

Device-independent certification of desirable properties with a confidence interval

Wan-Guan Chang,^{1,2,3} Kai-Chun Chen,¹ Kai-Siang Chen,¹ Shin-Liang Chen,^{4,3} and Yeong-Cherng Liang^{1,3,*}

¹*Department of Physics and Center for Quantum Frontiers of Research & Technology (QFort), National Cheng Kung University, Tainan 701, Taiwan*

²*Institute of Information Science, Academia Sinica, Taiwan*

³*Physics Division, National Center for Theoretical Sciences, Taipei 10617, Taiwan*

⁴*Department of Physics, National Chung Hsing University, Taichung 40227, Taiwan*

(Dated: May 6, 2024)

In the development of quantum technologies, a reliable means for characterizing quantum devices, be it a measurement device, a state-preparation device, or a transformation device, is crucial. However, the conventional approach based on, for example, quantum state tomography or process tomography relies on assumptions that are often not necessarily justifiable in a realistic experimental setting. While the device-independent approach to this problem bypasses the shortcomings above by making only minimal, justifiable assumptions, most of the theoretical proposals to date only work in the idealized setting where independent and identically distributed (i.i.d.) trials are assumed. Here, we provide a versatile solution for rigorous device-independent certification that does not rely on the i.i.d. assumption. Specifically, we describe how the prediction-based-ratio (PBR) protocol and martingale-based protocol developed for hypothesis testing can be applied in the present context to achieve a device-independent certification of desirable properties with confidence interval. To illustrate the versatility of these methods, we demonstrate how we can use them to certify—with finite data—the underlying negativity, Hilbert space dimension, entanglement depth, and fidelity to some target pure state. In particular, we give examples showing how the amount of certifiable negativity and fidelity scales with the number of trials, and how many experimental trials one needs to certify a qutrit state space, or the presence of genuine tripartite entanglement. Overall, we have found that the PBR protocol and the martingale-based protocol often offer similar performance, even though the former does have to presuppose any witness (Bell-like inequality). In contrast, our findings also show that the performance of the martingale-based protocol may be severely affected by one's choice of the Bell-like inequality. Intriguingly, a Bell function useful for self-testing does not necessarily give the optimal confidence-gain rate for certifying the fidelity to the corresponding target state.

I. INTRODUCTION

The proper analysis of quantum experiments is an indispensable part in the development of quantum technologies. However, it is not trivial to reliably characterize a quantum setup, which may include, e.g., measurement devices and state-preparation devices. Moreover, imperfections in the experimental setup can easily result in a mismatch [1–3] between the characterization tools developed for an idealized situation and an actual experimental situation. However, we can circumvent this problem by the so-called “device-independent approach” [4, 5]. In quantum information, the term “device-independent” (DI) was first coined [6] in the task of quantum key distribution [7–9], even though the idea was already conceived independently, but implicitly in [10, 11].

In a nutshell, the DI approach is a framework for analyzing physical systems without relying on any assumption about the degrees of freedom measured. Its basis is Bell-nonlocality [5, 12], which shows that no local-hidden-variable theory (LHV) can reproduce *all* quantum predictions, even though *no* further assumption is made about the details of such a theory. For example, it is known that the violation of Bell inequalities [12] obtained by locally measuring a shared state implies [13] shared entanglement [14], which is a powerful resource in many quantum information processing tasks. More generally, many other desirable properties

of the underlying state [15–23], measurements [22–28], and channel [21, 29, 30] may be derived directly from the observation of a Bell-inequality-violating correlation between measurement outcomes. Recently, the DI approach has been also been incorporated into the security analysis of quantum secure direct communication, see, e.g., [31] and references therein.

However, due to statistical fluctuations, even when the experimental trials are independent and identically distributed (*i.i.d.*), relative frequencies of the measurement outcomes obtained from a Bell experiment do not faithfully represent the underlying distribution. In particular, such raw distributions estimated from the experimental results typically [32–34] lead to a violation of the nonsignaling conditions [35, 36], which is a prerequisite for the analysis in [16–30]. In other words, statistical fluctuations render the many theoretical tools developed for such a purpose inapplicable. To address this issue, some *ad hoc* methods [32–34] have been proposed to regularize the relative frequencies obtained to ensure that the resulting distribution satisfy the nonsignaling conditions. In [37], a more in-depth discussion was provided and two better-motivated regularization methods were proposed.

While these more recent attempts do provide a point estimator that fits within the framework of the usual DI analysis, they are still problematic in two aspects. Firstly, they do not provide any confidence region associated with the estimate. However, any real experiment necessarily involves only a finite number of experimental trials. Therefore a useful analysis should provide not only an estimate but also an indication of the reliability of such an estimate. In many of the Bell experiments reported [38–41], this is achieved by reporting the

* ycliang@mail.ncku.edu.tw

standard deviations of Bell violations. However, for finite, especially a relatively small number of trials, the central limit theorem is not warranted, so the usual interpretation of standard deviations may become dubious. Secondly, these usual approaches and those that provide a DI point estimator [32–34, 37] implicitly assumes that the experimental trials are *i.i.d.*, and hence free of the memory effect [42, 43] (see more discussions in [5, 44–46]). Again, in a realistic experimental setting, the *i.i.d.* assumption may be difficult to justify.

For the tasks of DI randomness expansion [47, 48] and DI quantum key distribution [49, 50], specific tools [51–59] have been developed to overcome the above problems. Here, we are interested in providing a general solution to other device-independent certification tasks¹ that (1) can provide a confidence region and (2) does not *a priori* require the *i.i.d.* assumption. Our approach is inspired by the prediction-based ratio (PBR) protocol developed in [61] and the martingale-based method proposed by Gill [43, 62] for performing a hypothesis testing against the assumption of Bell-locality. Following [63], we further adapt these earlier methods and illustrate how they can be used for the device-independent certification of various properties of interest, including the underlying amount of entanglement and its fidelity with respect to some target quantum state.

To this end, we structure the rest of this paper as follows. In Section II, we explain the basic concepts relevant to the understanding of DI certification in the ideal setting. After that, we introduce in Section III our adapted statistical tools for performing a rigorous device-independent certification. Results obtained from these tools are then presented in Section IV. Finally, we give some concluding remarks and future directions in Section V.

II. PRELIMINARIES

A. Correlations and Bell inequalities

The starting point of the DI approach is a Bell test. To this end, consider a bipartite Bell scenario, where two observers, Alice and Bob, can choose, respectively, their measurements labeled by $x, y \in \{0, 1, \dots\}$ and register outcomes $a, b \in \{0, 1, \dots\}$.² In the *i.i.d.* setting, one can estimate the underlying correlation between measurement outcomes, i.e., $\vec{P} = \{P(ab|xy)\}$ from the registered empirical frequencies. Interestingly, as Bell first showed in [12], highly nontrivial conclusions can be drawn by inspecting \vec{P} alone.

For example, correlations that can be produced in an LHV theory have to satisfy a Bell inequality:

$$\sum_{x,y,a,b} \beta_{xy}^{ab} P(ab|xy) \stackrel{\mathcal{L}}{\leq} B_{\mathcal{L}}(\vec{\beta}) \quad (1)$$

where the *Bell coefficients* $\beta_{xy}^{ab} \in \mathbb{R}$, $\vec{\beta} := \{\beta_{xy}^{ab}\}$, and $B_{\mathcal{L}}(\vec{\beta})$ is the so-called local (upper) bound. Here, we use \mathcal{L} to signify that the inequality holds under the assumption that \vec{P} is compatible with an LHV theory. Explicitly, the nature of such a theory demands that \vec{P} is factorizable in the form of [5, 12]

$$P(ab|xy) \stackrel{\mathcal{L}}{=} \sum_{\lambda} q_{\lambda} P_A(a|x\lambda) P_B(b|y\lambda) \quad (2)$$

where $q_{\lambda} \geq 0$ for all λ , $\sum_{\lambda} q_{\lambda} = 1$, and $P_A(a|x\lambda), P_B(b|y\lambda) \in [0, 1]$ are local response functions.

In an actual Bell test, the measurement settings ought to be chosen randomly according to some predetermined distribution P_{xy} . To manifest this fact, one may write Eq. (1) using the unconditional joint distribution $P(abxy) = P(ab|xy)P_{xy}$ such that $P_{xy} = \sum_{a,b} P(abxy)$. In turn, we can then write a Bell inequality as a bound on the expectation value of a *Bell function* $I(v)$, defined in terms of $\vec{\beta}$ and P_{xy} , i.e.,

$$\langle I(v) \rangle := \frac{\langle \beta_{xy}^{ab} \rangle}{P_{xy}} \stackrel{\mathcal{L}}{\leq} B_{\mathcal{L}}(\vec{\beta}) \quad (3)$$

where $v = (a, b, x, y)$ is the quadruple of random variables for the measurement outcomes (a, b) and settings (x, y) . As an example, the famous Clauser-Horne-Shimony-Holt (CHSH) Bell inequality [64] may be specified via:

$$I_{\text{CHSH}} : \beta_{xy}^{ab} = (-1)^{xy+a+b} \quad \text{and} \quad B_{\mathcal{L}} = 2, \quad (4)$$

or equivalently, in terms of the correlator $E_{xy} := \sum_{a,b=0,1} (-1)^{a+b} P(ab|xy)$, as:

$$\mathcal{S}_{\text{CHSH}} = \sum_{x,y=0,1} (-1)^{xy} E_{xy} \stackrel{\mathcal{L}}{\leq} 2, \quad (5)$$

where $\mathcal{S}_{\text{CHSH}} = \langle I_{\text{CHSH}}(v) \rangle$.

In contrast, quantum theory allows correlations that cannot be cast in the form of Eq. (2). In fact, in a bipartite Bell test, general quantum correlations read as:

$$P(ab|xy) \stackrel{\mathcal{Q}}{=} \text{tr}(\rho M_{a|x}^{(A)} \otimes M_{b|y}^{(B)}) \quad (6)$$

where $\{M_{a|x}^{(A)}\}$ and $\{M_{b|y}^{(B)}\}$ are, respectively, the local positive-operator-valued measure (POVM) describing Alice and Bob's local measurements. For the benefits of subsequent discussions, it is also worth noting that both LHV and quantum correlations satisfy the nonsignaling conditions [35, 36]:

$$\begin{aligned} \sum_a P(ab|xy) &= \sum_a P(ab|x'y) \quad \forall x, x', \\ \sum_b P(ab|xy) &= \sum_b P(ab|xy') \quad \forall y, y'. \end{aligned} \quad (7)$$

For the CHSH Bell function, cf. Eq. (4), quantum theory dictates the upper bound

$$\langle I_{\text{CHSH}}(v) \rangle \stackrel{\mathcal{Q}}{\leq} B_{\mathcal{Q}} = 2\sqrt{2}, \quad (8)$$

which can be seen as a Bell-like inequality. Other Bell and Bell-like inequalities relevant to this work will be presented in the corresponding sections below.

¹ Note that the same task is called device-independent verification in [60].

² If a third party is involved in the Bell test, as in the case of Sections II B 2 and IV B, we denote by z and c , respectively, its label for the measurement setting and outcome. All other notations generalize accordingly.

B. Examples of properties to be certified

1. Negativity and dimension

As mentioned above, with local measurements on a quantum system, a Bell-inequality-violating correlation $\vec{P} \notin \mathcal{L}$ necessarily originates [13] from an entangled state ρ . Interestingly, the entanglement of the underlying ρ can also be lower bounded [17, 18, 20, 23] directly from the observed correlation \vec{P} . In this work, we focus on negativity [65] but it is worth noting that DI entanglement quantification can be also achieved, e.g., for the linear entropy of entanglement [18], generalized robustness of entanglement [23], and one-shot distillable entanglement [20].

For a bipartite density operator ρ , let ρ^{TA} be its partial transposition [66] with respect to subsystem A . Then, the negativity for a bipartite density operator ρ is defined as [65] $\mathcal{N}(\rho) := \sum_{\lambda_i < 0} |\lambda_i(\rho^{TA})|$, i.e., the sum of the absolute value of all negative eigenvalues $\lambda_i < 0$ of ρ^{TA} . Using a variational characterization of negativity provided in [65], it was shown in [17] that $\mathcal{N}(\rho)$ is lower bounded by the optimum value of the following semidefinite program (SDP):

$$\min \chi_\ell[\sigma_-]_{\text{tr}} \quad (9a)$$

$$\text{s.t. } \chi_\ell[\rho] = \chi_\ell[\sigma_+] - \chi_\ell[\sigma_-], \quad \chi_\ell[\sigma_\pm]^{TA} \succeq 0, \quad (9b)$$

$$\chi_\ell[\rho] \succeq 0, \quad \chi_\ell[\rho]_{\text{tr}} = 1, \quad (9c)$$

where $\chi_\ell[\rho]$ is a moment matrix that can be obtained by applying a particular local map on ρ (see [17] for details), \bar{A} is the output Hilbert space of the local map on A , $\chi_\ell[\sigma]_{\text{tr}} = \text{tr}[\sigma]$ represents the trace of the underlying operator σ . It is worth noting that for every integer $\ell \geq 1$, the constraints of Eq. (9c) provide a superset characterization of the quantum set \mathcal{Q} of correlations, analogous to those considered in [67–69]. Indeed, all entries from \vec{P} appear somewhere in the moment matrix $\chi_\ell[\rho]$, see [17].

As an explicit example, note that an observed *violation* of the CHSH Bell inequality of Eq. (5) gives the following non-trivial negativity lower bound of the underlying state ρ :

$$\mathcal{N}(\rho) \geq \frac{\mathcal{S}_{\text{CHSH}} - 2}{4(\sqrt{2} - 1)}. \quad (10)$$

Also worth noting is that if ρ acts on $\mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B}$ with $d = \min\{d_A, d_B\}$, then the maximal possible negativity $\mathcal{N}(\rho)$ is upper bounded by $\mathcal{N}_{\text{max}}^d := \frac{d-1}{2}$. Consequently, the observation of a large enough negativity also provides a non-trivial lower bound on the local Hilbert space dimension of the underlying system. More precisely, if the lower bound on $\mathcal{N}(\rho)$ obtained from Eq. (9) exceeds $\mathcal{N}_{\text{max}}^d$, one immediately deduces that ρ must act on a local Hilbert space of dimension $\geq d + 1$, thereby giving a dimension witness [15].

From Eqs. (5), (8) and (10), nonetheless, we see that a violation of the CHSH Bell inequality can never witness a local Hilbert space dimension > 2 . Instead, witnessing a local Hilbert space beyond qubits can be achieved by observing a reasonably strong violation of the 3-outcome Collins-Gisin-Linden-Massar-Popescu (CGLMP) Bell inequality [70] (see

also [71]), defined by

$$I_{\text{CGLMP3}} : \beta_{xy}^{ab} = (-1)^{x(y-1)} \{ \delta_{a-b}^{(2)} - [1 - \delta_x^{(2)} \delta_{y-1}^{(2)}] \delta_{b-a-1}^{(3)} \} \\ - \delta_x^{(2)} \delta_{y-1}^{(2)} \delta_{b-a+1}^{(3)} \quad \text{and} \quad B_{\mathcal{L}}(\vec{\beta}) = 2, \quad (11)$$

where $\delta_f^{(d)} = 1$ if $\text{mod}(f, d) = 0$ and vanishes otherwise. Denoting the corresponding expectation value by $\mathcal{S}_{\text{CGLMP3}} = \langle I_{\text{CGLMP3}}(v) \rangle$, the results from [17, 72, 73] suggest a negativity lower bound that increases linearly with $\mathcal{S}_{\text{CGLMP3}}$ from $\frac{1}{2}$ whenever $\mathcal{S}_{\text{CGLMP3}} \geq \frac{3}{\sqrt{2}} + \frac{1}{2}$.

2. Entanglement depth

In a many-body system, entanglement can come in various forms or structures [74]. In particular, an n -partite quantum state that is not fully separable is *not* necessarily genuinely n -partite entangled either. To witness the latter, one could rely on the demonstration of so-called genuine multipartite non-locality [75]. However, as remarked in [16], it is possible to witness genuine multipartite entanglement without relying on this strong form of multipartite nonlocality. In fact, using the SDP introduced in [17], one can even systematically construct DI witnesses of this kind, starting from a given multipartite Bell function, say $\vec{\beta} = \{\beta_{xyz}^{abc}\}$. Later, it was further shown in [19] (see also [76]) that the extent to which a multipartite Bell inequality is violated can be used to witness (lower-bound) the underlying entanglement depth [77, 78], i.e., the extent to which a many-body entanglement is needed to prepare the given multipartite state.

For illustration, consider the expectation value of the Mermin Bell function [79] $I_{\text{Mermin}}(v)$ with $v = (a, b, c, x, y, z)$:

$$\mathcal{S}_{\text{Mermin}} = \langle I_{\text{Mermin}}(v) \rangle = \frac{\langle \beta_{xyz}^{abc} \rangle}{P_{xyz}} = \sum_{x,y,z}' (-1)^{xyz} E_{xyz} \quad (12)$$

where $E_{xyz} := \sum_{a,b,c=0}^1 (-1)^{a+b+c} P(abc|xyz)$ is the tripartite correlator, the restricted sum \sum' is over all combinations of $x, y, z \in \{0, 1\}$ such that $\text{mod}(x+y+z, 2) = 1$, $P_{xyz} = \frac{1}{4}$ for the same combinations of x, y, z , and the Bell coefficients are

$$\beta_{xyz}^{abc} = (-1)^{xyz+a+b+c} \delta_{x+y+z-1}^{(2)}. \quad (13)$$

Then, it is known that the following Bell-like inequalities hold, respectively, for fully-separable states, 2-producible [77] tripartite quantum states (i.e., quantum states that can be generated using only 2-body entanglement), and general tripartite quantum states:

$$\mathcal{S}_{\text{Mermin}} \stackrel{\mathcal{L}}{\leq} 2, \quad \mathcal{S}_{\text{Mermin}} \stackrel{2\text{-prod.}}{\leq} 2\sqrt{2}, \quad \mathcal{S}_{\text{Mermin}} \stackrel{\mathcal{Q}}{\leq} 4. \quad (14)$$

3. State fidelity

The strongest form of device certification one can hope for within a DI paradigm is called *self-testing* [80], first proposed

in [10]. The key observation behind this feat is that the quantum strategy compatible with *certain* extremal quantum correlations \vec{P}_Q is essentially unique. Hence, with the observation of \vec{P}_Q in a Bell test, we can conclude, unambiguously that some degree of freedom (DOF) of the measured system *must* match a specific target state $|\psi\rangle$. Often, one can also self-test the underlying measurements alongside the state (see, however, [81, 82] for some examples of exceptions).

For instance, it is long known [83–86] that the maximal CHSH Bell-inequality violation of $\mathcal{S}_{\text{CHSH}} = 2\sqrt{2}$ can only obtained (up to local isometry) by measuring the following observables on a shared maximally entangled state (MES):

$$|\psi_{\text{MES}}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle), \quad (15a)$$

$$A_0 = \sigma_z, \quad A_1 = \sigma_x, \quad (15b)$$

$$B_y = \frac{1}{\sqrt{2}}[\sigma_z + (-1)^y \sigma_x], \quad (15c)$$

where the respective POVM elements (with $x, y = 0, 1$) are

$$M_{a|x}^{(A)} = \frac{1 + (-1)^a A_x}{2}, \quad M_{b|y}^{(B)} = \frac{1 + (-1)^b B_y}{2}. \quad (16)$$

Moreover, to obtain the maximal CHSH Bell-inequality violation for a partially entangled two-qubit state,

$$|\psi(\theta)\rangle = \cos \theta |00\rangle + \sin \theta |11\rangle, \quad \theta \in \left(0, \frac{\pi}{4}\right] \quad (17a)$$

it suffices [73] to consider A_x of Eq. (15c) but generalize B_y to [87]:

$$B_y = \cos \mu \sigma_z + (-1)^y \sin \mu \sigma_x, \quad \tan \mu = \sin(2\theta), \quad (17b)$$

thereby giving

$$\mathcal{S}_{\text{CHSH}} = 2\sqrt{1 + \sin^2 2\theta}. \quad (18)$$

Interestingly, the resulting correlation also self-tests [87, 88] the corresponding quantum strategy of Eq. (17) and maximally violate the family of tilted CHSH Bell inequalities for $\alpha = 2\sqrt{\frac{\cos^2 2\theta}{1 + \sin^2 2\theta}}$:

$$\mathcal{S}_{\text{CHSH}}^{\text{Tilted}}(\alpha) = \mathcal{S}_{\text{CHSH}} + \alpha \sum_{a,b=0}^1 (-1)^a P(ab|0y) \stackrel{\mathcal{L}}{\leq} 2 + \alpha, \quad (19)$$

giving $\mathcal{S}_{\text{CHSH}}^{\text{Tilted}}(\alpha) = \sqrt{8 + 2\alpha^2}$. Note that in Eq. (19), thanks to the nonsignaling [35, 36] property of \vec{P} , the expression for $\mathcal{S}_{\text{CHSH}}^{\text{Tilted}}(\alpha)$ is in fact independent of whether $y = 0$ or 1.

In practice, however, due to various imperfections, one can, at best attain a correlation close to the ideal one \vec{P}_Q . In other words, in a realistic experimental setting, one can only hope to lower bound the similarity of the measured state ρ with respect to the target state $|\psi\rangle$ via a fidelity measure. To this end, a powerful numerical technique known as the SWAP method has been introduced in [89] (see also [87]) for exactly this purpose. More precisely, for any observed quantum correlation \vec{P} , the method allows one to lower bound the fidelity:

$$\mathcal{F} = \langle \psi | \rho_{\text{SWAP}} | \psi \rangle \quad (20)$$

with the help of an SDP outer approximation of the quantum set \mathcal{Q} (e.g., due to [17, 67, 68]). Here, ρ_{SWAP} is the “swapped” state:

$$\rho_{\text{SWAP}} = \text{tr}_{AB}[\Phi \rho_{AB} \otimes (|00\rangle\langle 00|)_{A'B'} \Phi^+] \quad (21)$$

extracted from the underlying quantum state ρ via some local extraction map Φ , which is a function of the actual POVM elements. Consequently, \mathcal{F} is a function of the entries of the moment matrix $\chi_\ell[\rho]$, discussed below Eq. (9). For details of the method, we refer the readers to [87].

C. Some general remarks

At this point, it is worth noting that for all the three properties \mathcal{P} discussed above—negativity (and hence dimension), entanglement depth, and reference-state fidelity—their DI certification can be achieved via the characterization of some *convex set* $\mathcal{C}_\mathcal{P}$ in the space of correlation vectors $\{\vec{P}\}$. More precisely, for negativity, by turning the objective function of Eq. (9) into the constraint [17]

$$\chi_\ell[\sigma_-]_{\text{tr}} \leq \mathcal{N}_0, \quad (22)$$

we obtain an SDP that characterizes of the set of correlations attainable by quantum states having a negativity upper-bounded by \mathcal{N}_0 . Then, Eq. (10) can be understood as a separating hyperplane relevant for witnessing a negativity larger than \mathcal{N}_0 .

On the other hand, if we drop the constraint in Eq. (9) but imposes additional positive-partial-transposition constraints, then we get an SDP characterization of the set $\mathcal{C}_\mathcal{P}$ having a bounded amount of entanglement depth [19] [see constraints of Eq. (44) below]. In this case, the first two inequalities of Eq. (14) serve as the corresponding witness for entanglement depth. Finally, by demanding $\langle \psi | \rho_{\text{SWAP}} | \psi \rangle \leq \mathcal{F}_0$ together with Eq. (9c), we obtain an SDP characterization of the set $\mathcal{C}_\mathcal{P}$ associated with a swapped state [87] with a $|\psi\rangle$ -fidelity upper bounded by \mathcal{F}_0 . In fact, an SDP characterization can also be obtained for a number of other properties, including genuine negativity [17], steering robustness [22], entanglement robustness [23], (measurement) incompatibility robustness [22, 28], and so on.

III. METHODOLOGIES FOR HYPOTHESIS TESTING

Having understood how DI certification can be achieved from a given correlation \vec{P} , we now proceed to discuss the more realistic setting involving only a finite number of experimental trials. For concreteness, the following presentation assumes an analysis based on the data collected from N trials in a Bell test. Below, we explain our approaches to the problem based on *hypothesis testing*. Our first step is to formulate a *null hypothesis* \mathcal{H} based on the desired property to be certified. For example, to certify that the underlying state has a negativity larger than \mathcal{N}_0 , we formula the (converse) null hypothesis:

Null Hypothesis 1. $\mathcal{H}_{\mathcal{N}(\rho) \leq N_0}$: In every experimental trial, the underlying state has a negativity less than or equal to N_0 .

Since such a hypothesis involves a *set* of (rather than a single) compatible distribution \vec{P} , it is called a *composite hypothesis* [90].

Then, we apply appropriate methods for this kind of hypothesis testing on the collected data to determine an upper bound \mathbf{p} on the p -value associated with the hypothesis \mathcal{H} . Since a p -value quantifies the plausibility of observing the given data when \mathcal{H} holds, a small value of \mathbf{p} , say, less than 5% provides a strong indication that \mathcal{H} is falsified. It then follows that the desired feature corresponding to the negation of \mathcal{H} is certified with a confidence γ of at least $1 - \mathbf{p}$.

Of course, one may also be interested to understand how quickly statistical evidence (against a hypothesis \mathcal{H}) can be gathered when we increase the number of trials. To this end, we also consider the so-called (asymptotic) confidence-gain rate [61], defined by

$$G^{(\text{prot})} := - \lim_{N_{\text{tot}} \rightarrow \infty} \frac{\log_2 p_{N_{\text{tot}}}^{(\text{prot})}}{N_{\text{tot}}}, \quad (23)$$

where $p^{(\text{prot})}$ is the p -value (upper bound) deduced from some protocol (abbreviated as “prot”). From the definition, it is evident that asymptotically, and in the *i.i.d.* setting, a fewer number of trials is required to achieve the same level of statistical confidence if the corresponding $G^{(\text{prot})}$ is higher. Next, let us elaborate on the two hypothesis-testing protocols considered in this work.

A. Martingale-based protocol

We shall start with the martingale-based protocol, pioneered by Gill in [43, 62] for testing against LHV theories, and further developed in [61, 91]. The protocol relies on the observation of the (super)martingale structure in some random variable of interest. To employ the martingale-based protocol, one has to fix a Bell function $I(v)$ in advance. Ideally, $I(v)$ should be chosen such that the Bell-like inequality

$$\langle I(v) \rangle := \frac{\langle \beta_{xy}^{ab} \rangle}{P_{xy}} = \sum_{a,b,x,y} \beta_{xy}^{ab} P(ab|xy) \stackrel{\mathcal{H}}{\leq} B_{\mathcal{H}}(\vec{\beta}) \quad (24)$$

may be violated by some quantum correlation $\vec{P} = \vec{P}_Q$ [cf. Eq. (6)] to be prepared in an experiment.

Let $v_j = (a_j, b_j, x_j, y_j)$ be the value realized for the random variables of the measurement outcomes and settings at the j -th experimental trial, and $I(v_j)$ the corresponding value of Bell function for that trial. Moreover, let $\mathbf{v} = \{v_1, \dots, v_i, \dots, v_N\}$. Then, from the observed average value of $I(v)$ over N trials, i.e., $\hat{I}(\mathbf{v}) = \sum_{j=1}^N \frac{I(v_j)}{N}$, the following p -value upper bound is known [91] to hold whenever $\hat{I} \geq B_{\mathcal{H}}$:

$$p^{(\text{mart})} \leq \left[\left(\frac{\mathbf{b}_+ - B_{\mathcal{H}}}{\mathbf{b}_+ - \hat{I}} \right)^{\frac{\mathbf{b}_+ - \hat{I}}{\mathbf{b}_+ - \mathbf{b}_-}} \left(\frac{B_{\mathcal{H}} - \mathbf{b}_-}{\hat{I} - \mathbf{b}_-} \right)^{\frac{\hat{I} - \mathbf{b}_-}{\mathbf{b}_+ - \mathbf{b}_-}} \right]^N, \quad (25)$$

where, for simplicity, we have suppressed the dependency of $B_{\mathcal{H}}$ on $\vec{\beta}$ (and \hat{I} on \mathbf{v}), while the minimum and maximum value of $I(v)$ over all possible values of $v = (a, b, x, y)$ are

$$\mathbf{b}_- := \inf_v I(v) < B_{\mathcal{H}} \leq \hat{I} < \mathbf{b}_+ := \sup_v I(v). \quad (26)$$

It is worth noting that the martingale-based p -value upper bound of Eq. (25) improves over the one given in [48, 61, 62].

Let I_Q be the expectation value of $I(v)$ when we replace \vec{P} by some \vec{P}_Q capable of violating the Bell-like inequality in Eq. (24). Then, in the *i.i.d.* setting where the experimental data follows the distributions given by \vec{P}_Q , the corresponding asymptotic confidence-gain rate can be deduced from Eq. (23) and Eq. (25) as:

$$G^{(\text{mart})} = \frac{\mathbf{b}_+ - I_Q}{\mathbf{b}_+ - \mathbf{b}_-} \log_2 \frac{\mathbf{b}_+ - I_Q}{\mathbf{b}_+ - B_{\mathcal{H}}} + \frac{I_Q - \mathbf{b}_-}{\mathbf{b}_+ - \mathbf{b}_-} \log_2 \frac{I_Q - \mathbf{b}_-}{B_{\mathcal{H}} - \mathbf{b}_-}. \quad (27)$$

B. The prediction-based-ratio (PBR) protocol

The other hypothesis-testing protocol that we consider in this work is based on the so-called PBR protocol proposed in [61] (see also [91]). In contrast with a martingale-based protocol, the PBR protocol does *not* need to presuppose any Bell-like inequality for determining a p -value bound. Instead, for the data \mathbf{v} collected in N trials, one may start by using the first $N_{\text{est}} < N$ trials from $i = 1, 2, \dots, N_{\text{est}}$ to estimate the relative frequency

$$f(ab|xy) = \frac{N_{\text{est}}(a, b, x, y)}{N_{\text{est}}(x, y)}, \quad (28)$$

where $N_{\text{est}}(x, y) = \sum_{a,b} N_{\text{est}}(a, b, x, y)$ and $N_{\text{est}}(a, b, x, y)$ counts among these N_{est} trials the total number of times the *specific* combination of measurement settings and outcomes (x, y, a, b) occurs.

The key idea of the PBR protocol is to use this relative frequency $\vec{f} = \{f(ab|xy)\}$ to obtain an *optimized Bell-like inequality*³ and apply that to v_i from $i = \{N_{\text{est}} + 1, N_{\text{est}} + 2, \dots, N_{\text{est}} + N_{\text{test}}\}$. To this end, we minimize the Kullback-Leibler (KL) divergence [92] from a regularized relative frequency \vec{f}_{reg} (explained below) to the set $\mathcal{S}_{\mathcal{H}}$ of correlations compatible with \mathcal{H} :

$$D_{\text{KL}}(\vec{f}_{\text{reg}} || \mathcal{S}_{\mathcal{H}}) := \inf_{\vec{P} \in \mathcal{S}_{\mathcal{H}}} \sum_{a,b,x,y} P_{xy} f_{\text{reg}}(ab|xy) \log \frac{f_{\text{reg}}(ab|xy)}{P(ab|xy)}. \quad (29)$$

An important point to note now is that if the composite null hypothesis $\mathcal{S}_{\mathcal{H}}$ is associated with a convex set that admits an

³ Here, the inequality is optimized in the sense that it provides the largest possible asymptotic confidence-gain rate, cf. Eq. (23).

SDP characterization (as discussed in Section II C) like the kind proposed in [17, 22, 67–69], then Eq. (29) is a conic program (see [37]), and thus efficiently solvable using a solver like MOSEK [93].

The unique [37] minimizer $\vec{P}_* \in \mathcal{S}_{\mathcal{H}}$ can then be used to define the *non-negative* prediction-based-ratio (PBR)

$$R(a, b, x, y) := \frac{f_{\text{reg}}(ab|xy)}{P_*(ab|xy)}, \quad (30)$$

which gives the optimized Bell-like inequality $\langle R(v) \rangle \stackrel{\mathcal{H}}{\leq} 1$. Next, we compute the test statistic

$$t(\mathbf{v}) = \prod_j R(a_j, b_j, x_j, y_j) \quad (31)$$

where the product is only carried out over the remaining N_{test} trials. Using arguments completely analogous to those given in [61] for $\mathcal{H} = \mathcal{L}$, it can then be shown that the following upper bound on the p -value holds:

$$p^{(\text{pbr})} \leq \min \left\{ \frac{1}{t(\mathbf{v})}, 1 \right\}. \quad (32)$$

Several remarks are now in order. Firstly, if none of the entries in \vec{f} vanishes, one could also use \vec{f} directly in the optimization problem of Eq. (29). However, for a small N_{est} , a vanishing entry in \vec{f} is almost bound to happen, we thus follow [61] and mix \vec{f} with the *uniform distribution* \vec{P}_{\perp} to obtain:

$$\vec{f} \rightarrow \vec{f}^{\dagger} := \frac{N_{\text{est}}}{N_{\text{est}} + 1} \vec{f} + \frac{1}{N_{\text{est}} + 1} \vec{P}_{\perp}. \quad (33)$$

Next, notice that \vec{f}^{\dagger} typically cannot be cast in the form of Eq. (6). Consequently, we observe empirically that the R obtained by solving Eq. (29) with \vec{f}^{\dagger} in place of \vec{f}_{reg} gives evidently suboptimal performance (see, e.g., Figs. 10, 13 and 15 in Appendix A for some explicit examples). As such, we shall first regularize [37] \vec{f}^{\dagger} to some outer approximation of the quantum set \mathcal{Q}_{ℓ} by solving Eq. (29) with $\mathcal{S}_{\mathcal{H}}$ replaced by \mathcal{Q}_{ℓ} . In our work, \mathcal{Q}_{ℓ} is the level- ℓ outer approximation of the quantum set \mathcal{Q} introduced in [17]. However, one may also consider other approximations [22, 67]. Since all these outer approximations admit an SDP characterization, this regularization process is a conic program (see also [37]). The resulting minimizer, which we call the regularized relative frequency \vec{f}_{reg} is then fed into Eq. (29) to obtain the desired PBR.

Another important feature of the PBR protocol is that the optimized inequality characterized by $\vec{R} = \{R(a, b, x, y)\}$ can be updated as more data is incorporated into the analysis. In principle, one can update \vec{R} as frequently as one desires. However, this is neither necessary nor efficient. As such, we work with blocks of N_{blk} trials. The first block of data is used exclusively for producing the first regularized relative frequency, the first PBR \vec{R}_1 , and by applying to the second block of \mathbf{v} , we get the first test statistic

$$t_1 = \prod_{i=N_{\text{est}}^{(1)}+1}^{N_{\text{est}}^{(1)}+N_{\text{test}}^{(1)}} R_1(a_i, b_i, x_i, y_i), \quad (34)$$

where $N_{\text{test}}^{(k)} = N_{\text{blk}}$ for all k . In the next iteration, we determine the PBR \vec{R}_2 by solving Eq. (29) using \mathbf{v} from the first two blocks, and apply this updated PBR to the third block of \mathbf{v} to get, for $k = 2$,

$$t_k = t_{k-1} \times \prod_{i=N_{\text{est}}^{(k)}+1}^{N_{\text{est}}^{(k)}+N_{\text{test}}^{(k)}} R_k(a_i, b_i, x_i, y_i) \quad (35)$$

where $N_{\text{est}}^{(k)} = kN_{\text{blk}}$. These steps may then repeated iteratively until all the data \mathbf{v} has been consumed in one way or another in the computation of t_k for $k = 3, \dots, \frac{N_{\text{tot}}}{N_{\text{blk}}} - 1$. For a schematic illustration of this procedure, see Fig. 1. Importantly, once the test statistic t_k for each iteration is determined, we can obtain the corresponding p -value bound using Eq. (32).

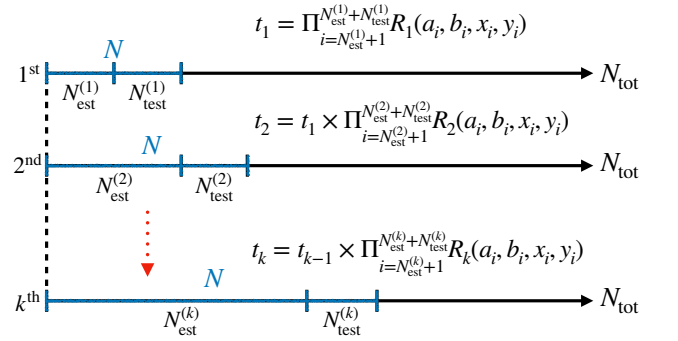


FIG. 1: Instruction for the PBR method.

Finally, note that for an ideal Bell test giving the correlation $\vec{P}_{\mathcal{Q}}$ and a composite hypothesis associated with \mathcal{H} , the PBR protocol has the asymptotic confidence-gain rate:

$$G^{(\text{pbr})} = D_{\text{KL}}(\vec{P}_{\mathcal{Q}} || \mathcal{S}_{\mathcal{H}}), \quad (36)$$

which may be obtained by solving Eq. (29) with \vec{f}_{reg} replaced by $\vec{P}_{\mathcal{Q}}$. The proof is again completely analogous to that given for $\mathcal{H} = \mathcal{L}$ in [61] and is thus omitted.

IV. DEVICE-INDEPENDENT CERTIFICATION WITH A CONFIDENCE INTERVAL

We are now ready to present our simulations results involving a finite number of trials. Throughout this section, the results presented for finite trials consist of an average over 30 complete Bell tests, each involving $N_{\text{tot}} = 10^5$ trials, with the trials partitioned into blocks of size $N_{\text{blk}} = 500$. Moreover, we always consider a uniform distribution for (possibly a restricted set of) measurement settings. In each Bell test, we then simulate the raw data $\mathbf{v} = \{v_i = (a_i, b_i, x_i, y_i)\}_{i=1}^{N_{\text{tot}}}$ using the function `sample.hist` from the `Lightspeed` Matlab toolbox [94]. For the certification with finite data, we set a confidence level of $\gamma = 0.99$. We also present some related confidence-gain rates in the respective subsections.

A. Negativity and dimension certification

1. Negativity certification

Our first example consists of a Bell test based on the quantum strategy presented in Eq. (15), which leads to a CHSH Bell value of $S_{\text{CHSH}} = 2\sqrt{2}$. Using Eq. (10), we know that the resulting quantum correlation \vec{P}_{CHSH} gives a tight negativity lower bound of $\frac{1}{2}$ for a Bell state. From the numerically simulated data, we then perform composite hypothesis testing for Null Hypothesis 1 with $\mathcal{N}_0 \in \{0, 0.01, \dots, 0.50\}$.

Specifically, for the martingale-based protocol, we use Eq. (25) with the CHSH Bell expression of Eq. (4). In this case, $b_{\pm} = \pm 4$ for the chosen P_{xy} while it follows from Eqs. (4), (10) and (24) that

$$\sum_{a,b,x,y} (-1)^{xy+a+b} 4 P(abxy) \stackrel{\mathcal{N} \leq \mathcal{N}_0}{\leq} 2 + 4\mathcal{N}_0(\sqrt{2} - 1). \quad (37)$$

On the other hand, for the PBR protocol, the optimizing distribution $P_{\star}^{(k)}(a, b|x, y)$ for the k -iteration can be obtained by solving [cf. Eq. (29)]

$$\underset{\vec{P}}{\text{argmin}} \quad - \sum_{a,b,x,y} P_{xy} f_{\text{reg}}^{(k)}(ab|xy) \log P(ab|xy), \quad (38a)$$

$$\text{s.t.} \quad \chi_{\ell}[\rho] = \chi_{\ell}[\sigma_{+}] - \chi_{\ell}[\sigma_{-}] \geq 0, \quad \chi_{\ell}[\sigma_{\pm}]^{\text{T}_A} \geq 0, \quad (38b)$$

$$\chi_{\ell}[\rho]_{\text{tr}} = 1, \quad \chi_{\ell}[\sigma_{-}]_{\text{tr}} \leq \mathcal{N}_0 \quad (38c)$$

where $\text{argmin}_{\vec{P}}$ seeks for the argument minimizing the expression in Eq. (38a), $\tilde{f}_{\text{reg}}^{(k)}$ is the regularized frequency obtained for the same iteration, and each $P(ab|xy)$ also appears as an optimization variable in the moment matrix $\chi_{\ell}[\rho]$. Then, the PBR used in the computation of t_k can be evaluated by replacing $f_{\text{reg}}(ab|xy)$ and $P_{\star}(ab|xy)$ in Eq. (30), respectively, by $\tilde{f}_{\text{reg}}^{(k)}(ab|xy)$ and $\tilde{P}_{\star}^{(k)}(ab|xy)$.

In Fig. 2, we show the average amount of certifiable negativity from these two methods as a function of the number of trials N employed. From the figure, it is clear that for certifying the underlying negativity using the data arising from \vec{P}_{CHSH} , the performance of the two protocols are similar. In fact, even though the martingale-based protocol appears to have a slight advantage over the PBR protocol for this certification task for small N 's, our computations of the asymptotic gain-rates $G^{(\text{pbr})}$ and $G^{(\text{mart})}$ show that they in fact agree (for all these values of \mathcal{N}_0 that we have considered), up to a numerical precision of 10^{-7} . Also, in both cases, we see that with about 5×10^3 and 2×10^4 trials, we can already certify, respectively, more than 80% and 90% of the underlying negativity with a confidence $\gamma \geq 0.99$. In Appendix A 1, we provide some additional plots showing how the p -value bound changes with N for several values of \mathcal{N}_0 .

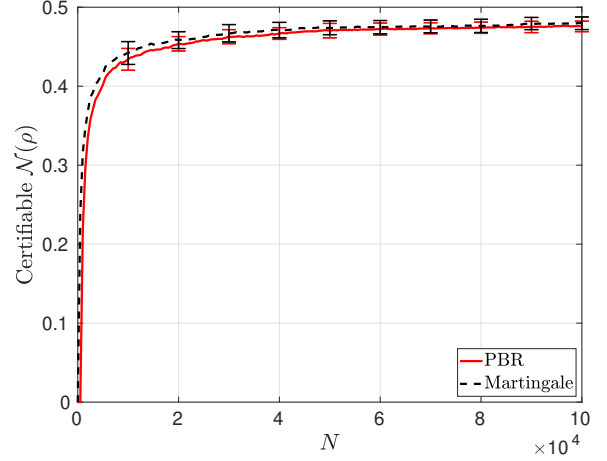


FIG. 2: Negativity certifiable from the data observed in a Bell test generating \vec{P}_{CHSH} . For the martingale-based protocol and any given \mathcal{N}_0 among $\mathcal{N}_0 = \{0, 0.01, \dots, 0.49\}$, we use Eq. (37) in Eq. (25) to upper-bound $p^{(\text{mart})}$ after every block of $N_{\text{blk}} = 500$ trials, thereby generating 200×50 upper bounds on $p^{(\text{mart})}$ for a complete Bell test. For the PBR protocol and a given \mathcal{N}_0 from \mathcal{N}_0 , we solve Eq. (38) by considering the same block size and the level-3 outer approximation of \mathcal{Q} introduced in [17]. Then, we obtain 199×50 upper bounds on $p^{(\text{pbr})}$ from Eqs. (32), (34) and (35). To determine the lower bound on the underlying $\mathcal{N}(\rho)$ with the desired confidence of $\gamma \geq 99\%$, we look for the largest \mathcal{N}_0 in \mathcal{N}_0 such that $\mathcal{H}_{\mathcal{N}(\rho) \leq \mathcal{N}_0}$ is rejected with a p -value bound being less than or equal to 0.01. Each data point shown in the plot is an average over 30 such lower bounds, and the error bar (standard deviation) gives an indication of the spread of the certifiable negativity. To avoid cluttering the plots, in each line, we show only a small number of markers.

These results clearly suggest that the CHSH Bell function of Eq. (4) is optimal for certifying the underlying negativity of $|\psi_{\text{MES}}\rangle$ using the martingale-based protocol. Indeed, a separate computation of Eq. (29) and Eq. (30) using \vec{P}_{CHSH} in place of \tilde{f}_{reg} show that, within a precision of 10^{-4} , the optimized Bell-like inequality for $\mathcal{N}_0 = 0, 0.05, \dots, 1$ is equivalent to Eq. (10). How would things change if we perform a DI negativity certification using the data generated from the partially entangled state $|\psi(\theta)\rangle$, Eq. (17a)? To this end, consider the quantum strategy of Eq. (17), whose resulting correlation \vec{P}_{θ} gives the maximal Bell CHSH violation for $|\psi(\theta)\rangle$, as well as the maximal violation of the tilted CHSH Bell inequality of Eq. (19). Then, instead of repeating the same analysis, we show in Fig. 3 the confidence-gain rates due to both protocols for certifying several *given fractions* of the underlying negativity. From the plots shown, it is evident that asymptotically, the martingale-based protocol employing the CHSH Bell function is far from optimal for certifying the underlying negativity of $|\psi(\theta)\rangle$. Indeed, the PBR protocol could identify some other Bell-like inequality that gives a much better confidence-gain rate, especially for the correlations arising from $|\psi(\theta)\rangle$ that is weakly entangled (small θ). To a large ex-

tent, this can be understood by noting that the negativity lower bound of Eq. (10) due to its CHSH Bell violation is generally far from tight for these states, see Fig. 12 in Appendix A.

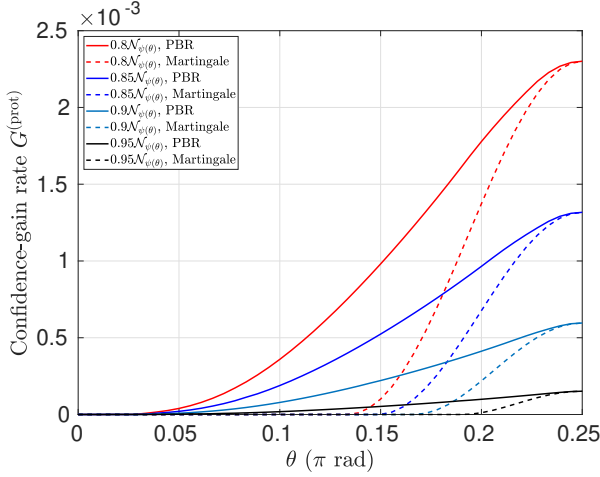


FIG. 3: Asymptotic confidence-gain rate $G^{(\text{prot})}$ based on the family of quantum correlations \vec{P}_θ derived from Eq. (17), where $\theta = \frac{k\pi}{180}$ rad, $k = \{1, 2, \dots, 45\}$, parametrizes the two-qubit entangled state $|\psi(\theta)\rangle = \cos \theta |00\rangle + \sin \theta |11\rangle$. Here, we again consider the composite hypothesis 1, with $\mathcal{N}_0 = \{0.8\mathcal{N}_{\psi(\theta)}, 0.85\mathcal{N}_{\psi(\theta)}, 0.9\mathcal{N}_{\psi(\theta)}, 0.95\mathcal{N}_{\psi(\theta)}\}$ and $\mathcal{N}_{\psi(\theta)}$ being the negativity of $\psi(\theta)$, see Fig. 12. The gain rate for the martingale-based protocol is computed from Eq. (27) using the CHSH Bell-like inequality of Eq. (37) whereas that for the PBR protocol is evaluated from Eq. (36) using the correlation derived from Eq. (17).

2. Dimension certification via negativity certification

As mentioned in Section II B 1, the correlation \vec{P}_{CHSH} is insufficient to demonstrate any nontrivial dimension bound. Let us consider, instead, a correlation \vec{P}_{CGLMP} derived by local measuring the partially entangled two-qutrit state

$$|\Psi\rangle = \frac{1}{\sqrt{2 + \zeta^2}}(|00\rangle + \zeta|11\rangle + |22\rangle), \quad \zeta = \frac{1}{2}(\sqrt{11} - \sqrt{3}), \quad (39a)$$

with the local measurements

$$M_{a|x}^{(A)} = |a\rangle_{A,x} \langle a|, \quad M_{b|y}^{(B)} = |b\rangle_{B,y} \langle b|, \quad (39b)$$

$$|a\rangle_{A,x} = \sum_{j=0}^2 \frac{\omega^{j(\varphi_x^A + a)}}{\sqrt{3}} |j\rangle, \quad |b\rangle_{B,y} = \sum_{j=0}^2 \frac{\omega^{j(\varphi_y^B - b)}}{\sqrt{3}} |j\rangle,$$

where $\varphi_x^A = \frac{\pi}{2}$, $\varphi_y^B = (-1)^y \frac{1}{4}$, and $\{|j\rangle\}$ is the set of computational basis states. It is known [68, 72] that this strategy gives the maximal CGLMP Bell-inequality violation of $\mathcal{S}_{\text{CGLMP3}} = 1 + \sqrt{\frac{11}{3}} \cong 2.91485$. Moreover, the negativity of $|\Psi\rangle$ can be easily evaluated to give $\cong 0.98358$.

Next, we use the numerically data simulated from \vec{P}_{CGLMP} to perform a hypothesis testing for Null Hypothesis 1, but

now with $\mathcal{N}_0 \in \mathcal{N}_0 = \{0.5, 0.51, \dots, 0.98\}$. For the PBR protocol, the computation proceeds in exactly the same way as described above [see the paragraph containing Eq. (38)]. However, for the martingale-based protocol, since we do not have an explicit expression like that shown in Eq. (37) for the CGLMP Bell expression, we compute an upper bound on $B_{\mathcal{H}}$ for each given value of $\mathcal{N}_0 \in \mathcal{N}_0$ according to:

$$\max \sum_{a,b,x,y} \beta_{xy}^{ab} P(ab|xy) \quad (40a)$$

$$\text{s.t. } \chi_\ell[\rho] = \chi_\ell[\sigma_+] - \chi_\ell[\sigma_-], \quad \chi_\ell[\sigma_\pm]^T \bar{A} \succeq 0, \quad (40b)$$

$$\chi_\ell[\rho] \succeq 0, \quad \chi_\ell[\rho]_{\text{tr}} = 1, \quad \chi_\ell[\sigma_-]_{\text{tr}} \leq \mathcal{N}_0, \quad (40c)$$

where the CGLMP Bell coefficients β_{xy}^{ab} are defined in Eq. (11). Meanwhile, since $P_{xy} = \frac{1}{4}$ and $\beta_{xy}^{ab} \in \{-1, 0, 1\}$, we again have $b_\pm = \pm 4$ for I_{CGLMP3} .

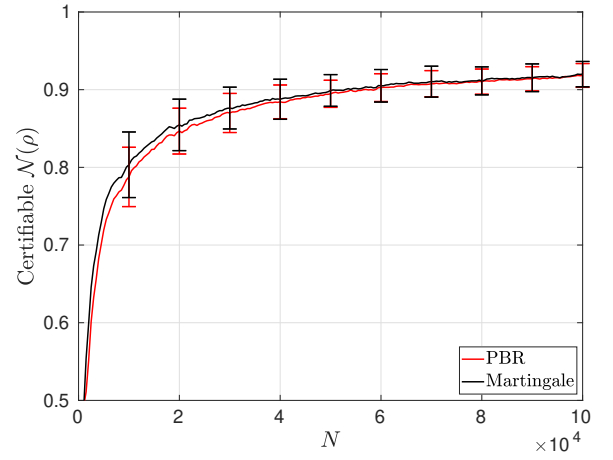


FIG. 4: Negativity certifiable from the data observed in a Bell test generating \vec{P}_{CGLMP} . For details on how the plot is generated, see the caption of Fig. 2, but bearing in mind that for the martingale-based protocol, we now use in Eq. (25) the $B_{\mathcal{H}}$ determined from Eq. (40).

From Fig. 4, we see that with about 6×10^4 trials we can already certify a negativity lower bound of 0.9. On the other hand, if we want to certify that we need at least a *two-qutrit state* to produce the observed data (arising from \vec{P}_{CGLMP}), it suffices to certify that the underlying negativity is *strictly* larger than 0.50, which happens already with approximately 1500 trials. Could other two-qutrit states provide a more favorable correlation in this regard? To gain insight into the problem, we consider the following one-parameter family of two-qutrit states

$$|\Psi(\tilde{\zeta})\rangle = \frac{1}{\sqrt{2 + \tilde{\zeta}^2}}(|00\rangle + \tilde{\zeta}|11\rangle + |22\rangle) \quad (41)$$

and numerically maximize their CGLMP Bell-inequality violation using the heuristic algorithm given in [95]. We denote the corresponding correlation by $\vec{P}_{\tilde{\zeta}}$, compute the corresponding asymptotic confidence-gain rate for both protocols, and plot the results in Fig. 5.

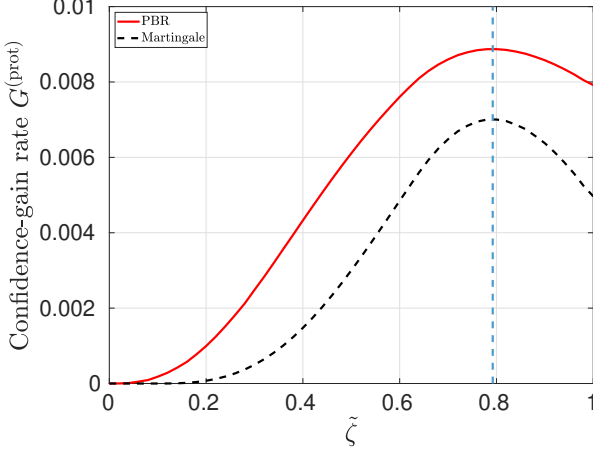


FIG. 5: Asymptotic confidence-gain rate $G^{(\text{prot})}$ based on the quantum correlation $\vec{P}_{\tilde{\zeta}}$ derived from maximizing the CGLMP Bell-inequality-violation of $|\Psi(\tilde{\zeta})\rangle$, Eq. (41). Here, we consider $\tilde{\zeta} = \{0, 0.02, \dots, 1\}$ and the composite hypothesis 1, with $N_0 = 0.5$. The gain rate for the martingale-based protocol is computed from Eq. (27) using the CGLMP Bell function and the bound $B_{\mathcal{H}}$ determined in Eq. (40) whereas that for the PBR protocol is evaluated from Eq. (36) and setting $\vec{P}_{\mathcal{Q}}$ to $\vec{P}_{\tilde{\zeta}}$. The dashed blue line corresponds to the value $\tilde{\zeta} = \zeta$, cf. Eq. (39a).

Interestingly, even though Fig. 4 suggests that the CGLMP Bell function is very effective in providing a good p -value bound against hypothesis 1, Fig. 5 clearly show that, asymptotically, it is not optimal. The results shown in Fig. 5 further suggest that among the family of two-qutrit states given in Eq. (41), the qutrit signature of $|\Psi(\zeta)\rangle$, cf. Eq. (39a), could even be the most prominent, when it comes to its DI certification using these hypothesis-testing techniques.

B. Entanglement depth certification

Next, we consider the tripartite correlation \vec{P}_{GHZ} that results from locally measuring the ± 1 -eigenvalue observables

$$A_0 = B_0 = C_0 = \sigma_y, \quad A_1 = B_1 = C_1 = -\sigma_x \quad (42a)$$

on the Greenberger-Horne-Zeilinger (GHZ) state [96, 97]:

$$|\text{GHZ}\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle). \quad (42b)$$

It is easy to verify that \vec{P}_{GHZ} leads to a violation of the Mermin Bell inequality, Eqs. (12) and (14), giving the algebraic maximum of $\mathcal{S}_{\text{Mermin}} = 4$. For our simulations, we assume a uniform distribution $P_{xyz} = \frac{1}{4}$ over all measurement settings $x, y, z \in \{0, 1\}$ that satisfy $\text{mod}(x + y + z, 2) = 1$. Then, we test the data against the following composite hypotheses.

Null Hypothesis 2. \mathcal{H}_{Sep} : In every experimental trial, the underlying state is separable (having an entanglement depth of 1).

Null Hypothesis 3. $\mathcal{H}_{2\text{-prod}}$: In every experimental trial, the underlying state is 2-producible, i.e., having an entanglement depth of 2 or less.

For the martingale-based method, we use Eq. (25) with the Mermin Bell expression of Eq. (12) and the bounds given in Eq. (14), i.e., $B_{\mathcal{H}} = 2$ for hypothesis 2 and $B_{\mathcal{H}} = 2\sqrt{2}$ for hypothesis 3. Since $\beta_{xyz}^{abc} \in \{-1, 0, 1\}$, we again have $b_{\pm} = \pm 4$. Note that separable states can only generate Bell-local correlations [5], cf. Eq. (2). Thus, for the PBR protocol with hypothesis 2, the optimizing distribution $P_{\star}^{(k)}(abc|xyz)$ for the k -iteration can be obtained by solving [cf. Eq. (29)]

$$\underset{\vec{P}}{\text{argmin}} - \sum_{a,b,c,x,y,z} P_{xyz} f_{\text{reg}}^{(k)}(abc|xyz) \log P(abc|xyz), \quad (43a)$$

$$\text{s.t. } \vec{P} = \sum_i q_i \vec{D}_i, \quad q_i \geq 0, \quad \sum_i q_i = 1 \quad (43b)$$

where \vec{D}_i is the i -th (local deterministic) extreme points of the set of tripartite Bell-local distributions.

On the other hand, notice that 2-producibility [77] is equivalent to biseparability [14] in the tripartite scenario. Hence, for hypothesis 3, we obtain the corresponding optimizing distribution by solving

$$\underset{\vec{P}}{\text{argmin}} - \sum_{a,b,c,x,y,z} P_{xyz} f_{\text{reg}}^{(k)}(abc|xyz) \log P(abc|xyz), \quad (44a)$$

$$\text{s.t. } \chi_{\ell}[\rho] = \chi_{\ell}[\rho_1] + \chi_{\ell}[\rho_2] + \chi_{\ell}[\rho_3], \quad \chi_{\ell}[\rho] \geq 0, \quad (44b)$$

$$\chi_{\ell}[\rho]_{\text{tr}} = 1, \quad \chi_{\ell}[\rho_i] \geq 0, \quad \forall i \in \{1, 2, 3\}, \quad (44c)$$

$$\chi_{\ell}[\rho_1]^{\text{T}_A} \geq 0, \chi_{\ell}[\rho_2]^{\text{T}_B} \geq 0, \chi_{\ell}[\rho_3]^{\text{T}_C} \geq 0, \quad (44d)$$

where ρ_i , with $i = \{1, 2, 3\}$ are meant to represent, respectively, the constituent of ρ that is separable with respect to the $A|BC$, $B|AC$, and $C|AB$ bipartitions. In evaluating Eq. (44), we use level $\ell = 1$ of the hierarchy introduced in [17]. For both hypotheses, we then evaluate

$$R(a, b, c, x, y, z) = \frac{f_{\text{reg}}^{(k)}(abc|xyz)}{P_{\star}(abc|xyz)} \quad (45)$$

for the computation of the test statistic t_k .

For \vec{P}_{GHZ} and hypothesis 2, the confidence-gain rate $G^{(\text{pbr})}$ is already known (see Table I of [90]) to be approximately 0.415037; our computation reproduces this and further shows that for hypothesis 3, this is approximately 0.228446. Moreover, to six decimal places, $G^{(\text{pbr})}$ and $G^{(\text{mart})}$ agree for both hypotheses. What about finite data? Based on the average results from 30 simulations, we find that the p -value bounds, or more precisely, $\mathfrak{P} = -\log_2 p^{(\text{prot})}$ can be very well fitted into the following straight lines:⁴

$$\mathfrak{P}_{\mathcal{H}_{\text{Sep}}}^{(\text{pbr})} = 0.414958 N - 204.978, \quad N \in [10^3, 10^5], \quad (46)$$

$$\mathfrak{P}_{\mathcal{H}_{\text{Sep}}}^{(\text{mart})} = 0.415037 N, \quad N \in [0, 10^5] \quad (47)$$

⁴ In all these fits, the coefficient of determination R^2 is 1 even if we keep up to 7 significant digits.

for the separable hypothesis \mathcal{H}_{Sep} of 2, and

$$\mathfrak{P}_{\mathcal{H}_{2\text{-prod.}}}^{(\text{pbr})} = 0.22838 N - 115.22, \quad N \in [10^3, 10^5], \quad (48)$$

$$\mathfrak{P}_{\mathcal{H}_{2\text{-prod.}}}^{(\text{mart})} = 0.228447 N, \quad N \in [0, 10^5] \quad (49)$$

for the 2-producible hypothesis $\mathcal{H}_{2\text{-prod.}}$ of 3. Consequently, based on this interpolation, even if we only run the Bell test using the strategy of Eq. (42) for 100 trials, there is already sufficient data to certify genuine tripartite-entanglement with a confidence of at least $1 - 10^{-6}$.

C. Fidelity certification

Our last examples concern the DI certification of a lower bound on the fidelity of the swapped state ρ_{SWAP} with respect to the target state $|\psi(\theta)\rangle$ of Eq. (17a). To this end, we use the same set of data generated for the analysis in Section IV A 1 and consider the following null hypothesis.

Null Hypothesis 4. $\mathcal{H}_{\mathcal{F}_\theta(\rho_{\text{SWAP}}) \leq \mathcal{F}_0}$: In every experimental trial, the swapped state ρ_{SWAP} extractable from the underlying state ρ has a $|\psi(\theta)\rangle$ -fidelity upper bounded by \mathcal{F}_0 , i.e.,

$$\mathcal{F}_\theta(\rho_{\text{SWAP}}) := \langle \psi(\theta) | \rho_{\text{SWAP}} | \psi(\theta) \rangle \leq \mathcal{F}_0. \quad (50)$$

Then, for any given θ and $\mathcal{F}_0 \geq \cos \theta$, to apply the PBR protocol, we solve the optimizing distribution $P_\star^{(k)}(ab|xy)$ for the k -iteration [cf. Eq. (29)] via:

$$\underset{\vec{P}}{\text{argmin}} \quad - \sum_{a,b,x,y} P_{xy} f_{\text{reg}}^{(k)}(ab|xy) \log P(ab|xy), \quad (51a)$$

$$\text{s.t.} \quad \chi_\ell[\rho] \succeq 0, \quad \chi_\ell[\rho]_{\text{tr}} = 1, \quad \mathcal{F}_\theta(\rho_{\text{SWAP}}) \leq \mathcal{F}_0, \quad (51b)$$

where the left-hand side of the last inequality in Eq. (51b) consists of some specific linear combination of entries of $\chi_\ell[\rho]$, see [87] for details. Then, as with negativity certification, we can evaluate the PBR used in the computation of t_k by replacing $f_{\text{reg}}(ab|xy)$ and $P_\star(ab|xy)$, respectively, by $f_{\text{reg}}^{(k)}(ab|xy)$ and $P_\star^{(k)}(ab|xy)$ in Eq. (30). As for the martingale-based protocol, we first solve

$$\max \sum_{a,b,x,y} \beta_{xy}^{ab} P(ab|xy) \quad (52a)$$

$$\text{s.t.} \quad \chi_\ell[\rho] \succeq 0, \quad \chi_\ell[\rho]_{\text{tr}} = 1, \quad \mathcal{F}_\theta(\rho_{\text{SWAP}}) \leq \mathcal{F}_0 \quad (52b)$$

to determine $B_{\mathcal{H}}$ for hypothesis 4 and then apply Eq. (25) to determine the corresponding p -value upper bound.

Let us start with the self-testing of a Bell state, corresponding to $\theta = \frac{\pi}{4}$ in Eq. (17a). In this case, we use the CHSH Bell function specified in Eq. (4) and consider $\mathcal{F}_0 \in \mathbf{F}_0 = \{0.5, 0.51, \dots, 0.99\}$. For both protocols, by systematically evaluating the p -value bounds from the data for each of these \mathcal{F}_0 's, we determine a lower bound on $\mathcal{F}_{\theta=\frac{\pi}{4}}(\rho_{\text{SWAP}})$ with the desired confidence of at least 99%. The results obtained from both hypothesis-testing protocols are shown in Fig. 6.

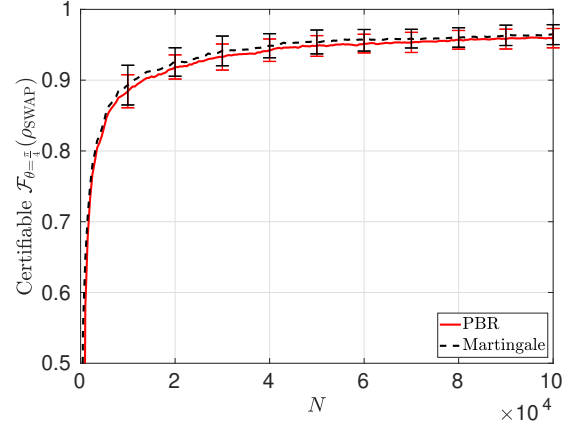


FIG. 6: Certifiable fidelity $\mathcal{F}_{\theta=\frac{\pi}{4}}(\rho_{\text{SWAP}})$ from the data observed in a Bell test generating \vec{P}_{CHSH} . For the martingale-based protocol and any given \mathcal{F}_0 among $\mathbf{F}_0 = \{0.5, 0.51, \dots, 0.99\}$, we use the $B_{\mathcal{H}}$ determined from Eq. (52) in Eq. (25) to upper-bound $p^{(\text{mart})}$ after every block of $N_{\text{blk}} = 500$ trials, thereby generating 200×50 upper bounds on $p^{(\text{mart})}$ for a complete Bell test. For the PBR protocol and a given \mathcal{F}_0 from \mathbf{F}_0 , we solve Eq. (51) by considering the same block size and the level-2 outer approximation of \mathcal{Q} introduced in [17]. Then, we obtain 199×50 upper bounds on $p^{(\text{pbr})}$ from Eqs. (32), (34) and (35). To determine the lower bound on the underlying $\mathcal{F}_{\theta=\frac{\pi}{4}}(\rho_{\text{SWAP}})$ with the desired confidence of $\gamma \geq 99\%$, we look for the largest \mathcal{F}_0 in \mathbf{F}_0 such that $\mathcal{H}_{\mathcal{F}_\theta(\rho_{\text{SWAP}}) \leq \mathcal{F}_0}$ is rejected with a p -value upper bound less than or equal to 0.01.

Interestingly, our results show that the martingale-based protocol with the CHSH Bell function of Eq. (4) again performs very well for the self-testing of a Bell state with finite statistics, even though our computation of the corresponding asymptotic confidence-gain rate for $\mathcal{F}_0 = 0.5$ clearly shows that it is suboptimal even for the Bell state, see Fig. 7. What about other partially entangled states? To answer this question, we evaluate the confidence-gain rate derived from both protocols for⁵ $\mathcal{F}_0 = \cos^2 \theta$ with $\theta = \{0^\circ, 1^\circ, 2^\circ, \dots, 45^\circ\}$. This time around, for the martingale-based protocol, we switch to the Bell function of the tilted CHSH Bell inequality of Eq. (19), which is known to facilitate the self-testing of all entangled $|\psi(\theta)\rangle$. The corresponding results are shown in Fig. 7.

⁵ For $|\psi(\theta)\rangle$ defined in Eq. (17a), a fidelity of $\cos^2 \theta$ is always achievable even if Alice and Bob does not share any entanglement; they merely have to prepare $|00\rangle$ using local operations and classical communication before the Bell test.

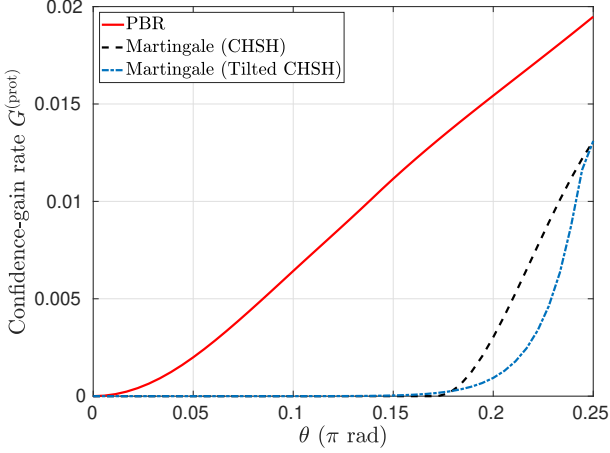


FIG. 7: Asymptotic confidence-gain rate $G^{(\text{prot})}$ based on the family of quantum correlations \vec{P}_θ derived from Eq. (17), where $\theta = \frac{k\pi}{180}$ rad, $k = \{1, 2, \dots, 45\}$. Here, we consider Null Hypothesis 4, with $\mathcal{F}_\theta(\rho_{\text{SWAP}}) = \cos^2 \theta$, the trivial fidelity achievable without shared entanglement. The gain rate for the martingale-based protocol is computed from Eq. (27) using the CHSH Bell function of Eq. (4) (dashed line, black) and the tilted CHSH Bell function of Eq. (19) (dashed-dotted line, blue) whereas that for the PBR protocol is evaluated from Eq. (36).

D. Properties certification via Bell-value certification

The advantage of a fidelity certification based on the SWAP method [87, 89] is that the technique is applicable to a general Bell scenario. However, in the simplest CHSH Bell scenario, it is known that a much tighter lower bound on the Bell-state fidelity can be obtained by considering a more general extraction map. Specifically, Kaniewski showed in [98] that

$$\max_{\Lambda_A, \Lambda_B} \min_{\rho_{AB}} \mathcal{F}[(\Lambda_A \otimes \Lambda_B)(\rho_{AB}), |\psi_{\text{MES}}\rangle\langle\psi_{\text{MES}}|] \geq \frac{1}{2} + \frac{1}{2} \frac{\mathcal{S}_{\text{CHSH}} - \beta^*}{2\sqrt{2} - \beta^*}, \quad (53)$$

where Λ_A, Λ_B are local extraction maps acting, respectively, on Alice's and Bob's subsystem, while $\beta^* := \frac{16+14\sqrt{2}}{17} \approx 2.1058$ is the threshold CHSH value for which the fidelity bound becomes trivial.

To take advantage of Eq. (53), we can first perform a hypothesis testing based on the following null hypothesis.

Null Hypothesis 5. $\mathcal{H}_{\mathcal{S}_{\text{CHSH}}(\rho) \leq \mathcal{S}_0}$: In every experimental trial, the underlying state and measurements give a CHSH value $\mathcal{S}_{\text{CHSH}}$ less than or equal to \mathcal{S}_0 .

Specifically, using the same set of data generated for the analysis in Section IV A 1 and Section IV C, we perform composite hypothesis testing for Null Hypothesis 5 with $\mathcal{S}_0 \in \{2, 2+\Delta\mathcal{S}, 2+2\Delta\mathcal{S}, \dots, 2\sqrt{2}-\Delta\mathcal{S}\}$, where $\Delta\mathcal{S} = \frac{2(\sqrt{2}-1)}{50}$.

In particular, for the martingale-based protocol, we can simply use Eq. (25) with $b_\pm = \pm 4$ and $B_{\mathcal{H}} = \mathcal{S}_0$. Could one also employ the PBR protocol, which does not usually presuppose any Bell-like inequality, for the current hypothesis testing? This is indeed possible. To this end, one may solve the optimizing distribution $P_\star^{(k)}(ab|xy)$ for the k -iteration [cf. Eq. (29)] of the PBR protocol via:

$$\underset{\vec{P}}{\text{argmin}} - \sum_{a,b,x,y} P_{xy} f_{\text{reg}}^{(k)}(ab|xy) \log P(ab|xy), \quad (54a)$$

$$\text{s.t.} \quad \sum_{a,b,x,y} (-1)^{xy+a+b} P(ab|xy) \leq \mathcal{S}_0, \quad (54b)$$

with or without imposing the SDP constraints of Eq. (9c). The results obtained from these tests are shown in Fig. 8.

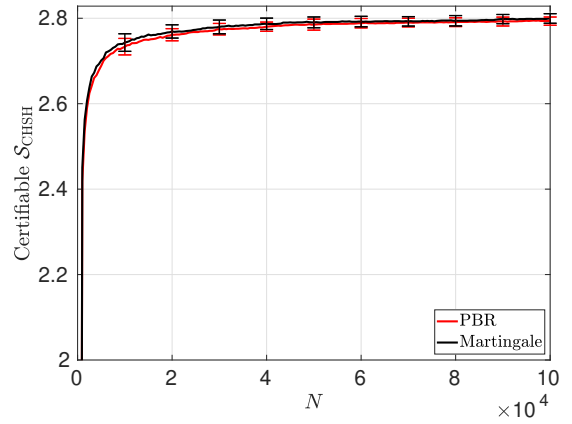


FIG. 8: Certifiable Bell-CHSH violation $\mathcal{S}_{\text{CHSH}}$ from the data observed in a Bell test generating \vec{P}_{CHSH} . For the martingale-based protocol and any given \mathcal{S}_0 among $\mathcal{S}_0 = \{2+k\Delta\mathcal{S}\}_{k=0}^{49}$, we use Eq. (54b) in Eq. (25) to upper-bound $p^{(\text{mart})}$ after every block of $N_{\text{blk}} = 500$ trials, thereby generating 200×50 upper bounds on $p^{(\text{mart})}$ for a complete Bell test. For the PBR protocol and a given \mathcal{S}_0 from \mathcal{S}_0 , we solve Eq. (54) by considering the same block size. Then, we obtain 199×50 upper bounds on $p^{(\text{pbr})}$ from Eqs. (32), (34) and (35). To determine the lower bound on the underlying $\mathcal{S}_{\text{CHSH}}$ with the desired confidence of $\gamma \geq 99\%$, we look for the largest \mathcal{S}_0 in \mathcal{S}_0 such that $\mathcal{H}_{\mathcal{S}_{\text{CHSH}} \leq \mathcal{S}_0}$ is rejected with a p -value bound being less than or equal to 0.01. A separate calculation shows that if we impose Eq. (9c) in addition to Eq. (54b), one may find a visually indistinguishable difference ($< 5 \times 10^{-4}$) in the certifiable $\mathcal{S}_{\text{CHSH}}$.

Using each lower bound on $\mathcal{S}_{\text{CHSH}}$ certified from the data, Eq. (53) immediately translates to a lower bound on the Bell-state fidelity with the desired confidence. For a direct comparison with the efficacy of the SWAP-based approach adopted in Section IV C, we plot in Fig. 9 the Bell-state fidelity certifiable using the two approaches. As expected, the tighter Bell-state fidelity lower bound provided by Eq. (53) also facilitates a considerably tighter lower bound when one has access to only a finite amount of data.

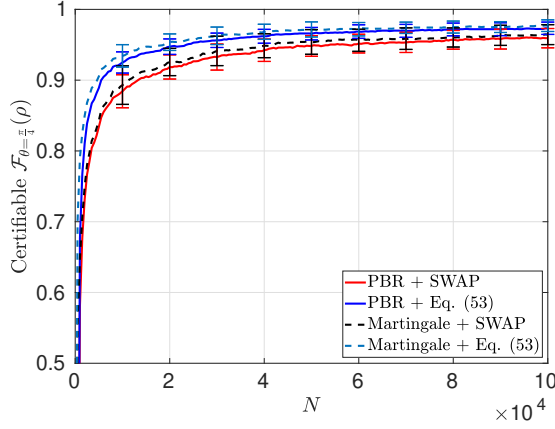


FIG. 9: Comparison of the Bell-state fidelity certifiable via the SWAP-based approach presented in Section IV C and that via Eq. (54) and $\mathcal{S}_{\text{CHSH}}$ -value certification.

It is also worth noting that in computing these PBR bounds, the computation may be further simplified by regularizing the relative frequency \hat{f}_{reg} using *only* the nonsignaling constraint of Eq. (7), instead of the quantum approximation via Eq. (9c). For the lower bounds on $\mathcal{S}_{\text{CHSH}}$ presented in Fig. 8, this further simplification was found to give, unsurprisingly, a worse lower bound, but with a deviation bounded by 8×10^{-3} . Of course, the lower bounds on $\mathcal{S}_{\text{CHSH}}$ can also be used to bound other desired properties. For example, Fig. 2 can equivalently be obtained by combining Eq. (10) with the results shown in Fig. 8.

V. DISCUSSION

Tomography and witnesses are two commonly employed toolkits for certifying the desirable properties of quantum devices [99]. In recent years, the device-independent (DI) paradigm has offered an appealing alternative to these conventional means as it involves only a minimal set of assumptions. Nonetheless, many DI certification schemes, e.g., [16–19, 21–27, 29, 30], implicitly assumes that the underlying quantum correlation \vec{P}_Q (or the actual Bell-inequality violation due to \vec{P}_Q) is known. In practice, this is unrealistic for two reasons: (1) we always have access to only a finite amount of experimental data, and (2) actual experimental trials are typically *not* independent and identically distributed (*i.i.d.*).

To this end, very specialized tools have been developed for the task of randomness generation, quantum key distributions, and the self-testing [87, 100, 101] of quantum states. Among them, the possibility of using hypothesis testing (based on the PBR protocol [61]) for self-testing with finite data was first discussed in [87] (see also [100] for a different approach). Meanwhile, it is long known [43, 62, 91] that hypothesis testing in a Bell test can also be carried out using a martingale-based protocol. Here, we demonstrate the viability and versatility of such hypothesis-testing-based approaches for the general problem of DI certification.

Central to our finding is the observation that many desirable quantum properties \mathcal{P} that one wishes to certify can be characterized by (the *complement* of) some convex set $\mathcal{C}_{\mathcal{P}}$ in the space of correlation vectors $\{\vec{P}\}$. In other words, if a given \vec{P}_Q lies outside $\mathcal{C}_{\mathcal{P}}$, a Bell-like inequality can be provided to witness this fact. This separating hyperplane then provides the basis for our martingale-based protocol for DI certification. On the other hand, if $\mathcal{C}_{\mathcal{P}}$ itself admits a semidefinite programming characterization like the kind proposed in [17, 22, 67–69], then the problem of minimizing the statistical distance to $\mathcal{C}_{\mathcal{P}}$ can be cast as a conic program, which can be readily solved using existing solvers, such as MOSEK [93]. In turn, the PBR protocol provides an *optimized Bell-like inequality* that facilitates the corresponding hypothesis testing.

In this paper, we explain in details how the two aforementioned hypothesis-testing protocols can be adapted for DI certification of desirable properties. Specifically, we illustrate how we can use them to perform DI certification of the underlying negativity [65], local Hilbert space dimension [15], entanglement depth [77, 78], and fidelity to some target two-qubit entangled pure state $|\psi(\theta)\rangle$. In each of these examples, we further demonstrate how the certifiable property (with a confidence of 99%) varies with the number of experimental trials involved, see Fig. 2, Fig. 4, Fig. 6. Even though we have focused on certifying desirable properties of quantum states, as explained above, it should be clear that the protocols can also be applied to certify desirable properties of the measurement devices [22, 23, 28], etc.

In the *i.i.d.* setting, the PBR protocol is known to be asymptotically optimal (in terms of its confidence-gain rate). However, we see from Figs. 2, 4 and 6 that for a relatively small number of trials and with the *right choice* of the Bell function, the martingale-based protocol performs equally well, if not better. A similar observation was also noted in [102] where the authors therein compare the PBR method with the Chernoff-Hoeffding bound in determining the success probability of Bernoulli trials. In our case, this is not surprising as the PBR method does not presuppose a Bell-like inequality but rather sacrifices some of the data to determine one. Indeed, if we equip the PBR protocol with the optimized Bell-like inequality right from the beginning, its performance is, as expected, no worse than the martingale-based protocol. See Figs. 10, 11 and 13 to 16 in Appendix A for some explicit examples.

Meanwhile, we also see from Figs. 3, 5 and 7 that for several cases that we have investigated, one’s intuitive choice of the Bell function for the martingale-based method can lead to a relatively poor confidence-gain rate, and hence impairs its efficiency to produce a good p -value bound, see Figs. 14 and 16. For example, even though the titled CHSH Bell inequality of Eq. (19) is known to self-test *all* entangled two-qubit pure states $|\psi(\theta)\rangle$, this choice of the Bell function in the martingale-based method leads to a worse performance (for bounding the target-state-fidelity) compared with using the CHSH Bell function, which, in turn, gives a suboptimal performance compared with that derived from the PBR protocol, see Fig. 7. At this point, it is worth reiterating that both protocols do *not* require the assumption that the experimental trials are *i.i.d.*, even though we have only given, for simplicity,

examples with *i.i.d.* trials.

Several research directions naturally follow from the present work. Firstly, even though our hypothesis-testing-based approaches enable rigorous DI certification with a confidence interval, by virtue of the techniques involved, one can only make a rather *weak* certification: out of the many experimental trials, we can be sure that *at least one* involves a setup that exhibits the desired property (say, with 99% confidence). This is evidently far from satisfactory. A preferable certification scheme should allow one to comment on the general or *average* behavior of all the measured samples, as has been achieved in [60, 101] for self-testing.

Given that self-testing with a high fidelity is technically challenging, it is still of interest to devise a *general recipe* for certifying the average behavior of other more specific properties (such as entanglement, steerability, etc.), which may already be sufficient for the specific information processing task at hand. However, note that the rejection of a null hypothesis on the *average behavior* (e.g., average negativity $\overline{\mathcal{N}(\rho)} \leq \mathcal{N}_0$) necessarily entails the rejection of the corresponding null hypothesis *for all trials* (e.g., $\mathcal{N}(\rho) \leq \mathcal{N}_0$ in *every trial*). Thus, we may expect a tradeoff when switching from the current kind of hypothesis testing to that for an average behavior.

Also worth noting that if the *i.i.d.* assumption is somehow granted, then our protocols also certify the quality of the setup for every single runs, including those that have not been measured. In this case, once a sufficiently small *p*-value bound is obtained, one can stop measuring the rest of the systems and use them, instead, for the information processing tasks of interest. Of course, since the *i.i.d.* assumption is generally not warranted, a protocol that achieves certification for some fraction of the copies while leaving the rest useful for subsequent tasks will be desirable. This has been considered for one-shot distillable entanglement in [20] and the self-testing fidelity in [60]. Again, a general treatment will be more than welcome (see, e.g. [103]).

ACKNOWLEDGMENTS

We are grateful to Jean-Daniel Bancal, Gelo Tabia, and Yanbao Zhang for many enlightening discussions. This work is supported by the National Science and Technology Council, Taiwan (Grants No. 107-2112-M-006-005-MY2 and 109-2112-M-006-010-MY3, 111-2112-M-005-007-MY4, 112-2119-M-001-004, 112-2119-M-001-006, 112-2124-M-002-003, 112-2628-M-006-007-MY4) and the Foxconn Research Institute, Taipei, Taiwan.

Appendix A: Miscellaneous results

In this Appendix, we provide some supplemental results to further illustrate the relative strength of the two hypothesis-testing protocols. For that matter, we extend some of the Figures shown in the main text to include two other plots in each of them. The first of this, dubbed “PBR-simplified” consists of the simplified implementation of the PBR protocol, where

we use in Eq. (29) and Eq. (30) the relative frequency \vec{f} computed from the raw data, instead of the regularized frequency \vec{f}_{reg} .

On the other hand, to see the best possible performance that one could hope for via any kind of implementation of the PBR protocol, we also replace \vec{f}_{reg} in Eq. (29) and Eq. (30) by the ideal quantum correlation \vec{P}_Q used in the generation of the data. This gets rid of the statistical fluctuations involved in the estimation of \vec{P}_Q right from the beginning. We call this “PBR-ideal” as it amounts to using the optimal Bell-like inequality right from the beginning.

1. Negativity

a. Certification based on the data generated from \vec{P}_{CHSH}

We start with Fig. 10, which extends Fig. 2 by including the two plots mentioned above.

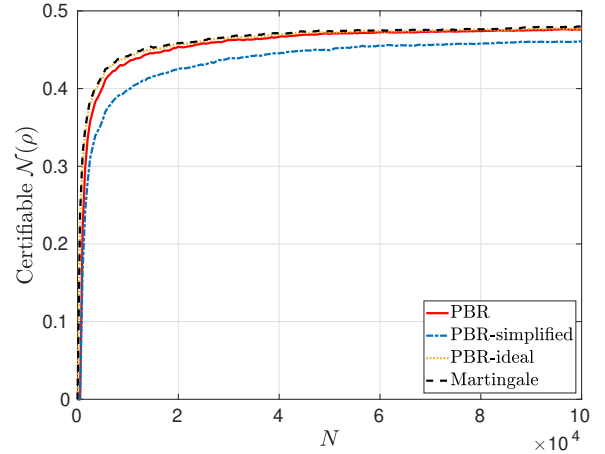


FIG. 10: Extended plot from Fig. 2 where we also include the results from the simplified implementation of the PBR protocol “PBR-simplified” and the best that one could hope for in implementing the PBR protocol “PBR-ideal”.

Clearly, as we can see in Fig. 10, the performance of the PBR-simplified protocol is far worse than the PBR protocol. For example, after $N = 10^5$ trials, the certifiable negativity is only about the same as that achieved with the PBR protocol with $N = 2 \times 10^4$ trials. Similarly, we see in Fig. 11 that with this protocol, the rate at which the *p*-value bound for Null Hypothesis 1 with $\mathcal{N}_0 = 0.3$ and 0.4 decreases at a rate that is far slower than the other protocols. On the other hand, we also see from these figures that the performance of PBR-ideal is essentially the same as the one given by the martingale-based protocol, an observation consistent with what one would expect from Fig. 3 (for $\theta = \frac{\pi}{4}$).

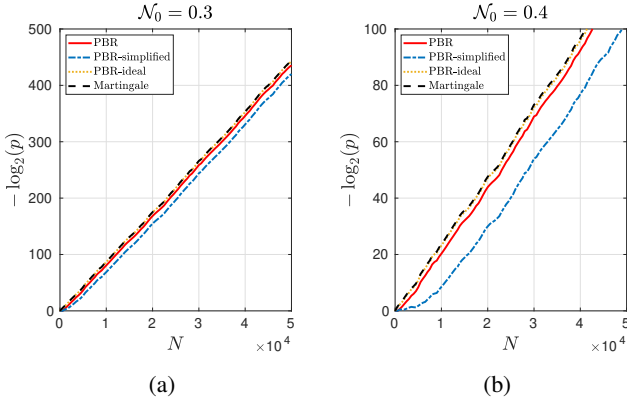


FIG. 11: Plot of $-\log_2 p^{(\text{prot})}$ vs N for Null Hypothesis 1 with $\mathcal{N}_0 = 0.3$ and 0.4 . The parameters are the same as those described in Fig. 2.

b. Certification based on the data generated from \vec{P}_θ

In the main text, we mention, in relation to Fig. 3, a possible cause of the noticeable discrepancy in the confidence-gain rates obtained between the two protocols. In Fig. 12, we see that, indeed, for any partially entangled two-qubit state $|\psi(\theta)\rangle$, the negativity lower bound obtained directly from the CHSH Bell violation, cf. Eq. (10), is far from tight compared to that obtained by solving Eq. (9) for, say, level-3 of the hierarchy introduced in [17]. Interestingly, we also see from Fig. 12 that negativity lower bound continues to improve even at level-12 of the hierarchy. See [104] for closely related studies on the convergence of this and other SDP hierarchies.

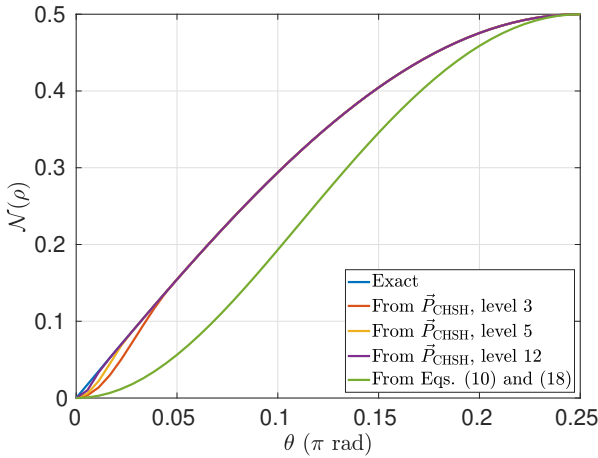


FIG. 12: Comparison of the actual negativity (top) of $\psi(\theta)$, Eq. (17a), and their various device-independent lower bounds. The computation for the negativity lower bound based on the underlying correlation \vec{P}_θ , derived from Eq. (17), is obtained by solving Eq. (9) whereas the bottom (dashed) line follows directly from the CHSH Bell inequality violation of these states, as given in Eqs. (10) and (18).

c. Certification based on the data generated from \vec{P}_{CGLMP}

The results analogous to Fig. 10 for the extension from Fig. 4 can be found in Fig. 13.

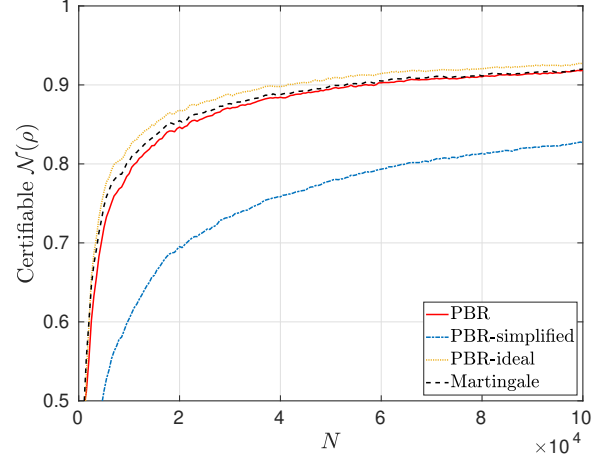


FIG. 13: Extended plot from Fig. 4 where we also include the results from the simplified implementation of the PBR protocol “PBR-simplified” and the ideal protocol “PBR-ideal”.

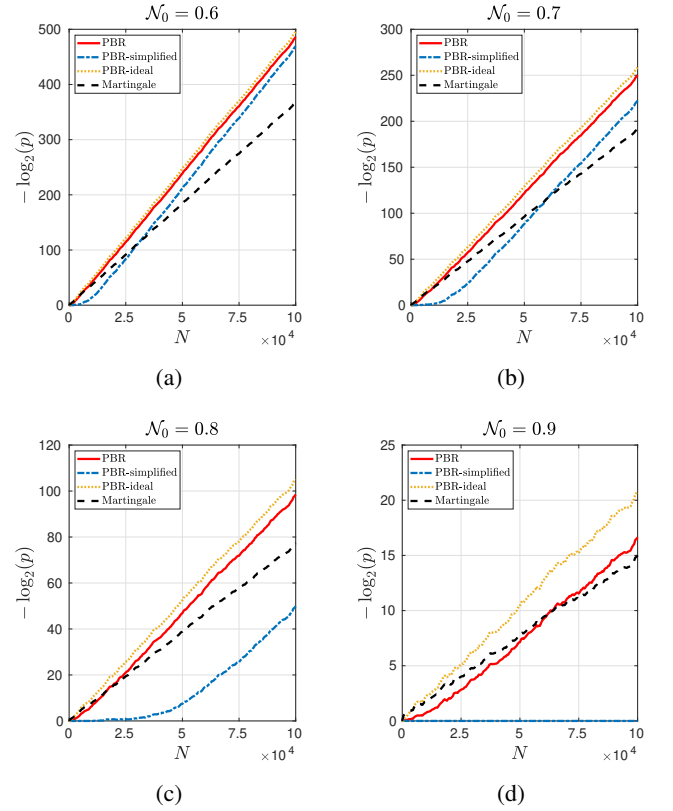


FIG. 14: Plot of $-\log_2 p^{(\text{prot})}$ vs N for hypothesis 1 with $\mathcal{N}_0 = 0.6, 0.7, 0.8$, and 0.9 . The parameters are the same as those described in Fig. 4.

In this case, interestingly, we see that for $\mathcal{N}_0 = 0.6, 0.7$ and sufficiently large N , even the PBR-simplified protocol could overtake the martingale-based protocol in its p -value upper bound. As \mathcal{N}_0 increases, we also see that the difference in the performance between the PBR protocol and the PBR-ideal protocol becomes more pronounced. Still, for sufficiently large N , the PBR protocol eventually surpasses the martingale-based protocol in its p -value bound. This last observation is consistent with our observation in Fig. 13 and the confidence gain rate shown in Fig. 5.

2. Fidelity

Next, let us include also the plots for PBR-ideal and PBR-simplified in Fig. 6, as shown in Fig. 15.

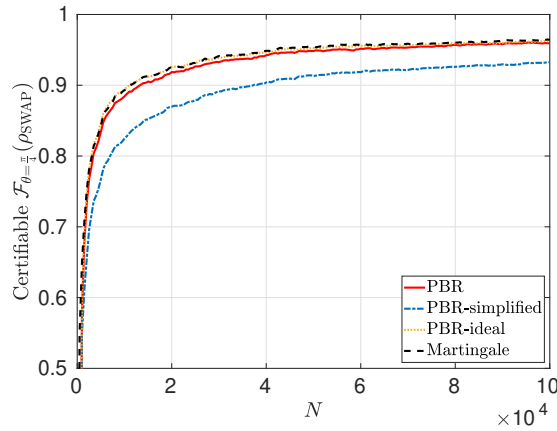


FIG. 15: Extended plot from Fig. 6 where we also include the results from the simplified implementation of the PBR protocol “PBR-simplified” and the best that one could hope for in implementing the PBR protocol “PBR-ideal”.

Again, we see that the performance of PBR is considerably impaired if we switch to PBR-simplified. Meanwhile, even though the difference between PBR-ideal and PBR is relatively insignificant for $\mathcal{F}_0 = 0.6, 0.7$, and 0.8 in Fig. 16, we see that the difference is significant enough to be manifested in Fig. 15.

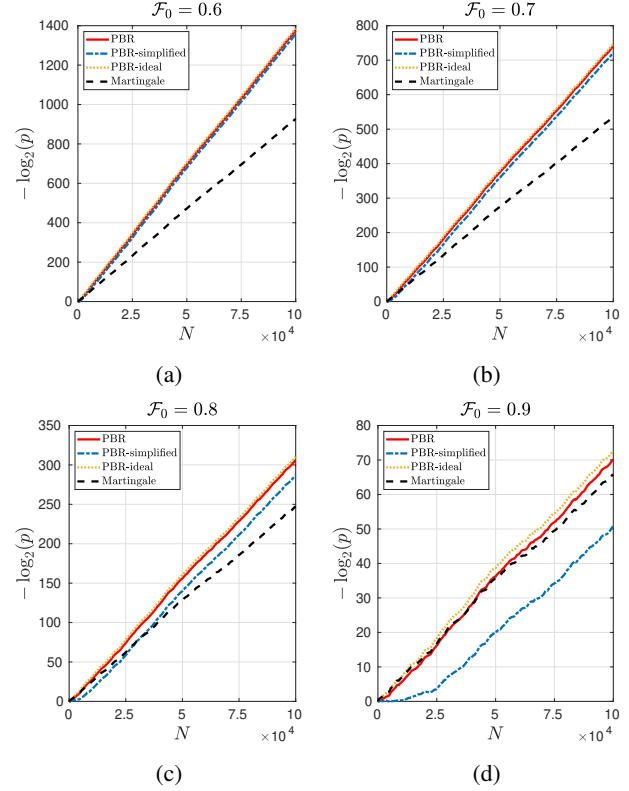


FIG. 16: Plot of $-\log_2 p^{(\text{prot})}$ vs N for hypothesis 4 with $\mathcal{F}_0 = 0.6, 0.7, 0.8$, and 0.9 . The parameters are the same as those described in Fig. 6.

[1] D. Rosset, R. Ferretti-Schöbitz, J.-D. Bancal, N. Gisin, and Y.-C. Liang, Imperfect measurement settings: Implications for quantum state tomography and entanglement witnesses, *Phys. Rev. A* **86**, 062325 (2012).
[2] T. Moroder, M. Kleinmann, P. Schindler, T. Monz, O. Gühne, and R. Blatt, Certifying systematic errors in quantum experiments, *Phys. Rev. Lett.* **110**, 180401 (2013).
[3] S. J. van Enk and R. Blume-Kohout, When quantum tomography goes wrong: drift of quantum sources and other errors, *New J. Phys.* **15**, 025024 (2013).
[4] V. Scarani, The device-independent outlook on quantum physics, *Acta Phys. Slovaca* **62**, 347 (2012).
[5] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Bell nonlocality, *Rev. Mod. Phys.* **86**, 419 (2014).
[6] A. Acín, N. Gisin, and L. Masanes, From Bell’s theorem to secure quantum key distribution, *Phys. Rev. Lett.* **97**, 120405 (2006).
[7] C. H. Bennett and G. Brassard, Quantum cryptography: Public

key distribution and coin tossing, *Theor. Comput. Sci.* **560**, 7 (2014).
[8] A. K. Ekert, Quantum cryptography based on Bell’s theorem, *Phys. Rev. Lett.* **67**, 661 (1991).
[9] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, Quantum cryptography, *Rev. Mod. Phys.* **74**, 145 (2002).
[10] D. Mayers and A. Yao, Quantum cryptography with imperfect apparatus, in *Proceedings 39th Annual Symposium on Foundations of Computer Science (Cat. No.98CB36280)* (1998) pp. 503–509.
[11] D. Mayers and A. Yao, Self testing quantum apparatus, *Quantum Info. Comput.* **4**, 273 (2004).
[12] J. S. Bell, On the Einstein Podolsky Rosen paradox, *Physics* **1**, 195 (1964).
[13] R. F. Werner, Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model, *Phys. Rev. A* **40**, 4277 (1989).
[14] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki,

- Quantum entanglement, *Rev. Mod. Phys.* **81**, 865 (2009).
- [15] N. Brunner, S. Pironio, A. Acín, N. Gisin, A. A. Méthot, and V. Scarani, Testing the dimension of Hilbert spaces, *Phys. Rev. Lett.* **100**, 210503 (2008).
- [16] J.-D. Bancal, N. Gisin, Y.-C. Liang, and S. Pironio, Device-independent witnesses of genuine multipartite entanglement, *Phys. Rev. Lett.* **106**, 250404 (2011).
- [17] T. Moroder, J.-D. Bancal, Y.-C. Liang, M. Hofmann, and O. Gühne, Device-independent entanglement quantification and related applications, *Phys. Rev. Lett.* **111**, 030501 (2013).
- [18] G. Tóth, T. Moroder, and O. Gühne, Evaluating convex roof entanglement measures, *Phys. Rev. Lett.* **114**, 160501 (2015).
- [19] Y.-C. Liang, D. Rosset, J.-D. Bancal, G. Pütz, T. J. Barnea, and N. Gisin, Family of Bell-like inequalities as device-independent witnesses for entanglement depth, *Phys. Rev. Lett.* **114**, 190401 (2015).
- [20] R. Arnon-Friedman and J.-D. Bancal, Device-independent certification of one-shot distillable entanglement, *New J. Phys.* **21**, 033010 (2019).
- [21] S.-L. Chen, H.-Y. Ku, W. Zhou, J. Tura, and Y.-N. Chen, Robust self-testing of steerable quantum assemblages and its applications on device-independent quantum certification, *Quantum* **5**, 552 (2021).
- [22] S.-L. Chen, C. Budroni, Y.-C. Liang, and Y.-N. Chen, Natural framework for device-independent quantification of quantum steerability, measurement incompatibility, and self-testing, *Phys. Rev. Lett.* **116**, 240401 (2016).
- [23] S.-L. Chen, C. Budroni, Y.-C. Liang, and Y.-N. Chen, Exploring the framework of assemblage moment matrices and its applications in device-independent characterizations, *Phys. Rev. A* **98**, 042127 (2018).
- [24] J.-D. Bancal, N. Sangouard, and P. Sekatski, Noise-resistant device-independent certification of Bell state measurements, *Phys. Rev. Lett.* **121**, 250506 (2018).
- [25] M. O. Renou, J. Kaniewski, and N. Brunner, Self-testing entangled measurements in quantum networks, *Phys. Rev. Lett.* **121**, 250507 (2018).
- [26] M. T. Quintino, C. Budroni, E. Woodhead, A. Cabello, and D. Cavalcanti, Device-independent tests of structures of measurement incompatibility, *Phys. Rev. Lett.* **123**, 180401 (2019).
- [27] S. Wagner, J.-D. Bancal, N. Sangouard, and P. Sekatski, Device-independent characterization of quantum instruments, *Quantum* **4**, 243 (2020).
- [28] S.-L. Chen, N. Miklin, C. Budroni, and Y.-N. Chen, Device-independent quantification of measurement incompatibility, *Phys. Rev. Res.* **3**, 023143 (2021).
- [29] P. Sekatski, J.-D. Bancal, S. Wagner, and N. Sangouard, Certifying the building blocks of quantum computers from Bell's theorem, *Phys. Rev. Lett.* **121**, 180505 (2018).
- [30] P. Sekatski, J.-D. Bancal, M. Ioannou, M. Afzelius, and N. Brunner, Toward the device-independent certification of a quantum memory, *Phys. Rev. Lett.* **131**, 170802 (2023).
- [31] L. Zhou, Y.-B. Sheng, and G.-L. Long, Device-independent quantum secure direct communication against collective attacks, *Sci. Bull.* **65**, 12 (2020).
- [32] C. Bernhard, B. Bessire, A. Montina, M. Pfaffhauser, A. Stefanov, and S. Wolf, Non-locality of experimental qutrit pairs, *J. Phys. A: Math. Theo.* **47**, 424013 (2014).
- [33] J.-D. Bancal, L. Sheridan, and V. Scarani, More randomness from the same data, *New J. Phys.* **16**, 033011 (2014).
- [34] S. Schwarz, B. Bessire, A. Stefanov, and Y.-C. Liang, Bipartite Bell inequalities with three ternary-outcome measurements—from theory to experiments, *New J. Phys.* **18**, 035001 (2016).
- [35] S. Popescu and D. Rohrlich, Quantum nonlocality as an axiom, *Found. Phys.* **24**, 379 (1994).
- [36] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts, Nonlocal correlations as an information-theoretic resource, *Phys. Rev. A* **71**, 022101 (2005).
- [37] 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).
- [38] A. Aspect, J. Dalibard, and G. Roger, Experimental test of Bell's inequalities using time-varying analyzers, *Phys. Rev. Lett.* **49**, 1804 (1982).
- [39] W. Tittel, J. Brendel, H. Zbinden, and N. Gisin, Violation of Bell inequalities by photons more than 10 km apart, *Phys. Rev. Lett.* **81**, 3563 (1998).
- [40] G. Weihs, T. Jennewein, C. Simon, H. Weinfurter, and A. Zeilinger, Violation of Bell's inequality under strict einstein locality conditions, *Phys. Rev. Lett.* **81**, 5039 (1998).
- [41] M. Rowe, D. Kielpinski, V. Meyer, C. A. Sackett, W. M. Itano, C. Monroe, and D. J. Wineland, Experimental violation of a Bell's inequality with efficient detection, *Nature* **409**, 791 (2001).
- [42] J. Barrett, D. Collins, L. Hardy, A. Kent, and S. Popescu, Quantum nonlocality, Bell inequalities, and the memory loophole, *Phys. Rev. A* **66**, 042111 (2002).
- [43] R. D. Gill, Accardi contra Bell (cum mundi): The impossible coupling, *Lecture Notes-Monograph Series* **42**, 133 (2003).
- [44] B. Hensen, H. Bernien, A. E. Dreau, A. Reiserer, N. Kalb, M. S. Blok, J. Ruitenberg, R. F. L. Vermeulen, R. N. Schouten, C. Abellan, W. Amaya, V. Pruneri, M. W. Mitchell, M. Markham, D. J. Twitchen, D. Elkouss, S. Wehner, T. H. Taminiau, and R. Hanson, Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres, *Nature (London)* **526**, 682 (2015).
- [45] L. K. Shalm, E. Meyer-Scott, B. G. Christensen, P. Bierhorst, M. A. Wayne, M. J. Stevens, T. Gerrits, S. Glancy, D. R. Hamel, M. S. Allman, K. J. Coakley, S. D. Dyer, C. Hodge, A. E. Lita, V. B. Verma, C. Lambrocco, E. Tortorici, A. L. Migdall, Y. Zhang, D. R. Kumor, W. H. Farr, F. Marsili, M. D. Shaw, J. A. Stern, C. Abellán, W. Amaya, V. Pruneri, T. Jennewein, M. W. Mitchell, P. G. Kwiat, J. C. Bienfang, R. P. Mirin, E. Knill, and S. W. Nam, Strong loophole-free test of local realism, *Phys. Rev. Lett.* **115**, 250402 (2015).
- [46] M. Giustina, M. A. M. Versteegh, S. Wengerowsky, J. Handsteiner, A. Hochrainer, K. Phelan, F. Steinlechner, J. Kofler, J.-A. Larsson, C. Abellán, W. Amaya, V. Pruneri, M. W. Mitchell, J. Beyer, T. Gerrits, A. E. Lita, L. K. Shalm, S. W. Nam, T. Scheidl, R. Ursin, B. Wittmann, and A. Zeilinger, Significant-loophole-free test of Bell's theorem with entangled photons, *Phys. Rev. Lett.* **115**, 250401 (2015).
- [47] R. Colbeck, *Quantum And Relativistic Protocols For Secure Multi-Party Computation*, Ph.D. thesis, University of Cambridge (2006).
- [48] S. Pironio, A. Acín, S. Massar, A. B. de la Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, and C. Monroe, Random numbers certified by Bell's theorem, *Nature* **464**, 1021 (2010).
- [49] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, Device-independent security of quantum cryptography against collective attacks, *Phys. Rev. Lett.* **98**, 230501 (2007).
- [50] S. Pironio, A. Acín, N. Brunner, N. Gisin, S. Massar, and V. Scarani, Device-independent quantum key distribution se-

- cure against collective attacks, *New J. Phys.* **11**, 045021 (2009).
- [51] S. Pironio and S. Massar, Security of practical private randomness generation, *Phys. Rev. A* **87**, 012336 (2013).
- [52] O. Nieto-Silleras, C. Bamps, J. Silman, and S. Pironio, Device-independent randomness generation from several Bell estimators, *New J. Phys.* **20**, 023049 (2018).
- [53] P. Bierhorst, E. Knill, S. Glancy, Y. Zhang, A. Mink, S. Jordan, A. Rommal, Y.-K. Liu, B. Christensen, S. W. Nam, M. J. Stevens, and L. K. Shalm, Experimentally generated randomness certified by the impossibility of superluminal signals, *Nature* **556**, 223 (2018).
- [54] B. Bourdoncle, P.-S. Lin, D. Rosset, A. Acín, and Y.-C. Liang, Regularising data for practical randomness generation, *Quantum Sci. Technol.* **4**, 025007 (2019).
- [55] Y. Zhang, H. Fu, and E. Knill, Efficient randomness certification by quantum probability estimation, *Phys. Rev. Res.* **2**, 013016 (2020).
- [56] E. Knill, Y. Zhang, and P. Bierhorst, Generation of quantum randomness by probability estimation with classical side information, *Phys. Rev. Res.* **2**, 033465 (2020).
- [57] T. Metger, O. Fawzi, D. Sutter, and R. Renner, Generalised entropy accumulation, in *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)* (2022) pp. 844–850.
- [58] R. Arnon-Friedman, R. Renner, and T. Vidick, Simple and tight device-independent security proofs, *SIAM Journal on Computing* **48**, 181 (2019).
- [59] F. Dupuis, O. Fawzi, and R. Renner, Entropy accumulation, *Commun. Math. Phys.* **379**, 867 (2020).
- [60] A. Gočanin, I. Šupić, and B. Dakić, Sample-efficient device-independent quantum state verification and certification, *PRX Quantum* **3**, 010317 (2022).
- [61] Y. Zhang, S. Glancy, and E. Knill, Asymptotically optimal data analysis for rejecting local realism, *Phys. Rev. A* **84**, 062118 (2011).
- [62] R. D. Gill, Time, finite statistics, and Bell’s fifth position, in *Proc. Foundations of Probability and Physics-2*, Math. Modelling in Phys. Engi., and Cog. Sc., Vol. 5 (Växjö University Press, Växjö, 2003) pp. 179–206.
- [63] Y.-C. Liang and Y. Zhang, Bounding the plausibility of physical theories in a device-independent setting via hypothesis testing, *Entropy* **21**, 185 (2019).
- [64] 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).
- [65] G. Vidal and R. F. Werner, Computable measure of entanglement, *Phys. Rev. A* **65**, 032314 (2002).
- [66] A. Peres, Separability criterion for density matrices, *Phys. Rev. Lett.* **77**, 1413 (1996).
- [67] M. Navascués, S. Pironio, and A. Acín, Bounding the set of quantum correlations, *Phys. Rev. Lett.* **98**, 010401 (2007).
- [68] M. Navascués, S. Pironio, and A. Acín, A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations, *New J. Phys.* **10**, 073013 (2008).
- [69] A. C. Doherty, Y.-C. Liang, B. Toner, and S. Wehner, The quantum moment problem and bounds on entangled multiprover games, in *23rd Annu. IEEE Conf. on Comput. Comp.*, 2008, CCC’08 (Los Alamitos, CA, 2008) pp. 199–210.
- [70] D. Collins, N. Gisin, N. Linden, S. Massar, and S. Popescu, Bell inequalities for arbitrarily high-dimensional systems, *Phys. Rev. Lett.* **88**, 040404 (2002).
- [71] D. Kaszlikowski, L. C. Kwek, J.-L. Chen, M. Żukowski, and C. H. Oh, Clauser-Horne inequality for three-state systems, *Phys. Rev. A* **65**, 032118 (2002).
- [72] A. Acín, T. Durt, N. Gisin, and J. I. Latorre, Quantum nonlocality in two three-level systems, *Phys. Rev. A* **65**, 052325 (2002).
- [73] Y.-C. Liang, *Correlations, Bell Inequality Violation & Quantum Entanglement*, Ph.D. thesis, University of Queensland (2008).
- [74] H. Lu, Q. Zhao, Z.-D. Li, X.-F. Yin, X. Yuan, J.-C. Hung, L.-K. Chen, L. Li, N.-L. Liu, C.-Z. Peng, Y.-C. Liang, X. Ma, Y.-A. Chen, and J.-W. Pan, Entanglement structure: Entanglement partitioning in multipartite systems and its experimental detection using optimizable witnesses, *Phys. Rev. X* **8**, 021072 (2018).
- [75] J.-D. Bancal, J. Barrett, N. Gisin, and S. Pironio, Definitions of multipartite nonlocality, *Phys. Rev. A* **88**, 014102 (2013).
- [76] F. J. Curchod, N. Gisin, and Y.-C. Liang, Quantifying multipartite nonlocality via the size of the resource, *Phys. Rev. A* **91**, 012121 (2015).
- [77] O. Gühne, G. Tóth, and H. J. Briegel, Multipartite entanglement in spin chains, *New J. Phys.* **7**, 229 (2005).
- [78] A. S. Sørensen and K. Mølmer, Entanglement and Extreme Spin Squeezing, *Phys. Rev. Lett.* **86**, 4431 (2001).
- [79] N. D. Mermin, Extreme quantum entanglement in a superposition of macroscopically distinct states, *Phys. Rev. Lett.* **65**, 1838 (1990).
- [80] I. Šupić and J. Bowles, Self-testing of quantum systems: a review, *Quantum* **4**, 337 (2020).
- [81] C. Jebarathinam, J.-C. Hung, S.-L. Chen, and Y.-C. Liang, Maximal violation of a broad class of Bell inequalities and its implication on self-testing, *Phys. Rev. Res.* **1**, 033073 (2019).
- [82] J. Kaniewski, Weak form of self-testing, *Phys. Rev. Res.* **2**, 033420 (2020).
- [83] S. J. Summers and R. Werner, Maximal violation of Bell’s inequalities is generic in quantum field theory, *Commun. Math. Phys.* **110**, 247 (1987).
- [84] S. Popescu and D. Rohrlich, Which states violate Bell’s inequality maximally?, *Phys. Lett. A* **169**, 411 (1992).
- [85] S. L. Braunstein, A. Mann, and M. Revzen, Maximal violation of Bell inequalities for mixed states, *Phys. Rev. Lett.* **68**, 3259 (1992).
- [86] B. S. Tsirelson, Some results and problems on quantum Bell-type inequalities, *Hadronic J. Suppl.* **8**, 329 (1993).
- [87] J.-D. Bancal, M. Navascués, V. Scarani, T. Vértesi, and T. H. Yang, Physical characterization of devices from nonlocal correlations, *Phys. Rev. A* **91**, 022115 (2015).
- [88] T. H. Yang and M. Navascués, Robust self-testing of unknown quantum systems into any entangled two-qubit states, *Phys. Rev. A* **87**, 050102 (2013).
- [89] T. H. Yang, T. Vértesi, J.-D. Bancal, V. Scarani, and M. Navascués, Robust and versatile black-box certification of quantum devices, *Phys. Rev. Lett.* **113**, 040401 (2014).
- [90] W. van Dam, R. Gill, and P. Grunwald, The statistical strength of nonlocality proofs, *IEEE Trans. Inf. Theory* **51**, 2812 (2005).
- [91] Y. Zhang, S. Glancy, and E. Knill, Efficient quantification of experimental evidence against local realism, *Phys. Rev. A* **88**, 052119 (2013).
- [92] S. Kullback and R. A. Leibler, On Information and Sufficiency, *Ann. Math. Stat.* **22**, 79 (1951).
- [93] Mosek conic optimization, *MOSEK Modeling Cookbook*.
- [94] T. Minka, The Lightspeed Matlab toolbox, <https://github.com/tminka/lightspeed>.
- [95] Y.-C. Liang and A. C. Doherty, Bounds on quantum correlations in Bell-inequality experiments, *Phys. Rev. A* **75**, 042103 (2007).

- (2007).
- [96] D. M. Greenberger, M. A. Horne, and A. Zeilinger, Bell's Theorem, *Quantum Theory and Conceptions of the Universe* (Kluwer, Dordrecht, 1989) Chap. Going Beyond Bell's Theorem, pp. 69–72.
 - [97] N. D. Mermin, Quantum mysteries revisited, *Am. J. Phys.* **58**, 731 (1990).
 - [98] J. Kaniewski, Analytic and nearly optimal self-testing bounds for the Clauser-Horne-Shimony-Holt and Mermin inequalities, *Phys. Rev. Lett.* **117**, 070402 (2016).
 - [99] J. Eisert, D. Hangleiter, N. Walk, I. Roth, D. Markham, R. Parekh, U. Chabaud, and E. Kashefi, Quantum certification and benchmarking, *Nat. Rev. Phys.* **2**, 382 (2020).
 - [100] T. R. Tan, Y. Wan, S. Erickson, P. Bierhorst, D. Kienzler, S. Glancy, E. Knill, D. Leibfried, and D. J. Wineland, Chained Bell inequality experiment with high-efficiency measurements, *Phys. Rev. Lett.* **118**, 130403 (2017).
 - [101] J.-D. Bancal, K. Redeker, P. Sekatski, W. Rosenfeld, and N. Sangouard, Self-testing with finite statistics enabling the certification of a quantum network link, *Quantum* **5**, 401 (2021).
 - [102] P. Wills, E. Knill, K. Coakley, and Y. Zhang, Performance of test supermartingale confidence intervals for the success probability of Bernoulli trials, *J. Res. Natl. Inst. Stan.* **125**, 125003 (2020).
 - [103] Y. Zhang, A. Seshadri, and E. Knill, Confidence-interval construction with non-i.i.d. spot-checking trials & its application in quantum information, in *Optica Quantum 2.0 Conference and Exhibition*, Technical Digest Series (Optica Publishing Group, Denver, Colorado, 2023) p. QTu3A.20.
 - [104] P.-S. Lin, T. Vértesi, and Y.-C. Liang, Naturally restricted subsets of nonsignaling correlations: typicality and convergence, *Quantum* **6**, 765 (2022).