IOP Publishing Quantum Sci. Technol. 10 (2025) 035056 https://doi.org/10.1088/2058-9565/ade911

Quantum Science and Technology

OPEN ACCESS

RECEIVED
6 April 2025

REVISED
9 June 2025

ACCEPTED FOR PUBLICATION
27 June 2025

PUBLISHED
8 July 2025

PAPER

Almost device-independent calibration beyond Born's rule: Bell tests for cross-talk detection

Gelo Noel M Tabia[1,2,3,*], **Alex Yueh-Ting Shih**[2], **Jin-Yuan Zheng**[2] and **Yeong-Cherng Liang**[2,3,4,*]

[1] Hon Hai (Foxconn) Research Institute, Taipei, Taiwan
[2] Department of Physics and Center for Quantum Frontiers of Research & Technology (QFort), National Cheng Kung University, Tainan 701, Taiwan
[3] Physics Division, National Center for Theoretical Sciences, Taipei 106319, Taiwan
[4] Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada
[*] Authors to whom any correspondence should be addressed.

E-mail: gelo.tabia@foxconn.com and ycliang@mail.ncku.edu.tw

Keywords: device-independent, Bell test, finite statistics, quantum computer, signaling, cross-talk effects, outcome correlation

Abstract

In quantum information, device-independent (DI) protocols offer a new approach to information processing tasks, making minimal assumptions about the devices used. Typically, since these protocols draw conclusions directly from the data collected in a meaningful Bell test, the no-signaling conditions, and often even Born's rule for local measurements, are taken as premises of the protocol. Here, we demonstrate how to test such premises in an (almost) DI setting, i.e. directly from the raw data and with minimal assumptions. In particular, for IBM's quantum computing cloud services, we implement the prediction-based ratio protocol to characterize how well the qubits can be accessed locally and independently. More precisely, by performing a variety of Clauser–Horne–Shimony–Holt-type experiments on these systems and carrying out rigorous hypothesis tests on the collected data, we provide compelling evidence showing that some of these qubits suffer from measurement cross-talks, i.e. their measurement statistics are affected by the choice of measurement bases on another qubit. Unlike standard randomized benchmarking, our approach does not rely on assumptions such as gate-independent Markovian noise. Moreover, despite the relatively small number of experimental trials, the direction of 'signaling' may also be identified in some cases. Our approach thus serves as a complementary tool for benchmarking the local addressability of quantum computing devices.

1. Introduction

The device-independent (DI) [1] approach to physics can be traced back to Bell [2] when he proved that local-hidden-variable (LHV) theories necessarily fail to reproduce some predictions of quantum theory. His proof relies only on the correlations among measurement outcomes conditioned on the chosen measurement settings. Thus, it requires no further knowledge about how the devices function. Since then, a few other no-go theorems based on the violation of Bell-like inequalities have also been obtained (see, e.g. [3–5]).

Apart from quantum foundational issues, the DI methodology also finds applications in several cryptographic tasks, such as randomness expansion [6–8] and key distribution [9–11]. In these DI protocols, it is crucial that the correlations obtained from the Bell experiment satisfy the so-called no-signaling (NS) [12] conditions. Often, the security analysis further assumes that quantum theory is correct, in particular, that the outcome probabilities are specified by Born's rule for local measurements (see [13, 14] for a recent review).

In this work, we focus on applications of the DI approach to the characterization of quantum devices (see, e.g. [15–25]). One of the requirements for the proper functioning of quantum computers is the ability to protect fragile quantum states from noise [26]. However, in some quantum computers, due to the proximity of the qubits and their high level of interconnectivity, it is conceivable that the interaction with a

targeted qubit could simultaneously affect the state of the neighboring qubits. To correct the errors from such cross-talk [27] and other unwanted effects, we need some way to identify and quantify the noise in a quantum device. The most widely used approaches for this task are based on randomized benchmarking (RB) [28–31] or gate-set tomography (GST) [32, 33] (see also [34]).

In a typical RB method, we measure the error rate of a particular set of quantum gates by applying a sequence of random gates that would ideally correspond to an identity operation if the gates were perfect. Meanwhile, GST is a method that incorporates elements of quantum process tomography into a procedure that also deals directly with state preparation and measurement (SPAM) errors. GST inherits some of the problems of tomographic methods, particularly the need for large samples to estimate noise parameters. To achieve sample efficiency, one can turn a GST protocol into a randomized scheme and use classical shadow estimation techniques [35, 36] that allow one to deduce various linear functionals of the gate-set noise [37].

However, both RB and GST often involve the assumption of temporally uncorrelated noise. In RB, the exponential decay in the average gate-sequence fidelity assumes that the noise is Markovian, and one can even identify the presence of non-Markovian noise by the failure of the exponential model [38–40]. Similarly, in GST, a Markovian noise model is used so that the contributions of SPAM and gate-set errors can be estimated separately. While there have been recent attempts to incorporate non-Markovian noise [41, 42], it is natural to wonder whether one can identify unwanted cross-talks using only *minimal assumptions*.

Here, building on earlier studies [43], we show that it is indeed possible to certify—in an *almost* DI manner—the presence of cross-talks directly from the raw measurement data obtained from a quantum computer. By 'almost' DI, we mean that we consider standard DI assumptions on our Bell tests but with the usual assumption of measurement independence [4] (more commonly known as the 'freedom of choice' assumption) replaced by the *assumption* that the pseudo-random string of inputs does *not* alter the behavior of the individual qubits, even though we generate the inputs and feed them to the device preparing the qubits *before* their preparation.

Importantly, experimental trials are not necessarily independent and identically distributed (*i.i.d.*). In particular, assuming that the trials are *i.i.d.* when they are not may open the so-called memory loophole [44]. Even if the trials are *i.i.d.*, statistical fluctuations may still render the relative frequencies of the measurement outcomes—taken as a proxy of the underlying correlation—incompatible with the NS constraints. To cope with this complication in the context of DI certification, various methods for regularizing the relative frequencies to the set of correlations compatible with the NS constraints [45, 46] (or even outer approximations of the quantum set [47, 48]) have been proposed. However, one can also adopt a more rigorous approach based on hypothesis testing.

Indeed, in statistical inference, it is customary to report the p-value for a null hypothesis to be correct. Here, we follow [43, 49, 50] and consider the prediction-based-ratio (PBR) protocol [51, 52] for upper bounding the p-value on the plausibility of a given null hypothesis in producing the data observed in a Bell test. The PBR protocol was originally introduced as a rigorous statistical tool for rejecting the null hypothesis associated with LHV theories. In [49], it was adapted to perform DI certification of desirable quantum properties (e.g. those discussed in [15, 17, 18, 53]) with a confidence interval. Notice that these certification tasks presuppose Born's rule for local measurements and, hence, compatibility with the NS constraints. In this work, we illustrate how the PBR protocol can be used to reveal a violation of these premises and, consequently, the presence of cross-talks in real quantum devices with a relatively small sample size.

We structure the rest of this paper as follows. Section 2 introduces our notations and recalls the background knowledge required for analyzing the data collected in a Bell test. Then, in section 3, we explain how we apply the PBR protocol to the data collected from 'Bell tests' performed on IBM quantum (IBMQ) devices. We then present our results in section 4 and end with further discussions in section 5.

2. Preliminaries

2.1. No-signaling conditions and the no-signaling set

For simplicity, we consider only the simplest, bipartite Bell scenario where two parties, Alice and Bob, with two inputs and two outputs each. If we denote Alice's (Bob's) inputs/ settings by $x \in X$ ($y \in Y$) and outputs/ outcomes by $a \in A$ ($b \in B$), then a Bell correlation $\vec{P} := \{P(a,b|x,y)\}$ is the collection of joint conditional probability distributions of measurement outcomes given the choice of settings.

If we require that Bob cannot signal his input choice (y or y') to Alice, then her marginal probabilities must satisfy

$$P(a|x) = \sum_b P(a,b|x,y) = \sum_b P(a,b|x,y')\,, \forall\, a,x,y,y'. \tag{1a}$$

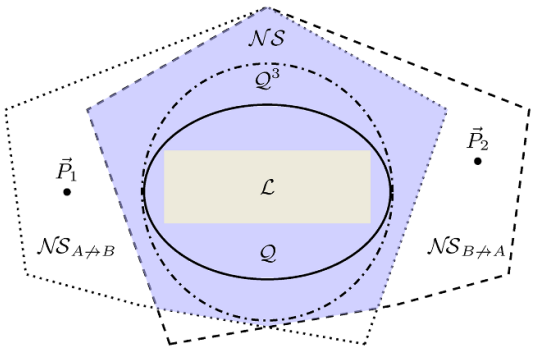

Figure 1. Schematic illustrating the inclusion relations of equation (4). A correlation or relative frequency such as $\vec{P}_1$ lies outside $\mathcal{NS}_{B\not\to A}$ (dashed-boundary polygon), and, hence, also outside $\mathcal{NS}$ (filled polygon), $\mathcal{Q}^3$ (dashed-dotted ellipse), $\mathcal{Q}$ (solid ellipse), and $\mathcal{L}$ (shaded rectangle). Similarly, a correlation or relative frequency such as $\vec{P}_2$ lies outside $\mathcal{NS}_{A\not\to B}$ (dotted-boundary polygon), and, hence, also outside $\mathcal{NS}$, $\mathcal{Q}^3$, $\mathcal{Q}$, and $\mathcal{L}$. Note that the figure is not meant to capture *all* geometrical aspects of these sets. For example, $\mathcal{L}$ and $\mathcal{NS}$, and hence all sets sandwiched between them, cf equation (4), share common boundaries [1], but this and various other geometrical features [54, 55] are not depicted in the figure.

In this case, we say that $\vec{P}$ is one-way no-signaling (OWNS) from Bob to Alice, and we denote the set of all such correlations by $\mathcal{NS}_{B\not\to A}$. On the other hand, if, instead, we require that Alice cannot signal her input choice (x or x') to Bob, then we have

$$P(b|y) = \sum_a P(a,b|x,y) = \sum_a P(a,b|x',y)\,, \forall\, b,y,x,x'. \tag{1b}$$

We refer to $\vec{P}$ satisfying equation (1*b*) as being OWNS from Alice to Bob, and we denote the set of such correlations accordingly by $\mathcal{NS}_{A\not\to B}$.

The set $\mathcal{NS}$ of (two-way) NS correlations, defined by equation (1), is the intersection of the two OWNS sets $\mathcal{NS}_{A\not\to B}$ and $\mathcal{NS}_{B\not\to A}$. Originally, the NS conditions of equation (1) were inspired by the notion of relativistic causality from special relativity [12], which prohibits a causal influence between spacelike separated parties. In our work, we provide an alternative interpretation of the NS conditions in the context of measurement cross-talk effects: if there is no unintended measurement cross-talk between the qubits, the choice of measurement basis on one qubit will have no impact on the marginal measurement statistics of any other qubit. In this case, the NS conditions of equation (1) follow. In other words, the violation of any constraint from equation (1) is a signature of measurement cross-talks, modulo the assumption mentioned in the Introduction.

In a Bell test, we are also often interested in two particular subsets of $\mathcal{NS}$: the set $\mathcal{L}$ of (Bell-)local [1] correlations and the set $\mathcal{Q}$ of quantum correlations. We have $\vec{P} \in \mathcal{L}$ if there exists an LHV λ satisyfing a normalized distribution $p(\lambda) \geqslant 0$ and local deterministic response functions $P_A(a|x,\lambda), P_B(b|y,\lambda) = 0,1$ with $\sum_a P_A(a|x,\lambda) = 1 = \sum_b P_B(b|y,\lambda)$ such that for all a,b,x,y, we can write [1, 2]

$$P(a,b|x,y) \overset{\mathcal{L}}{=} \sum_\lambda p(\lambda)\, P_A(a|x,\lambda)\, P_B(b|y,\lambda)\,. \tag{2}$$

Otherwise, we say that $\vec{P}$ is (Bell-)nonlocal. Meanwhile, we have $\vec{P} \in \mathcal{Q}$ if it can be obtained from local measurements performed by Alice and Bob on a shared quantum state ρ_{AB}; then Born's rule dictates that

$$P(a,b|x,y) \overset{\mathcal{Q}}{=} \mathrm{tr}\left(\rho_{AB} M^A_{a|x} \otimes M^B_{b|y}\right), \tag{3}$$

where $M^A_{a|x}(M^B_{b|y})$ denotes the positive operator-valued measure element associated with outcome a (b) of Alice's (Bob's) xth (yth) measurement setting. It is easy to verify that $\mathcal{L}$, $\mathcal{Q}$, and $\mathcal{NS}$ are convex sets that satisfy the strict inclusion

$$\mathcal{L} \subset \mathcal{Q} \subset \mathcal{Q}^k \subset \mathcal{NS} = \mathcal{NS}_{B\not\to A} \cap \mathcal{NS}_{A\not\to B} \quad \forall\, k, \tag{4}$$

where $\mathcal{Q}^k$ is an outer NS approximation of $\mathcal{Q}$ (more on this below). See figure 1 for a diagrammatic representation.

When the cardinalities of X, Y, A, and B are finite, $\mathcal{L}$ is a convex polytope, i.e. the convex hull of a finite set of extreme points. In contrast, since the quantum set $\mathcal{Q}$ is not a polytope, there is generally no simple

criterion to test whether a correlation $\vec{P}$ belongs in $\mathcal{Q}$. Nevertheless, various outer approximations of $\mathcal{Q}$ (see, e.g. [17, 56–58]) facilitate its membership test via a sequence of supersets $\mathcal{Q}^k$ such that $\mathcal{NS} \supset \mathcal{Q}^1 \supset \mathcal{Q}^2 \supset \cdots \supset \mathcal{Q}$. In the following, we use the level-3 of the Moroder hierarchy [17], denoted by $\mathcal{Q}^3$, as our outer approximation of $\mathcal{Q}$. This choice is motivated by the observation in [59] (see, e.g. table 2 and the top left subplot of figure 3 therein) that the lowest level of the hierarchy from either [17, 56] is visibly not tight, but going to a level even higher than $\mathcal{Q}^3$ may not be worth the extra computation time. However, from the analysis of [59], we expect similar results to hold if we adopt other outer approximations with similar computational complexity.

2.2. Hypothesis testing and the prediction-based-ratio method

Often, we perform an experiment to test a particular (null) hypothesis, such as that derived from a theoretical prediction. In statistical hypothesis testing, one effective way of determining the plausibility of a null hypothesis $\mathcal{H}$ from experimental data is to compute a p-value upper bound from some real function of the data called a test statistic T. The p-value then represents the tail probability for the observed value of T conditioned on $\mathcal{H}$, i.e. if the observed value of T is t, then

$$p\text{-value} = \text{Prob}\left(T \geqslant t | \mathcal{H} \text{ holds}\right), \tag{5}$$

which tells us how likely the data can be explained by the hypothesis $\mathcal{H}$.

Historically, Bell tests were introduced to determine if Nature is compatible with the description of LHV theories. However, any real Bell test necessarily involves only a finite number of trials where we obtain the counts of events involving different combinations of inputs and outputs. To cope with this limitation, the PBR protocol—motivated by an earlier work of Gill [60]—was introduced to provide a systematic, efficient method for upper bounding the corresponding p-value. In [43], it was noted that the PBR protocol can be straightforwardly adapted to test the plausibility of other physical theories, including a general NS theory.

For concreteness, suppose we conduct a Bell test with a total of N trials. In each trial, the inputs x and y are chosen randomly according to some fixed distribution $P(x,y)$. Thus, the data generated in each trial is a set of four numbers (a,b,x,y). For definiteness, consider now the hypothesis that the data observed is generated by an underlying NS process describable by some correlation $\vec{P} \in \mathcal{NS}$, which may vary from one trial to the next.

Even if the experimental trials are *i.i.d.*, the data alone will not allow us to identify $\vec{P}$ exactly. Nonetheless, we can estimate $\vec{P}$ by computing the relative frequencies $\vec{f} := \{f(a,b|x,y)\}$ for each outcome pair (a,b) given the choice of input pair (x,y),

$$f(a,b|x,y) := N_{a,b,x,y}/N_{x,y}, \tag{6}$$

where $N_{a,b,x,y}$ is the number of trials where the input-output combination (a,b,x,y) occurs, $N_{x,y} := \sum_{a,b} N_{a,b,x,y}$, and

$$\sum_{x,y} N_{x,y} = N. \tag{7}$$

In the asymptotic limit where $N \to \infty$, statistical fluctuations vanish, and therefore $\vec{f}$ approaches $\vec{P}$. For (finite) *i.i.d.* trials, the amount of statistical evidence in the data contrary to our hypothesis can be measured [61] in terms of the Kullback–Leibler (KL) divergence.

More precisely, if we believe the NS hypothesis to be true, the 'best-fitting' NS correlation would be given by the minimizer of the following optimization problem:

$$D_{\text{KL}}\left(\vec{f}||\mathcal{NS}\right) = \min_{\vec{P}\in\mathcal{NS}} \sum_{a,b,x,y} P(x,y) f(a,b|x,y) \times \log\left[\frac{f(a,b|x,y)}{P(a,b|x,y)}\right]. \tag{8}$$

Importantly, this optimization can be efficiently solved using a numerical solver such as MOSEK [62]. In [63], we provide an implementation of this optimization in MATLAB via YALMIP [64][5]. Since $\mathcal{NS}$ is a convex set and the KL divergence is a *strictly* convex function of $\vec{P}$, the minimizer $\vec{P}^{\text{NS}}_{\star}$ of the above optimization problem is unique [48].

[5] For the results presented in section 4, we also use a somewhat more accurate implementation of equation (8) via PENLAB [65] (courtesy of Denis Rosset), which generally gives a tighter p-value upper bound.

However, in a real experiment, it would be hard to justify that the trials are *i.i.d.*, since this entails running every trial under the exact same conditions, which would be impractical with imperfect devices. The key observation of the PBR protocol is that even for non-*i.i.d.* trials, the following Bell-like inequality remains valid [43, 51] for all $\vec{P} \in \mathcal{NS}$:

$$\sum_{a,b,x,y} R_{abxy} P(x,y)\, P(a,b|x,y) \overset{\mathcal{NS}}{\leqslant} 1 \tag{9a}$$

where the coefficients $R_{abxy} \geqslant 0$ for all a, b, x, y are the so-called prediction-based ratios (PBRs), defined as[6],

$$R_{abxy} := \frac{f(a,b|x,y)}{P_\star^{\mathrm{NS}}(a,b|x,y)}. \tag{9b}$$

Note that equation (9) is an optimized Bell-like inequality for witnessing the violation of the $\mathcal{NS}$ hypothesis by data that follows the distribution governed by $\vec{f}$ (see [51] for a discussion based on the hypothesis of LHV theories associated with $\mathcal{L}$). Hence, if the subsequent trials follow a distribution significantly different from the $\vec{f}$ used in defining equation (9*b*), even if the data violates the $\mathcal{NS}$ hypothesis, it may not be reflected by the corresponding p-value bound determined from the above PBRs.

For the purposes of hypothesis testing via the PBR protocol, we only use part of the data to establish the Bell-like inequality of equation (9), while the remaining part is used to compute a test statistic from its coefficients. Suppose we take the first $N_{\mathrm{est}} < N$ trials of the data to obtain R_{abxy} via equations (6), (8) and (9*b*), i.e. the right-hand side of equation (7) is now N_{est}. Then, we have the remaining $N_{\mathrm{test}} := N - N_{\mathrm{est}}$ rounds of data for computing a p-value upper bound. Let (x_i, y_i) denote the settings and (a_i, b_i) the outcomes observed in the ith trial. The PBR for this round would be $r_i := R_{a_i b_i x_i y_i}$, which corresponds to the value of R_{abxy} for the combination of inputs and outputs seen in the trial. In the PBR protocol, we consider a test statistic given by the product of all r_i's from the N_{test} remaining trials:

$$t = \prod_{i=N_{\mathrm{est}}+1}^{N} r_i = \prod_{a,b,x,y} R_{abxy}^{N_{a,b,x,y}}, \tag{10}$$

where $N_{a,b,x,y}$ is now the number of times the combination (a, b, x, y) occurs in the N_{test} hypothesis-testing trials.

Let T_m denote the random variable obtained from the product of the PBRs of m trials. It can be shown [51] that if each r_i satisfies equation (9*b*), then we have that $\mathbb{E}(T_{i+1}|H_{\leqslant i}) \leqslant \mathbb{E}(T_i)$, where $\mathbb{E}$ denotes the expectation value and $H_{\leqslant i}$ denotes all past information obtained until the ith trial. This means the probability that T_m exceeds a particular value t can be upper bounded using Markov's inequality, and the upper bound itself is our p-value upper bound p_U:

$$\Pr\left[T_{N_{\mathrm{test}}} \geqslant t\right] \leqslant \min\left(t^{-1}, 1\right) =: p_U. \tag{11}$$

A small p_U, and hence a small p-value, would represent a large value of t, which would only occur if we had sufficiently many $r_i > 1$. Note that this argument relies only on the supermartingale property of T_m; thus anything we conclude from the hypothesis testing is valid even with non-*i.i.d.* trials. In contrast, if $t^{-1} > 1$, we have the *trivial* p-value bound $p_U = 1$, which does not provide any useful information about the validity of the null hypothesis.

Before presenting our results and analysis, let us briefly comment on one final subtlety regarding the detection of NS violation via a Bell-like inequality violation. Clearly, the NS constraints of equation (1) consist of a collection of equality constraints. To see their connection with an inequality like equation (9), it suffices to remember that any equality constraint ($=$) is *equivalent* to the conjunction of two inequality constraints ($\geqslant$ and $\leqslant$). In other words, violating any of the NS conditions must also imply a violation of at least one inequality constraint analogous to those shown in equation (1).

[6] Due to numerical imprecisions, the solver may only find a correlation close to the true minimizer $\vec{P}_\star^{\mathrm{NS}}$. Then, the Bell-like inequality of equation (9) only holds approximately, with the maximum of the left-hand-side of equation (9*a*) over all $\vec{P} \in \mathcal{NS}$ being $1 + \epsilon$, for some tiny $\epsilon > 0$. In this case, we ought to renormalize (i.e. divide) the PBRs obtained from equation (9*b*) by $1 + \epsilon$ to ensure that the p-value bound obtained thereafter is valid.

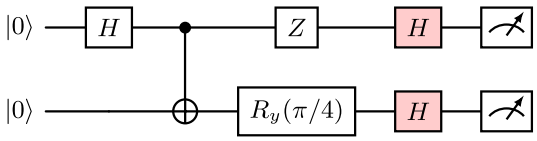

Figure 2. Quantum circuits generating (ideally) the maximal CHSH-Bell-inequality-violating correlation. H is the Hadamard gate, $R_y(\theta) = \cos\frac{\theta}{2}\mathbb{1} - i\sin\frac{\theta}{2}Y$, where $\mathbb{1}$ and Y are, respectively, the identity and the Pauli-Y operator, and the meter symbol represents the computational basis measurement. The pink shading indicates that the Hadamard gate is implemented before the measurement only when the input for the top (bottom) qubit is $x = 1$ ($y = 1$).

Table 1. List of qubit pairs in each IBMQ device where we perform the two types of CHSH Bell tests. For example, Washington(12,17) means the qubit pair (12, 17) of the IBMQ device Washington. For the topology of the qubit connections in these devices and their calibration data, see appendix A.

IBMQ device	Qubit Pairs
Washington	(12,17) (38,39) (79,91) (91,98)
Geneva	(7,10) (14,16) (21,23)
Cairo	(0,1) (7,10) (13,14) (23,24)
Hanoi	(5,8) (6,7) (11,14) (19,20)
Mumbai	(5,8) (16,19) (23,24)

3. CHSH Bell tests in IBM quantum computers

As mentioned at the beginning of section 2.1, in this work, we focus on the case where $|X| = |Y| = |A| = |B| = 2$. In this case, it is known [1] that $\mathcal{L}$ can be equivalently specified as the intersection of positivity facets and eight different versions of the Clauser–Horne–Shimony–Holt (CHSH) [66] Bell inequality. In what follows, we explain how the PBR protocol can be applied to the data collected in this simplest Bell scenario in conjunction with various hypotheses that allow us to identify measurement cross-talks. Note, however, that the analysis can be easily adapted to other certification tasks when the NS conditions of equation (1) hold and more complicated Bell scenarios, as illustrated in [49].

For the CHSH Bell test on an IBMQ device, we consider the setting where Alice and Bob share a two-qubit state and they perform a local measurement in two possible bases on each qubit. In the actual implementation, this means we first choose the pair of qubits representing Alice and Bob. Then, each round of the Bell test goes as follows: First, we apply the quantum gates needed to prepare the initial shared state. Next, according to the pair of inputs (x, y) with $x, y \in \{0, 1\}$, we perform one of the four possible quantum circuits that implement the local measurements to obtain the pair of outcomes (a, b). We record the data (a, b, x, y) in each round to facilitate subsequent analysis.

In a typical Bell test, one enforces the NS conditions in one way or another and seeks to demonstrate Bell nonlocality. Here, we test whether the observations are consistent with a quantum model *assuming* local, independent measurements, equation (3), or more generally, the NS constraints of equation (1). To this end, we focus on the commonly encountered Bell test that aims to produce a Bell-nonlocal correlation maximally violating the CHSH Bell inequality. We also fix Alice's and Bob's two measurements to be the ones in the computational (Pauli-Z) and Hadamard (Pauli-X) bases for $x, y = 0, 1$, respectively. The Bell test can then be conveniently described using the quantum circuits we implement on the IBMQ devices.

Specifically, the circuits $\mathcal{C}_{\rm NL}$ to generate a Bell-nonlocal correlation are given in figure 2. Ideally, this circuit prepares the maximally entangled state

$$|\psi\rangle = \frac{1}{\sqrt{2}}\left[\cos\frac{\pi}{8}\left(|00\rangle - |11\rangle\right) + \sin\frac{\pi}{8}\left(|01\rangle + |10\rangle\right)\right], \tag{12}$$

and measures Pauli-Z and Pauli-X on both qubits, thereby giving the maximal-CHSH-violating nonlocal correlation.

Now that we have specified the Bell test, it remains to choose the two specific qubits in the IBMQ device to represent Alice and Bob. To demonstrate the viability of the PBR protocol, we use the information about the average CNOT gate errors reported around the last week of April 2023 in several IBMQ devices to select those pairs of qubits with relatively high errors, see appendix A for details. The pairs of qubits chosen are indicated in table 1 using the device name and qubit numbers.

A few remarks on the data acquisition process are now in order. In an IBMQ device, a task consists of specifying the quantum circuit to be implemented and the number of shots, i.e. how many times we repeat the experiment. However, for a proper Bell test, the inputs (x, y) must be generated randomly and

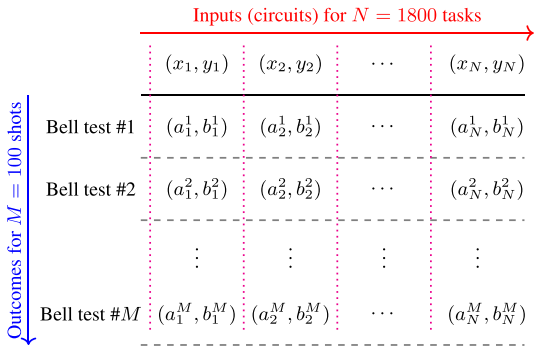

Figure 3. Schematic showing how the data for hypothesis testing is collected in each IBMQ device. For each device, we submit $N = 1800$ tasks with $M = 100$ shots each. The kth task is specified by the two input bits (x_k, y_k), corresponding to one of the four circuits shown in figure 2. All results (a_k^i, b_k^i) for the ith shot across the different tasks are then consolidated as results from the ith Bell test on each device.

uncorrelated with the state of the qubits to be tested. To this end, one may first generate a (pseudo)random sequence of input pairs (x, y) and submit a task defined by the sequence of circuits corresponding to these pairs while setting the number of shots to *unity* for each circuit.

Even then, the issue remains that various shots may become correlated since we must specify the entire input bit strings $(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)$ when submitting the task. In other words, from the perspective of a loophole-free Bell test [67–70], this potentially allows the leakage of inputs across parties. For cross-talk detection, we shall *assume* that this potential leakage does not alter the behavior of the individual qubits. Even though this *assumption* renders our protocol *non-fully-DI*, and hence our choice of the term *almost DI*[7], its violation would again imply some kind of cross-talks that should be addressed to improve local addressability. Moreover, to collect statistically significant sets of data more efficiently, instead of measuring one shot for each circuit, we carry out multiple shots for each circuit but assign each shot to a different Bell test. This means if we want to run M Bell tests where each Bell test consists of N rounds (experimental trials), we submit N tasks to an IBMQ device where each task consists of M shots. Then, the data produced by the ith shot of every task is treated as the data for the ith Bell test. See figure 3 for a schematic explanation of the data acquisition process.

Finally, to make the comparisons across different IBMQ devices relatively fair, we standardize each CHSH Bell test of figure 2 to have $N = 1800$ trials, and we perform $M = 100$ tests for each pair of qubits chosen. Moreover, for the PBR analysis, we use the data from the first $N_{\text{est}} = 600$ trials to obtain the empirical frequencies $\vec{f}$, and the remaining $N - N_{\text{est}} = 1200$ trials for computing the p-value upper bounds. Importantly, one can equally well make other choices of N_{est}. The general principle here is that we need a sufficient amount of data to get a reasonably good estimate of the general behavior (via $\vec{f}$), and hence a good PBR via equation (9*b*), but we also need a sufficient amount of data from a *different* set of trials for performing the actual hypothesis testing (via the test statistic). In our analysis, we adopt a significance level of $\alpha = 0.05$, which means we reject the null hypothesis if the p-value bound is less than α.

4. Results

After collecting the data from the Bell tests described in the previous section, we perform various PBR analyses by testing the data against different null hypotheses.

4.1. PBR protocol for revealing the violation of Born's rule for local measurements

Since we are interested in the local addressability of these devices, we start by employing a PBR analysis to check for signatures that the measurement statistics violate Born's rule for local measurements, cf equation (3). Due to statistical fluctuations, empirical frequencies $\vec{f}$ of equation (6) typically do not satisfy the NS constraints of equation (1). While this does not compromise the PBR protocol in computing valid p-value bounds, previous studies [49, 51, 52] have suggested that the quality of p-value bounds may be improved by first transforming $\vec{f}$ into an initial estimate $\vec{G} \in \mathcal{NS}$. For example, we can set [48] $\vec{G}$ as the

[7] The term almost DI was also used very differently in [71] to refer to a situation where only one of the parties in a multipartite scenario is trusted.

Table 2. Summary of nontrivial hypothesis-testing results based on the PBR protocol applied to the data collected in Bell tests performed on various IBMQ devices via the circuits $\mathcal{C}_{\mathrm{NL}}$ of figure 2 during 2023-04 to 2023-05 (see table 5 in appendix A for details). For each qubit pair, we implement 1800 tasks with 100 shots each, which means we conduct $M = 100$ separate Bell tests with $N = 1800$ trials each. For each Bell test, we run the PBR protocol for various hypotheses $\mathcal{H} \in \{\mathcal{Q}^3, \mathcal{NS}, \mathcal{NS}_{A\not\to B}, \mathcal{NS}_{B\not\to A}, \mathcal{L}\}$ with $N_{\mathrm{est}} = 600$ at a significance level of $\alpha = 0.05$. The integers from the third to the rightmost column show the number of Bell tests where we observe a signature, with a confidence of at least 95%, for the violation of various hypotheses: a relaxation of Born's rule for local measurements '$\mathcal{Q}^3$', (two-way) no-signaling '$\mathcal{NS}$', no-signaling from A to B '$\mathcal{NS}_{A\not\to B}$', no-signaling from B to A '$\mathcal{NS}_{B\not\to A}$', and LHV '$\mathcal{L}$'. Qubit A (B) corresponds to the first (second) integer entry in the second column. Only the combinations of device and qubit-pair where at least one of the entries from the third to the sixth column is nonzero is listed. For the corresponding results with the significance level tightened to $\alpha = 0.01$, see table 7 in appendix B.

Device	Qubits [Circuit]	$\mathcal{Q}^3$	$\mathcal{NS}$	$\mathcal{NS}_{A\not\to B}$	$\mathcal{NS}_{B\not\to A}$	$\mathcal{L}$
Washington	12,17 [$\mathcal{C}_{\mathrm{NL}}$]	1	1	0	1	0
	38,39 [$\mathcal{C}_{\mathrm{NL}}$]	2	2	0	1	2
	79,91 [$\mathcal{C}_{\mathrm{NL}}$]	1	1	0	0	1
	91,98 [$\mathcal{C}_{\mathrm{NL}}$]	2	2	1	1	1
Geneva	14, 16 [$\mathcal{C}_{\mathrm{NL}}$]	0	0	0	1	0
	21, 23 [$\mathcal{C}_{\mathrm{NL}}$]	19	19	5	30	17
Cairo	13,14 [$\mathcal{C}_{\mathrm{NL}}$]	1	1	1	0	100
Hanoi	5,8 [$\mathcal{C}_{\mathrm{NL}}$]	2	2	0	1	47
	11,14 [$\mathcal{C}_{\mathrm{NL}}$]	0	0	0	2	100
	19,20 [$\mathcal{C}_{\mathrm{NL}}$]	1	2	0	0	99
Mumbai	23,24 [$\mathcal{C}_{\mathrm{NL}}$]	1	1	0	1	64

minimizer of the KL-divergence from $\vec{f}$ to $\mathcal{NS}$:

$$\vec{G} := \underset{\vec{P}\in\mathcal{NS}}{\operatorname{argmin}} D_{\mathrm{KL}}\left(\vec{f}||\vec{P}\right). \tag{13}$$

The initial estimate $\vec{G}$ then plays the role of the frequencies $\vec{f}$ in the subsequent PBR analysis in equations (8) and (9).

More precisely, in determining the PBR for the hypothesis of equation (3) via equation (8), we use $\mathcal{Q}^3$, level-3 of the Moroder hierarchy [17] as a proxy for the local quantum constraints of equation (3), see the last paragraph of section 2.1. Hence, our null hypothesis, in fact, corresponds to the set $\mathcal{Q}^3$, which is a strict outer approximation of $\mathcal{Q}$. Still, a small p-value bound for $\mathcal{Q}^3$ signifies the violation of equation (3), in the sense that a rejection of $\vec{P} \in \mathcal{Q}^3$ must entail a rejection of $\vec{P} \in \mathcal{Q}$ since $\vec{P} \notin \mathcal{Q}^3 \implies \vec{P} \notin \mathcal{Q}$, see figure 1.

After the estimation stage, we obtain the PBRs

$$R_{abxy} = \frac{G(a,b|x,y)}{P_\star^{\mathcal{Q}^3}(a,b|x,y)}, \tag{14}$$

where $\vec{P}_\star^{\mathcal{Q}^3}$ represents the minimizer of the KL-divergence from $\vec{G}$ to $\mathcal{Q}^3$, i.e.

$$\vec{P}_\star^{\mathcal{Q}^3} := \underset{\vec{Q}\in\mathcal{Q}^3}{\operatorname{argmin}} D_{\mathrm{KL}}\left(\vec{G}||\vec{Q}\right). \tag{15}$$

Next, we proceed to the hypothesis testing stage using the PBRs R_{abxy} from equation (14) in the usual way. That is, we compute the test statistic in equation (10) from the number of occurrences $N_{a,b,x,y}$ of the input-output combination (a,b,x,y) in the hypothesis testing trials.

Finally, for the IBMQ devices listed in table 1 and each of the $M = 100$ Bell tests performed, we compare the p-value upper bound obtained against the significance level $\alpha = 0.05$ to decide if the null hypothesis corresponding to (a relaxation of) equation (3) should be rejected. In table 2, we list the IBMQ devices (alongside the qubit pairs) where at least one instance of rejection is recommended by the PBR protocol for this significance level. In most cases, we only observe one or two such instances. However, for our implementation of the circuits of figure 2 using qubit pair Geneva(21,23), cf figure 6, we find, with a confidence of at least 95%, incompatibility with equation (3) for 19 out of 100 instances of the conducted Bell tests. A histogram showing the distribution of these p-value bounds can be found in figure 4. These results reveal a strong signature for the inappropriateness of using equation (3) to model the measurement statistics on these two qubits of this particular IBMQ device.

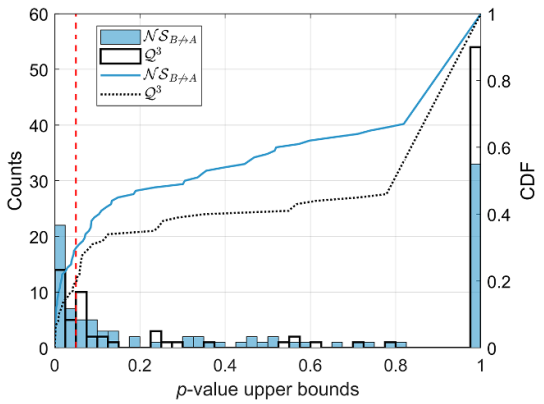

Figure 4. Histogram (H) and cumulative distribution function (CDF) of p-values upper bounds obtained from the PBR protocol for qubits 21 and 23 of the IBMQ device Geneva (see figure 6) for hypotheses $\mathcal{H} = \mathcal{Q}^3$ (H: unfilled, thicker edged boxes; CDF: dotted black line) and $\mathcal{H} = \mathcal{NS}_{B\not\to A}$ (H: filled blue boxes; CDF: solid blue line). With the inclusion relations of equation (4), see also figure 1, rejecting the latter hypothesis for a particular test must also entail a rejection of the $\mathcal{NS}$, and hence the $\mathcal{Q}^3$ hypothesis for the same set of data. Each bin spreads over a p-value bound of 0.025. The dashed red vertical line indicates the significance level of $\alpha = 0.05$.

4.2. PBR protocol for revealing signaling effects

Since the measurement results analyzed above are those generated from the circuits of figure 2, their incompatibility with equation (3) even when non-*i.i.d.* behavior is allowed, i.e. the state ρ_{AB} in equation (3) is allowed to vary from one trial to another, is already a strong indication that cross-talks are present in some of these IBMQ devices. A more direct evidence of this undesired aspect follows if we can show that the measurement statistics exhibit signaling effects, i.e. violate *one or more* of the NS conditions given in equation (1). To this end, instead of following the analysis presented in section 4.1, we proceed according to the illustration given in section 2.2 to obtain p-value upper bounds according to the NS hypothesis. The corresponding results are listed under column $\mathcal{NS}$ of table 2.

When we compare these results with the ones obtained above for the $\mathcal{Q}^3$ hypothesis (see also table 3), we see that they are almost identical, except for the Bell test corresponding to the 21st shots on Hanoi(19,20) (figure 9), where we find evidence for the violation of $\mathcal{NS}$ but not $\mathcal{Q}^3$. In this case, we find a p-value bound of 0.046 for the former but only the trivial bound of 1 for the latter. Since $\mathcal{Q}^3 \subset \mathcal{NS}$, i.e. $\mathcal{NS}$ is a less-constraining set of correlations than $\mathcal{Q}^3$, the above observation may appear counterintuitive at first glance, as one may expect to see fewer, rather than more, instances of violation of the $\mathcal{NS}$ hypothesis. To understand the origin of this discrepancy, we remind that our protocol involves an estimation of the PBR (i.e. a Bell-like inequality) optimized for the relative frequencies $\vec{f}_{\text{est}}$ deduced from the data collected during the first N_{est} trials in each Bell test. However, for finite and non-*i.i.d.* trials, there is no guarantee that such an estimate is again optimal for the data collected during the remaining trials, see [51]. Indeed, for this specific Bell test, if we had used the PBR obtained for $\mathcal{NS}$—also valid for the $\mathcal{Q}^3$ hypothesis—for our test against $\mathcal{Q}^3$, we would have also concluded a rejection of $\mathcal{Q}^3$.

In other words, if the underlying process is always described by *a fixed* $\vec{P} \notin \mathcal{NS}$, we must also have $\vec{P} \notin \mathcal{Q}^3$. However, for non-*i.i.d.* trials, as remarked in the paragraph below equation (9*b*), if the empirical frequencies $\vec{f}_{\text{est}}$ do not reflect well the behavior of subsequent trials, the PBR derived therefrom for $\mathcal{Q}^3$ may fail to manifest the incompatibility between the hypothesis-testing trials data and $\mathcal{Q}^3$. In contrast, even if $\vec{f}_{\text{test}}$ differ considerably from $\vec{f}_{\text{est}}$, so long as the main signaling direction is preserved, it is conceivable that the PBR derived for $\mathcal{NS}$ remains effective for the testing trials.

Apart from this one exceptional instance with Hanoi(19,20), the compatibility of every other Bell test's data with the two hypotheses (i.e. whether the p-value bound is less than α) is the same. In fact, even though the two hypotheses are not the same, the difference in their p-value bounds is typically not large enough to alter their distribution in a significant manner. For example, for the p-value upper bounds shown in figure 4, the corresponding p-value bounds for the $\mathcal{NS}$ hypothesis differ from the former by at most 0.0017 and are thus not visibly different from the histogram of figure 4 for $\mathcal{Q}^3$ (unfilled, thicker edge).

While small p-values indicate strong evidence against the $\mathcal{NS}$ hypothesis, they do not tell us anything about how the NS constraints of equation (1) are violated. One possibility is that including a Hadamard or not before the top (bottom) qubit measurement in figure 2 indeed results in different measurement statistics on the other qubit, which, of course, goes against the assumption of equation (1), and hence equation (3). To this end, it will also be useful to check if the cross-talk has a specific directionality by running a PBR protocol

Table 3. Further details about the instances of Bell tests giving the results reported in table 2. Under the third column to the rightmost, we list the Bell test number implemented on the respective device and qubit pair that shows a violation of the corresponding null hypothesis. To simplify the presentation, we have put the almost identical results for $\mathcal{Q}^3$ and $\mathcal{NS}$ under the same column, with the additional instance for $\mathcal{NS}$ in a bracket. Moreover, we denote the common instances for $\mathcal{Q}^3$ ($\mathcal{NS}$) and $\mathcal{NS}_{B\not\to A}$ in the case of Geneva(21,23) by $\mathfrak{C} = 3, 16, 18, 19, 31, 38, 55, 62, 65, 81, 97$.

Device	Qubits [Circuit]	$\mathcal{Q}^3$ / $\mathcal{NS}$	$\mathcal{NS}_{A\not\to B}$	$\mathcal{NS}_{B\not\to A}$
Washington	12,17 [$\mathcal{C}_{\mathrm{NL}}$]	56	—	49
	38,39 [$\mathcal{C}_{\mathrm{NL}}$]	88, 100	—	88
	79,91 [$\mathcal{C}_{\mathrm{NL}}$]	41	—	—
	91,98 [$\mathcal{C}_{\mathrm{NL}}$]	15, 50	15	15
Geneva	14, 16 [$\mathcal{C}_{\mathrm{NL}}$]	—	—	99
	21, 23 [$\mathcal{C}_{\mathrm{NL}}$]	$\mathfrak{C}$, 9, 17, 26, 32, 51, 66, 73, 99	9, 29, 38, 66, 73	$\mathfrak{C}$, 6, 33, 39 41–44, 47 56–59, 68 72, 86, 89 90, 92, 94
Cairo	13,14 [$\mathcal{C}_{\mathrm{NL}}$]	8	8	—
Hanoi	5,8 [$\mathcal{C}_{\mathrm{NL}}$]	23, 52	—	23
	11,14 [$\mathcal{C}_{\mathrm{NL}}$]	—	—	65, 75
	19,20 [$\mathcal{C}_{\mathrm{NL}}$]	75 (21)	—	—
Mumbai	23,24 [$\mathcal{C}_{\mathrm{NL}}$]	70	—	70

assuming the hypothesis $\mathcal{H} = \mathcal{NS}_{A\not\to B}$ ($\mathcal{H} = \mathcal{NS}_{B\not\to A}$) of one-way NS from Alice to Bob (Bob to Alice). Our results for these tests can be found in their respective columns in table 2.

From tables 2 and 3, we observe several instances—namely, Hanoi(11,14), Geneva(14,16), and Geneva(21,23)—where more violation of the *less* constraining OWNS hypothesis (either $\mathcal{NS}_{A\not\to B}$ or $\mathcal{NS}_{B\not\to A}$) is observed, but via the PBR protocol described in section 4.1, fewer or *no* violation of the *more* constraining $\mathcal{NS}$ hypothesis is picked up. This anomaly can again be understood from the non-*i.i.d.* nature of the experimental trials, where the main direction of signaling (estimated from $\vec{f}_{\mathrm{est}}$ and $\vec{f}_{\mathrm{test}}$) changes from the first 600 trials to the remaining 1200 trials.

In fact, we can 'utilize' this discrepancy to our advantage in our hypothesis-testing tasks. By recalling from equation (4) and figure 1 the strict inclusions of the various sets of correlations, we note that $\mathcal{Q}$ is the most constraining hypothesis among all those discussed above, while the OWNS hypothesis is the weakest. In other words, if we reject the plausibility of any of the hypotheses from $\{\mathcal{NS}_{A\not\to B}, \mathcal{NS}_{B\not\to A}\}$ in explaining the data observed for a particular Bell test, we must also reject the plausibility of $\mathcal{NS}$ (and hence $\mathcal{Q}$) in explaining the same set of data. Using this observation, we conclude from table 3 that of the 100 tests performed on Geneva(21,23) 39 are deemed incompatible with the NS hypothesis $\mathcal{NS}$ (or $\mathcal{Q}$). See table 4 for a complete summary of such results on all the IBMQ devices we have tested.

5. Discussion

In recent years, due to the widespread availability of quantum computers through the cloud, we have seen a surging interest in running various quantum tasks on these devices. Naturally, given the proximity of the qubits arranged in some of these platforms—such as those offered by IBMQ—one may wonder about the extent to which they exhibit cross-talks and whether such effects can detected with minimal assumptions, like other DI certification tasks. To this end, it is worth noting that the NS conditions of equation (1) are usually separately enforced and taken as a premise for DI protocols.

In this work, we show under a mild assumption that measurement cross-talks or incompatibility with Born's rule for local measurements can again be certified in an essentially DI manner via the PBR protocol (initially developed in [51, 52] for testing LHV theories but later generalized in [43]). More precisely, we use the protocol to obtain p-value upper bounds on the plausibility of the NS assumption or the natural assumption that Born's rule for local measurements holds. Note that an analysis of the first kind has previously been applied as a consistency check in the loophole-free Bell test performed with superconducting circuits [72], where no evidence for signaling is found.

Similarly, from our analysis of the data obtained across five different IBMQ systems, we see, in most cases, very little evidence for a strong violation of either the $\mathcal{Q}^3$ or any of the $\mathcal{NS}$ hypotheses. Although we observe a small p-value upper bound p_U in a few instances (see table 4 for a summary), it should be noted

Table 4. Summary of the number of (nonzero) instances of Bell tests found to be incompatible with the no-signaling hypothesis, either via the rejection of the $\mathcal{NS}$ null hypothesis, or indirectly via one of the OWNS hypotheses for a significance level of 5% (third column) and 1% (fourth column). We find the same results for rejecting the hypothesis of Born's rule for local measurements, equation (3). Results listed on top and bottom are, respectively, those based on the circuits $\mathcal{C}_{\rm NL}$ of figure 2 (meant for generating a Bell-nonlocal correlation) and $\mathcal{C}_{\rm L}$ of figure 11(meant for generating a Bell-local correlation, see appendix C for details).

Device	Qubits[Circuit]	$\alpha = 0.05$	$\alpha = 0.01$
Washington	12,17 [$\mathcal{C}_{\rm NL}$]	2	1
	38,39 [$\mathcal{C}_{\rm NL}$]	2	1
	79,91 [$\mathcal{C}_{\rm NL}$]	1	0
	91,98 [$\mathcal{C}_{\rm NL}$]	2	1
Geneva	14, 16 [$\mathcal{C}_{\rm NL}$]	1	1
	21, 23 [$\mathcal{C}_{\rm NL}$]	39	22
Cairo	13,14 [$\mathcal{C}_{\rm NL}$]	1	1
Hanoi	5,8 [$\mathcal{C}_{\rm NL}$]	2	1
	11,14 [$\mathcal{C}_{\rm NL}$]	2	0
	19,20 [$\mathcal{C}_{\rm NL}$]	2	0
Mumbai	23,24 [$\mathcal{C}_{\rm NL}$]	1	0
Washington	12,17 [$\mathcal{C}_{\rm L}$]	1	0
	38,39 [$\mathcal{C}_{\rm L}$]	2	1
	79,91 [$\mathcal{C}_{\rm L}$]	2	1
	91,98 [$\mathcal{C}_{\rm L}$]	2	1
Cairo	13,14 [$\mathcal{C}_{\rm L}$]	2	0
	23,24 [$\mathcal{C}_{\rm L}$]	3	1
Hanoi	5,8 [$\mathcal{C}_{\rm L}$]	2	0
	6,7 [$\mathcal{C}_{\rm L}$]	1	0
	11,14 [$\mathcal{C}_{\rm L}$]	2	1
	19,20 [$\mathcal{C}_{\rm L}$]	1	0

that even when the null hypothesis holds, there remains a small chance ($< p_U$) of observing a false positive [51]. In contrast, for measurements on qubits 21 and 23 of the IBMQ-Geneva device, we have stumbled upon 39 instances of these tests where the PBR protocol would end up rejecting the $\mathcal{NS}$, and hence $\mathcal{Q}$ hypothesis, either directly, or indirectly via a weaker hypothesis. This shows that, despite the relatively small number of samples (1800 trials for each test) and allowing non-*i.i.d.* trials (cf the approach by [73] with *i.i.d.* assumption), the PBR protocol is capable of detecting (measurement) cross-talks in a real quantum computer.

Note further that when we check the same set of data from Geneva(21,23) against the $\mathcal{L}$ hypothesis of LHV theories[8], we also find several instances that result in rejecting the $\mathcal{L}$ hypothesis. However, given the observed signaling effects, the relevance of this violation becomes questionable. For comparison, we have also implemented several trivial 'Bell tests' using the circuits $\mathcal{C}_{\rm L}$ of figure 11 in appendix C, which are only expected to produce Bell-local correlations. Then, for ideal devices, we anticipate many small p-values for Bell tests involving $\mathcal{C}_{\rm NL}$, figure 2, and none for those involving $\mathcal{C}_{\rm L}$. The results shown in tables 2 and 8 clearly do not follow this intuition. In fact, from tables 8 and 9, we even observe a few instances of rejection of $\mathcal{L}$ alongside $\mathcal{Q}^3$ and $\mathcal{NS}$ with $\mathcal{C}_{\rm L}$, suggesting that these violations of equation (2) are merely an artifact of the cross-talks present in the system. Even though we have not seen overwhelming instances of rejections of the $\mathcal{NS}$ or any of the OWNS hypotheses for the $\mathcal{C}_{\rm L}$ circuit, cf tables 4 and 8, their presence, nonetheless, support the idea that these cross-talks show up even without implementing any nonlocal unitary gate.

Let us make a few final remarks about our general methodology. Our certification protocol can be seen as assuming the causal structure of figure 5(a) and applying the PBR method to show that the observed data is incompatible with this assumption. Strictly, a refutation of the assumed causal structure does not necessarily entail signaling, and hence a measurement cross-talk. For example, the causal structure depicted in figure 5(b)—which allows signaling (dashed arrows) and dependence of C on X, Y (dotted arrows)—facilitates the generation of all possible $\vec{P}$ in this scenario, cf lemma 1 of [74].

[8] To check against $\mathcal{L}$ using the PBR protocol, we replace, in the definition of the PBRs of equation (14), the denominator by $\text{argmin}_{\vec{Q}\in\mathcal{L}} D_{\rm KL}(\vec{G}||\vec{Q})$, i.e. the minimizer of the KL divergence from $\vec{G}$ to $\mathcal{L}$.

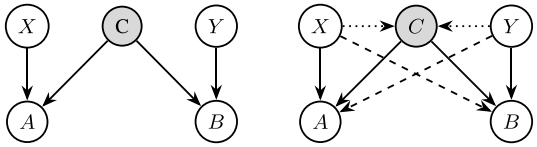

Figure 5. Causal structures relevant to the current work. (a) The Bell directed acyclic graph [76] (DAG) where the random variable A (B) is allowed to depend on both X (Y) and the latent variable C. When C is a classical or a quantum common cause, the resulting correlation $\vec{P}$ satisfies, respectively, equations (2) and (3). Even if C is a more powerful no-signaling common cause, $\vec{P}$ must still respect equation (1). (b) A natural extension of the Bell DAG allowing signaling (dashed arrows) and 'measurement dependence' [4] (dotted arrows).

In an ordinary Bell test, one invokes the freedom of choice assumption to remove these dependencies between C and X, Y, making $P(X, Y|C) = P(X, Y)$, which is equivalent to $P(C|X, Y) = P(C)$. In our case, however, the inputs X, Y are not only generated before C, but are even fed into the device used to prepare C. Thus, although we make the same independence assumption, some may find it more difficult to justify in the present context. To this end, one might find a random permutation of the bit strings (and qubits) before each submission helpful for breaking any accidental correlations between the input pattern and the underlying time-dependent noise. Alternatively, one can follow the approach of [4] and try to develop a more general type of NS conditions that hold even if we allow these undesired dependencies.

On the other hand, most of our tests admittedly involve qubit pairs exhibiting relatively high error rates. However, even among those pairs where the error rates seem low, including Washington(38, 39) (see figure 7), Cairo(13, 14) and Cairo(23, 24) (see figure 8), and Hanoi(6,7) (see figure 9), our protocol has also identified instances showing signatures of cross-talk, see table 4. For future reference, it would be helpful to perform a comprehensive investigation involving a control set of low-error pairs and compare the results obtained against IBMQ's calibration data. In particular, this will shed light on the effectiveness of our tests—which involve *far fewer assumptions*—even in those cases that may not be flagged via the conventional approach.

Note also that our results (see table 2, 3, 8 and 9) clearly suggest that, once we have done the much faster computation checking against the $\mathcal{NS}$ hypothesis, the computation using any approximation of $\mathcal{Q}$ may well be redundant for detecting cross-talk. A natural question that follows is whether this observation holds in general. Another obvious question that follows is: for the same number of trials, whether one can obtain—for the sake of detecting cross talks—a tighter p-value bound for refuting the hypothesis of Born's rule for local measurements, equation (3), or even NS of equation (1). To this end, we remind the readers that the PBR protocol is only known to be optimal in the asymptotic setting (and when the trials are *i.i.d.*). For example, is there a way to adopt the analysis from [75] to the present setting by considering the conjunction of all inequalities equivalent to equation (1)? Evidently, it is also relevant to understand if adapting the present analysis can give a useful quantification of cross-talks. Finally, given the current findings, one may ask whether other, more general (almost) DI certification or calibration tasks can be developed to detect additional undesirable behaviors of quantum devices.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://github.com/ycliangTW/AlmostDICertCrossTalks.

Acknowledgments

We thank Yi-Te Huang for his help in implementing some of the earlier computations at IBMQ and are very grateful to him, Marina Maciel Ansanelli, and Yanbao Zhang for many helpful discussions. We are also very grateful to anonymous reviewers for providing very helpful comments and suggestions on an earlier version of this paper. This work is supported by the National Science and Technology Council, Taiwan (Grant Nos. 109-2112-M-006-010-MY3,112-2628-M006-007-MY4,113-2918-I-006-001), the Foxconn Research Institute, Taipei, Taiwan, and in part by the Perimeter Institute for Theoretical Physics. Research at Perimeter Institute is supported by the Government of Canada through the Department of Innovation, Science, and Economic Development, and by the Province of Ontario through the Ministry of Colleges and Universities.

Appendix A. Miscellaneous details

Here, we provide further details about the data acquisition period, table 5, and the IBMQ devices investigated in this work. These include IBM Cairo (figure 8), the exploratory—now retired—IBM Geneva (figure 6), IBM Hanoi (figure 9), IBM Mumbai (figure 10), and IBM Washington (figure 7).

A.1. Data acquisition period

Table 5. Period for which we collected the data at IBMQ devices.

Device	Data collection period
Washington	24 April 2023–26 April 2023
Geneva	26 April 2023–27 April 2023
Cairo	24 April 2023–30 April 2023
Hanoi	24 April 2023–17 May 2023
Mumbai	3 May 2023–6 May 2023

A.2. Topology of qubit connections in each IBMQ device and their calibration data

For each device listed in table 5, we provide below the topology map showing its qubit connection, calibration data taken around the time the computation data was collected, the range and median of its readout assignment error and CNOT error, and the qubit pairs investigated in this work.

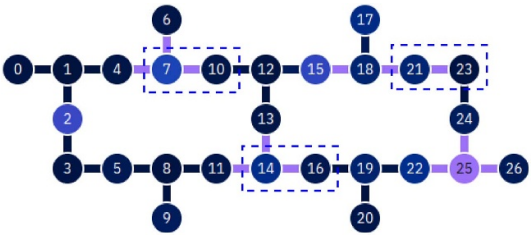

Figure 6. Topology of the 27-qubit exploratory IBM Geneva device and its calibration data on 26 April 2023: the readout (CNOT) assignment error ranges from 7.300×10^{-3} to 3.683×10^{-1} (3.872×10^{-3} to 1.000) with a median of 2.930×10^{-2} (5.457×10^{-2}). Here and below, the highest (lowest) error is associated with the brightest (darkest) color; qubit pairs analyzed in this work are enclosed in dashed blue rectangles.

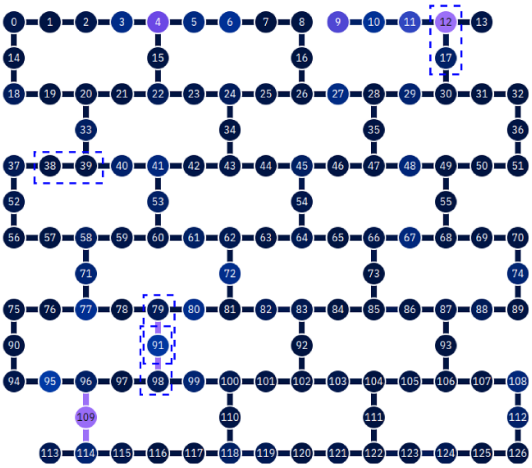

Figure 7. Topology of the 127-qubit IBM Washington device and its calibration data on 24 April 2023: the readout assignment (CNOT) error ranges from 1.900×10^{-3} to 4.854×10^{-1} (5.999×10^{-3} to 1.000) with a median of 1.290×10^{-2} (1.234×10^{-2}).

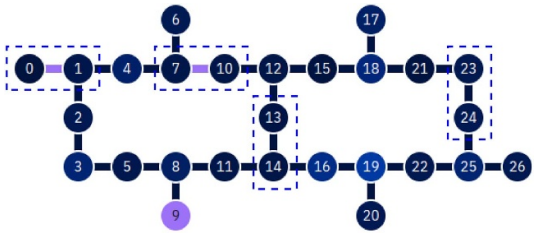

Figure 8. Topology of the 27-qubit IBM Cairo device and its calibration data on 24 April 2023: the readout assignment (CNOT) error ranges from 6.000×10^{-3} to 1.221×10^{-1} (4.436×10^{-3} to 1.000) with a median of 1.190×10^{-2} (9.815×10^{-3}).

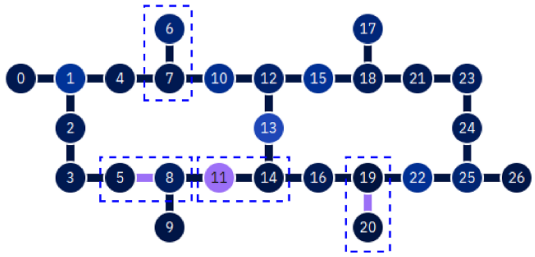

Figure 9. Topology of the 27-qubit IBM Hanoi device and its calibration data on 24 April 2023: the readout assignment (CNOT) error ranges from 5.800×10^{-3} to 8.690×10^{-2} (2.982×10^{-3} to 1.000) with a median of 1.150×10^{-2} (7.465×10^{-3}).

Table 6. Median and range of the readout assignment error and CNOT error for each IBMQ device.

	Readout error		CNOT error	
Device	Median	Range	Median	Range
Washington	1.290×10^{-2}	1.900×10^{-3}–4.854×10^{-1}	1.234×10^{-2}	5.999×10^{-3}–1.000
Geneva	2.930×10^{-2}	7.300×10^{-3}–3.683×10^{-1}	5.457×10^{-2}	3.872×10^{-3}–1.000
Cairo	1.190×10^{-2}	6.000×10^{-3}–1.221×10^{-1}	9.815×10^{-3}	4.436×10^{-3}–1.000
Hanoi	1.150×10^{-2}	5.800×10^{-3}–8.690×10^{-2}	7.465×10^{-3}	2.982×10^{-3}–1.000
Mumbai	1.860×10^{-2}	1.150×10^{-2}–1.115×10^{-1}	7.804×10^{-3}	4.201×10^{-3}–1.888×10^{-2}

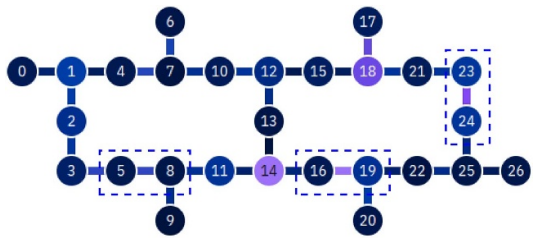

Figure 10. Topology of the 27-qubit IBM Mumbai device and its calibration data on 8 May 2023: the readout assignment (CNOT) error ranges from 1.150×10^{-2} to 1.115×10^{-1} (4.201×10^{-3} to 1.888×10^{-2}) with a median of 1.860×10^{-2} (7.804×10^{-3}).

Appendix B. Other miscellaneous results

We give here the results analogous to table 2, but with a more stringent (smaller) significance level.

Table 7. Summary of results parallel to those presented in table 2 but with the significance level set at the more stringent value of $\alpha = 0.01$.

Device	Qubits[Circuit]	$\mathcal{Q}^3$	$\mathcal{NS}$	$\mathcal{NS}_{A\not\to B}$	$\mathcal{NS}_{B\not\to A}$	$\mathcal{L}$
Washington	12,17 [$\mathcal{C}_{\mathrm{NL}}$]	0	0	0	1	0
	38,39 [$\mathcal{C}_{\mathrm{NL}}$]	0	0	0	1	0
	91,98 [$\mathcal{C}_{\mathrm{NL}}$]	1	1	0	0	1
Geneva	14, 16 [$\mathcal{C}_{\mathrm{NL}}$]	0	0	0	1	0
	21, 23 [$\mathcal{C}_{\mathrm{NL}}$]	10	10	1	17	8
Cairo	13,14 [$\mathcal{C}_{\mathrm{NL}}$]	1	1	1	0	100
Hanoi	5,8 [$\mathcal{C}_{\mathrm{NL}}$]	1	1	0	0	41

Appendix C. Results for a product-state generating circuit

The circuits $\mathcal{C}_{\mathrm{L}}$ for generating a Bell-local correlation are given in figure 11. The ideal correlation resulting from this circuit is that obtained by measuring Pauli-Z and Pauli-X on the state $|00\rangle$. Note that while the T gate is irrelevant in theory, the fact that it is performed in the circuit can still have a nontrivial consequence in the experiment.

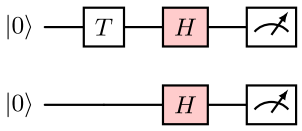

Figure 11. Quantum circuits generating a non-Bell-inequality-violating correlation. $T = \mathrm{diag}(1, e^{i\pi/4})$ refers to the $\pi/8$-phase gate. The symbols and shading carry the same meaning as those in figure 2.

Table 8. Summary of results analogous to those presented in table 2 but with the circuits considered being those given in figure 11.

Device	Qubits [Circuit]	$\mathcal{Q}^3$	$\mathcal{NS}$	$\mathcal{NS}_{A\not\to B}$	$\mathcal{NS}_{B\not\to A}$	$\mathcal{L}$
Washington	12,17 [$\mathcal{C}_{\mathrm{L}}$]	0	0	1	0	0
	38,39 [$\mathcal{C}_{\mathrm{L}}$]	1	1	2	0	1
	79,91 [$\mathcal{C}_{\mathrm{L}}$]	1	1	0	2	1
	91,98 [$\mathcal{C}_{\mathrm{L}}$]	2	2	0	2	1
Cairo	13,14 [$\mathcal{C}_{\mathrm{L}}$]	0	0	2	0	0
	23,24 [$\mathcal{C}_{\mathrm{L}}$]	1	1	1	1	1
Hanoi	5,8 [$\mathcal{C}_{\mathrm{L}}$]	2	2	0	1	1
	6,7 [$\mathcal{C}_{\mathrm{L}}$]	1	1	0	0	0
	11,14 [$\mathcal{C}_{\mathrm{L}}$]	2	2	2	0	0
	19,20 [$\mathcal{C}_{\mathrm{L}}$]	0	0	0	1	0

Table 9. Further details about the instances of Bell tests giving the results reported in table 8. Under the third column to the rightmost, we list the Bell test number implemented on the respective device and qubit pair that shows a violation of the corresponding null hypothesis. To simplify the presentation, we have put the identical results for $\mathcal{Q}^3$ and $\mathcal{NS}$ under the same column.

Device	Qubits [Circuit]	$\mathcal{Q}^3/\mathcal{NS}$	$\mathcal{NS}_{A\not\to B}$	$\mathcal{NS}_{B\not\to A}$	$\mathcal{L}$
Washington	12,17 [$\mathcal{C}_{\mathrm{L}}$]	—	49	—	—
	38,39 [$\mathcal{C}_{\mathrm{L}}$]	98	57, 98	—	98
	79,91 [$\mathcal{C}_{\mathrm{L}}$]	78	—	30, 78	78
	91,98 [$\mathcal{C}_{\mathrm{L}}$]	38, 64	—	38, 64	38
Cairo	13,14 [$\mathcal{C}_{\mathrm{L}}$]	—	64, 76	—	—
	23,24 [$\mathcal{C}_{\mathrm{L}}$]	30	59	77	30
Hanoi	5,8 [$\mathcal{C}_{\mathrm{L}}$]	11, 61	—	11	61
	6,7 [$\mathcal{C}_{\mathrm{L}}$]	58	—	—	—
	11,14 [$\mathcal{C}_{\mathrm{L}}$]	27, 68	27, 68	—	27
	19,20 [$\mathcal{C}_{\mathrm{L}}$]	—	—	96	—

References

[1] Brunner N, Cavalcanti D, Pironio S, Scarani V and Wehner S 2014 Bell nonlocality *Rev. Mod. Phys.* **86** 419
[2] Bell J S 1964 On the Einstein Podolsky Rosen paradox *Physics* **1** 195
[3] Bancal J-D, Pironio S, Acín A, Liang Y-C, Scarani V and Gisin N 2012 Quantum non-locality based on finite-speed causal influences leads to superluminal signalling *Nat. Phys.* **8** 867
[4] Pütz G, Rosset D, Barnea T J, Liang Y-C and Gisin N 2014 Arbitrarily small amount of measurement independence is sufficient to manifest quantum nonlocality *Phys. Rev. Lett.* **113** 190402
[5] Bong K-W, Utreras-Alarcón A, Ghafari F, Liang Y-C, Tischler N, Cavalcanti E G, Pryde G J and Wiseman H M 2020 A strong no-go theorem on the Wigner's friend paradox *Nat. Phys.* **16** 1199
[6] Colbeck R 2009 Quantum and relativistic protocols for secure multi-party computation *PhD Thesis* University of Cambridge (arXiv:0911.3814)
[7] Pironio S *et al* 2010 Random numbers certified by Bell's theorems theorem *Nature* **464** 1021
[8] Colbeck R and Kent A 2011 Private randomness expansion with untrusted devices *J. Phys. A: Math. Theor.* **44** 095305
[9] Barrett J, Hardy L and Kent A 2005 No signaling and quantum key distribution *Phys. Rev. Lett.* **95** 010503
[10] Acín A, Brunner N, Gisin N, Massar S, Pironio S and Scarani V 2007 Device-independent security of quantum cryptography against collective attacks *Phys. Rev. Lett.* **98** 230501
[11] Vazirani U and Vidick T 2014 Fully device-independent quantum key distribution *Phys. Rev. Lett.* **113** 140501
[12] Popescu S and Rohrlich D 1994 Quantum nonlocality as an axiom *Found. Phys.* **24** 379
[13] Zapatero V, van Leent T, Arnon-Friedman R, Liu W-Z, Zhang Q, Weinfurter H and Curty M 2023 Advances in device-independent quantum key distribution *npj Quantum Inf.* **9** 10
[14] Primaatmaja I W, Goh K T, Tan E Y-Z, Khoo J T-F, Ghorai S and Lim C C-W 2023 Security of device-independent quantum key distribution protocols: a review *Quantum* **7** 932
[15] Brunner N, Pironio S, Acin A, Gisin N, Méthot A A and Scarani V 2008 Testing the dimension of Hilbert spaces *Phys. Rev. Lett.* **100** 210503
[16] Bancal J-D, Gisin N, Liang Y-C and Pironio S 2011 Device-independent witnesses of genuine multipartite entanglement *Phys. Rev. Lett.* **106** 250404
[17] Moroder T, Bancal J-D, Liang Y-C, Hofmann M and Gühne O 2013 Device-independent entanglement quantification and related applications *Phys. Rev. Lett.* **111** 030501
[18] Liang Y-C, Rosset D, Bancal J-D, Pütz G, Barnea T J and Gisin N 2015 Family of Bell-like inequalities as device-independent witnesses for entanglement depth *Phys. Rev. Lett.* **114** 190401
[19] Chen S-L, Budroni C, Liang Y-C and Chen Y-N 2016 Natural framework for device-independent quantification of quantum steerability, measurement incompatibility and self-testing *Phys. Rev. Lett.* **116** 240401
[20] Chen S-L, Budroni C, Liang Y-C and Chen Y-N 2018 Exploring the framework of assemblage moment matrices and its applications in device-independent characterizations *Phys. Rev. A* **98** 042127
[21] Bancal J-D, Sangouard N and Sekatski P 2018 Noise-resistant device-independent certification of Bell state measurements *Phys. Rev. Lett.* **121** 250506
[22] Quintino M T, Budroni C, Woodhead E, Cabello A and Cavalcanti D 2019 Device-independent tests of structures of measurement incompatibility *Phys. Rev. Lett.* **123** 180401
[23] Wagner S, Bancal J-D, Sangouard N and Sekatski P 2020 Device-independent characterization of quantum instruments *Quantum* **4** 243
[24] Chen S-L, Miklin N, Budroni C and Chen Y-N 2021 Device-independent quantification of measurement incompatibility *Phys. Rev. Res.* **3** 023143
[25] Sekatski P, Bancal J-D, Wagner S and Sangouard N 2018 Certifying the building blocks of quantum computers from Bell's theorem *Phys. Rev. Lett.* **121** 180505
[26] DiVincenzo D P 2000 The physical implementation of quantum computation *Fortschr. Phys.* **48** 771
[27] Sarovar M, Proctor T, Rudinger K, Young K, Nielsen E and Blume-Kohout R 2020 Detecting crosstalk errors in quantum information processors *Quantum* **4** 321

[28] Emerson J, Alicki R and Życzkowski K 2005 Scalable noise estimation with random unitary operators *J. Opt. B: Quantum Semiclass. Opt.* **7** S347
[29] Lévi B, López C C, Emerson J and Cory D G 2007 Efficient error characterization in quantum information processing *Phys. Rev. A* **75** 022314
[30] Knill E, Leibfried D, Reichle R, Britton J, Blakestad R B, Jost J D, Langer C, Ozeri R, Seidelin S and Wineland D J 2008 Randomized benchmarking of quantum gates *Phys. Rev. A* **77** 012307
[31] Magesan E, Gambetta J M and Emerson J 2011 Scalable and robust randomized benchmarking of quantum processes *Phys. Rev. Lett.* **106** 180504
[32] Blume-Kohout R, Gamble J K, Nielsen E, Rudinger K, Mizrahi J, Fortier K and Maunz P 2017 Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography *Nat. Commun.* **8** 14485
[33] Nielsen E, Gamble J K, Rudinger K, Scholten T, Young K and Blume-Kohout R 2021 Gate set tomography *Quantum* **5** 557
[34] Merkel S T, Gambetta J M, Smolin J A, Poletto S, Córcoles A D, Johnson B R, Ryan C A and Steffen M 2013 Self-consistent quantum process tomography *Phys. Rev. A* **87** 062119
[35] Huang H-Y, Kueng R and Preskill J 2020 Predicting many properties of a quantum system from very few measurements *Nat. Phys.* **16** 1050
[36] Paini M, Kalev A, Padilha D and Ruck B 2021 Estimating expectation values using approximate quantum states *Quantum* **5** 413
[37] Helsen J, Ioannou M, Kitzinger J, Onorati E, Werner A H, Eisert J and Roth I 2023 Shadow estimation of gate-set properties from random sequences *Nat. Commun.* **14** 5039
[38] Epstein J M, Cross A W, Magesan E and Gambetta J M 2014 Investigating the limits of randomized benchmarking protocols *Phys. Rev. A* **89** 062321
[39] Fogarty M A, Veldhorst M, Harper R, Yang C H, Bartlett S D, Flammia S T and Dzurak A S 2015 Nonexponential fidelity decay in randomized benchmarking with low-frequency noise *Phys. Rev. A* **92** 022326
[40] Wallman J J 2018 Randomized benchmarking with gate-dependent noise *Quantum* **2** 47
[41] Figueroa-Romero P, Modi K, Harris R J, Stace T M and Hsieh M-H 2021 Randomized benchmarking for non-markovian noise *PRX Quantum* **2** 040351
[42] Figueroa-Romero P, Modi K and Hsieh M-H 2022 Towards a general framework of Randomized Benchmarking incorporating non-Markovian Noise *Quantum* **6** 868
[43] Liang Y-C and Zhang Y 2019 Bounding the plausibility of physical theories in a device-independent setting via hypothesis testing *Entropy* **21** 185
[44] Barrett J, Collins D, Hardy L, Kent A and Popescu S 2002 Quantum nonlocality, Bell inequalities and the memory loophole *Phys. Rev. A* **66** 042111
[45] Bancal J-D, Sheridan L and Scarani V 2014 More randomness from the same data *New J. Phys.* **16** 033011
[46] Bernhard C, Bessire B, Montina A, Pfaffhauser M, Stefanov A and Wolf S 2014 Non-locality of experimental qutrit pairs *J. Phys. A: Math. Theor.* **47** 424013
[47] Schwarz S, Bessire B, Stefanov A and Liang Y-C 2016 Bipartite Bell inequalities with three ternary-outcome measurements - from theory to experiments *New J. Phys.* **18** 035001
[48] Lin P-S, Rosset D, Zhang Y, Bancal J-D and Liang Y-C 2018 Device-independent point estimation from finite data and its application to device-independent property estimation *Phys. Rev. A* **97** 032309
[49] Chang W-G, Chen K-C, Chen K-S, Chen S-L and Liang Y-C 2024 Device-independent certification of desirable properties with a confidence interval *Front. Phys.* **12** 1434095
[50] Patra S and Bierhorst P 2024 Strength of statistical evidence for genuine tripartite nonlocality *Phys. Rev. A* **110** 062411
[51] Zhang Y, Glancy S and Knill E 2011 Asymptotically optimal data analysis for rejecting local realism *Phys. Rev. A* **84** 062118
[52] Zhang Y, Glancy S and Knill E 2013 Efficient quantification of experimental evidence against local realism *Phys. Rev. A* **88** 052119
[53] Bancal J-D, Navascués M, Scarani V, Vértesi T and Yang T H 2015 Physical characterization of devices from nonlocal correlations *Phys. Rev. A* **91** 022115
[54] Goh K T, Kaniewski J, Wolfe E, Vértesi T, Wu X, Cai Y, Liang Y-C and Scarani V 2018 Geometry of the set of quantum correlations *Phys. Rev. A* **97** 022104
[55] Chen K-S, Tabia G N M, Jebarathinam C, Mal S, Wu J-Y and Liang Y-C 2023 Quantum correlations on the no-signaling boundary: self-testing and more *Quantum* **7** 1054
[56] Navascués M, Pironio S and Acín A 2007 Bounding the set of quantum correlations *Phys. Rev. Lett.* **98** 010401
[57] Navascués M, Pironio S and Acín A 2008 A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations *New J. Phys.* **10** 073013
[58] Doherty A C, Liang Y-C, Toner B and Wehner S 2008 The quantum moment problem and bounds on entangled multi-prover games *23rd Annu. IEEE Conf. on Comput. Comp, 2008, CCC'08* (https://doi.org/10.1109/CCC.2008.26) pp 199–210
[59] Lin P-S, Vértesi T and Liang Y-C 2022 Naturally restricted subsets of nonsignaling correlations: typicality and convergence *Quantum* **6** 765
[60] Gill R 2003 Time, finite statistics and Bell's fifth position *Proc. Foundations of Probability and Physics-2* (*Ser. Math. Modelling in Phys., Engin. and Cogn. Sc.* vol 5) ed A Khrennikov (Växjö University Press) (https://doi.org/10.1109/CCC.2004.1313847) pp 179–206
[61] van Dam W, Gill R D and Grunwald P D 2005 The statistical strength of nonlocality proofs *IEEE Trans. Inf. Theory* **51** 2812
[62] MOSEK ApS 2025 The MOSEK optimization toolbox for MATLAB manual. Version 11.0.9 (available at: https://docs.mosek.com/latest/toolbox/index.html)
[63] Liang Y-C 2025 Almost DI certification of cross talks (available at: https://github.com/ycliangTW/AlmostDICertCrossTalks)
[64] Löfberg J 2004 YALMIP: a toolbox for modeling and optimization in MATLAB (available at: https://yalmip.github. io/)
[65] Fiala J, Kočvara M and Michael S 2013 PENLAB: a MATLAB solver for nonlinear semidefinite optimization (arXiv:1311.5240)
[66] Clauser J F, Horne M A, Shimony A and Holt R A 1969 Proposed experiment to test local hidden-variable theories *Phys. Rev. Lett.* **23** 880
[67] Hensen B *et al* 2015 Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres *Nature* **526** 682
[68] Shalm L K *et al* 2015 Strong loophole-free test of local realism *Phys. Rev. Lett.* **115** 250402
[69] Giustina M *et al* 2015 Significant-loophole-free test of Bell's theorem with entangled photons *Phys. Rev. Lett.* **115** 250401

[70] Rosenfeld W, Burchardt D, Garthoff R, Redeker K, Ortegel N, Rau M and Weinfurter H 2017 Event-ready Bell test using entangled atoms simultaneously closing detection and locality loopholes *Phys. Rev. Lett.* **119** 010402
[71] Sarkar S, Orthey A C Jr, Sharma G and Augusiak R 2024 Almost device-independent certification of GME states with minimal measurements (arXiv:2402.18522 [quant-ph])
[72] Storz S *et al* 2023 Loophole-free Bell inequality violation with superconducting circuits *Nature* **617** 265
[73] Rybotycki T, Białecki T, Batle J and Bednorz A 2025 Violation of no-signaling on a public quantum computer *Adv. Quantum Technol.* **8** 2400661
[74] Evans R J 2016 Graphs for margins of Bayesian networks *Scand. J. Stat.* **43** 625
[75] Elkouss D and Wehner S 2016 (Nearly) optimal P values for all Bell inequalities *npj Quantum Inf.* **2** 16026
[76] Wood C J and Spekkens R W 2015 The lesson of causal discovery algorithms for quantum correlations: causal explanations of bell-inequality violations require fine-tuning *New J. Phys.* **17** 033002