analysis implements a computation that simulates a *local* classical postprocessing of the primary statistics $p^{\text{pri}}(ab|xy)$, the result of which is termed the *secondary* statistics, denoted $p^{\text{sec}}(ab|xy)$. The postprocessing is chosen to ensure that the secondary statistics satisfy *exactly* a specific set of operational identities. The noncontextuality inequalities derived from the specific set of operational identities can then be applied directly to $p^{\text{sec}}(ab|xy)$.

Since the secondary statistics are obtained from the primary ones through a simulation of local classical processing, which can be conceptualized as injecting noise, something that can only make the operational statistics easier to model noncontextually. Hence, any violation of a noncontextuality inequality by the secondary statistics is sufficient to imply the NCOM-nonrealizability of the primary statistics, and, hence, it is also sufficient to certify entanglement in the bipartite state [57]. Note that such certification depends only on the observed data and the choice of target operational identities. This approach is a generalization of the secondary-procedures technique introduced in Ref. [54] for testing NCOM-realizability in the prepare-measure circuit. Here, we summarize the main steps and leave the details to Appendix C.

1. Derive noncontextuality inequalities

For the sets of operational identities $\mathcal{O}_{\text{all}}(\mathbf{N})$ and $\mathcal{O}_{\text{all}}(\mathbf{M})$ satisfied by the multi-measurement $\mathbf{N}$ and $\mathbf{M}$ in the targeted Bell circuit $\mathcal{B}$, determine the noncontextuality inequalities relative to $\mathcal{O}_{\text{aattt}}(\mathcal{B})$. Examples of these inequalities are

given in Sec. III A, and the tools for their derivation are detailed in Appendix B.

2. Construct secondary statistics

Given the primary statistics $p^{\text{pri}}(ab|xy)$ from self-consistent tomography and a set of targeted operational identities $\mathcal{O}_{\text{all}}(\mathbf{N})$ and $\mathcal{O}_{\text{all}}(\mathbf{M})$ from step 1, solve for the *secondary* multi-measurements, denoted $\mathbf{N}^{\text{sec}} = \{\{N_{a|x}^{A,\text{sec}}\}_a\}_x$ and $\mathbf{M}^{\text{sec}} = \{\{M_{b|y}^{B,\text{sec}}\}_b\}_y$ as follows:

- (i) Take each effect to be an element of the convex hull of the primary effects, i.e.,

$$N_{a|x}^{A,\text{sec}} = \sum_{a',x'} u_{a',x'}^{a,x} N_{a'|x'}^{A,\text{pri}}, \quad (21a)$$

$$M_{b|y}^{B,\text{sec}} = \sum_{b',y'} v_{b',y'}^{b,y} M_{b'|y'}^{B,\text{pri}},$$

$$\text{such that } \sum_a N_{a|x}^{A,\text{sec}} = \mathbb{1}^A \sum_b M_{b|y}^{B,\text{sec}} = \mathbb{1}^B \quad \forall x, y, \quad (21b)$$

with $\sum_{a',x'} u_{a',x'}^{a,x} = \sum_{b',y'} v_{b',y'}^{b,y} = 1 \quad \forall a, x, b, y$ and $u_{a',x'}^{a,x}, v_{b',y'}^{b,y} \geq 0$;

- (ii) Require the effects to satisfy the set of operational identities that are respected by the targeted multi-measurements, i.e., impose the constraints that $\mathcal{O}_{\text{all}}(\mathbf{N}^{\text{sec}}) = \mathcal{O}_{\text{all}}(\mathbf{N})$ and $\mathcal{O}_{\text{all}}(\mathbf{M}^{\text{sec}}) = \mathcal{O}_{\text{all}}(\mathbf{M})$, or, equivalently,

$$\sum_{ax} \alpha_{ax} N_{a|x}^{A,\text{sec}} = \mathbb{0}^A \quad \forall \vec{\alpha} \in \mathcal{O}_{\text{all}}(\mathbf{N}), \quad (22a)$$

$$\sum_{by} \beta_{by} M_{b|y}^{B,\text{sec}} = \mathbb{0}^B \quad \forall \vec{\beta} \in \mathcal{O}_{\text{all}}(\mathbf{M}); \quad (22b)$$

- (iii) In a variation over the secondary effects, find those that maximize a measure of closeness with the primary effects, namely, those that maximize the objective functions

$$C_N = \frac{1}{o\Delta_N} \sum_{ax} u_{a,x}^{a,x}, \quad (23a)$$

$$C_M = \frac{1}{o\Delta_M} \sum_{b,y} v_{b,y}^{b,y}, \quad (23b)$$

where $u_{a,x}^{a,x}$ and $v_{b,y}^{b,y}$ are defined in Eq. (21).

At first glance, it may seem to be the case that one needs to specify a choice of tomographic gauge for the primary measurements, to obtain the coefficients $\{u_{a',x'}^{a,x}\}$ and $\{v_{b',y'}^{b,y}\}$. However, as discussed in Appendix C 2, this is not the case. Indeed, assuming *tomographic completeness*,

all possible choices of gauge yield primary measurements with the same operational identities since they are related by an invertible linear map (i.e., operational identities are gauge independent). It follows that the coefficients $\{u_{a',x'}^{a,x}\}$ can be inferred purely from the *primary statistics*, using the following linear program:

$$\max C_N = \frac{1}{o\Delta_N} \sum_{ax} u_{a,x}^{a,x}$$

such that

$$\begin{aligned} \sum_{axa'x'} p^{\text{pri}}(a'b'|x'y') u_{a',x'}^{a,x} \alpha_{a,x} &= 0 \quad \forall b', y', \quad \forall \vec{\alpha} \in \mathcal{O}_{\text{all}}(\mathbf{N}), \\ \sum_{aax'} p^{\text{pri}}(a'b'|x'y') u_{a',x'}^{a,x} &= p^{\text{pri}}(b'|y') \quad \forall x, b', y', \\ \sum_{a'x'} u_{a',x'}^{a,x} &= 1, \quad u_{a',x'}^{a,x} \geq 0 \quad \forall a, x, a', x', \end{aligned} \quad (24)$$

where $p^{\text{pri}}(b'|y') := \sum_{a'} p^{\text{pri}}(a'b'|x'y')$ is independent of x' since the primary statistics are nonsignaling. One can obtain the coefficients $v_{b',y'}^{b,y}$ in an analogous fashion.

The *secondary statistics* $p^{\text{sec}}(ab|xy)$ are then defined as follows:

$$\begin{aligned} p^{\text{sec}}(ab|xy) &= \text{Tr}[N_{a|x}^{A,\text{sec}} \otimes M_{b|y}^{B,\text{sec}} \rho^{AB}] \\ &= \sum_{a'b'x'y'} v_{b',y'}^{b,y} u_{a',x'}^{a,x} p^{\text{pri}}(a'b'|x'y'). \end{aligned} \quad (25)$$

One can check that the secondary statistics are normalized and nonnegative by noting that the secondary measurements $\{\{N_{a|x}^{A,\text{sec}}\}_a\}_x$ and $\{\{M_{b|y}^{B,\text{sec}}\}_b\}_y$ are enforced to be valid POVMs.

3. Test the noncontextuality inequalities on the secondary statistics

Finally, we apply the noncontextuality inequalities from step 1 to the secondary statistics $p^{\text{sec}}(ab|xy)$ in step 2. A violation of any such inequality by the secondary statistics witnesses entanglement in the underlying bipartite state.

D. Comparing the two techniques

A disadvantage of the secondary-procedures technique, relative to the cut-to-your-cloth technique, is that the former may fail to witness NCOM-nonrealizability in cases where the latter succeeds. This is because, as noted above, postprocessing can only make statistics easier to model noncontextually. Consequently, it may happen that the secondary statistics are NCOM-realizable even though the primary statistics are not.

On the other hand, an advantage of the secondary-procedures technique over the cut-to-your-cloth technique is that it is much more straightforward to make

comparisons across different experimental runs, for instance, runs where the bipartite state is varying or where the primary measurements are varying (due to drift, for instance). In the secondary-procedures technique, one enforces a particular set of operational identities to be satisfied by the secondary multi-measurements, independent of the primary multi-measurements or the bipartite state, so the noncontextuality inequality one is testing is also independent of these. Moreover, in this case, one can resample the primary statistics to obtain an estimate of the statistical error in the degree of violation of the fixed noncontextuality inequality. Consequently, the magnitude of the violation of a given inequality can be used to compare different bipartite entangled states, as we demonstrated in Sec. VA for the family of two-qubit isotropic states.

By contrast, in the cut-to-your-cloth technique, variations in the primary multi-measurements across experimental runs imply variation in the set of operational identities satisfied by these and, hence, in the noncontextuality inequality being tested. Variations in the primary multi-measurements might arise from drift (for instance, drift between runs probing different bipartite states) or from fluctuations in the raw frequencies. This variation precludes an even-handed comparison of violation magnitudes across different bipartite states and makes it infeasible to assign statistical error bars to the violation, even for repeated trials on the same state. (We return to this issue in Sec. VB, where we introduce a noise-robustness-based quantifier for the degree of violation which allows a comparison between different experimental runs in the cut-to-your-cloth technique.)

V. EXPERIMENTAL RESULTS

Our experiment tests the noncontextuality inequality of Eq. (17) for various choices of the value of the visibility v in the one-parameter family of two-qubit states. It uses the photonic setup depicted in Fig. 4. As discussed in depth in Appendix D, we generate polarization-entangled photons via type-II spontaneous parametric down-conversion in a Sagnac interferometer. These photons are subsequently directed through a controlled depolarization channel, generating probabilistic mixtures of the four Bell states as outlined in Appendix D 1. This allows for the preparation of different mixed states closely approximating (with about 99% state fidelity) two-qubit isotropic states with different values of the visibility v [see Eq. (16)]. By varying v , we can prepare separable, entangled, steerable, and nonlocal states [6,8,47,48].

In the experiment, we prepared 12 different quantum states ρ_{exp} , each with a measured fidelity to the nearest isotropic state ρ_{iso}^v consistently near 99% and remain above 98% (see Table IX in Appendix D). The value of the visibility v will be used to represent the state we prepared. The NC inequality of Eq. (17) requires Alice's and Bob's multi-measurements to have the intrinsic geometries of an icosahedron and a dodecahedron, respectively. We target a regular icosahedron and a regular dodecahedron inscribed in the Bloch sphere for Alice and Bob's multi-measurements, respectively, as depicted in Fig. 3 [see Eq. (B5) in Appendix B for a precise definition]. This corresponds to implementing six projective binary-outcome measurements on Alice's photon and ten on Bob's. These measurements clearly satisfy the relevant operational identities and are

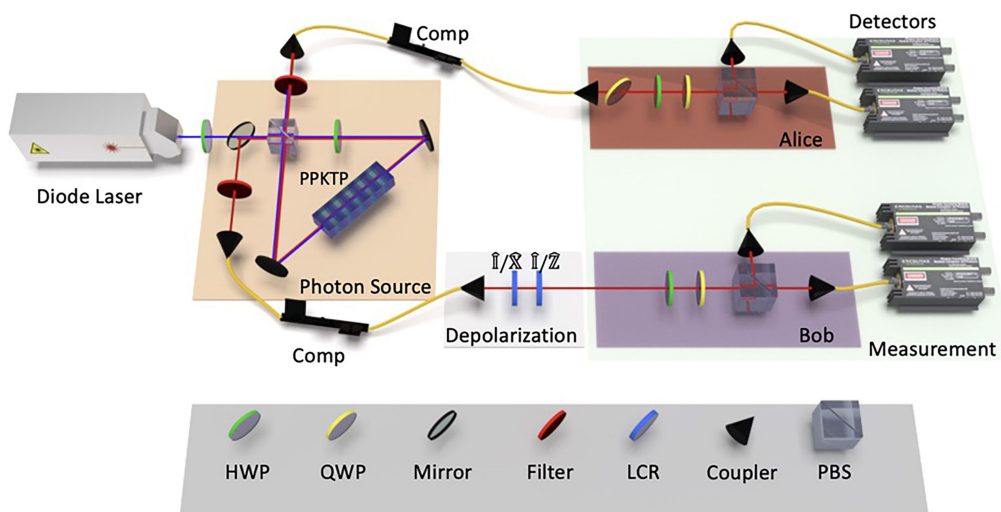


FIG. 4. Schematic of the experiment. Polarization-entangled photon pairs are generated via parametric down-conversion in a Sagnac interferometer [58]. Polarization controllers (Comp) compensate polarization rotations in the single-mode fibers. A pair of computer-controlled liquid crystal phase retarders (LCR) aims to implement arbitrary strength depolarizing channels on one of the photons. Each photon is then directed to separate polarization analyzers—Alice and Bob—where polarization measurements are performed using half-wave plates (HWPs) and quarter-wave plates (QWPs). Coincidence counts are recorded. PPKTP: periodically poled potassium titanyl phosphate; PBS: polarizing beam splitter; $\hat{I}$, $\hat{X}$, and $\hat{Z}$: Pauli gates.

predicted to violate the NC inequality of Eq. (17). They are implemented using a combination of half-wave plates (HWP), quarter-wave plates (QWP), and polarization beam splitters (PBS).

For each of the 12 states, we implemented 60 distinct pairs of measurement settings (six on Alice and ten on Bob). In 2 h of total run-time for the experiment, each pair of settings accumulated on the order of 2×10^6 twofold coincidence counts (consisting of four coincidence pairs), which are denoted as the raw count statistics $N_{ab|xy}$.

Imperfections in experimental implementation, such as polarization optics quality, fiber drift, and laser power drift, are unavoidable. Additionally, the raw count statistics $N_{ab|xy}$ represent finite samples from the true probability distribution, and directly normalizing them may yield a distribution that is not quantumly realizable (indeed, possibly not even realizable within a GPT [59,60]). For example, statistical fluctuations can lead to violations of the no-signaling conditions. To address this, the raw counts $N_{ab|xy}$ are converted to relative frequency $f_{ab|xy}$, and these are fitted to a probability distribution $p^{\text{pri}}(ab|xy)$ using

self-consistent tomography fit described in Sec. IV using a total weighted least squares method [61].

From $p^{\text{pri}}(ab|xy)$, entanglement certification can be achieved using either of the two different techniques discussed in Sec. IV. We consider each in turn.

A. Certifying entanglement using the secondary-procedures technique

Given the primary statistics obtained from the self-consistent tomography, we construct secondary statistics $p^{\text{sec}}(ab|xy)$ via the technique discussed in Sec. IV C. For the quantities $C_N = [1/(o\Delta_N)] \sum_{a,x} u_{a,x}^{a,x}$ and $C_M = [1/(o\Delta_M)] \sum_{b,y} v_{b,y}^{b,y}$, defined in Eq. (23), we achieve, experimentally, an average of 0.98 and 0.96, respectively (see Table XI), which confirms that the secondary statistics remain very close to the primary statistics (the simulated local postprocessing effectively introduces approximately 2% white noise to the measurements).

The values of the quantity $\mathcal{I}$ appearing on the left-hand side of the noncontextuality inequalities of Eq. (17), computed from the secondary statistics $p^{\text{sec}}(ab|xy)$ for

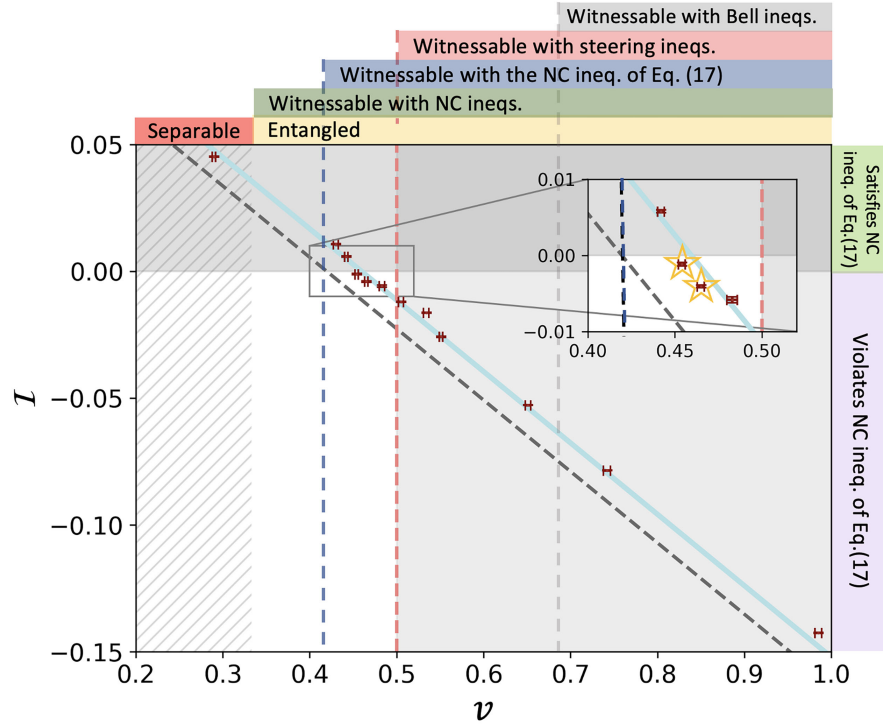


FIG. 5. Experimental violation of the noncontextuality (NC) inequality of Eq. (17), i.e., $\mathcal{I} \geq 0$, as a function of the isotropic-state visibility v . Our experimental data are shown as points with vertical and horizontal error bars, while the ideal case is shown as the diagonal dashed gray line, and the noisy case (i.e., the ideal case modified by the additional 2% white noise in the measurements) is shown as a solid blue line. On top of the diagram, we depict the separable-entangled divide, as well as the parameter regimes certifiable by various techniques: general noncontextuality inequalities, the specific noncontextuality inequality derived in Eq. (17), steering inequalities, and Bell inequalities. The region in white contains isotropic states that simultaneously (i) are entangled, (ii) are not certifiable by steering (or Bell) inequalities, and (iii) are certifiable by noncontextuality inequalities. Three of our experimental data points lie in this white region, and the enlarged inset distinguishes the two reconstructed states that are independently certified to be unsteerable in Appendix E (hollow yellow stars). (By contrast, the unsteerability of the third state was not confirmed.) Error bars for both $\mathcal{I}$ and v are obtained from 100 trials of Monte Carlo simulation and are depicted. (See Table IX for detailed data.).

different values of the visibility v of the prepared two-qubit isotropic states, are presented in Fig. 5. Error bars are obtained via Monte Carlo simulations assuming Poisson count statistics. In these simulations, Poissonian fluctuations in the raw counts are also propagated through the steps of the secondary-procedures technique so that different primary statistics lead to different secondary statistics, all of which are evaluated against the same set of exact operational identities and noncontextuality inequality, as noted in Sec. IV D.

Violations of the inequality in Eq. (17) are observed for all setups that correspond to isotropic states with $v > 0.43$, including those with $v < 0.5$, for which it is known that their entanglement cannot be certified by steering inequalities [37].

Our experimentally prepared states achieve fidelities near 99% with the nearest isotropic state. However, this guarantees only that, for instance, when $v \leq 1/2$, they are *close to* unsteerable states. It does not guarantee that they are unsteerable (because it remains an open question whether proximity to an unsteerable isotropic state implies unsteerability). To address this, we directly assess whether each bipartite state prepared in our experiment is unsteerable (i.e., admits a local-hidden-state model) by solving a semidefinite program (see Appendix E). This analysis showed that the state associated to parameter value $v = 0.483$ cannot be confirmed to be unsteerable but that the states associated to parameter values $v = 0.454$ and $v = 0.465$ are confirmed to be unsteerable (see discussion in Appendix E). This establishes that our noncontextuality-based technique can certify entanglement for two-qubit states that are unsteerable and hence whose entanglement cannot be certified by steering inequalities or Bell inequalities.

B. Certifying entanglement using the cut-to-your-cloth technique

As noted in Sec. IV D, because the primary statistics $p^{\text{pri}}(ab|xy)$ and the primary measurements typically vary across experimental runs due to small systematic drifts and (in our simulations) Monte Carlo statistical uncertainty, one cannot certify entanglement using a single fixed noncontextuality inequality in the cut-to-your-cloth technique. Consequently, we cannot plot the degree of violation of a fixed noncontextuality inequality as the bipartite state varies or report an error bar for the degree of violation. Using the cut-to-your-cloth technique, therefore, we cannot produce the sort of analysis provided in Fig. 5 or Sec. VA for witnessing entanglement.

To address this, we propose an alternative way of comparing the degree of NCOM-nonrealizability as we vary the bipartite state. Specifically, we quantify the *noise robustness* of $\mathbf{p}^{\text{pri}} = \{p^{\text{pri}}(ab|xy)\}_{abxy}$ by asking how much noise needs to be added to make the correlations compatible with a noncontextual ontological model.

Concretely, we define

$$\eta_c[\mathbf{p}^{\text{pri}}] := \min\{\eta \in [0, 1] | p_\eta^{\text{pri}}(ab|xy) \in \mathcal{NC}_{\text{aatt}}(\mathcal{B}^{\text{pri}})\},$$

$$p_\eta^{\text{pri}}(ab|xy) := (1 - \eta)p^{\text{pri}}(ab|xy) + \eta d^{\text{pri}}(ab|xy), \quad (26)$$

where $d^{\text{pri}}(ab|xy)$ is defined as the statistics obtained by averaging over all extreme vertices of $\mathcal{NC}_{\text{aatt}}(\mathcal{B}^{\text{pri}})$. We note that $d^{\text{pri}}(ab|xy)$ obtained from our operational statistics is very close to uniform but is not, in general, exactly uniform over outcomes, since the uniform distribution need not belong to $\mathcal{NC}_{\text{aatt}}(\mathcal{B}^{\text{pri}})$. It will be useful to also consider the noise robustness of $p^{\text{sec}}(ab|xy)$. We denote this as $\eta_c[\mathbf{p}^{\text{sec}}]$ and define it as the obvious analog of Eq. (26).

C. Comparing the cut-to-your-cloth and secondary-procedures techniques

Table VI demonstrates the improved noise robustness of the cut-to-your-cloth technique relative to the secondary-procedures technique, hence the greater sensitivity to NCOM-nonrealizability that is offered by the cut-to-your-cloth technique.

On the other hand, if one wishes to compare the degree of NCOM-nonrealizability across different experimental runs, e.g., runs implementing different two-qubit isotropic states, then, although noise robustness provides a means of

TABLE VI. Critical noise thresholds η_c for 12 states with estimated isotropic visibility v . The second and third columns give η_c obtained from the primary and secondary statistics, respectively. The error bar is calculated from 100 trials of Monte Carlo simulation with the experimental statistics. The second column reveals that noise robustness can indeed provide a means of comparing the degree of NCOM-nonrealizability of bipartite states with different values of v in the cut-to-your-cloth technique, in spite of the fact that the noncontextuality inequality being tested varies with v . A comparison of the second and third columns shows that the secondary statistics consistently yield a lower η_c (i.e., they are less noise robust) than the primary statistics, since the construction of secondary statistics from the primary statistics involves local classical postprocessing, which introduces additional noise.

Visibility v	$\eta_c[\mathbf{p}^{\text{pri}}]$	$\eta_c[\mathbf{p}^{\text{sec}}]$
0.290 ± 0.002	0	0
0.441 ± 0.002	0	0
0.430 ± 0.002	0	0
0.454 ± 0.002	0.0413 ± 0.0005	0.0045 ± 0.0009
0.465 ± 0.002	0.0589 ± 0.0004	0.0336 ± 0.0006
0.483 ± 0.003	0.0993 ± 0.0002	0.0472 ± 0.0014
0.505 ± 0.003	0.1397 ± 0.0003	0.0929 ± 0.0008
0.534 ± 0.002	0.1803 ± 0.0002	0.1213 ± 0.0004
0.551 ± 0.003	0.2166 ± 0.0003	0.1796 ± 0.0005
0.651 ± 0.003	0.3368 ± 0.0001	0.3092 ± 0.0004
0.742 ± 0.003	0.4168 ± 0.0001	0.3995 ± 0.0004
0.982 ± 0.004	0.5613 ± 0.0001	0.5470 ± 0.0002

doing so for both techniques (as illustrated in Table VI), it is arguably less natural than being able to compare the degree of violation of a single noncontextuality inequality (just as we compare the degree of violation of a single Bell inequality when witnessing entanglement via Bell inequality violation). The latter approach, the outcome of which is illustrated in Fig. 5, is available only for the secondary-procedures technique.

VI. LOOPHOLES IN THE NONCONTEXTUALITY TEST

The preceding sections describe how noncontextuality inequalities can certify entanglement and how the required operational identities can be inferred from experimental data. We now discuss in detail the conditions under which this inference is valid.

A. The tomographic incompleteness loophole in noncontextuality tests

The notion of *tomographic completeness* is the core condition behind our noncontextuality-based technique for entanglement certification. It plays the same role here that it does in *self-consistent* tomography, including GPT tomography [32] and gate-set quantum tomography [31]: One assumes an *effective finite-dimensional vector space* (in the case of GPT tomography, one assumes a finite-dimensional GPT state space), and one requires that the *implemented* preparations and measurement effects span this vector space and its dual, respectively.

1. From standard quantum state tomography to self-consistent quantum tomography

In standard quantum state tomography, one typically assumes that the measurements have already been characterized, in the sense of the POVMs representing them being known *a priori*. The set of measurement effects is termed *tomographically complete* if the linear span is the full operator space associated with the (assumed finite-dimensional) Hilbert space.

In self-consistent quantum tomography, no device is taken as characterized *a priori*. In a prepare-measure experiment on a quantum system with Hilbert space dimension D , one works in the operator space of dimension $K = D^2$ (e.g., qubit quantum theory has a model dimension $K = 4$). Tomographic completeness then means that the implemented set of states and the implemented set of measurement effects each span the K -dimensional vector space [32]. (The same idea underlies gate-set tomography: One needs a set of experimentally available preparations and measurements that are tomographically complete for the assumed model space [31].) When this holds, the reconstructed description is unique only up to an overall invertible linear change of representation (gauge freedom); when it fails, some directions in the vector space of

Hermitian operators cannot be accessed from the data, and the reconstruction becomes underdetermined.

2. Tomographic completeness in a Bell circuit

In a bipartite Bell circuit with a state ρ^{AB} on $\mathcal{H}^A \otimes \mathcal{H}^B$, local POVMs $N^A = \{N_{a|x}^A\}_x$ on A and $M^B = \{M_{b|y}^B\}_y$ on B , each party's measurements induce a family of *subnormalized steered states* $\{\tilde{\rho}_{a|x}^A\}_x$ and $\{\tilde{\rho}_{b|y}^B\}_y$ for the other party, as in Eq. (7). We can define the corresponding steered-state spans (real vector spaces of Hermitian operators)

$$V_A := \text{span}(\{\tilde{\rho}_{b|y}^A\}_{b,y}) \subseteq \text{Herm}(\mathcal{H}^A), \quad (27a)$$

$$V_B := \text{span}(\{\tilde{\rho}_{a|x}^B\}_{a,x}) \subseteq \text{Herm}(\mathcal{H}^B). \quad (27b)$$

One can then formally define the notion of tomographic completeness in a Bell circuit as follows.

Definition 5. The Bell circuit (N, M, P) is *tomographically complete* if Alice's effects separate all states in V_A and Bob's effects separate all states in V_B . Equivalently, for all $X^A \in V_A$ and $Y^B \in V_B$,

$$\text{Tr}[N_{a|x}^A X^A] = 0, \quad \forall a, x \Rightarrow X^A = 0^A \quad (28a)$$

$$\text{Tr}[M_{b|y}^B Y^B] = 0, \quad \forall b, y \Rightarrow Y^B = 0^B. \quad (28b)$$

In other words, Alice's effects, when restricted to V_A , span the dual space V_A^* , and Bob's effects, when restricted to V_B , span the dual space V_B^* .

In a Bell experiment, when the dimensions of the quantum systems are known, e.g., $\dim(\mathcal{H}^A) = \dim(\mathcal{H}^B) = D$, we always have $\dim(V_A) \leq D^2$ and $\dim(V_B) \leq D^2$. Moreover, a generic bipartite state ρ^{AB} defines steered-state spans V_A and V_B [by Eq. (27)] such that $\dim(V_A) = \dim(V_B) = D^2$. Only on a measure-zero subset of states is it the case that $\dim(V_A) < D^2$ or $\dim(V_B) < D^2$ [29,37]. It follows that, in practice, once a finite-dimensional model is specified, tomographic completeness of a Bell circuit can be assessed experimentally rather than assumed as part of a prior trusted characterization of the devices. For a generic bipartite state, this reduces to checking that the implemented local measurements are tomographically complete on the relevant local operator spaces.

3. Tomographic incompleteness loophole

The condition of tomographic completeness is used exactly as in self-consistent tomography: It ensures that operational identities among the implemented measurement effects can be inferred *from the observed statistics*, i.e., identities that hold among any primary measurements that realize the operational statistics are the same, and can,

thus, be obtained from the identities that hold among the operational statistics [see, e.g., Eq. (20)].

If the tomographic-completeness condition fails, then the inferred operational identities may fail to be true identities of the physical system. For example, two effects that appear operationally identical on the realized family of steered states may be distinguishable by some unrealized steered state. This is closely related to what has been called the tomographic incompleteness loophole [22,62]: Generalized noncontextuality is assessed relative to a choice of system and requires that the implemented preparations and measurements are tomographically complete for that system. If tomographic completeness fails, then a noncontextuality inequality need not witness nonclassicality. This loophole is the way in which our method differs most sharply from entanglement certification based on Bell-inequality violations.

4. How tomographic completeness can be checked in practice

In practice, an expectation about the dimension K cannot be verified by showing that the observed statistics when arranged as a data table have rank equal to K , since finite statistics and noise generically make the data table full rank. However, a hypothesis about the dimension K can be assessed using the standard train-and-test methodology used in GPT tomography [32]. This is done by ruling out dimensions less than K by showing that a lower-dimensional model cannot fit the statistics and disfavoring dimensions greater than K because they overfit the test data. In our experiment, we expect $K = 4$ (i.e., a two-qubit model) to hold. The experimental data are found to be consistent with this expectation; i.e., the operational statistics can be well fit by such a model, where the fitted steered states and measurements span the relevant operator space and, thus, are tomographically complete.

B. Absence of a detector inefficiency loophole in noncontextuality tests

Unlike Bell tests, our methods do not suffer from a detector inefficiency loophole. In a Bell test, inefficient detectors can open a loophole because whether the detector registers an outcome or fails to register any outcome (termed the “null outcome”) may depend on the value of the hidden variable (as well as the local measurement setting). Consequently, the full distribution, including null-outcome events, may admit a local-hidden-variable model even when the distribution conditioned on *not* seeing the null-outcome event violates a Bell inequality [63–65]. This is why loophole-free Bell tests include null-outcome events explicitly in the analysis and rely on achieving sufficiently high efficiency in the detectors to close the loophole, as in the detection-loophole-free and fully loophole-free Bell experiments of Refs. [66–68].

In contrast, Refs. [69,70] showed that generalized non-contextuality tests in prepare-and-measure circuits have no such detector inefficiency loophole. The basic reason is that detector inefficiency is represented operationally by attenuating the original effects and adding a null-outcome effect to the measurement, provided the efficiencies of all detected outcomes are nonzero. The set of attenuated effects is the same as in the ideal experiment, up to strictly positive rescalings. The null-outcome effect is then fixed by the requirement that the effects sum to the unit effect. Strictly positive rescalings do not change the positive cone generated by a set of effects. Thus, the sets of effects in the ideal and inefficient experiments are cone-equivalent in the sense of Refs. [69,70]. Finally, as shown in that work, cone equivalence preserves NCOM-realizability.

The same reasoning applies to the bipartite Bell circuit considered in this work. Specifically, once diagram preservation and tomographic completeness of the relevant Bell circuit are granted, detector inefficiency is a modification of the implemented measurement procedures. Let Alice’s and Bob’s ideal multi-measurements be described by $\mathbf{N} := \{\{N_{a|x}^A\}_a\}_x$ and $\mathbf{M} := \{\{M_{b|y}^B\}_b\}_y$, respectively, and let $\mathcal{B} = (\mathbf{N}, \mathbf{M}, \rho^{AB})$ denote the ideal Bell circuit. Detector inefficiency implies replacing the ideal multi-measurements with $\tilde{\mathbf{N}} := \{\{\tilde{N}_{\alpha|x}^A\}_\alpha\}_x$ and $\tilde{\mathbf{M}} := \{\{\tilde{M}_{\beta|y}^B\}_\beta\}_y$, where α and β range over the original outcome set, and the new “null” outcome labeled $\emptyset$. The attenuated effects, i.e., those corresponding to $\alpha \neq \emptyset$, $\beta \neq \emptyset$, are

$$\tilde{N}_{a|x}^A = \epsilon_{a|x} N_{a|x}^A, \quad (29a)$$

$$\tilde{M}_{b|y}^B = \epsilon_{b|y} M_{b|y}^B, \quad (29b)$$

where $0 < \epsilon_{a|x}, \epsilon_{b|y} \leq 1$ for all a, x, b, y , and the corresponding null-outcome effects are

$$\tilde{N}_{\emptyset|x}^A = \sum_a (1 - \epsilon_{a|x}) N_{a|x}^A, \quad (30a)$$

$$\tilde{M}_{\emptyset|y}^B = \sum_b (1 - \epsilon_{b|y}) M_{b|y}^B. \quad (30b)$$

These measurements, together with the bipartite state ρ^{AB} , define an inefficient Bell circuit $\mathcal{B}^e = (\tilde{\mathbf{N}}, \tilde{\mathbf{M}}, \rho^{AB})$ with statistics

$$p_e(\alpha, \beta | x, y) = \text{Tr}[\rho^{AB} \tilde{N}_{\alpha|x}^A \otimes \tilde{M}_{\beta|y}^B]. \quad (31)$$

The fact that there is no detector inefficiency loophole in noncontextuality tests follows from the following result.

Proposition 2. Let $\mathcal{B}^e$ be the inefficient Bell circuit defined above, with $0 < \epsilon_{a|x}, \epsilon_{b|y} \leq 1$ for all attenuated effects. Then $p(a, b | x, y)$ is NCOM-realizable relative to

$\mathcal{O}_{\text{aatt}}(\mathcal{B})$ if and only if $p_e(\alpha, \beta|x, y)$ is NCOM-realizable relative to $\mathcal{O}_{\text{aatt}}(\mathcal{B}^e)$.

The proof is given in Appendix C 3. The intuition is that the operational identities of the ideal and inefficient measurements are related by an invertible rescaling on the attenuated effects, and, consequently, so are the sets of response functions that represent these in the noncontextual ontological model. If the distribution $p(a, b|x, y)$ in the Bell circuit $\mathcal{B}$ is NCOM-realizable, then there exists a set of response functions that provides a noncontextual representation of $\mathbf{N}$ and one that does so for $\mathbf{M}$ (i.e., the response functions satisfy the same operational identities as the effects). But then by rescaling these response functions by the same rescaling factors that take $\mathbf{N}$ to $\tilde{\mathbf{N}}$ and $\mathbf{M}$ to $\tilde{\mathbf{M}}$, we obtain two new sets of response functions that provide noncontextual representations of $\tilde{\mathbf{N}}$ and $\tilde{\mathbf{M}}$ while realizing the statistics $p_e(\alpha, \beta|x, y)$. The converse follows by the same argument.

Therefore, the proposition shows that it suffices to have a nonzero detector efficiency to implement our noncontextuality test. The distribution $p_e(\alpha, \beta|x, y)$ can be analyzed directly, treating null outcomes as ordinary outcomes, while maintaining the same conclusion about NCOM-realizability [71].

VII. COMPARISON WITH ALTERNATIVE ENTANGLEMENT CERTIFICATION APPROACHES

We now compare our entanglement certification technique to preexisting techniques. A summary of the relative merits of the different techniques, including ours, can be found in Table I.

A. Separability test with quantum state tomography

Witnessing entanglement by implementing a separability test on a state obtained via quantum state tomography requires prior characterization of one's measurements. Such a prior characterization is also required for estimating other entanglement witnesses [15].

B. Separability test with GPT and gate-set tomography

By contrast, our approach resembles that of GPT tomography [32,73] and gate-set tomography [31,74], which do not require a prior characterization of any subset of the procedures in the experiment but rather infer characterizations of all the procedures self-consistently from how they interact, i.e., in a bootstrap fashion.

Both GPT tomography and gate-set tomography do not return a unique output, neither for the bipartite state nor for the local effects, but rather a *set* of possible triples of these with different possibilities being related by an invertible linear transformation. As noted earlier, following Ref. [31], we refer to this nonuniqueness as a tomographic *gauge freedom*. The most obvious way to certify entanglement of

the bipartite state in this case is to implement a separability test for all possible choices of gauge. Note that merely fixing a tomographic gauge arbitrarily (rather than optimizing over it) can lead to estimates of quantities of interest that may be off by orders of magnitude from the true value [74].

Like GPT tomography and gate-set tomography, our approach also begins with self-consistent tomography. However, it differs in how we use this output to certify entanglement. Rather than performing a separability test on the entangled state that is most “entangled” after an optimization over gauge, we derive the noncontextuality inequality using the gauge-independent operational identities, and we test the noncontextuality inequality on the primary statistics $p^{\text{pri}}(ab|xy)$ [or the secondary statistics $p^{\text{sec}}(ab|xy)$] that is also gauge independent. Our technique consequently provides the possibility of certifying entanglement without requiring any optimization over gauge and indeed without even requiring a description of the full gauge freedom of the state and local effects.

C. Bell inequality test

Another technique for certifying entanglement that is gauge independent and requires no prior characterization of the measurements is to use violations of Bell inequalities (e.g., in the device-independent paradigm [3]). As argued in Ref. [35], the rigorous implementation of this approach requires fitting one's raw data to the nearest correlations in the polytope of nonsignaling correlations and then testing Bell inequalities on these correlations. These device-independent tests have an advantage over the technique described here: They do not require the sets of local effects to be tomographically complete. However, the price Bell tests pay for this advantage is that they can certify entanglement only in a strict subset of all entangled states [6], whereas our technique can certify any entangled state. Furthermore, we demonstrated in Sec. III C that, for a given number of measurement settings at each wing, a noncontextuality test can certify entanglement in a significantly larger fraction of the set of entangled states than can be certified in a Bell test. Moreover, the computational cost of solving the membership problem in a Bell test grows rapidly with the number of measurement settings, quickly becoming infeasible, as also noted in Sec. III C, whereas the analogous membership problem for a noncontextuality test remains feasible.

D. Steering inequality test

An entanglement certification technique that is closely related to the Bell-based technique is to use steering inequalities [5,33,36]. Like a Bell test, such a test can certify entanglement only for a strict subset of all entangled states (those that are steerable) [6,8]. Furthermore, this technique is device independent only on one side; i.e., they require a full characterization of a tomographically complete set of measurements on one side. By contrast, our approach can certify all entangled states and can also be

adapted to certify entanglement via noncontextuality-based steering certification, without requiring prior characterization of the measurement devices in either task. Examples of steering certification are provided in Appendix B 1.

E. Inequality test in non-Bell causal network

A different device-independent approach to verifying entanglement is to test causal compatibility inequalities within more elaborate causal networks that use the bipartite state as one common cause. As shown in Ref. [34], such methods can, in principle, certify entanglement in *any* entangled state while retaining the advantages of being a gauge-independent technique requiring no prior characterization of the measurements. However, these approaches require network structures that are more experimentally complex than that of the Bell circuit. Moreover, they require *additional* entangled sources beyond the one being characterized and rely on an extra assumption of source independence.

F. Entanglement certification and Kochen-Specker contextuality

Finally, a closely related idea—certifying entanglement by linking local hidden-variable models to realizability by a Kochen-Specker (KS) noncontextuality model—was already noted by Werner in Ref. [6]. However, this route differs in essential ways from our framework, which is based on the notion of generalized contextuality rather than KS contextuality. First, KS contextuality arises only in Hilbert-space dimension $D \geq 3$ and, thus, does not, even in principle, provide a certification method applicable to the two-qubit states. Second, KS-type tests are typically formulated for, and experimentally interpreted through, the idealization of projective measurements, which limits their direct practical applicability. Most importantly for the present discussion, Werner’s argument was framed under the assumption of access to *all* projective measurements. It is, therefore, unclear how strong such a certification can be when one is restricted to only a finite set of projective measurements. In this sense, while this idea is conceptually aligned with the broader theme of using contextuality to witness entanglement, it does not by itself amount to a practical entanglement certification protocol.

G. Entanglement certification using our approach

Compared to a Bell test, which relies on the weakest assumptions among standard entanglement-certification protocols, our noncontextuality-based approach offers no advantage in regimes where Bell-based certification already succeeds. Indeed, as noted in Sec. III, in such regimes our noncontextuality test in the $\mathcal{O}_{\text{tttt}}$ scenario is equivalent to the Bell test. Moreover, all operational identities required for the noncontextuality test are then trivially guaranteed by

the normalization of the measurement effects and, therefore, require no additional justification.

As noted earlier, our noncontextuality-based approach can certify entanglement in regimes where Bell-based certification fails. The strength of the extra operational conditions that are sufficient for being able to do so depends on the scope of entangled states one wants to certify. Specifically, as discussed in the main text, all entangled states can be certified using noncontextuality inequalities in the $\mathcal{O}_{\text{aatt}}$ scenario, for which tomographic completeness of both Alice’s and Bob’s measurements is sufficient to infer the relevant operational identities. By contrast, as discussed in Appendix B 1, if one seeks to certify entanglement for just the set of steerable states, one may instead use noncontextuality inequalities in the $\mathcal{O}_{\text{attt}}$ or $\mathcal{O}_{\text{tatt}}$ scenarios, in which case tomographic completeness of only one party’s measurements, Alice’s or Bob’s, respectively, is sufficient to infer the relevant operational identities.

As discussed above, there do exist other entanglement-certification protocols that can, in principle, certify more entangled states than a Bell-based test, some of which can even certify *all* entangled states. For instance, one can do inequality tests in quantum networks that are more complex than a Bell circuit [17,34]. Our entanglement-certification protocol offers an advantage over these insofar as a Bell circuit presents less of an experimental burden than does a more general quantum network. Other protocols may involve an experimental burden comparable to realizing a Bell circuit but require prior characterization of (all or some of) the measurement devices, as in quantum state tomography, entanglement witnesses, and steering-based tests. Our entanglement-certification protocol is preferable to these insofar as one can regard tomographic completeness as a weaker condition than that of having a prior characterization of measurement devices.

If the dimension of the underlying physical system is known, one can use self-consistent tomography to directly determine whether the measurement operators span the relevant operator spaces. In this case, tomographic completeness is inferred, not assumed. Even if the dimension of the underlying system is not known, one can still perform a hypothesis test for the system dimension, as demonstrated in Ref. [32], and then assess tomographic completeness relative to that hypothesis. The fact that such an assessment might yield a negative verdict for tomographic completeness means that our approach provides a concrete opportunity for falsifying some prior assumption about the condition of tomographic completeness. Such a negative verdict then flags the fact that one cannot implement a noncontextuality-based entanglement-certification technique using the experimental procedures that were implemented.

VIII. DISCUSSION

On the theoretical and foundational side, we showed that the noncontextuality inequalities that witness entanglement,

steerability, and nonlocality can be derived from applying the principle of noncontextuality to a hierarchy of subsets of the full set of operational identities. This hierarchy also offers an alternative *definition* of entanglement, steering, and nonlocality in terms of NCOM-nonrealizability for the measurement-full Bell circuit.

One clear avenue for future theoretical work is to use our methods to identify a hierarchy of sufficient conditions for nonclassicality in bipartite Bell circuits, i.e., tractable noncontextuality inequalities whose violations can witness nonclassicality (in particular, entanglement or steerability). One can also explore the power of noncontextuality inequalities for certifying nonclassicality of a target state in other experimental scenarios beyond the bipartite Bell circuit. For instance, one could consider the certification of multipartite entanglement using multipartite Bell circuits or investigate possible improvements to entanglement-certification protocols in circuits with more general causal structures, such as quantum networks. Another avenue worth pursuing is to modify the bipartite Bell circuit itself by allowing entangled measurements or by including states other than the target state whose entanglement is to be certified, thereby making the target state part of a multi-state. Each such modification can introduce additional operational identities that constrain the noncontextual ontological model and can thereby lead to more stringent conditions for nonclassicality, stronger noncontextuality inequalities, and improved prospects for entanglement certification.

Another direction for future work is to adapt our methods to identify a hierarchy of sufficient conditions for the nonclassicality of more general types of processes, such as quantum channels and instruments, as initially outlined in Ref. [37]. This could lead not only to new sufficient conditions for witnessing the nonclassicality of various quantum processes, but also to novel ways of *quantifying* nonclassicality.

On the experimental and technical side, we have shown that our noncontextuality inequalities can be directly translated into a practical entanglement-certification protocol. Uniquely among techniques for doing so, ours offers the possibility of realizing all of the following features at once: requiring no prior characterization of the measurement devices, delivering a verdict based on quantities that are independent of the tomographic gauge, being implementable using only the causal structure of a standard bipartite Bell circuit, and being able, in principle, to witness the entanglement of any entangled state.

One interesting direction for future experimental work is to use a noncontextuality-based technique to test sufficient conditions for the nonclassicality of multipartite states, channels, and instruments. Another frontier is to examine whether the assumption of tomographic completeness typically required for experimental tests of noncontextuality inequalities can be relaxed [62,75].

ACKNOWLEDGMENTS

We thank Yilè Yīng, Emanuele Polino, Elie Wolfe, and Yeong-Cherng Liang for useful discussions. We thank three anonymous referees for helpful comments, including requests for clarification of the assumptions underlying our entanglement-certification protocol, which led us to make many improvements in the presentation. The research was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC), the NSERC-Alliance Hyperspace project, Industry Canada, the Canada Foundation for Innovation (CFI), Canada Excellence Research Chair (CERC) program, and the Canada First Research Excellence Fund (CFREF). D. S. and R. W. S. were supported by Perimeter Institute for Theoretical Physics. Research at Perimeter Institute is supported in part by the Government of Canada through the Department of Innovation, Science and Economic Development and by the Province of Ontario through the Ministry of Colleges and Universities.

DATA AVAILABILITY

The code needed to reproduce the main results is available [76].

APPENDIX A: PROOF OF THEOREM 1 AND PROPOSITION 1

Below, we provide a detailed proof of Theorem 1 and Proposition 1 from the main text. We begin by recalling the notion of a classical preparation (or classical set of states) introduced in Ref. [37].

Definition 6. [37]—In measurement-full prepare-and-measure circuit $\mathcal{P}^{\text{full}} = (\mathbf{M}^{\text{full}}, \mathbf{P} = \{\{\tilde{\rho}_{a|x}^B\}_a\}_x)$. The multi-source $\{\{\tilde{\rho}_{a|x}^B\}_a\}_x$ is said to be *classical* if the operational statistics arising in this circuit are NCOM-realizable relative to $\mathcal{O}_{\text{all}}(\mathbf{M}^{\text{full}}) \cup \mathcal{O}_{\text{all}}(\mathbf{P})$.

Theorem 1 and Proposition 1. Let $\mathcal{B}^{\text{full}} = (\mathbf{N}^{\text{full}}, \mathbf{M}^{\text{full}}, \mathbf{P})$ denote a measurement-full Bell circuit. For the bipartite state ρ , there is an equivalence between the two properties in each of the following pairs:

- (i) classicality relative to $\mathcal{O}_{\text{aattt}}(\mathcal{B}^{\text{full}})$ and separability,
- (ii) classicality relative to $\mathcal{O}_{\text{attt}}(\mathcal{B}^{\text{full}})$ and $B \rightarrow A$ unsteerability,
- (iii) classicality relative to $\mathcal{O}_{\text{tattt}}(\mathcal{B}^{\text{full}})$ and $A \rightarrow B$ unsteerability,
- (iv) classicality relative to $\mathcal{O}_{\text{tttt}}(\mathcal{B}^{\text{full}})$ and locality,
- (v) classicality relative to $\mathcal{O}_{\text{aatt}}(\mathcal{B}^{\text{full}})$ and the impossibility of steering to a nonclassical set of states on B , and
- (vi) classicality relative to $\mathcal{O}_{\text{aatat}}(\mathcal{B}^{\text{full}})$ and the impossibility of steering to a nonclassical set of states on A .

Proof. The core of the proof is to apply the generalized Gleason theorem [77] iteratively to the different components of the ontological model. We start with case (ii) of Theorem 1 and then treat case (i). Cases (iii) and (iv) follow

directly from these arguments, and cases (v) and (vi) will be proved at the end.

Case (ii)—If a bipartite state ρ^{AB} is classical relative to $\mathcal{O}_{\text{attt}}(\mathcal{B}^{\text{full}})$, by Definition 3, the operational statistics $p(ab|xy)$ it generates with all local POVMs $\{N_{a|x}^A\}_a \in \mathbf{N}^{\text{full}}$ and $\{M_{b|y}^B\}_b \in \mathbf{M}^{\text{full}}$ should be NCOM-realizable relative to $\mathcal{O}_{\text{attt}}(\mathcal{B}^{\text{full}})$ per Definition 1. More specifically, one can write

$$\begin{aligned} p(ab|xy) &= \text{Tr}[(N_{a|x}^A \otimes M_{b|y}^B) \rho^{AB}] \\ &= \sum_{\lambda_A \lambda_B} p(\lambda_A \lambda_B) p(a|x, \lambda_A) p(b|y, \lambda_B). \end{aligned} \quad (\text{A1})$$

The assumption of noncontextuality applied to $\mathcal{O}_{\text{all}}(\mathbf{N}^{\text{full}})$ implies that the response function $p(a|x, \lambda)$ is a linear function of $N_{a|x}^A$. The generalized Gleason's theorem then ensures the existence of a unique density operator $\rho_{\lambda_A}^A$ such that $p(a|x, \lambda_A) = \text{Tr}[\rho_{\lambda_A}^A N_{a|x}^A]$. Defining the steering assemblage $\{\{\tilde{\rho}_{b|y}^A = \text{Tr}[(1^A \otimes M_{b|y}^B) \rho^{AB}]\}_b\}_y$, one has

$$\begin{aligned} p(ab|xy) &= \text{Tr}[\tilde{\rho}_{b|y}^A N_{a|x}^A] \\ &= \sum_{\lambda_A \lambda_B} p(\lambda_A \lambda_B) p(b|y, \lambda_B) \text{Tr}[N_{a|x}^A \rho_{\lambda_A}^A] \\ \Rightarrow \tilde{\rho}_{b|y}^A &= \sum_{\lambda_A \lambda_B} p(\lambda_A \lambda_B) p(b|y, \lambda_B) \rho_{\lambda_A}^A, \end{aligned} \quad (\text{A2})$$

where, in the last line, we use the fact that the set of all measurements $\mathbf{N}^{\text{full}}$ is tomographically complete. This decomposition is a local hidden state model for the assemblage, so the state ρ^{AB} is unsteerable from B to A .

The reverse direction follows directly, if a bipartite state ρ^{AB} is unsteerable, by definition, the state assemblage $\{\{\tilde{\rho}_{b|y}^A = \text{Tr}[(1^A \otimes M_{b|y}^B) \rho^{AB}]\}_b\}_y$, $\forall \{M_{b|y}^B\}_b \in \mathbf{M}^{\text{full}}$ should admit a local hidden state model [5]:

$$\tilde{\rho}_{b|y}^A = \sum_{\lambda_A} p(\lambda_A) p(b|y, \lambda_A) \rho_{\lambda_A}^A, \quad (\text{A3})$$

with $\sum_{\lambda_A} p(\lambda_A) = \sum_b p(b|y, \lambda_A) = 1$, $p(\lambda_A), p(b|y, \lambda_A) \geq 0$, and $\rho_{\lambda_A}^A \geq 0$. Therefore, the operational statistics $p(ab|xy)$ it generates with all local POVMs $\{N_{a|x}^A\}_a \in \mathbf{N}^{\text{full}}$ and $\{M_{b|y}^B\}_b \in \mathbf{M}^{\text{full}}$ must be NCOM-realizable relative to $\mathcal{O}_{\text{attt}}(\mathcal{B}^{\text{full}})$, since one can define

$$p(ab|xy) := \text{Tr}[\tilde{\rho}_{b|y}^A N_{a|x}^A] = \sum_{\lambda_A} p(\lambda_A) p(b|y, \lambda_A) \text{Tr}[N_{a|x}^A \rho_{\lambda_A}^A],$$

where $\sum_b p(b|y, \lambda_A) = 1$ and satisfies the trivial operational identities $\mathcal{O}_{\text{triv}}(\mathbf{M}^{\text{full}})$, while one can define $p(a|x, \lambda_A) := \text{Tr}[N_{a|x}^A \rho_{\lambda_A}^A]$, which is a non-negative linear function of all $N_{a|x}^A \in \mathbf{N}^{\text{full}}$ that satisfies $\mathcal{O}_{\text{all}}(\mathbf{N}^{\text{full}})$.

Case (i)—If a bipartite state ρ^{AB} is classical relative to $\mathcal{O}_{\text{aattt}}(\mathcal{B}^{\text{full}})$, by Definition 3, the operational statistics $p(ab|xy)$ it generates with all local POVMs $\{N_{a|x}^A\}_a \in \mathbf{N}^{\text{full}}$ and $\{M_{b|y}^B\}_b \in \mathbf{M}^{\text{full}}$ should be NCOM-realizable relative to $\mathcal{O}_{\text{aattt}}(\mathcal{B}^{\text{full}})$ per Definition 1.

Specifically, its operational statistics $p(ab|xy)$ should be realizable by Eq. (A1), where the assumption of noncontextuality applied to $\mathcal{O}_{\text{all}}(\mathbf{N}^{\text{full}})$ and $\mathcal{O}_{\text{all}}(\mathbf{M}^{\text{full}})$ implies that the response function $p(a|x, \lambda_A)$ is a linear function of the effect $N_{a|x}^A$ and that $p(b|y, \lambda_B)$ is a linear function of $M_{b|y}^B$. The generalized Gleason's theorem then guarantees the existence of unique density operators $\rho_{\lambda_A}^A$ and $\sigma_{\lambda_B}^B$ such that we can write $p(a|x, \lambda_A) = \text{Tr}[\rho_{\lambda_A}^A N_{a|x}^A]$ and $p(b|y, \lambda_B) = \text{Tr}[\sigma_{\lambda_B}^B M_{b|y}^B]$. Plugging these into Eq. (A1), for every pair of POVMs $\{N_{a|x}^A\}_a$ and $\{M_{b|y}^B\}_b$, we have

$$\begin{aligned} &\text{Tr}[(N_{a|x}^A \otimes M_{b|y}^B) \rho^{AB}] \\ &= \sum_{\lambda_A \lambda_B} p(\lambda_A \lambda_B) \text{Tr}[N_{a|x}^A \rho_{\lambda_A}^A] \text{Tr}[M_{b|y}^B \sigma_{\lambda_B}^B] \\ \Rightarrow \rho^{AB} &= \sum_{\lambda_A \lambda_B} p(\lambda_A \lambda_B) \rho_{\lambda_A}^A \otimes \sigma_{\lambda_B}^B, \end{aligned} \quad (\text{A4})$$

where, in the last line, we used the fact that the set of all product measurements is tomographically complete for the space of quantum states. So the state is separable.

The reverse direction is straightforward, as one needs to show only that the operational statistics arising from a separable state $\rho^{AB} = \sum_{\lambda_A \lambda_B} p(\lambda_A \lambda_B) \rho_{\lambda_A}^A \otimes \sigma_{\lambda_B}^B$ along with all POVMs $\{N_{a|x}^A\}_a \in \mathbf{N}^{\text{full}}$ and $\{M_{b|y}^B\}_b \in \mathbf{M}^{\text{full}}$ are NCOM-realizable relative to $\mathcal{O}_{\text{aattt}}(\mathcal{B}^{\text{full}})$. This can be met by defining the response functions $p(a|x, \lambda_A) = \text{Tr}[\rho_{\lambda_A}^A N_{a|x}^A]$ and $p(b|y, \lambda_B) = \text{Tr}[\sigma_{\lambda_B}^B M_{b|y}^B]$, which clearly are linear maps from POVMs to their ontological representations and respect $\mathcal{O}_{\text{all}}(\mathbf{N}^{\text{full}})$ and $\mathcal{O}_{\text{all}}(\mathbf{M}^{\text{full}})$.

Now, we will prove case (v); case (vi) follows similarly.

Case (v)—If a bipartite state ρ^{AB} is classical relative to $\mathcal{O}_{\text{aattt}}(\mathcal{B}^{\text{full}})$, by Definition 3, the operational statistics $p(ab|xy)$ it generates with all local POVMs $\{N_{a|x}^A\}_a \in \mathbf{N}^{\text{full}}$ and $\{M_{b|y}^B\}_b \in \mathbf{M}^{\text{full}}$ should be NCOM-realizable relative to $\mathcal{O}_{\text{aattt}}(\mathcal{B}^{\text{full}})$ per Definition 1. Specifically, its operational statistics $p(ab|xy)$ should be realizable by Eq. (A1) and the assumption of noncontextuality applied to $\mathcal{O}_{\text{all}}(\mathbf{M}^{\text{full}})$ implies the existence of unique density operators $\sigma_{\lambda_B}^B$ such that we can write $p(b|y, \lambda_B) = \text{Tr}[\sigma_{\lambda_B}^B M_{b|y}^B]$. Moreover, the assumption of noncontextuality applied to $\mathcal{O}_{\text{all}}(\mathbf{N}^{\text{full}} \circ \mathbf{P})$ implies the existence of a Hermitian operator $H_{\lambda_B}^B$ such that $\sum_{\lambda_A} p(a|x, \lambda_A) p(\lambda_A \lambda_B) = \text{Tr}[(N_{a|x}^A \otimes H_{\lambda_B}^B) \rho^{AB}] = \text{Tr}[H_{\lambda_B}^B \tilde{\rho}_{a|x}^B] \geq 0$, where we use the definition of $\tilde{\rho}_{a|x}^B$ in

Eq. (7). Combining these facts, we have

$$\begin{aligned}
\text{Tr}[\tilde{\rho}_{a|x}^B M_{b|y}^B] &= \text{Tr}[(N_{a|x}^A \otimes M_{b|y}^B) \rho^{AB}] \\
&= \sum_{\lambda_A \lambda_B} p(\lambda_A \lambda_B) p(a|x, \lambda_A) p(b|y, \lambda_B) \\
&= \sum_{\lambda_B} \text{Tr}[H_{\lambda_B}^B \tilde{\rho}_{a|x}^B] \text{Tr}[M_{b|y}^B \sigma_{\lambda_B}^B] \\
\Rightarrow \tilde{\rho}_{a|x}^B &= \sum_{\lambda_B} \text{Tr}[H_{\lambda_B}^B \tilde{\rho}_{a|x}^B] \sigma_{\lambda_B}^B, \quad (\text{A5})
\end{aligned}$$

where, in the last line, we used the fact that the set of all measurements is tomographically complete for the space of quantum states. According to Theorem 2 in Ref. [37], decomposition in Eq. (A5) ensures that the assemblage $\{\{\tilde{\rho}_{a|x}^B\}_a\}_x$ is classical.

For the reverse direction, if the state assemblage $\{\{\tilde{\rho}_{a|x}^B := \text{Tr}[(N_{a|x}^A \otimes 1^B) \rho^{AB}]\}_a\}_x$ defined $\forall \{N_{a|x}^A\}_a \in \mathbf{N}^{\text{full}}$ is classical, according to Theorem 2 in Ref. [37], one can find a frame representation $\tilde{\rho}_{a|x}^B = \sum_{\lambda_B} \text{Tr}[H_{\lambda_B}^B \tilde{\rho}_{a|x}^B] \sigma_{\lambda_B}^B$ with $\sigma_{\lambda_B}^B \geq 0$ and $\text{Tr}[H_{\lambda_B}^B \tilde{\rho}_{a|x}^B] \geq 0$ for all λ_B .

Therefore, combining with all $\{M_{b|y}^B\}_b \in \mathbf{M}^{\text{full}}$, one has

$$\begin{aligned}
\text{Tr}[(N_{a|x}^A \otimes M_{b|y}^B) \rho^{AB}] &= \text{Tr}[\tilde{\rho}_{a|x}^B M_{b|y}^B] \\
&= \sum_{\lambda_B} \text{Tr}[H_{\lambda_B}^B \tilde{\rho}_{a|x}^B] \text{Tr}[M_{b|y}^B \sigma_{\lambda_B}^B] \\
&= \sum_{\lambda_B} \text{Tr}[(N_{a|x}^A \otimes H_{\lambda_B}^B) \rho^{AB}] \text{Tr}[M_{b|y}^B \sigma_{\lambda_B}^B] \\
&:= \sum_{\lambda_B} \text{Tr}[N_{a|x}^A \tau_{\lambda_B}^A] \text{Tr}[M_{b|y}^B \sigma_{\lambda_B}^B]. \quad (\text{A6})
\end{aligned}$$

In the last equality, we define $\tau_{\lambda_B}^A := \text{Tr}_B[(1^A \otimes H_{\lambda_B}^B) \rho^{AB}]$, where $\tau_{\lambda_B}^A \geq 0$ is ensured because $\text{Tr}[N_{a|x}^A \tau_{\lambda_B}^A] = \text{Tr}[H_{\lambda_B}^B \tilde{\rho}_{a|x}^B] \geq 0$ for all $\{N_{a|x}^A\}_a \in \mathbf{N}^{\text{full}}$. Hence, one can simply define $p(\lambda_A \lambda_B) = \text{Tr}[\tau_{\lambda_A}^A \delta_{\lambda_A \lambda_B}]$, $p(b|y, \lambda_B) = \text{Tr}[M_{b|y}^B \sigma_{\lambda_B}^B]$, and $p(a|x, \lambda_A) = \text{Tr}[N_{a|x}^A \tau_{\lambda_A}^A] / \text{Tr}[\tau_{\lambda_A}^A]$ [if $\text{Tr}[\tau_{\lambda_A}^A] = 0$, $p(a|x, \lambda_A)$ can be defined arbitrarily], respectively, as the ontological representation for each process in the ontological model.

Checking that this model satisfies every noncontextuality constraint in $\mathcal{O}_{\text{aatt}}(\mathcal{B}^{\text{full}})$ is then straightforward. The constraints in $\mathcal{O}_{\text{all}}(\mathbf{N}^{\text{full}})$ and $\mathcal{O}_{\text{all}}(\mathbf{M}^{\text{full}})$ hold trivially, since $\tau_{\lambda_B}^A, \sigma_{\lambda_B}^B \geq 0$. For $\mathcal{O}_{\text{all}}(\mathbf{N}^{\text{full}} \circ \mathbf{P})$, note that $\sum_{\lambda_A} p(\lambda_A \lambda_B) p(a|x, \lambda_A) = \sum_{\lambda_A} \delta_{\lambda_A \lambda_B} \text{Tr}[N_{a|x}^A \tau_{\lambda_A}^A] = \text{Tr}[(N_{a|x}^A \otimes H_{\lambda_B}^B) \rho^{AB}] = \text{Tr}[\tilde{\rho}_{a|x}^B H_{\lambda_B}^B] \geq 0$; hence, the composition map is linear in the assemblage $\{\{N_{a|x}^A \circ \mathbf{P}^{AB} = \tilde{\rho}_{a|x}^B\}_a\}_x$ for all $\{N_{a|x}^A\}_a \in \mathbf{N}^{\text{full}}$. Therefore, we have a valid ontological model that respects $\mathcal{O}_{\text{aatt}}(\mathcal{B}^{\text{full}})$. ■

APPENDIX B: DERIVING THE NONCONTEXTUALITY INEQUALITIES

Consider a Bell circuit $\mathcal{B} = (\mathbf{N}, \mathbf{M}, \mathbf{P})$ where Alice and Bob each perform $\Delta_{\mathbf{N}}$ and $\Delta_{\mathbf{M}}$ different o -outcome measurements, respectively. One can visualize the operational statistics $p(ab|xy)$ as a point $\mathbf{p} := \{p(ab|xy)\}_{a,b,x,y} \in \mathbb{R}^{o^2 \Delta_{\mathbf{N}} \Delta_{\mathbf{M}}}$. The set of all the operational statistics that are NCOM-realizable relative to $\mathcal{O}_{\text{aatt}}(\mathcal{B})$ forms a polytope in this space, which we denote as $\mathcal{NC}_{\text{aatt}}(\mathcal{B})$.

As stated in the main text (and illustrated by Example 1), a set of operational identities $\mathcal{O}_{\text{all}}(\mathbf{N})$ can be represented by a geometric object—specifically, a polytope [78]. Concretely, each operational identity is a set of coefficients $\mathcal{O}_{\text{all}}(\mathbf{N}) = \{\{\alpha_{a,x}\}_{a,x}\}$, so one can associate to it a polytope P whose vertices are indexed by $\{\vec{t}_{a,x}\}$ satisfying all and only the equalities $\sum_{a,x} \alpha_{a,x} \vec{t}_{a,x} = \vec{0}$ and no other equalities. Conversely, given a polytope P with vertices indexed by a, x , one can pick an arbitrary coordinate system to represent these vertices as vector $\vec{t}_{a,x}$ and define the set of operational identities via the equalities they satisfy: $\mathcal{O}_{\text{all}}(\mathbf{N}) = \{\{\alpha_{a,x}\} | \sum_{a,x} \alpha_{a,x} \vec{t}_{a,x} = \vec{0}\}$. (An explicit method for computing these coefficients from a given set of vectors is given in Appendix A in Ref. [79].) This duality allows us to visualize operational constraints in purely geometric terms (as we did in the main text).

The extreme points $\vec{v} \in \mathcal{NC}_{\text{aatt}}(\mathcal{B})$ are those that can be obtained from the extreme operational statistics $p(ab|xy) = \sum_{\lambda_A \lambda_B} p(\lambda_A \lambda_B) p(a|x, \lambda_A) p(b|y, \lambda_B)$ when $p(\lambda_A \lambda_B)$ and when the response functions $\{p(a|x, \lambda_A)\}_{a,x}$ and $\{p(b|y, \lambda_B)\}_{b,y}$ are also extremal within the *noncontextual-measurement-assignment polytope* [42,80]. The noncontextual-measurement-assignment polytope $\mathbb{P}_{\mathbf{N}}$ is formally defined as a collection of points $\vec{u} := \{p(a|x)\}_{a,x}$ such that

$$(i) \quad p(a|x) \geq 0 \quad \forall a, x, \quad (\text{B1a})$$

$$(ii) \quad \sum_a p(a|x) = 1 \quad \forall x, \quad (\text{B1b})$$

$$(iii) \quad \sum_{a,x} \alpha_{a,x} p(a|x) = 0 \quad \forall \{\vec{\alpha}\} \in \mathcal{O}_{\text{all}}(\mathbf{N}). \quad (\text{B1c})$$

Similarly, one can define the *measurement-assignment polytope* $\mathbb{D}_{\mathbf{N}}$ as the collection of points $\vec{u} := \{p(a|x)\}_{a,x}$ satisfying only Eqs. (B1a) and (B1b) above.

Therefore, all extreme points $\vec{v} \in \mathcal{NC}_{\text{aatt}}(\mathcal{B})$ can be obtained from the extreme points $\vec{u} \in \mathbb{P}_{\mathbf{N}}$ and $\vec{w} \in \mathbb{P}_{\mathbf{M}}$ via $\vec{v} = \vec{u} \otimes \vec{w}$, and, analogously, all extreme points $\vec{v} \in \mathcal{NC}_{\text{tatt}}(\mathcal{B})$ can be obtained from the extreme points $\vec{u} \in \mathbb{D}_{\mathbf{N}}$ and $\vec{w} \in \mathbb{P}_{\mathbf{M}}$ via $\vec{v} = \vec{u} \otimes \vec{w}$.

It is important to note that the noncontextual-measurement-assignment polytope $\mathbb{P}_N$ is a polytope in the space of conditional probability vectors $\{p(a|x)\}_{a,x} \in \mathbb{R}^{o\Delta_N}$ constrained by the normalization conditions in Eq. (B1b) and ontological identities in Eq. (B1c). Consequently, the affine dimension of $\mathbb{P}_N$ is $K - 1$, bounded solely by the operator space dimension K that the set of effects in N spans (assuming the effects in N are informationally complete for a D -dimensional Hilbert space, then $K = D^2$). More importantly, since $\mathbb{P}_N$ has only $o\Delta_N$ inequality constraints in Eq. (B1a), the number of vertices in $\mathbb{P}_N$ is upper bounded by a polynomial in $o\Delta_N$ as per the dual form of the upper bound theorem [81,82]. In contrast, the local response functions in the Bell test define the measurement-assignment polytope $\mathbb{D}_N$ based solely on the positivity and normalization constraints in Eq. (B1). This polytope lives in an $(o - 1)\Delta_N$ -dimensional affine subspace of $\mathbb{R}^{o\Delta_N}$ and has o^{Δ_N} deterministic vertices.

Similarly, one can verify that the noncontextual polytope $\mathcal{NC}_{\text{aattt}}(\mathcal{B})$ resides in a $(D^4 - 1)$ -dimensional subspace of the vector space $\mathbb{R}^{o^2\Delta_N\Delta_M}$ (assuming the products of effects in N and M are informationally complete for the D^2 -dimensional Hilbert space), which is independent of the number of outcomes o and settings Δ_N, Δ_M . Moreover, the number of vertices in $\mathcal{NC}_{\text{aattt}}(\mathcal{B})$ scales only polynomially with $o^2\Delta_N\Delta_M$, since they are determined by the number of vertices in $\mathbb{P}_N$ and $\mathbb{P}_M$. By comparison, the Bell polytope lives in an affine subspace of dimension $(o - 1)(\Delta_N + \Delta_M) + (o - 1)^2\Delta_N\Delta_M$ [3], and its vertex count scales exponentially as $o^{\Delta_N + \Delta_M}$. As a result, in contrast with characterizing the local polytope, characterizing $\mathcal{NC}_{\text{aattt}}(\mathcal{B})$ remains computationally feasible, even when the number of settings for each party is very large.

For simple cases, one can completely characterize the polytope $\mathcal{NC}_{\text{aattt}}(\mathcal{B})$ by enumerating all the extreme points $\{\vec{v} = \vec{u} \otimes \vec{w}\}$ with $\vec{u}$ being an extreme point of $\mathbb{P}_N$ and $\vec{w}$ being an extreme point of $\mathbb{P}_M$. Taking the class of (octahedron, cube)-Bell circuits, for instance: For multi-measurements $N = \{\{N_{a|x}^A\}_{a,x}\}$ and $M = \{\{M_{b|y}^B\}_{b,y}\}$, where $a, b \in [0, 1]$, $x \in [0, 1, 2]$, and $y \in [0, 1, 2, 3]$, having cubic and octahedral symmetry, respectively, we find that the $\mathcal{NC}_{\text{aattt}}(\mathcal{B})$ polytope lives in a 15-dimensional subspace of the full 48-dimensional probability space ($d = o^2\Delta_N\Delta_M = 48$), where the spanned dimension is reduced by equality constraints coming from normalization and noncontextuality.

In the 15-dimensional subspace, we find that the noncontextual polytope has 336 facet inequalities in total, which can be classified into three equivalence classes under permutations of values of the inputs and of the outputs that respect the cubic and octahedral symmetries, respectively. We, therefore, explicitly list a single representative inequality for each class, as in Table II. Importantly, one can show that, for a two-qubit isotropic

state ρ_{iso}^v with $v > (1/\sqrt{3})$, the last inequality can be violated using the following local qubit measurements that respect operational identities $\mathcal{O}_{\text{aattt}}(\text{Bell}_{oc})$ in the class of (octahedron, cube)-Bell circuits: $N_{a|x}^A = \frac{1}{2}[\mathbb{1} + (-1)^a \hat{n}_x \cdot \vec{\sigma}]$ and $M_{b|y}^B = \frac{1}{2}[\mathbb{1} + (-1)^b \hat{n}_y \cdot \vec{\sigma}]$, where

$$\begin{aligned} \hat{n}_0 &= [1, 0, 0], & \hat{n}_1 &= [0, 1, 0], & \hat{n}_2 &= [0, 0, 1], \\ \hat{m}_0 &= \frac{1}{\sqrt{3}}[1, 1, 1], & \hat{m}_1 &= \frac{1}{\sqrt{3}}[1, 1, -1], \\ \hat{m}_2 &= \frac{1}{\sqrt{3}}[1, -1, 1], & \hat{m}_3 &= \frac{1}{\sqrt{3}}[1, -1, -1]. \end{aligned} \quad (\text{B2})$$

In the Bloch ball, these have the geometries depicted in Fig. 3(a).

In a more complex scenario, the computational complexity of the facet enumeration problem increases with the number of measurement settings. Moreover, unlike the local polytope, where vertices are rational vectors (deterministic strategies), the noncontextual polytope generally contains vertices with irrational components, arising from the presence of irrational components in the operational identities. While this complicates the full characterization the polytope $\mathcal{NC}_{\text{aattt}}(\mathcal{B})$ as most standard facet-enumeration tools assume rational vertices (so they can perform Fourier-Motzkin elimination efficiently), the lower dimensionality and the moderate number of vertices of the noncontextual polytope $\mathcal{NC}_{\text{aattt}}(\mathcal{B})$ (relative to the local polytope) render *polytope membership certification* efficient. For scenarios with hundreds of settings, the membership problem can be solved exactly via standard linear programming, avoiding the need for iterative algorithms such as Frank-Wolfe [45] or column generation.

We now turn to the problem of deciding membership of a given correlation $p(ab|xy)$ in the noncontextual polytope $\mathcal{NC}_{\text{aattt}}(\mathcal{B})$ and how this problem compares with that of deciding membership in the local polytope.

Consider a Bell circuit $\mathcal{B}$ with setting variable cardinalities Δ_N and Δ_M and outcome variable cardinality o again. We can represent the k extreme points $\vec{v} \in \mathcal{NC}_{\text{aattt}}(\mathcal{B}) \subset \mathbb{R}^{o^2\Delta_N\Delta_M}$ as columns of a $o^2\Delta_N\Delta_M \times k$ matrix $\mathbf{T}$. A correlation $p(ab|xy)$, expressed as a probability vector $\mathbf{p} := \{p(ab|xy)\}_{ab,xy} \in \mathbb{R}^{o^2\Delta_N\Delta_M}$, lies within the noncontextual polytope $\mathcal{NC}_{\text{aattt}}(\mathcal{B})$ if and only if

$$\begin{aligned} &\text{there exists some } \mathbf{x} \\ &\text{such that } \mathbf{T}\mathbf{x} = \mathbf{p}, \\ &\text{where } \mathbf{x} \geq \mathbf{0}, \end{aligned} \quad (\text{B3})$$

with $\mathbf{x} \in \mathbb{R}^k$. We can also write up the dual LP as

$$\begin{aligned} & \min_{\mathbf{s}} \mathbf{s}^T \mathbf{p} \\ & \text{such that } \mathbf{0} \leq \mathbf{s}^T \mathbf{T} \leq \mathbf{1} \end{aligned} \quad (\text{B4})$$

with $\mathbf{s} \in \mathbb{R}^{o^2 \Delta_N \Delta_M}$.

If a correlation $\mathbf{p} = \{p(ab|xy)\}_{abxy}$ does not lie in the polytope, then this dual linear program will always find a real vector $\mathbf{s} \in \mathbb{R}^{o^2 \Delta_N \Delta_M}$ satisfying the condition $\mathbf{s}^T \mathbf{p} < 0$. This real vector $\mathbf{s}$ can then be used to define a facet-defining noncontextuality inequality that $\mathbf{p}$ violates.

Crucially, as noted earlier, the noncontextual polytope resides within a space of dimension $D^4 - 1 \ll o^2 \Delta_N \Delta_M$, where D is the Hilbert space dimension, which is much smaller than the dimension of the space in which the local polytope resides. Consequently, the linear program defined above can always be recast as an equivalent problem in its minimal affine representation (i.e., using reduced coordinates of dimension $D^4 - 1$). As demonstrated in our Example 4, the membership problem above can be efficiently solved even when the full probability space dimension $d = o^2 \Delta_N \Delta_M \gg 10^5$.

The inequality we gave in Eq. (17) is a facet inequality for the noncontextual polytope $\mathcal{NC}_{\text{aattt}}(\text{Bell}_{\text{id}})$, where Bell_{id} denotes the class of (icosahedron, dodecahedron)-Bell circuits, considered in Example 3. For an isotropic state ρ_{iso}^v with $v > \sqrt{[(1 + \phi^2)/3\phi^4]}$, the inequality can be violated using the following measurements [33,83] that have icosahedral and dodecahedral symmetries, respectively: $N_{a|x}^A = \frac{1}{2}[\mathbb{1} + (-1)^a \hat{n}_x \cdot \vec{\sigma}]$ and $M_{b|y}^B = \frac{1}{2}[\mathbb{1} + (-1)^b \hat{m}_y \cdot \vec{\sigma}]$ with $\hat{n}_x = (1/\sqrt{1 + \phi^2})\vec{n}_x$ and $\hat{m}_y = (1/\sqrt{3})\vec{m}_y$:

$$\begin{aligned} \vec{n}_0 &= [0, 1, \phi], & \vec{n}_1 &= [0, 1, -\phi], & \vec{n}_2 &= [1, \phi, 0], \\ \vec{n}_3 &= [1, -\phi, 0], & \vec{n}_4 &= [\phi, 0, 1], & \vec{n}_5 &= [\phi, 0, -1], \\ \vec{m}_0 &= [-1, -1, -1], & \vec{m}_1 &= \left[-\phi, \frac{1}{\phi}, 0\right], & \vec{m}_2 &= \left[-\phi, -\frac{1}{\phi}, 0\right], \\ \vec{m}_3 &= [-1, -1, 1], & \vec{m}_4 &= \left[0, -\phi, \frac{1}{\phi}\right], & \vec{m}_5 &= [1, -1, -1], \\ \vec{m}_6 &= \left[0, -\phi, -\frac{1}{\phi}\right], & \vec{m}_7 &= [-1, 1, -1], & \vec{m}_8 &= \left[-\frac{1}{\phi}, 0, -\phi\right], \\ \vec{m}_9 &= \left[\frac{1}{\phi}, 0, -\phi\right]. \end{aligned} \quad (\text{B5})$$

In the Bloch ball, these have the geometries depicted in Fig. 3(b).

1. Deriving the facet inequalities from the noncontextual polytope $\mathcal{NC}_{\text{tattt}}$

As noted in Sec. III, characterizing the facet inequalities of the noncontextual polytope $\mathcal{NC}_{\text{aattt}}(\mathcal{B})$, where one uses the nontrivial sets of operational identities for both Alice's and Bob's multi-measurements, offers distinct advantages

for entanglement certification over existing protocols. For instance, unlike quantum tomography, it does not require a prior characterization of the measurement. Moreover, it improves upon the Bell-inequality-based approach because, in principle, it can certify any entangled state.

A similar argument applies to characterizing the facet inequalities of the noncontextual polytope $\mathcal{NC}_{\text{tattt}}(\mathcal{B})$ for $A \rightarrow B$ steering, where one uses only the nontrivial sets of operational identities for Bob's multi-measurement. Our noncontextuality-based approach does not require characterization of the measurement devices on either side. In contrast, standard steering inequalities require full tomography of the measurement devices on one side while remaining fully device independent on the other side (often termed one-sided device independent). Moreover, while violating a Bell inequality is sufficient for device-independent steering certification [5], the noncontextuality inequalities derived from our noncontextual polytope $\mathcal{NC}_{\text{tattt}}(\mathcal{B})$ can outperform all known Bell inequalities, as demonstrated in Table VII.

In the main text, we have already presented two classes of Bell circuits (Examples 1 and 2) where $\mathcal{NC}_{\text{tattt}}(\mathcal{B}) = \mathcal{NC}_{\text{aattt}}(\mathcal{B})$. To derive more generic noncontextual inequalities for certifying quantum steering, one needs to construct the corresponding noncontextual polytope $\mathcal{NC}_{\text{tattt}}(\mathcal{B})$ by enumerating all its extreme points $\vec{v}$ from the extreme points $\vec{u} \in \mathbb{D}_N$ and $\vec{w} \in \mathbb{P}_M$ via $\vec{v} = \vec{u} \otimes \vec{w}$. While the subsequent analysis of $\mathcal{NC}_{\text{tattt}}(\mathcal{B})$ remains the same as that of $\mathcal{NC}_{\text{aattt}}(\mathcal{B})$ introduced in Appendix B, the linear program for its characterization is more complex in practice, as both the dimension and vertex complexity of the measurement-assignment polytope $\mathbb{D}_N$ are far greater than those of the noncontextual measurement-assignment polytope $\mathbb{P}_N$ (see Appendix B).

Consequently, to reduce the numerical complexity, we restrict our analysis to cases in which only the number of settings on Bob's multi-measurement increases, while

TABLE VII. Critical visibility v of the two-qubit isotropic state ρ_{iso}^v that can be certified as steerable using noncontextuality inequalities derived from the noncontextual polytope $\mathcal{NC}_{\text{tattt}}(\mathcal{B})$. Note that the result for row 3 differs from the corresponding entry in Table III, which uses inequalities derived from the noncontextual polytope $\mathcal{NC}_{\text{aattt}}(\mathcal{B})$ for entanglement certification.

No. of Alice's settings	No. of Bob's settings	Visibility v
2 (square)	2 (square)	≈ 0.7071
3 (octahedron)	4 (cube)	≈ 0.5773
6 (icosahedron)	10 (dodecahedron)	≈ 0.6787
6 (icosahedron)	40 (Goldberg $\nu = 2$)	≈ 0.5655
6 (icosahedron)	90 (Goldberg $\nu = 3$)	≈ 0.5493
6 (icosahedron)	160 (Goldberg $\nu = 4$)	≈ 0.5445
6 (icosahedron)	250 (Goldberg $\nu = 5$)	≈ 0.5425
6 (icosahedron)	360 (Goldberg $\nu = 6$)	≈ 0.5414
6 (icosahedron)	490 (Goldberg $\nu = 7$)	≈ 0.5408

Alice employs a fixed multi-measurement with six measurement settings. It is important to note that, because we fix the cardinality of Alice's setting variable, the resulting critical visibility v will not converge to the fundamental limit $v = 1/2$ as the cardinality of Bob's setting variable approaches infinity; however, our approach is shown to already outperform all known Bell-inequality-based steering certification [45].

We note again that our test requires only the standard assumption for testing noncontextuality in a prepare-measure circuit on Bob's side, namely, tomographic completeness of Bob's measurements and steered states on B . This suggests that our protocol could offer practical advantages relative to standard steering-certification methods. A more thorough analysis and comparison along these lines is left for future work.

APPENDIX C: DEALING WITH INEXACT OPERATIONAL IDENTITIES

1. Self-consistent tomography of raw data: Frequencies to probabilities

Experiments do not yield probabilities directly. Rather, they yield relative frequencies which, under ideal circumstances (of the samples being independent and identically distributed), would converge to the true underlying probabilities in the infinite-run limit. Interpreting relative frequencies as probabilities can lead to a misinterpretation of experimental results. For instance, in a typical Bell experiment, interpreting frequencies as probabilities without accounting for statistical fluctuations can result in an apparent violation of the no-signaling principle [35].

This issue is equally relevant in our experiment. In principle, statistical fluctuations in the relative frequencies could lead to a situation wherein interpreting these as probabilities yields an apparent violation of the predictions of quantum theory (for a given choice of dimensionality of systems). To address this issue, we adapted a well-motivated regularization method, similar to those used in previous works [32,35,54], to convert the experimentally observed relative frequencies into *estimates* of the true distribution. Specifically, we denote the frequency obtained directly from raw data in a bipartite Bell circuit as $f_{ab|xy}$ and its variance as $\Delta f_{ab|xy}$. We then consider the weighted χ^2 optimization, where

$$\chi^2 := \min_{p(ab|xy)} \sum_{abxy} \left| \frac{f_{ab|xy} - p(ab|xy)}{\Delta f_{ab|xy}} \right|^2. \quad (\text{C1})$$

Note that one could also replace the χ^2 function with the L^2 function or the KL divergence, as discussed in Ref. [35].

In our data analysis, we considered two types of theoretical models to fit the relative frequencies. The first assumes quantum mechanics with a fixed Hilbert space dimension D (hence, operator space dimension D^2) [84],

where $p(ab|xy) = \text{tr}[(N_{a|x}^A \otimes M_{b|y}^B) \rho^{AB}]$. The second allows for a realization in an arbitrary generalized probabilistic theory (GPT) with fixed GPT vector space dimension D^2 , where $p(ab|xy) = N_{a|x}^A \otimes M_{b|y}^B \cdot G^{AB}$. In this case, $N_{a|x}^A$, $M_{b|y}^B$, and G^{AB} are GPT vectors corresponding to unipartite effects on A and B and the bipartite state on AB , respectively.

As the two regularization methods yielded no statistically significant differences, we reported the quantum realization in the main text. We note that the associated optimization problem cannot be solved in a single step; instead, we employ a seesaw algorithm to iteratively optimize the bipartite state and local effects, i.e.,

- (1) given an initial guess for $\mathbf{N}$, $\mathbf{M}$, and ρ^{AB} ,
- (2) fix $\mathbf{N}$ and $\mathbf{M}$ and update states ρ^{AB} to be $\arg \min_{\rho^{AB}} \sum_{abxy} |f_{ab|xy} - \text{Tr}[(N_{a|x}^A \otimes M_{b|y}^B) \rho^{AB}]| / \Delta f_{ab|xy}|^2$,
- (3) fix $\mathbf{N}$ and state ρ^{AB} and update $\mathbf{M}$ to be $\arg \min_{\mathbf{M}} \sum_{abxy} |f_{ab|xy} - \text{Tr}[(N_{a|x}^A \otimes M_{b|y}^B) \rho^{AB}]| / \Delta f_{ab|xy}|^2$,
- (4) fix $\mathbf{M}$ and state ρ^{AB} and update $\mathbf{N}$ to be $\arg \min_{\mathbf{N}} \sum_{abxy} |f_{ab|xy} - \text{Tr}[(N_{a|x}^A \otimes M_{b|y}^B) \rho^{AB}]| / \Delta f_{ab|xy}|^2$, and
- (5) repeat steps 2–4 until it converges relative to a certain threshold.

Each step of the optimization above can be expressed as a quadratic program, which could be efficiently solved computationally. However, since the bipartite statistics $p(ab|xy)$ are invariant under invertible local transformations (the tomographic gauge freedom noted in Ref. [31] and described in the main text), the optimization does not yield a unique solution for the bipartite state ρ^{AB} and the sets of local effects $\mathbf{N}$ and $\mathbf{M}$.

These regularization procedures are closely related to GPT tomography [32] and gate-set tomography [31]. Consequently, any tools developed for such bootstrap tomography schemes can be directly applied here as well. However, we do not use this self-consistent tomography naively to certify entanglement, namely, by performing a separability test on the bipartite state obtained from the tomography scheme. We avoid this approach because this output ρ^{AB} from a self-consistent tomography is gauge dependent and would require an optimization over the choice of gauge. Instead, our entanglement-certification technique evaluates noncontextuality inequalities on the regularized statistics, and, since these statistics are gauge-independent quantities, no gauge optimization is required.

Furthermore, since characterizing the bipartite state is not our objective, a distinct yet equivalent approach can be taken to regularize the raw statistics. Specifically, rather than employing a seesaw algorithm to iteratively optimize the bipartite state on AB and the sets of local effects on A and on B , one can instead just iteratively

optimize the steering assemblage on A , $\{\{\tilde{\rho}_{b|y}^A\}\}_{b,y}$, where $\tilde{\rho}_{b|y}^A = \text{Tr}_B[(\mathbb{1}^A \otimes M_{b|y}^{B,\text{pri}})\rho^{AB}]$ and the set of local effects on A and B (or, equivalently, the steering assemblage and set of local effects on B). The GPT version of this regularization process is precisely the one studied in the prepare-measure circuit in Ref. [32].

It is worth noting again that, because of the tomographic gauge freedom described in the main text, there is a nonuniqueness in the output of the fitting procedure that yields the bipartite state and the primary measurements and, hence, a nonuniqueness in the secondary measurements as well. However, under the assumptions of tomographic completeness, the operational identities and the statistics are manifestly gauge independent. Because the noncontextuality inequalities are derived from those gauge-independent operational identities, and it is the gauge-independent statistics on which the noncontextuality inequalities are tested, our noncontextuality tests are gauge independent.

2. Secondary measurements and secondary statistics

As in the main text, we denote the output of the regularization step as the *primary statistics*, denoted as

$$p^{\text{pri}}(ab|xy) = \text{Tr}[N_{a|x}^{A,\text{pri}} \otimes M_{b|y}^{B,\text{pri}} \rho^{AB}], \quad (\text{C2})$$

where $\{N_{a|x}^{A,\text{pri}}\}$ and $\{M_{b|y}^{B,\text{pri}}\}$ are referred to as the primary measurements.

Here, we detail the method of secondary procedures. This method takes as input a set of processes that are obtained by fitting to the relative frequencies (i.e., self-consistent tomography) and that do not satisfy the required operational identities used in deriving the inequality test and outputs a set of processes in their convex hull that satisfy these identities exactly.

Consider again a multi-measurement on A , $\mathbf{N} = \{\{N_{a|x}^A\}_a\}_x$, satisfying operational identities $\mathcal{O}_{\text{all}}(\mathbf{N}) = \{\{\alpha_{ax}\}_{ax} | \sum_{ax} \alpha_{ax} N_{a|x}^A = \mathbb{0}\}$, describing the ideal identities that one is targeting in the experiment. Any set of operational identities of this form can be reduced to a finite *generating set* [79], meaning that all other operational identities can be deduced from this generating set, and NCOM-realizability relative to this generating set is equivalent to NCOM-realizability relative to the full set. Therefore, we consider generating only sets of operational identities of minimal cardinality. This cardinality, which we denote by $|T|$, coincides with the number of linearly independent operational identities in $\mathcal{O}_{\text{all}}(\mathbf{N})$. A given operational identity can be associated to a vector $\vec{\alpha} = (\alpha_{ax})_{ax}$. Consequently, a minimal generating set can be associated to a set of vectors $\{\vec{\alpha}^{(t)}\}_{t \in T}$. Similar comments apply to the measurements on B .

The procedure for finding the best-fit probabilities to the relative frequencies obtained in the experiment [35,54], i.e., the regularization procedure, is detailed in Appendix C 1.

As stated above, we refer to the measurements that come out of the regularization procedure as *primary measurements* and denote them by $\mathbf{N}^{\text{pri}} = \{\{N_{a|x}^{A,\text{pri}}\}_a\}_x$ and $\mathbf{M}^{\text{pri}} = \{\{M_{b|y}^{B,\text{pri}}\}_b\}_y$, respectively.

In general, these measurements do not satisfy the targeted operational identities. Consider the case of the measurements on A . In general, $\sum_{ax} \alpha_{ax}^{(t)} N_{a|x}^{A,\text{pri}} \neq \mathbb{0}$ for one or more $\vec{\alpha}^{(t)}$ in a minimal generating set for the targeted set of operational identities, $\mathcal{O}_{\text{all}}(\mathbf{N})$. However, it is possible to find a set of *secondary measurements*, denoted $\mathbf{N}^{\text{sec}} = \{\{N_{a|x}^{A,\text{sec}}\}_a\}_x$, such that each of the secondary effects lies in the convex hull of the set of primary effects $\mathbf{N}^{\text{pri}}$ and where the secondary effects exactly satisfy the operational identities $\mathcal{O}_{\text{all}}(\mathbf{N})$, i.e.,

$$\begin{aligned} N_{a|x}^{A,\text{sec}} &= \sum_{a',x'} u_{a',x'}^{a,x} N_{a'|x'}^{A,\text{pri}} \quad \forall a, x, \\ \sum_{ax} \alpha_{ax}^{(t)} N_{a|x}^{A,\text{sec}} &= \mathbb{0} \quad \forall t \in T, \\ \sum_a N_{a|x}^{A,\text{sec}} &= \mathbb{1}^A \quad \forall x, \\ \sum_{a',x'} u_{a',x'}^{a,x} &= 1, \quad u_{a',x'}^{a,x} \geq 0 \quad \forall a, x, a', x'. \end{aligned} \quad (\text{C3})$$

In particular, we solve an optimization problem that selects the closest approximation to the primary measurements while satisfying these constraints, by maximizing the objective function $C_N = 1/(o\Delta_N) \sum_{a,x} u_{a,x}^{a,x}$, that is, through the following linear programming:

$$\begin{aligned} \max C_N &= \frac{1}{o\Delta_N} \sum_{ax} u_{ax}^{ax} \\ \text{such that } \sum_{axa',x'} u_{a',x'}^{a,x} N_{a'|x'}^{A,\text{pri}} \alpha_{ax}^{(t)} &= \mathbb{0} \quad \forall t \in T, \\ \sum_{aa',x'} u_{a',x'}^{a,x} N_{a'|x'}^{A,\text{pri}} &= \mathbb{1}^A \quad \forall x, \\ \sum_{a',x'} u_{a',x'}^{a,x} &= 1, \quad u_{a',x'}^{a,x} \geq 0 \quad \forall a, x, a', x', \end{aligned} \quad (\text{C4})$$

where the value of C_N will be close to 1 when the primary measurements are close to the ideal measurements being targeted. We define the secondary measurements on B , denoted $\mathbf{M}^{\text{sec}}$, similarly. That is, we define $M_{b|y}^{B,\text{sec}} = \sum_{b',y'} v_{b',y'}^{b,y} M_{b'|y'}^{B,\text{pri}}$ and maximize the objective function $C_M = \sum_{b,y} [1/(o\Delta_M)] v_{b,y}^{b,y}$ while satisfying the analogs of the constraints in Eq. (C3).

With the tomographic completeness condition already made in the self-consistent tomography step, on

Alice's side, the assemblage on A induced by Bob's primary measurement, i.e., the set of state $\{\tilde{\rho}_{b|y}^A := \text{Tr}_B[(\mathbb{1}^A \otimes M_{b|y}^{B,\text{pri}})\rho^{AB}]\}_{b,y}$, will be tomographically complete (relative to the operator subspace spanned by $\mathbf{N}^{\text{pri}}$) so that the operational identities of $\mathbf{N}^{\text{pri}}$ can be inferred from the operational statistics [85]. And similarly, with the roles of $\mathbf{N}^{\text{pri}}$ and $\mathbf{M}^{\text{pri}}$ reversed. This condition is significantly milder than those requiring a prior characterization of the measurement devices, but it already allows us to infer the coefficients $u_{a',x'}^{a,x}$ purely from the *primary statistics*, using the following linear program:

$$\begin{aligned} \max C_N &= \frac{1}{o\Delta_N} \sum_{ax} u_{ax}^{ax} \\ \text{such that } \sum_{axa'x'} p^{\text{pri}}(a'b'|x'y') u_{a',x'}^{a,x} \alpha_{a,x}^{(t)} &= 0 \quad \forall b', y', \forall t \in T, \\ \sum_{aa'x'} p^{\text{pri}}(a'b'|x'y') u_{a',x'}^{a,x} &= p^{\text{pri}}(b'|y') \quad \forall x, b', y', \\ \sum_{a'x'} u_{a',x'}^{a,x} &= 1, \quad u_{a',x'}^{a,x} \geq 0 \quad \forall a, x, a', x'. \end{aligned} \quad (\text{C5})$$

We note that the second condition resembles the normalization condition in Eq. (C4), and here $p^{\text{pri}}(b'|y') := \sum_{a'} p^{\text{pri}}(a'b'|x'y')$, which is independent of x' since the primary statistics are defined by Eq. (C2), and this form of statistics is necessarily nonsignaling.

One can obtain the coefficients $v_{b',y'}^{b,y}$ in an analogous fashion. In the end, the *statistics* defined by the secondary measurements, denoted $p^{\text{sec}}(ab|xy)$ and termed the *secondary statistics*, are given by

$$\begin{aligned} p^{\text{sec}}(ab|xy) &= \text{Tr}[N_{a|x}^{A,\text{sec}} \otimes M_{b|y}^{B,\text{sec}} \rho^{AB}] \\ &= \sum_{a'b'x'y'} v_{b',y'}^{b,y} u_{a',x'}^{a,x} p^{\text{pri}}(a'b'|x'y'), \end{aligned} \quad (\text{C6})$$

where $p^{\text{pri}}(a'b'|x'y')$ are the statistics defined by the set of primary measurements.

3. Proof of Proposition 2

Proof. We prove the two directions explicitly.

First suppose that the operational statistics $p(a, b|x, y)$ from the ideal Bell circuit $\mathcal{B} = (\mathbf{N}, \mathbf{M}, \rho^{AB})$ is NCOM-realizable relative to $\mathcal{O}_{\text{aattt}}(\mathcal{B})$. Then there exists an ontic state space $\Lambda_A \times \Lambda_B$, a distribution $p(\lambda_A \lambda_B)$, and local response functions $p(a|x, \lambda_A)$ and $p(b|y, \lambda_B)$ that reproduce $p(a, b|x, y)$ and respect the operational identities of $\mathcal{O}_{\text{all}}(\mathbf{N})$ and $\mathcal{O}_{\text{all}}(\mathbf{M})$ as in Eqs. (4) and (6).

One can then define response functions for the inefficient multi-measurements $\tilde{\mathbf{N}}$ on Alice by

$$p_\epsilon(a|x, \lambda_A) = \epsilon_{a|x} p(a|x, \lambda_A), \quad (\text{C7a})$$

$$p_\epsilon(\emptyset|x, \lambda_A) = \sum_a (1 - \epsilon_{a|x}) p(a|x, \lambda_A), \quad (\text{C7b})$$

and analogously for Bob. Since all efficiencies $\epsilon_{a|x}$ and $\epsilon_{b|y}$ are strictly positive and less than or equal to 1, these response functions are nonnegative and normalized. One can easily verify that they also respect the operational identities in $\mathcal{O}_{\text{all}}(\tilde{\mathbf{N}})$ and $\mathcal{O}_{\text{all}}(\tilde{\mathbf{M}})$. To see this, consider an arbitrary operational identity among Alice's measurement effects:

$$\sum_{\alpha, x} \tilde{\gamma}_{\alpha, x} \tilde{N}_{a|x}^A = 0, \quad (\text{C8})$$

where α ranges over the original outcome set together with the null outcome $\emptyset$. Using Eqs. (29) and (30), this identity can be equivalently rewritten as

$$\sum_{\alpha, x} [\epsilon_{a|x} \tilde{\gamma}_{\alpha, x} + (1 - \epsilon_{a|x}) \tilde{\gamma}_{\emptyset, x}] N_{a|x}^A = 0. \quad (\text{C9})$$

Thus, every identity among the attenuated effects and null effect pulls back to an identity among the ideal effects. By assumption, the ideal response functions $p(a|x, \lambda_A)$ respect all operational identities in $\mathcal{O}_{\text{all}}(\mathbf{N})$, so we have

$$\sum_{\alpha, x} [\epsilon_{a|x} \tilde{\gamma}_{\alpha, x} + (1 - \epsilon_{a|x}) \tilde{\gamma}_{\emptyset, x}] p(a|x, \lambda_A) = 0. \quad (\text{C10})$$

Using Eq. (C7), this is precisely

$$\sum_{\alpha, x} \tilde{\gamma}_{\alpha, x} p_\epsilon(a|x, \lambda_A) = 0. \quad (\text{C11})$$

Thus, the inefficient response functions respect all the operational identities in $\mathcal{O}_{\text{all}}(\tilde{\mathbf{N}})$. The same argument applies to Bob. Finally, it is straightforward to check that using the same distribution $p(\lambda_A \lambda_B)$, the inefficient response functions in Eq. (C7) reproduce the inefficient statistics. Hence, $p_\epsilon(\alpha, \beta|x, y)$ is NCOM-realizable relative to $\mathcal{O}_{\text{aattt}}(\mathcal{B}^\epsilon)$.

Conversely, suppose that $p_\epsilon(\alpha, \beta|x, y)$ from the inefficient Bell circuit $\mathcal{B}^\epsilon = (\tilde{\mathbf{N}}, \tilde{\mathbf{M}}, \rho^{AB})$ is NCOM-realizable relative to $\mathcal{O}_{\text{aattt}}(\mathcal{B}^\epsilon)$. Then there exist $p_\epsilon(\lambda_A \lambda_B)$ and local response functions $p_\epsilon(a|x, \lambda_A)$ and $p_\epsilon(b|y, \lambda_B)$ reproducing the inefficient statistics and respecting the operational identities of $\tilde{\mathbf{N}}$ and $\tilde{\mathbf{M}}$. Since all efficiencies $\epsilon_{a|x}$ and $\epsilon_{b|y}$ are strictly positive, we can define response functions for the ideal multi-measurement $\mathbf{N}$ and $\mathbf{M}$ by

$$p(a|x, \lambda_A) = \frac{p_\epsilon(a|x, \lambda_A)}{\epsilon_{a|x}}, \quad (\text{C12a})$$

$$p(b|y, \lambda_B) = \frac{p_\epsilon(b|y, \lambda_B)}{\epsilon_{b|y}}. \quad (\text{C12b})$$

These functions are clearly non-negative. They are normalized because the operational identity defining the null-outcome effect $\tilde{N}_{\emptyset|x}^A = \sum_a [(1 - \epsilon_{a|x})/\epsilon_{a|x}] \tilde{N}_{a|x}^A$ in Eqs. (29) and (30) implies

$$p_\epsilon(\emptyset|x, \lambda_A) = \sum_a \frac{1 - \epsilon_{a|x}}{\epsilon_{a|x}} p_\epsilon(a|x, \lambda_A). \quad (\text{C13})$$

Combining this with the normalization condition

$$\sum_a p_\epsilon(a|x, \lambda_A) + p_\epsilon(\emptyset|x, \lambda_A) = 1 \quad (\text{C14})$$

gives

$$\begin{aligned} 1 &= \sum_a p_\epsilon(a|x, \lambda_A) + \sum_a \frac{1 - \epsilon_{a|x}}{\epsilon_{a|x}} p_\epsilon(a|x, \lambda_A) \\ &= \sum_a \frac{p_\epsilon(a|x, \lambda_A)}{\epsilon_{a|x}} = \sum_a p(a|x, \lambda_A), \end{aligned} \quad (\text{C15})$$

where the first equality follows from Eqs. (C14) and (C13) and the last equality follows from Eq. (C12).

The reconstructed response functions also respect the operational identities in $\mathcal{O}_{\text{all}}(\mathbf{N})$ and $\mathcal{O}_{\text{all}}(\mathbf{M})$. This is shown as follows. First, note that

$$\sum_{a,x} \gamma_{a,x} N_{a|x}^A = 0 \Rightarrow \sum_{a,x} \frac{\gamma_{a,x}}{\epsilon_{a|x}} \tilde{N}_{a|x}^A = 0. \quad (\text{C16})$$

Since the inefficient response functions respect the identity $\sum_{a,x} (\gamma_{a,x}/\epsilon_{a|x}) p_\epsilon(a|x, \lambda_A) = 0$, it follows that

$$\sum_{a,x} \gamma_{a,x} p(a|x, \lambda_A) = 0. \quad (\text{C17})$$

The same argument holds for Bob.

Finally, it is straightforward to check that, using the same distribution $p_\epsilon(\lambda_A \lambda_B)$, the reconstructed ideal response functions in Eq. (C12) reproduce the ideal statistics. Hence, $p(a, b|x, y)$ is NCOM-realizable relative to $\mathcal{O}_{\text{aattt}}(\mathcal{B})$. ■

APPENDIX D: EXPERIMENTAL DETAILS

1. Generating two-qubit isotropic states experimentally

We generate polarization-entangled photons using type-II spontaneous parametric down-conversion inside a Sagnac interferometer consisting of a 404 nm continuous-wave diode laser and a 10 mm PPKTP crystal. The source aims to generate the Bell state $|\Phi^+\rangle = (1/\sqrt{2})(|HH\rangle + |VV\rangle)$ and is implemented with measured state fidelity greater than 97.5% at a rate of 40 000

TABLE VIII. Isotropic state generation.

Gate	Probability	Output
$\hat{1} \otimes \hat{1} \hat{1}$	$v^{\text{set}} + \frac{1-v^{\text{set}}}{4}$	$ \Phi^+\rangle \propto HH\rangle + VV\rangle$
$\hat{1} \otimes \hat{1} \hat{Z}$	$\frac{1-v^{\text{set}}}{4}$	$ \Phi^-\rangle \propto HH\rangle - VV\rangle$
$\hat{1} \otimes \hat{X} \hat{1}$	$\frac{1-v^{\text{set}}}{4}$	$ \Psi^+\rangle \propto HV\rangle + VH\rangle$
$\hat{1} \otimes \hat{X} \hat{Z}$	$\frac{1-v^{\text{set}}}{4}$	$ \Psi^-\rangle \propto HV\rangle - VH\rangle$

coincidences per second for a laser power of 4 mW. The photons are then sent through a depolarization channel composed of two liquid crystal retarders (LCR), which depolarize the maximally entangled state to a high-fidelity isotropic state by producing a probabilistic mixture of Bell states as in Table VIII.

The produced isotropic state typically has a value of the visibility v that deviates slightly from the value v^{set} that was targeted due to the initial state's nonunit fidelity with $|\Phi^+\rangle$ and the LCR's limited switching speed (80 ms hold time per setting, with 5 ms rise time and 20 ms fall time). Instead of relying solely on the targeted value v^{set} , we estimated the realized value using the measured fidelity to the nearest isotropic state. In Table IX, we compare the targeted and estimated values, showing that the fidelity between the tomographically reconstructed state and the nearest isotropic state is consistently near 99%.

2. Measurement technique

After preparing each isotropic state, we record coincidence events from both the transmitted and reflected output ports using single-photon avalanche diodes (SPCM-AQRH-11-FC, Excelitas) with a 1 ns coincidence window, and a 5 s measurement time per measurement configuration. We measure coincidence counts of the two photons in measurement configurations corresponding to the vertices of the shapes associated to the set of measurements being targeted. It is important to note that the polarizing beam splitters have different efficiencies for their transmitted and reflected ports, which has the potential to add nonisotropic noise to our state. To mitigate this noise, we follow the measurement scheme outlined in Table X, where, instead of assigning specific ports to specific effects, we iterate over the four permutations of ports and effects by adjusting the wave plates accordingly. Next, for each effect, we take the four trials (one per port configuration) and sum the coincidence counts for the desired effect. Finally, we divide each coincidence by the sum of the four to convert counts into relative frequencies for one measurement configuration. In total, 240 distinct measurement configurations are implemented in the same way, resulting in a total integration time of approximately 2 h. These raw relative frequencies are then converted to the raw frequency corresponding to the 60 measurement settings (effectively,

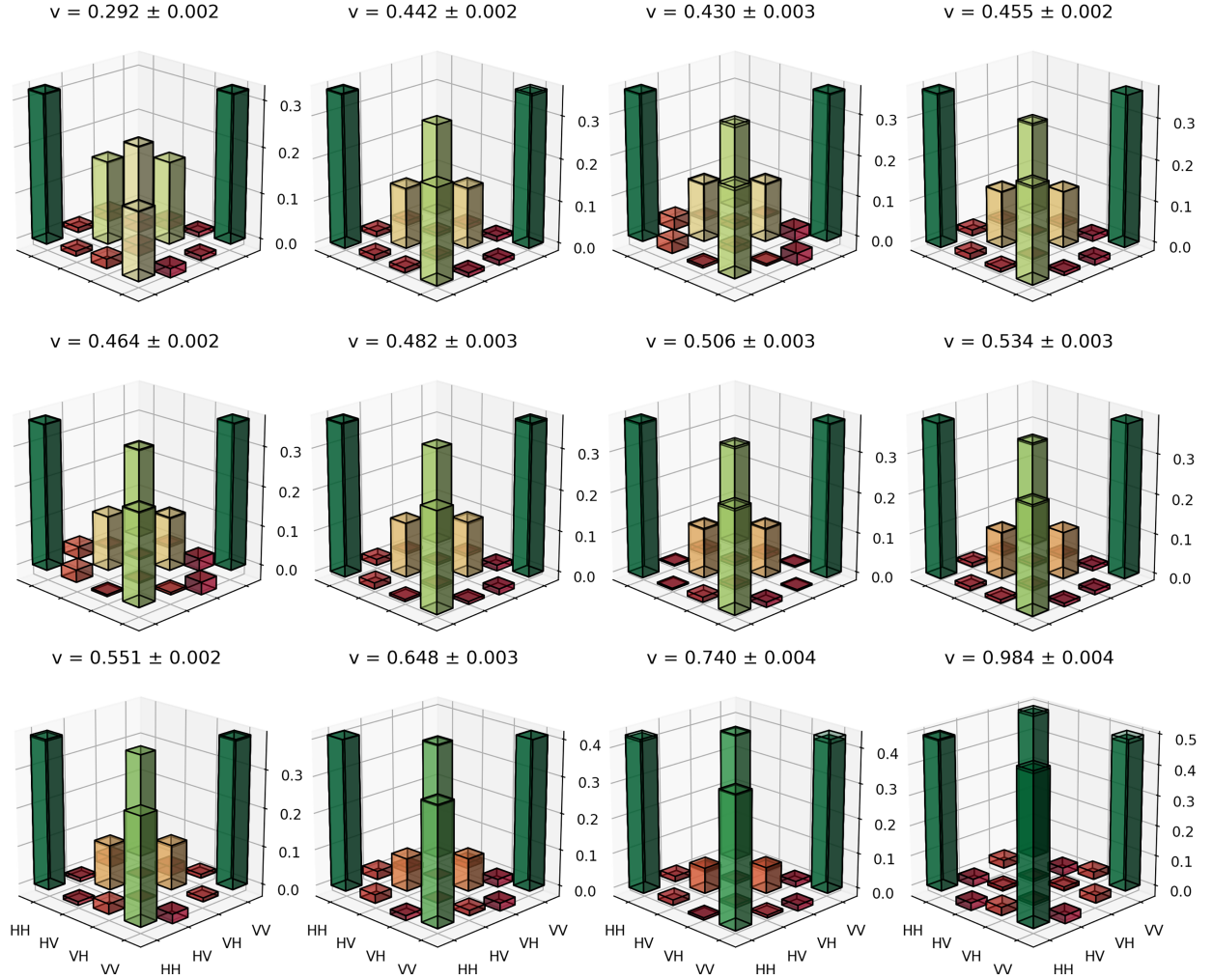


FIG. 6. Quantum state tomography results: the real part of the reconstructed density matrix is shown, while the imaginary components are negligible, with values below 0.02. The visibility v denotes the weight of the closest isotropic state, with error bars estimated from 100 Monte Carlo trials accounting for both statistical and systematic uncertainties (see the text below).

TABLE IX. Experimental results for 12 tomographically reconstructed quantum states for which the parameter value being targeted is v^{set} and the parameter of the closest isotropic state is v . In bold: unsteerable states that violate the inequality $\mathcal{I} \geq 0$ in the main text; in italics: states that violate the inequality in the main text but for which the entanglement could, in principle, be certified by violation of a steering inequality. The uncertainties in the violations arise from Poissonian statistics, whereas the uncertainties in v and in fidelities are calculated from Monte Carlo simulations of the different state tomography results, as described in the text.

v^{set}	$\mathcal{F}(\rho_{\text{iso}}^{v^{\text{set}}}, \rho^{\text{tomo}})$	v	$\mathcal{F}(\rho_{\text{iso}}^v, \rho^{\text{tomo}})$	$\mathcal{I}$
0.270	0.9949 ± 0.0007	0.290 ± 0.002	0.9950 ± 0.0006	0.04513 ± 0.00015
0.420	0.9965 ± 0.0009	0.441 ± 0.002	0.9967 ± 0.0009	0.00578 ± 0.00021
0.430	0.9892 ± 0.0010	0.430 ± 0.003	0.9892 ± 0.0010	0.01068 ± 0.00019
0.440	0.9966 ± 0.0008	0.454 ± 0.002	0.9963 ± 0.0008	-0.00117 ± 0.00022
0.450	0.9934 ± 0.0011	0.465 ± 0.002	0.9936 ± 0.0011	-0.00409 ± 0.00018
<i>0.480</i>	<i>0.9938 ± 0.0012</i>	<i>0.483 ± 0.003</i>	<i>0.9938 ± 0.0012</i>	<i>-0.00585 ± 0.00039</i>
<i>0.500</i>	<i>0.9962 ± 0.0011</i>	<i>0.505 ± 0.003</i>	<i>0.9962 ± 0.0011</i>	<i>-0.01207 ± 0.00018</i>
<i>0.520</i>	<i>0.9935 ± 0.0013</i>	<i>0.534 ± 0.003</i>	<i>0.9936 ± 0.0013</i>	<i>-0.01633 ± 0.00016</i>
<i>0.540</i>	<i>0.9957 ± 0.0010</i>	<i>0.551 ± 0.002</i>	<i>0.9958 ± 0.0010</i>	<i>-0.02583 ± 0.00022</i>
<i>0.650</i>	<i>0.9911 ± 0.0017</i>	<i>0.651 ± 0.003</i>	<i>0.9911 ± 0.0017</i>	<i>-0.05280 ± 0.00011</i>
<i>0.750</i>	<i>0.9869 ± 0.0029</i>	<i>0.742 ± 0.004</i>	<i>0.9870 ± 0.0030</i>	<i>-0.07849 ± 0.00012</i>
<i>1.000</i>	<i>0.9689 ± 0.0036</i>	<i>0.982 ± 0.004</i>	<i>0.9802 ± 0.0043</i>	<i>-0.14256 ± 0.00013</i>

TABLE X. Measurement iteration scheme for setting xy .

Iter no.	Alice transmission	Alice reflection ports	Bob transmission	Bob reflection ports
1	$N_{0 x}$	$N_{1 x}$	$M_{0 y}$	$M_{1 y}$
2	$N_{1 x}$	$N_{0 x}$	$M_{1 y}$	$M_{0 y}$
3	$N_{0 x}$	$N_{1 x}$	$M_{1 y}$	$M_{0 y}$
4	$N_{1 x}$	$N_{0 x}$	$M_{0 y}$	$M_{1 y}$

each setting is measured in four different experimental configurations).

3. Experimental error analysis

The experimental uncertainty in the noncontextual bound $\mathcal{I}$ is estimated using a Monte Carlo simulation with 100 independent trials, where each trial samples from a Poisson distribution based on the observed raw counts. To further quantify the uncertainties in the estimated state visibility v and to certify the unsteerability of the states, we also account for systematic errors in the wave plate settings by drawing these from a normal distribution, similarly as reported in Ref. [86].

Specifically, we consider two primary sources of systematic error: (i) miscalibration or model imperfection of the measurement wave plates, modeled as a Gaussian distribution over phase shifts with a standard deviation of 1° , and (ii) phase shifts introduced by model imperfections also assumed to have an uncertainty of 1° . For each trial, the reconstructed density matrix is used to estimate the isotropic state visibility v , and the standard deviation across all trials is reported as the final uncertainty. Additional details and data are summarized in Appendix D.

TABLE XI. Examples of experimental results on optimal secondary measurements: for each quantum state involved in the test, we report the optimal values of the parameters C_N and C_M obtained during the construction of the secondary measurements.

v^{set}	C_M	C_N
0.270	0.959	0.937
0.420	0.958	0.980
0.430	0.946	0.972
0.440	0.950	0.982
0.450	0.960	0.982
0.480	0.955	0.957
0.500	0.956	0.969
0.520	0.956	0.972
0.540	0.959	0.971
0.650	0.966	0.981
0.750	0.972	0.985
1.000	0.970	0.985

4. Detailed experimental results

The validity of our experimental outcomes—and the subsequent violation of the inequality under study—is inherently linked to the precision of our isotropic state preparation. Recall that the fidelity $\mathcal{F}$ between states ρ and σ is defined as follows:

$$\mathcal{F}(\rho, \sigma) = \text{Tr} \left(\sqrt{\sqrt{\rho} \sigma \sqrt{\rho}} \right)^2. \quad (\text{D1})$$

While the main text highlights a fidelity between our realized state and the best-matched isotropic state that is near 99%, this section provides a detailed evaluation of our measured fidelity with respect to both the best-matched isotropic state and the state we targeted. To rigorously assess the quality of our state preparation, we performed quantum state tomography using the same experimental setup described earlier. Measurements were conducted across all 12×20 basis settings, with the corresponding Bloch vectors arranged to realize the geometries of the icosahedron and dodecahedron, and the quantum state tomography was done by minimizing the L^2 norm

$$L^2 = |f_i - \langle \psi_i | \rho | \psi_i \rangle|^2, \quad (\text{D2})$$

where $|\psi_i\rangle$ represents the i th measurement setting and f_i is the corresponding relative frequency. The result of this optimization yields a density operator, which is denoted ρ^{tomo} .

To address potential mischaracterizations of ρ^{tomo} due to statistical and systematic errors, we conducted a comprehensive Monte Carlo simulation comprising 100 trials. In this simulation, (i) raw count data were modeled using Poissonian statistics; (ii) the calibration of the measurement wave-plate angles was simulated using a normal distribution over phase shifts with a standard deviation of 1° ; and (iii) variations in wave-plate thickness—leading to phase-shift errors—were also represented by a normal distribution with 1° of phase-shift error.

After reconstructing the density operator via tomography, we computed its fidelity with the target isotropic state (defined by visibility v^{set}) and the closest isotropic state (defined by visibility v).

The results, summarized in Table IX, confirm that, for each of the 12 values of v^{set} we considered, our realized state had a fidelity with our targeted state and a fidelity with the best-matched isotropic state of near 99%.

APPENDIX E: CERTIFYING UNSTEERABILITY FOR STATES

In our experiment, the prepared state is never exactly an isotropic state (for which unsteerability and locality are well studied [6,8]). Therefore, in addition to noting the high fidelity of the states we prepare with an unsteerable state,

we provide a direct demonstration of the unsteerability of some of the states we realized in the experiment.

To achieve this, we employ recent findings in Refs. [47,48], which show that the isotropic state $\rho_{\text{iso}}^v = v|\Phi^+\rangle\langle\Phi^+| + (1-v)[(\mathbb{1} \otimes \mathbb{1})/4]$ is unsteerable for $v \leq \frac{1}{2}$. We make use of each of the following facts: (i) any state $\rho = (\mathbb{1} \otimes U)\rho_{\text{iso}}^v(\mathbb{1} \otimes U^\dagger)$ with $v \leq \frac{1}{2}$ remains unsteerable for any unitary U ; (ii) all separable states are unsteerable; (iii) any convex combination of unsteerable states is unsteerable.

We can witness the unsteerability of our realized state ρ if we find feasible solutions of the following semidefinite program:

$$X \succeq 0, \quad X^{T_B} \succeq 0, \quad p_i \geq 0 \quad \forall i, \\ \sum_{i=1}^n p_i \rho_i + X = \rho,$$

where $\rho_i = (\mathbb{1} \otimes U_i)\rho_{\text{iso}}^v(\mathbb{1} \otimes U_i^\dagger)$ with $v \leq 1/2$ for some U_i sampled from the Haar measure of $\text{SU}(2)$ and X is any two-qubit quantum state with a positive partial transpose (and so is separable). That is, the feasibility of this program is a sufficient condition for unsteerability.

Moreover, following an argument similar to that in Ref. [86], the fact (i) can be replaced with a more general statement for any positive map; that is, any state $\rho = (\mathbb{1} \otimes \Lambda)(\rho_{\text{iso}}^v) \geq 0$, where $v \leq 1/2$ remains unsteerable for any positive map Λ [86]. Using this condition leads to a slightly different semidefinite program (SDP) that is generally capable of certifying a broader class of unsteerable states.

The numerics show that two of the states we realized in our experiment, those with estimated isotropic visibilities $v = 0.454$ and $v = 0.465$, yield feasible solutions to the SDP (even after Monte Carlo simulations accounting for small measurement imperfections), thereby confirming their unsteerability. However, the state with estimated isotropic parameter $v = 0.483$ fails this test and, thus, cannot be concluded to be unsteerable even though it is close to an unsteerable isotropic state.

-
- [1] J. S. Bell, *On the Einstein Podolsky Rosen paradox*, *Phys. Phys. Fiz.* **1**, 195 (1964).
 - [2] C. J. Wood and R. W. Spekkens, *The lesson of causal discovery algorithms for quantum correlations: Causal explanations of Bell-inequality violations require fine-tuning*, *New J. Phys.* **17**, 033002 (2015).
 - [3] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, *Bell nonlocality*, *Rev. Mod. Phys.* **86**, 419 (2014).
 - [4] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, *Quantum entanglement*, *Rev. Mod. Phys.* **81**, 865 (2009).
 - [5] R. Uola, Ana C. S. Costa, H. C. Nguyen, and O. Gühne, *Quantum steering*, *Rev. Mod. Phys.* **92**, 015001 (2020).

- [6] R. F. Werner, *Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model*, *Phys. Rev. A* **40**, 4277 (1989).
- [7] M. Piani, S. Gharibian, G. Adesso, J. Calsamiglia, P. Horodecki, and A. Winter, *All nonclassical correlations can be activated into distillable entanglement*, *Phys. Rev. Lett.* **106**, 220403 (2011).
- [8] H. M. Wiseman, S. J. Jones, and A. C. Doherty, *Steering, entanglement, nonlocality, and the Einstein-Podolsky-Rosen paradox*, *Phys. Rev. Lett.* **98**, 140402 (2007).
- [9] D. Schmid, T. C. Fraser, R. Kunjwal, A. B. Sainz, E. Wolfe, and R. W. Spekkens, *Understanding the interplay of entanglement and nonlocality: Motivating and developing a new branch of entanglement theory*, *Quantum* **7**, 1194 (2023).
- [10] M. T. Quintino, T. Vértesi, D. Cavalcanti, R. Augusiak, M. Demianowicz, A. Acín, and N. Brunner, *Inequivalence of entanglement, steering, and Bell nonlocality for general measurements*, *Phys. Rev. A* **92**, 032107 (2015).
- [11] N. Gisin, *Bell's inequality holds for all non-product states*, *Phys. Lett. A* **154**, 201 (1991).
- [12] N. Gisin and A. Peres, *Maximal violation of Bell's inequality for arbitrarily large spin*, *Phys. Lett. A* **162**, 15 (1992).
- [13] J. Barrett, *Nonsequential positive-operator-valued measurements on entangled mixed states do not always violate a Bell inequality*, *Phys. Rev. A* **65**, 042302 (2002).
- [14] 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).
- [15] O. Gühne and G. Tóth, *Entanglement detection*, *Phys. Rep.* **474**, 1 (2009).
- [16] C. Branciard, D. Rosset, Y.-C. Liang, and N. Gisin, *Measurement-device-independent entanglement witnesses for all entangled quantum states*, *Phys. Rev. Lett.* **110**, 060405 (2013).
- [17] F. Buscemi, *All entangled quantum states are nonlocal*, *Phys. Rev. Lett.* **108**, 200401 (2012).
- [18] R. W. Spekkens, *Contextuality for preparations, transformations, and unsharp measurements*, *Phys. Rev. A* **71**, 052108 (2005).
- [19] Note that, throughout this work, we use the term *noncontextual* to refer to the notion of generalized noncontextuality defined in Ref. [18] rather than the Kochen-Specker notion [20,21]. We use the term *NCOM-nonrealizable* in preference to *contextual* to emphasize that, in the face of a no-go result, one has the option to abandon the framework of ontological models rather than the principle of noncontextuality.
- [20] S. Kochen and E. P. Specker, *The problem of hidden variables in quantum mechanics*, in *The Logico-Algebraic Approach to Quantum Mechanics: Volume I: Historical Evolution*, edited by C. A. Hooker (Springer Netherlands, Dordrecht, 1975), pp. 293–328.
- [21] C. Budroni, A. Cabello, O. Gühne, M. Kleinmann, and J.-A. Larsson, *Kochen-Specker contextuality*, *Rev. Mod. Phys.* **94**, 045007 (2022).
- [22] D. Schmid, J. H. Selby, and R. W. Spekkens, *Addressing some common objections to generalized noncontextuality*, *Phys. Rev. A* **109**, 022228 (2024).
- [23] R. W. Spekkens, *Negativity and contextuality are equivalent notions of nonclassicality*, *Phys. Rev. Lett.* **101**, 020401 (2008).

- [24] D. Schmid, J. H. Selby, M. F. Pusey, and R. W. Spekkens, *A structure theorem for generalized-noncontextual ontological models*, *Quantum* **8**, 1283 (2024).
- [25] D. Schmid, J. H. Selby, E. Wolfe, R. Kunjwal, and R. W. Spekkens, *Characterization of noncontextuality in the framework of generalized probabilistic theories*, *PRX Quantum* **2**, 010331 (2021).
- [26] F. Shahandeh, *Contextuality of general probabilistic theories*, *PRX Quantum* **2**, 010330 (2021).
- [27] A. Tavakoli and R. Uola, *Measurement incompatibility and steering are necessary and sufficient for operational contextuality*, *Phys. Rev. Res.* **2**, 013011 (2020).
- [28] V. J. Wright and M. Farkas, *Invertible map between Bell nonlocal and contextuality scenarios*, *Phys. Rev. Lett.* **131**, 220202 (2023).
- [29] M. Plávala and O. Gühne, *Contextuality as a precondition for quantum entanglement*, *Phys. Rev. Lett.* **132**, 100201 (2024).
- [30] A. G. Kofman and A. N. Korotkov, *Bell-inequality violation versus entanglement in the presence of local decoherence*, *Phys. Rev. A* **77**, 052329 (2008).
- [31] E. Nielsen, J. K. Gamble, K. Rudinger, T. Scholten, K. Young, and R. Blume-Kohout, *Gate set tomography*, *Quantum* **5**, 557 (2021).
- [32] M. D. Mazurek, M. F. Pusey, K. J. Resch, and R. W. Spekkens, *Experimentally bounding deviations from quantum theory in the landscape of generalized probabilistic theories*, *PRX Quantum* **2**, 020302 (2021).
- [33] D. J. Saunders, S. J. Jones, H. M. Wiseman, and G. J. Pryde, *Experimental EPR-steering using Bell-local states*, *Nat. Phys.* **6**, 845 (2010).
- [34] J. Bowles, I. Šupić, D. Cavalcanti, and A. Acín, *Device-independent entanglement certification of all entangled states*, *Phys. Rev. Lett.* **121**, 180503 (2018).
- [35] P.-S. Lin, D. Rosset, Y. Zhang, J.-D. Bancal, and Y.-C. Liang, *Device-independent point estimation from finite data and its application to device-independent property estimation*, *Phys. Rev. A* **97**, 032309 (2018).
- [36] E. G. Cavalcanti, S. J. Jones, H. M. Wiseman, and M. D. Reid, *Experimental criteria for steering and the Einstein-Podolsky-Rosen paradox*, *Phys. Rev. A* **80**, 032112 (2009).
- [37] Y. Zhang, D. Schmid, Y. Yīng, and R. W. Spekkens, *Reassessing the boundary between classical and nonclassical for individual quantum processes*, *Phys. Rev. X* **16**, 021050 (2026).
- [38] Strictly speaking, the notion of noncontextuality can be applied only to ontological models of theories that are not quotiented relative to operational equivalences, whereas in this work we treat quantum theory as a theory wherein this quotienting procedure has been applied [24,39]. References [24,25] introduce more precise terminology, but we eschew it here for the sake of maintaining familiar terminology.
- [39] G. Chiribella, G. M. D’Ariano, and P. Perinotti, *Probabilistic theories with purification*, *Phys. Rev. A* **81**, 062348 (2010).
- [40] P. Skrzypczyk, M. Navascués, and D. Cavalcanti, *Quantifying Einstein-Podolsky-Rosen steering*, *Phys. Rev. Lett.* **112**, 180404 (2014).
- [41] D. Rosset, D. Schmid, and F. Buscemi, *Type-independent characterization of spacelike separated resources*, *Phys. Rev. Lett.* **125**, 210402 (2020).
- [42] Y. Zhang, Y. Yīng, and D. Schmid, *Quantifiers and witnesses for the nonclassicality of measurements and of states*, *Quantum* **10**, 2180 (2026).
- [43] For readers who would prefer a terminology that is less generic than “classical,” these bipartite states can be referred to as *Leibniz-classical*, a terminology introduced in Ref. [37].
- [44] The Royal Swedish Academy of Sciences, Scientific Background on the Nobel Prize in Physics 2022: Experiments with Entangled Photons, Establishing the Violation of Bell Inequalities and Pioneering Quantum Information Science, <https://www.nobelprize.org/uploads/2022/10/advanced-physicsprize2022.pdf> (2022) (Accessed: November 5, 2024).
- [45] S. Designolle, G. Iommarzo, M. Besançon, S. Knebel, P. Gelß, and S. Pokutta, *Improved local models and new Bell inequalities via Frank-Wolfe algorithms*, *Phys. Rev. Res.* **5**, 043059 (2023).
- [46] A. Acín, N. Gisin, and B. Toner, *Grothendieck’s constant and local models for noisy entangled quantum states*, *Phys. Rev. A* **73**, 062105 (2006).
- [47] Y. Zhang and E. Chitambar, *Exact steering bound for two-qubit Werner states*, *Phys. Rev. Lett.* **132**, 250201 (2024).
- [48] M. J. Renner, *Compatibility of generalized noisy qubit measurements*, *Phys. Rev. Lett.* **132**, 250202 (2024).
- [49] M. J. Wenninger, *Spherical Models* (Cambridge University Press, Cambridge, England, 1979).
- [50] H.-Y. Huang, R. Kueng, and J. Preskill, *Predicting many properties of a quantum system from very few measurements*, *Nat. Phys.* **16**, 1050 (2020).
- [51] A. Elben, S. T. Flammia, H.-Y. Huang, R. Kueng, J. Preskill, B. Vermersch, and P. Zoller, *The randomized measurement toolbox*, *Nat. Rev. Phys.* **5**, 9 (2023).
- [52] R. Horodecki, P. Horodecki, and M. Horodecki, *Violating Bell inequality by mixed spin-1/2 states: Necessary and sufficient condition*, *Phys. Lett. A* **200**, 340 (1995).
- [53] A. Peres, *Separability criterion for density matrices*, *Phys. Rev. Lett.* **77**, 1413 (1996).
- [54] M. D. Mazurek, M. F. Pusey, R. Kunjwal, K. J. Resch, and R. W. Spekkens, *An experimental test of noncontextuality without unphysical idealizations*, *Nat. Commun.* **7**, ncomms11780 (2016).
- [55] S. T. Merkel, J. M. Gambetta, J. A. Smolin, S. Poletto, A. D. Córcoles, B. R. Johnson, C. A. Ryan, and M. Steffen, *Self-consistent quantum process tomography*, *Phys. Rev. A* **87**, 062119 (2013).
- [56] D. Greenbaum, *Introduction to quantum gate set tomography*, [arXiv:1509.02921](https://arxiv.org/abs/1509.02921).
- [57] Suppose $\mathcal{B}^{\text{pri}}$ denotes the realized Bell circuit and $\mathcal{B}^{\text{sec}}$ denotes the image of this under the local classical post-processing; if the primary statistics are NCOM-realizable relative to $\mathcal{O}_{\text{aattt}}(\mathcal{B}^{\text{pri}})$, then the secondary statistics must be NCOM-realizable relative to $\mathcal{O}_{\text{aattt}}(\mathcal{B}^{\text{sec}})$, since local classical processing can be absorbed into a redefinition of the response functions in Eq. (4).
- [58] T. Kim, M. Fiorentino, and Franco N. C. Wong, *Phase-stable source of polarization-entangled photons using a*

- polarization Sagnac interferometer, *Phys. Rev. A* **73**, 012316 (2006).
- [59] L. Hardy, *Quantum theory from five reasonable axioms*, arXiv:quant-ph/0101012.
- [60] J. Barrett, *Information processing in generalized probabilistic theories*, *Phys. Rev. A* **75**, 032304 (2007).
- [61] M. Krystek and M. Anton, *A weighted total least-squares algorithm for fitting a straight line*, *Meas. Sci. Technol.* **18**, 3438 (2007).
- [62] D. Schmid, J. H. Selby, V. P. Rossi, R. D. Baldijão, and A. B. Sainz, *Shadows and subsystems of generalized probabilistic theories: When tomographic incompleteness is not a loophole for contextuality proofs*, *Quantum* **9**, 1880 (2025).
- [63] N. Gisin and B. Gisin, *A local hidden variable model of quantum correlation exploiting the detection loophole*, *Phys. Lett. A* **260**, 323 (1999).
- [64] S. Massar and S. Pironio, *Violation of local realism versus detection efficiency*, *Phys. Rev. A* **68**, 062109 (2003).
- [65] J.-Å. Larsson, *Loopholes in Bell inequality tests of local realism*, *J. Phys. A* **47**, 424003 (2014).
- [66] B. G. Christensen, K. T. McCusker, J. B. Altepeter, B. Calkins, T. Gerrits, A. E. Lita, A. Miller, L. K. Shalm, Y. Zhang, S. W. Nam, N. Brunner, C. C. W. Lim, N. Gisin, and P. G. Kwiat, *Detection-loophole-free test of quantum non-locality, and applications*, *Phys. Rev. Lett.* **111**, 130406 (2013).
- [67] M. Giustina *et al.*, *Significant-loophole-free test of Bell's theorem with entangled photons*, *Phys. Rev. Lett.* **115**, 250401 (2015).
- [68] L. K. Shalm *et al.*, *Strong loophole-free test of local realism*, *Phys. Rev. Lett.* **115**, 250402 (2015).
- [69] J. H. Selby, D. Schmid, E. Wolfe, A. B. Sainz, R. Kunjwal, and R. W. Spekkens, *Contextuality without incompatibility*, *Phys. Rev. Lett.* **130**, 230201 (2023).
- [70] J. H. Selby, D. Schmid, E. Wolfe, A. B. Sainz, R. Kunjwal, and R. W. Spekkens, *Accessible fragments of generalized probabilistic theories, cone equivalence, and applications to witnessing nonclassicality*, *Phys. Rev. A* **107**, 062203 (2023).
- [71] Note that certification of EPR steering is also possible in a manner that has no detection inefficiency loophole [72].
- [72] A. J. Bennet, D. A. Evans, D. J. Saunders, C. Branciard, E. G. Cavalcanti, H. M. Wiseman, and G. J. Pryde, *Arbitrarily loss-tolerant Einstein-Podolsky-Rosen steering allowing a demonstration over 1 km of optical fiber with no detection loophole*, *Phys. Rev. X* **2**, 031003 (2012).
- [73] M. J. Grabowecky, Christopher A. J. Pollack, A. R. Cameron, R. W. Spekkens, and K. J. Resch, *Experimentally bounding deviations from quantum theory for a photonic three-level system using theory-agnostic tomography*, *Phys. Rev. A* **105**, 032204 (2022).
- [74] T. Proctor, K. Rudinger, K. Young, M. Sarovar, and R. Blume-Kohout, *What randomized benchmarking actually measures*, *Phys. Rev. Lett.* **119**, 130502 (2017).
- [75] M. F. Pusey, L. del Rio, and B. Meyer, *Contextuality without access to a tomographically complete set*, arXiv:1904.08699.
- [76] Y. Zhang, *Bipartite-nonclassicality: Data and code (2025)*, GitHub repository. commit 1932f03, <https://github.com/yujie4phy/Bipartite-nonclassicality> (accessed 12, April 2025).
- [77] P. Busch, *Quantum states and generalized observables: A simple proof of Gleason's theorem*, *Phys. Rev. Lett.* **91**, 120403 (2003).
- [78] The polytope describing the intrinsic geometry of N should not be confused with the noncontextual polytope of NCOM-realizable distributions.
- [79] D. Schmid, R. D. Baldijão, J. H. Selby, A. B. Sainz, and R. W. Spekkens, *Noncontextuality inequalities for prepare-transform-measure scenarios*, arXiv:2407.09624.
- [80] D. Schmid, R. W. Spekkens, and E. Wolfe, *All the non-contextuality inequalities for arbitrary prepare-and-measure experiments with respect to any fixed set of operational equivalences*, *Phys. Rev. A* **97**, 062103 (2018).
- [81] P. McMullen, *The maximum numbers of faces of a convex polytope*, *Mathematika* **17**, 179 (1970).
- [82] N. Alon and G. Kalai, *A simple proof of the upper bound theorem*, *Eur. J. Combinatorics* **6**, 211 (1985).
- [83] Y. Zhang, J. Zhang, and E. Chitambar, *Cost of simulating entanglement in steering scenarios*, *Quantum* **9**, 1902 (2025).
- [84] This step assumes the dimension of the underlying quantum system. However, one can also perform the fit in a way that infers an effective quantum dimension from the experimental data using a train-and-test methodology rather than fixing D a priori [32,73].
- [85] We note that, for cases like our Example 1, tomographic completeness is not even required, since the operational identities on measurements that one needs to enforce in the test are automatically satisfied by the condition that the effects of each measurement sum to the unit effect. However, for generality, we always assume tomographic completeness in our entanglement-certification protocols.
- [86] L. Villegas-Aguilar, E. Polino, F. Ghafari, M. T. Quintino, K. T. Laverick, I. R. Berkman, S. Rogge, L. K. Shalm, N. Tischler, E. G. Cavalcanti, S. Slussarenko, and G. J. Pryde, *Nonlocality activation in a photonic quantum network*, *Nat. Commun.* **15**, 3112 (2024).