Ab initio experimental violation of Bell inequalities

Davide Poderini,[1] Emanuele Polino ⬤,[1] Giovanni Rodari,[1] Alessia Suprano,[1] Rafael Chaves,[2] and Fabio Sciarrino ⬤[1,*]

[1]*Dipartimento di Fisica, Sapienza Università di Roma, Piazzale Aldo Moro 5, I-00185 Roma, Italy*
[2]*International Institute of Physics & School of Science and Technology,*
Federal University of Rio Grande do Norte, 59078-970, P.O. Box 1613, Natal, Brazil

The violation of a Bell inequality is the paradigmatic example of device-independent quantum information: The nonclassicality of the data is certified without the knowledge of the functioning of devices. In practice, however, all Bell experiments rely on the precise understanding of the underlying physical mechanisms. Given that, it is natural to ask: Can one witness nonclassical behavior in a truly black-box scenario? Here, we propose and implement, computationally and experimentally, a solution to this *ab initio* task. It exploits a robust automated optimization approach based on the stochastic Nelder-Mead algorithm. Treating preparation and measurement devices as black boxes, and relying on the observed statistics only, our adaptive protocol approaches the optimal Bell inequality violation after a limited number of iterations for a variety photonic states, measurement responses, and Bell scenarios. In particular, we exploit it for randomness certification from unknown states and measurements. Our results demonstrate the power of automated algorithms, opening a venue for the experimental implementation of device-independent quantum technologies.

I. INTRODUCTION

The experimental guidance is quintessential in science. Empirical evidence not only is crucial to the development of models but also conversely allows one to test and improve theories. To interpret the data, however, one needs well-understood and well-calibrated instruments. For that, theoretical assumptions are always needed, creating a seemingly unstoppable circular argument. It is thus surprising that in the device-independent framework [1], quantum information tasks can be achieved simply from the data and without the need of precise knowledge of the devices.

The paradigmatic example of device independence is the violation of a Bell inequality [2]. It not only provides the most radical departure of quantum theory from classical concepts but also paves the way for applications ranging from cryptography [3], randomness certification [4], self-testing [5], and communication complexity [6,7]. In principle, all of that is achieved simply by imposing the causal structure of the experiment but without any knowledge of either the quantum states being prepared or the measurements performed on them. In practice, however, all Bell experiments performed to date exploit the precise knowledge of the physical platform under use to maximize the Bell inequality violation (see, for instance, Refs. [8–11]). Otherwise, how could one optimize the experimental setting in order to extract its most nonclassi-

cal features? If our device is really treated as a black box, how can we ever prove its quantum nature?

Apart from its relevance in Bell scenarios and related topics, solving such a question might find application in a range of fields such as quantum gravity [12,13] and quantum biology [14], scenarios in which the detection of quantum effects can be a crucial task. However, in all these cases, the quantum mechanisms at play, if any, might not be well understood. That is, simply by probing the system in different but not fully understood ways and looking at its response we should be able to witness its nonclassical nature.

Here, we propose an adaptive automated optimization protocol exactly to solve this *ab initio* task, in particular, focusing on optimizing the violation of Bell inequalities without any prior knowledge of the quantum system and measurements—that is, in a fully black-box scenario. We exploit a stochastic Nelder-Mead algorithm [15] that is an efficient, noise-resistant, and gradient-free algorithm for the search of global optima of functions. By repeatedly tuning the measurement parameters and collecting statistics, the adaptive algorithm is able to optimize the Bell inequality violation after a limited number of measurements. Remarkably, the protocol is also robust against Poissonian fluctuations. We provide the proof of the applicability of our method by performing simulations and photonic experiments involving different bipartite states and measurement responses as well as a range of different Bell inequalities. After a few hundreds of measurements the algorithm is able to approach the maximum possible violation for a given entangled state. In this way, we take a step forward for the crucial task of testing automated protocols in experimental noisy scenarios.

It is worth pointing out that machine learning techniques are spreading out as powerful tools for quantum information tasks, in particular, in the study of Bell nonlocality [16–20].

---

*fabio.sciarrino@uniroma1.it

For instance, reinforcement learning has been used for finding the optimal quantum violations of Bell inequalities for many-body systems with fixed known measurement settings [21] and for the design of optical experiments optimizing Bell inequality violations [22]. In such cases, however, a precise quantum description was required. The implemented method is a way to detect and optimize Bell nonlocality in a truly black-box situation. To showcase its applicability, we exploit it for maximizing the certified randomness extraction from an unknown system and unknown measurements, thus opening a fruitful venue for the device-independent quantum information framework.

II. VIOLATION OF BELL INEQUALITIES AS A BLACK-BOX OPTIMIZATION PROBLEM

Virtually any experiment in quantum physics can be understood as an instance of a prepare-and-measure scenario. Physical systems described by a quantum state are prepared, and measurements are used to reveal its statistical properties. Depending on the application, different levels of control over and characterization of the preparation and measurement devices can be allowed. In quantum tomography [23], for instance, the unknown quantum state being prepared can be reconstructed if we trust and know our measurement apparatus. In other cases [24], the preparation can be assumed to be known, while the measurements cannot. In the context of quantum information, the more we assume, the more open is the way to malicious attacks [25] or wrong conclusions [26].

In this sense, the violation of a Bell inequality provides the ultimate security level, since by assuming only the causal structure imposed on the experiment, but no knowledge of the preparation and measurement devices, one can infer a number of properties of the physical system under test [27]. More precisely, in a Bell test a number of distant parties receive shares of a quantum system prepared by an uncharacterized device. Upon receiving their share, they can locally manipulate and measure them using again unknown devices. That is, each of the parties has a black box with knobs that can be controlled, but the effect of these knobs within each device is unknown.

Consider the bipartite case involving the observers Alice and Bob. At each run of their experiment, input variables x and y decide how the knobs will be turned, leading to the corresponding measurement outcomes a and b, respectively (see Fig. 1). The experimental data are thus encoded into the conditional probability distribution $p(a, b|x, y)$.

In a quantum description of such an experiment, their observations are described as $p(a, b|x, y) = \text{Tr}[(M_a^x \otimes M_b^y)\rho]$, where ρ describes the quantum state being prepared and M_a^x and M_b^y describe the measurement operators. Classically, however, the assumption of local realism [27] implies that the experimental data can be decomposed as

$$p(a, b|x, y) = \sum_\lambda p(\lambda) \, p(a|x, \lambda) \, p(b|y, \lambda). \quad (1)$$

Bell's theorem [2] implies that some quantum predictions are incompatible with the classical prescription (1). This is the phenomenon known as quantum nonlocality that can be witnessed by the violation of a Bell inequality, generally written
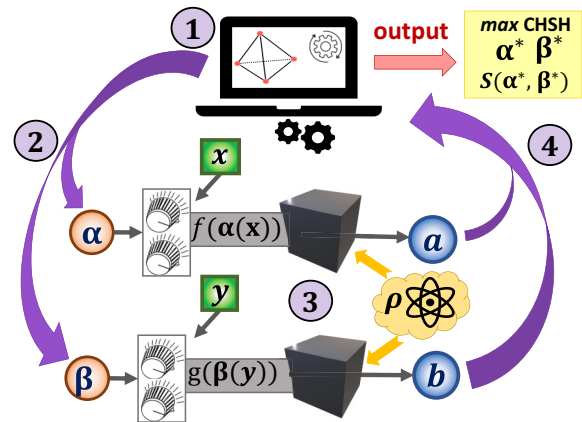

FIG. 1. Bell inequality optimization through the stochastic Nelder-Mead algorithm. An initial simplex in the parameter space is chosen. At each iteration the algorithm performs tests of the Bell inequality $S \leqslant L$ using measurement parameters chosen according to previous results. The numbers indicate the steps performed at each iteration: (1) choice of the parameters based on the history of the process, (2) setting of the chosen parameters, (3) evaluation of the cost function, and (4) transmission of the result to the computer running the stochastic Nelder-Mead algorithm. At the end of the protocol we reach values close to the optimal measurement parameters $\boldsymbol{\alpha}^*$ and $\boldsymbol{\beta}^*$ maximizing the Bell inequality. Measurements and states are treated as black boxes; thus one has to choose functions $f(\cdot)$ and $g(\cdot)$ mapping the parameters to the empirically observed value of S.

as a linear constraint over the observed probabilities given by

$$S = \sum_{a, b, x, y} \alpha_{a, b, x, y} \, p(a, b|x, y) \leqslant L, \quad (2)$$

where $\alpha_{a, b, x, y}$ are integer coefficients and L is the bound arising from the classical description (1).

Checking whether $p(a, b|x, y)$ violates a Bell inequality, we can conclude, for example, that the shared state has to be entangled and that measurements being performed are incompatible [5]. That is, we can probe some properties of the system under test in a device-independent manner. In practice, however, the quantum state ρ and the measurement operators M_a^x and M_b^y have to be tuned precisely in order to obtain and maximize a Bell inequality violation. In a number of situations, such knowledge contradicts the basic black-box assumption of the scenario. In what follows we will propose a solution to deal with that.

The inputs x and y tell the parties how to turn their knobs in their measurement devices. For instance, in a photonic implementation using the polarization degree of freedom of photons, the measurements are performed by changing the orientation of half- and quarter-wave plates. Similarly, the degree of entanglement of the source that we exploit can be tuned by changing the pump polarization. In a black-box approach, each of these changes in the preparation and measurements are simply described by a knob that can be turned. However, what this turning of knobs is doing inside the devices we cannot speak of. The knobs for Alice and Bob are described by a set of (continuous or discrete) variables $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, each a function of the inputs x and y, respectively. In turn, the source

is described by $\boldsymbol{\gamma}$. That is, $M_a^x = M_a(\boldsymbol{\alpha}(\boldsymbol{x}))$, $M_b^y = M_b(\boldsymbol{\beta}(\boldsymbol{y}))$, and $\rho = \rho(\boldsymbol{\gamma})$.

Given a Bell inequality described by $S = S(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$, our objective will be to maximize S, particularly obtaining a value that surpasses the local bound L. Importantly, we want to achieve that without knowing how function S depends on the knob parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$. That is, the optimization is performed in a truly black-box scenario. As we will show, changing iteratively the value of the parameters and observing, in real time, the change in S, we will be able to reach its optimal value after a remarkably low number of iterations.

We employ a gradient-free and direct search algorithm, the Nelder-Mead simplex method [28] and its stochastic variant [15], able to efficiently optimize a multidimensional function by simply comparing some of its values. This is achieved even in the presence of noise, unavoidable in an experimental implementation. In the following we will fix the parameter γ describing the source, but the method can also be used to optimize over it.

The algorithm adaptively evolves a simplex [29] living in the space of the parameters $\boldsymbol{t} = (\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \mathcal{T}$. The Bell function $S(\boldsymbol{t})$ can be calculated for each of the n simplex points: $\{\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n\} \subseteq \mathcal{T} \longrightarrow \{S(\boldsymbol{t}_1), \ldots, S(\boldsymbol{t}_n)\}$. As input of the optimization protocol an initial simplex $\Omega_0 = \{\boldsymbol{t}_1^0, \ldots, \boldsymbol{t}_n^0\}$ is generated through *Latin hypercube sampling* [15]. Starting from the initial simplex, the algorithm repeats an optimization cycle that updates the simplex (Fig. 1). Each cycle of the adaptive algorithm is composed of three main steps: (1) The points of the simplex Ω_{k-1} are sorted, based on the values of the associated cost function such that the worse point is deleted; (2) from the barycenter of the simplex, a new point $\boldsymbol{t}^{\text{ref}}$ is generated through a reflection rule; and (3) according to the cost of $\boldsymbol{t}^{\text{ref}}$, the simplex is updated by including $\boldsymbol{t}^{\text{ref}}$ itself or geometrically generating a further point. If such point is not promising enough, an *adaptive random search* is employed to generate a point $\boldsymbol{t}^{\text{ARS}}$ [30]. A more detailed account of the above steps is provided in Appendix A, while a comparison between the the devised algorithm and standard nonadaptive approaches is provided in Appendix B.

III. SIMULATION AND NUMERICAL TESTS ON THE CHSH INEQUALITY

To illustrate the main features of our approach, in the following we will focus on the Clauser, Horne, Shimony, and Holt (CHSH) scenario [31] consisting of dichotomic inputs and outputs. The only class of Bell inequalities in this scenario is given by

$$S = \langle A_0 B_0 \rangle + \langle A_0 B_1 \rangle + \langle A_1 B_0 \rangle - \langle A_1 B_1 \rangle \leqslant 2, \quad (3)$$

where $\langle A_x B_y \rangle \equiv \sum_{a,b} a\, b\, p(a, b | x, y)$ is the expectation value (given the inputs x and y) of Alice and Bob outcomes $(a, b) \in \{+1, -1\}$. Moreover, the employed approach can be applied for general optimization, and we consider a range of different scenarios such as the chained [32], tilted [33], and nonlinear Tsirelson-Landau-Masanes inequalities [34].

We performed simulated optimizations and experiments covering a range of pure and noisy quantum states. Here, we

focus on quantum states given by

$$|\psi^\gamma\rangle = \cos\gamma |HV\rangle + \sin\gamma |VH\rangle. \quad (4)$$

In turn, the most general projective measurement on a qubit state is given by $\hat{P}(\theta, \phi) = \vec{r}(\theta, \phi) \cdot \vec{\sigma}$, where $\vec{r}(\theta, \phi)$ represents a unit vector in polar coordinates and $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ are Pauli operators. Considering that the input of the CHSH test defines the angles θ, ϕ of such measurements (noticing that this knowledge is never used by the algorithm), we have at least a total number of eight measurement parameters (two for each measurement of Alice and Bob). This defines the parameter space where the algorithm searches for the CHSH optimization.

It is worth pointing out that, since we are in a fully black-box scenario, the optimization process does not require the exact form of the quantum state or measurements, which are only chosen for simulation purposes. As a matter of fact, in the optimization, one has to choose how the algorithm will map the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ to the observable measured at a particular iteration. As discussed in Appendix C, we have tested different nonlinear response functions, such as the hyperbolic
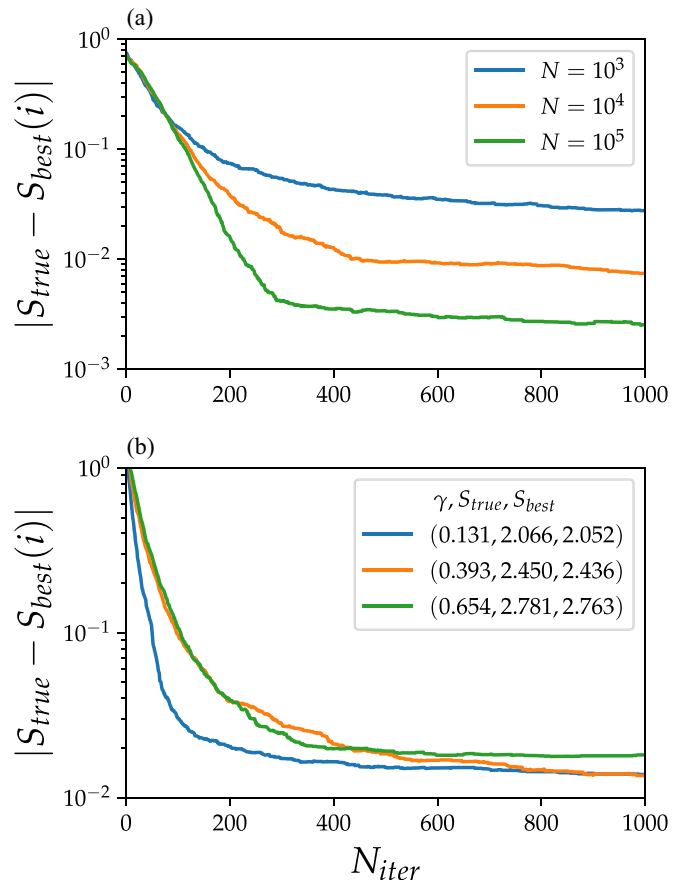

FIG. 2. Simulation results. The difference $|S_{\text{true}} - S_{\text{best}}|$ is plotted as a function of the algorithm's iteration. For all plots the curves are obtained by averaging over 100 simulation runs. (a) Singlet state with different numbers of events N. (b) Pure unbalanced states (parametrized by γ in radians). The label S_{true} specifies the optimal violation of the CHSH inequality for the corresponding states. The mean number of Poissonian events relative to each measurement is 10^5.
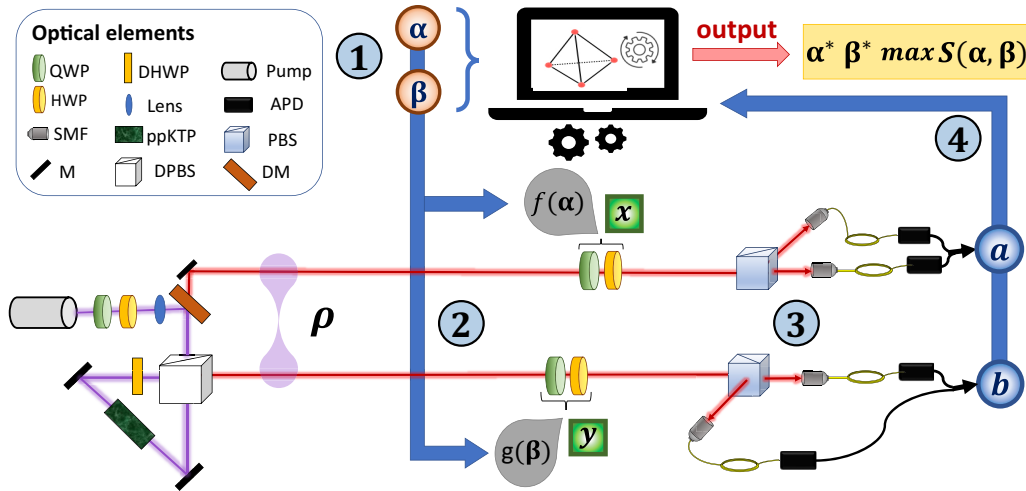
FIG. 3. Experimental setup. A Sagnac-based source generates photonic pair states of the form $\cos \gamma |HV\rangle + e^{i\phi} \sin \gamma |VH\rangle$ unknown to the optimization algorithm. The polarization of the two photons is measured by two wave plates whose angles are determined by unknown functions $f(\cdot)$ and $g(\cdot)$ of known parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. New parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are chosen based on the results a and b of the previous iteration. The numbers indicate the steps performed at each iteration. At the end of the process, optimal values α^* and β^* are chosen in order to optimize the Bell inequality violation. QWP, quarter-wave plate; HWP, half-wave plate; DHWP, dual-wavelength half-wave plate; PBS, polarizing beam splitter; DPBS, dual-wavelength polarizing beam splitter; SMF, single-mode fiber; APD, avalanche photodiode single-photon detector; ppKTP, periodically poled potassium titanyl phosphate; M, mirror; DM, dichroic mirror.

sine or a logistic function. Finally, to mimic the experimental situation, Poissonian fluctuations are added in the measurements of operators. Such fluctuations are characterized by the parameter N, corresponding to the total number of Poissonian events used to calculate the CHSH parameter at each iteration.

The figure of merit of the optimization process is given by $\Delta = |S_{\text{true}} - S_{\text{best}}(i)|$, where $S_{\text{best}}(i)$ corresponds to the best value of the CHSH parameter inside the simplex at the ith iteration step, while S_{true} is the maximum CHSH value achievable by the considered state and that can be computed resorting to the Horodecki criterion [35].

In Fig. 2(a) one can see that increasing N (thus reducing the Poissonian fluctuations) leads to faster convergence to values comparable to the expected error (i.e., $\Delta \sim 1/\sqrt{N}$). As shown in Fig. 2(b) (considering $N = 10^4$), for quantum states with different levels of entanglement, already with 200 iterations the error reaches values close to $\Delta = 1/\sqrt{N} = 10^{-2}$.

IV. EXPERIMENTAL *ab initio* OPTIMIZATION TESTS

To demonstrate the described protocol, we exploited polarization states of pairs of photons. Entangled states of the form (4) were generated through a type-II spontaneous parametric downconversion process inside a periodically poled potassium titanyl phosphate (ppKTP) crystal, pumped by a continuous-wave $\lambda = 404$ nm laser in a Sagnac interferometric geometry [36,37].

The photons exiting the source are measured through two polarization analysis stages, each composed of a quarter-wave plate (QWP) followed by a half-wave plate (HWP) and a polarizing beam splitter (PBS). The measurement parameters that are tuned determine the rotation angles of the wave plates by means of functions $f(\boldsymbol{\alpha})$ for Alice's and $g(\boldsymbol{\beta})$ for Bob's station (Fig. 3).

We performed experimental optimization tests varying the unknown states, number of parameters, and measurement responses. In the experimental case we compare the highest reached violations $S_{\text{best}}(i)$ at the ith iteration with the maximum violation S_{tomo} obtained by applying the Horodecki criterion [35] to the density matrix reconstructed with quantum state tomography [38]. Considering the maximally entangled state, Fig. 4(a) shows that we reach $\Delta < 10^{-2}$ after only 250 iterations, even considering an eight-parameter optimization. Experiments with different initial entangled states [Eq. (4)] show a similar fast convergence [see Fig. 4(b)]. Remarkably, we achieve experimental Bell violations even for states with little entanglement and optimal CHSH violation as low as $S_{\text{tomo}} = 2.161 \pm 0.026$ (see Table I for details).

V. BEYOND THE CHSH INEQUALITY

In the previous sections we have focused on the CHSH inequality. In what follows, we will show how our approach can be applied to a variety of different scenarios.

TABLE I. Experimental results. The average Bell violations (S_{best}) for different states, identified by γ as in (4), are listed alongside the reference violation found by tomographic reconstruction of the state (S_{tomo}). Column σ represents the standard deviation associated with distribution of the single optimization runs.

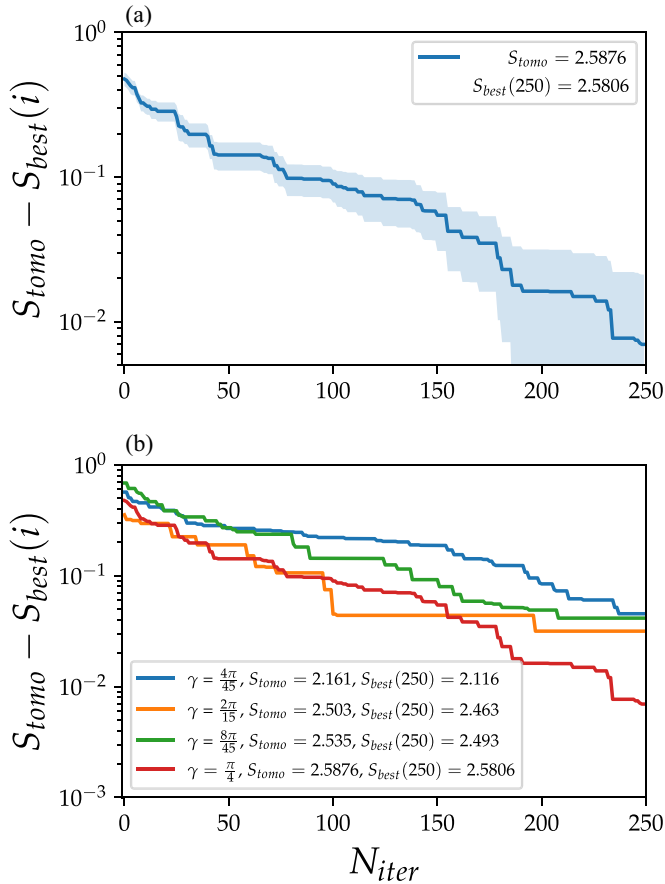| γ	S_{tomo}	S_{best}	σ
$\frac{4\pi}{45}$	2.161 ± 0.026	2.116 ± 0.025	0.056 ± 0.020
$\frac{2\pi}{15}$	2.503 ± 0.008	2.463 ± 0.024	0.054 ± 0.019
$\frac{8\pi}{45}$	2.535 ± 0.012	2.493 ± 0.031	0.069 ± 0.024
$\frac{\pi}{4}$	2.588 ± 0.006	2.581 ± 0.014	0.044 ± 0.010

FIG. 4. Experimental results. Average difference $S_{tomo} - S_{best}$ between the value S_{tomo}, which is the maximum CHSH violation achievable by the state from the quantum state tomography, and S_{best}, which is the best value obtained by the algorithm as function of the iteration of the algorithm. (a) The shaded blue area represents the standard deviation of the mean. The mean number of events for each CHSH measurement is $\sim 4 \times 10^4$. $S_{tomo} = 2.588 \pm 0.006$, and $S_{best} = 2.581 \pm 0.014$. (b) Different values of the γ parameter relative to the state are chosen: $4/45\pi$, $2/15\pi$, $8/45\pi$, and $\pi/4$. For all the data the number of parameters is 8 (four for each station), and the mean number of collected photons for each CHSH measurement is $\sim 4 \times 10^4$. Each curve is the average of five repetitions of the optimization (see Table I).

A. Chained Bell inequalities

An important class of bipartite Bell inequalities are the so-called chained Bell chained inequalities [32], given by

$$S^k = \sum_{i=1}^{k} |I_i| \leqslant k - 1, \quad \text{where } I_i = \frac{1}{2} \sum_{x=i-1}^{i} \langle A^x B^{i-1} \rangle, \quad (5)$$

with $A^k = -A^0$. Differently from the CHSH case, Alice and Bob can perform an arbitrary number of k measurements each. This class of inequalities has found many applications, ranging from randomness amplification [39] to self-testing of quantum states [40].

Our simulation and experimental results for the chain inequality are shown in Fig. 5. Numerical simulations [Fig. 5(a)] show that even for increasing k (up to $k = 12$) the black-box optimization algorithm is still able to reach values very

close to the maximum possible Bell inequality violation. Figure 5(b) shows that the number of iterations required to obtain a value lower than 2% for the distance between the maximum reachable violation and the violation achieved by the optimization scales approximately linearly with k. Thus the approach remains very efficient even by increasing the number of measurements and complexity of the Bell scenario. Finally, Fig. 5(c) shows our experimental results for $k = 3$ (involving 12 measurement parameters in total), also comparing them with a numerical simulation.

B. Tilted Bell inequality

Even though in the CHSH scenario the CHSH inequality is the only tight Bell inequality (a facet of the local set), it is known that nontight Bell inequalities might play an important role for quantum information processing. For instance, the maximal violation of the CHSH inequality allows the generation of 1.23 bits [41] of certified randomness. For optimal randomness certification in a Bell scenario, even using non-maximally entangled states of the form $\cos \gamma |00\rangle + \sin \gamma |11\rangle$, tilted Bell inequalities have been introduced in Ref. [33]. The general form of the inequality for local classical models is given by

$$I_\beta^\alpha = \beta \langle A_1 \rangle + \alpha \langle A_1 B_1 \rangle + \alpha \langle A_1 B_2 \rangle + \langle A_2 B_1 \rangle - \langle A_2 B_2 \rangle$$

$$\leqslant 2\alpha + \beta, \quad (6)$$

where $\alpha \geqslant 1$ and $\beta \geqslant 0$.

To gauge the employed approach in the violation of this inequality, we consider the case where $\alpha = 1$ and $\beta = 2 - 2\sin^2 \gamma$ are optimized to maximize the violation of Eq. (6), whose maximum quantum value is $I_\beta^\alpha = 2\sqrt{(1 + \alpha^2)(1 + \beta^2/4)}$. As displayed in Fig. 6, computational simulations considering different quantum states (parametrized by γ) show that the black-box optimization is able to approach the optimal value even with only 10^4 Poissonian events per measurement. This shows that the black-box optimization might find applications in device-independent randomness certification as well. It is worth pointing out that in this simulation we are already choosing the optimal inequality for the given state. In the future, one might consider the case where the underlying state is unknown and thus not only the measurements have to be optimized in a black-box scenario but also the objective function (the Bell inequality) can be optimized over.

C. Black-box optimization of quantum inequalities

So far, we have considered only the case of classical Bell inequalities. In some cases, however, one might also be interested in exploring the boundary of the set of quantum correlations [11,42,43].

To illustrate the applicability of the approach also in this case, we consider the black-box optimization of the so-called Tsirelson-Landau-Masanes (TLM) inequality [34] to bound the set of quantum correlations in the CHSH scenario. We consider a specific symmetry of these quantum inequalities given by

$$-\pi \leqslant -\sin^{-1} x_1 + \sin^{-1} x_2 + \sin^{-1} x_3 + \sin^{-1} x_4 \leqslant \pi, \quad (7)$$

where $\boldsymbol{x} = (\langle A_0 B_0 \rangle, \langle A_0 B_1 \rangle, \langle A_1 B_0 \rangle, \langle A_1 B_1 \rangle)$.
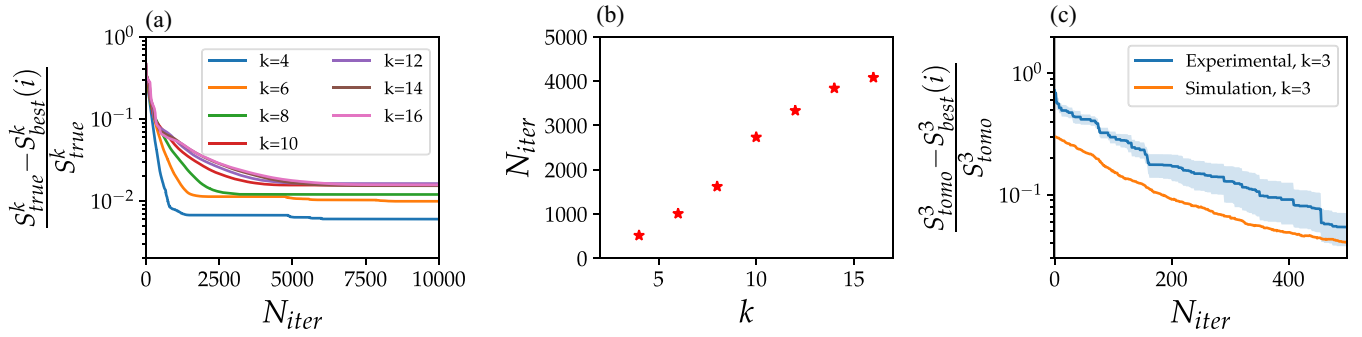
FIG. 5. Chained Bell inequalities optimization (simulation and experiment). (a) Mean results over 50 optimization runs of the chained Bell inequalities, computed considering a singlet state, increasing k measurement per party, and Poissonian measurement uncertainty with $N = 10^6$. Since the achievable maximal violation S_{true}^k varies with k, we report the rate $\frac{S_{\text{true}}^k - S_{\text{best}}^k}{S_{\text{true}}^k}$ as a function of the algorithm's iteration. The number of optimization parameters increases linearly and is given by $4k$. In all cases the algorithm is able to approach the expected minimum within $\approx 1\%$ after ≈ 5000 noisy evaluations. (b) Median (over 50 repetitions) number of measurements needed to reach a distance lower than 2% between the *ab initio* evaluation of the chained parameter S^k and its maximum reachable value, as a function of the number of settings k. As can be seen, the number of iterations grows linearly with the number of optimization parameters showing the efficiency of the employed method. (c) Experimental optimization of the chain inequality with $k = 3$; hence 12 control parameters are necessary. We report the rate $\frac{S_{\text{tomo}}^3 - S_{\text{best}}^3}{S_{\text{tomo}}^3}$, averaged over five experimental runs, as a function of the algorithm iteration. The number of events for each measurement is $\sim 3 \times 10^4$. Here, S_{tomo}^3 is again the maximum violation of S^3 achievable by the state from the quantum state tomography, computed through numerical optimization. The blue line indicates the average simulation (repeated 100 times) for a Werner state with $p = 0.90$, employing $N = 3 \times 10^4$ Poissonian events with added Gaussian noise ($\sigma = 0.05$) to simulate the real experimental conditions, where $S_{\text{tomo}}^3 = 2.399 \pm 0.003$ and for the average over the five runs we have $S_{\text{best}}^3 = 2.28 \pm 0.02$ with a variance of 0.002.

We simulated the optimization of this quantum inequality considering a singlet state and different numbers of Poissonian events [Fig. 7(a)]. The goal here is to reach the optimal value of the function of correlations in Eq. (7), possibly getting as close as possible to the quantum bounds, which in the case of maximally entangled states is equal to π. Moreover, we have also performed an experimental test to demonstrate the performance of the method also in a real noisy setup [Fig. 7(b)]. The black-box optimization of the quantum inequality follows a similar trend, approaching its optimum value after a reasonably small number of iterations. Thus our approach can also find relevant applications in testing the limits of quantum predictions for possible deviations from quantum theory.
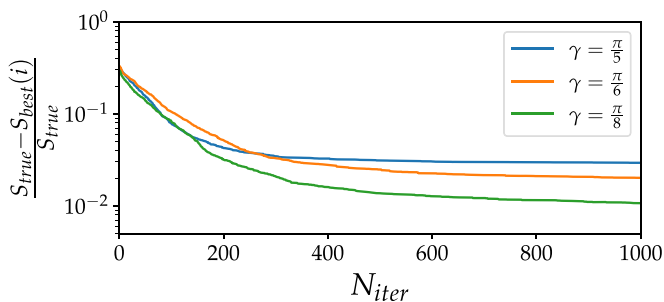

FIG. 6. Tilted Bell inequality black-box optimization. Computational simulation considering a mean (over 100 repetitions) of the tilted Bell inequality in Eq. (6) with parameters $\alpha = 1$ and $\beta = 2 - 2\sin^2\gamma$ optimized on the source state, simulating a Poissonian statistic with $N = 10^4$ events for each measurement.

VI. *Ab Initio* RANDOMNESS CERTIFICATION

Now, we show how the presented approach can be exploited to certify and maximize randomness, the paradigmatic application of Bell nonlocality [4,44–47]. While randomness can be certified via the violation of Bell's inequalities [4,48–55] and the algorithm is able to find them, our aim here is to maximize the certified randomness directly from the data, instead of an *a priori* known Bell inequality [56]. Following Ref. [57], we have approached this problem by constraining the guessing probability of the adversary (Eve) directly on the observed behavior. Denoting as e the outcome associated with Eve, randomness can be quantified via the guessing probability [58] given by

$$p_{\text{guess}}(x, y) = \max_{\{p \in \mathcal{Q}\}} \sum_{a,b} p(a, b, e = (a, b)|x, y), \quad (8)$$

where $p(a, b, e|x, y) \in \mathcal{Q}$ is the unknown global behavior including Eve, constrained to belong to the set of quantum correlations $\mathcal{Q}$. The distribution p_{guess} represents the amount of knowledge that she can extract on the outcomes of Alice and Bob, and it must be compatible with the observed distribution $p^*(a, b|x, y) = \text{Tr}(A_{a|x} \otimes B_{b|y} \, \rho)$. Combining the stochastic Nelder-Mead algorithm with the Navascués-Pironio-Acín (NPA) hierarchy [59], we are able to optimize the upper bound on p_{guess} over unknown measurements and states constrained on being compatible with p^*, i.e., we are solving the following maximization problem:

$$\max \; p_{\text{guess}} \; \text{such that} \; p(a, b, e|x, y) \in \mathcal{Q}_k$$
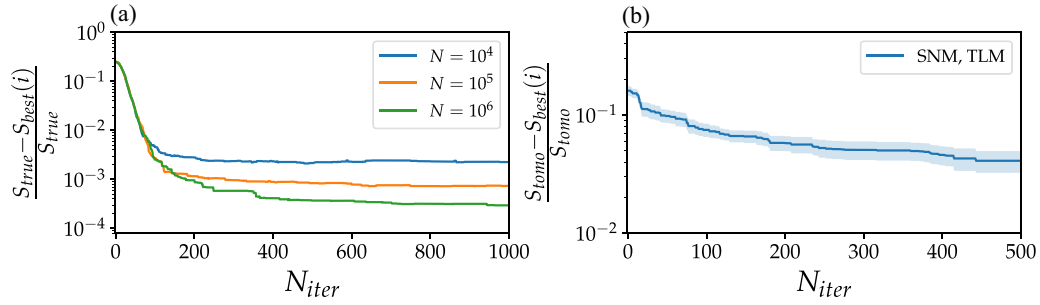$$\times \sum_e p(a, b, e|x, y) = p^*(a, b|x, y), \quad (9)$$

FIG. 7. Optimization of quantum Bell inequalities (simulation and experiment). (a) The rate $\frac{S_{\text{true}}-S_{\text{best}}}{S_{\text{true}}}$ is plotted as a function of the algorithm's iteration. We report the mean results obtained by averaging 50 simulation runs of the stochastic Nelder-Mead (SNM) optimization of the TLM inequality [34] considering a pure singlet state. The different curves represent the distinct trends for three separate values of the Poissonian measurement uncertainty N. (b) Experimental results of the optimization of the TLM inequality proposed in Ref. [34]. The difference $\frac{S_{\text{tomo}}-S_{\text{best}}}{S_{\text{tomo}}}$ has been averaged over eight experimental repetitions with eight measurement parameters, and it is reported as a function of the algorithms' iteration. The average of the optimal value found is $S_{\text{best}} = 2.862 \pm 0.007$, with a variance of 0.00039, which is comparable to the one found by performing the full quantum state tomography $S_{\text{tomo}} = 2.98 \pm 0.01$. The average number of events for each measurement is $\approx 4.4 \times 10^4$.

where $\mathcal{Q}_k$ represent the relaxation of quantum set at the k order of the NPA hierarchy [59]. In Fig. 8 we show, fixing a state of the form (4) with a specific value of the parameter γ, how the algorithm is able, in less than 500 iterations, to approach the optimal value of the p_{guess} for pure states [56], optimizing on all the possible measurements allowed by our setup. Considering experimental density matrices, we are able to approach the higher value of $p_{\text{guess}} \approx 0.7$ as is expected due to the noise of experimentally reconstructed states.

The simulation results are reported in Fig. 9 and clearly show that our approach is able to directly maximize the randomness extraction in a fully black-box approach.
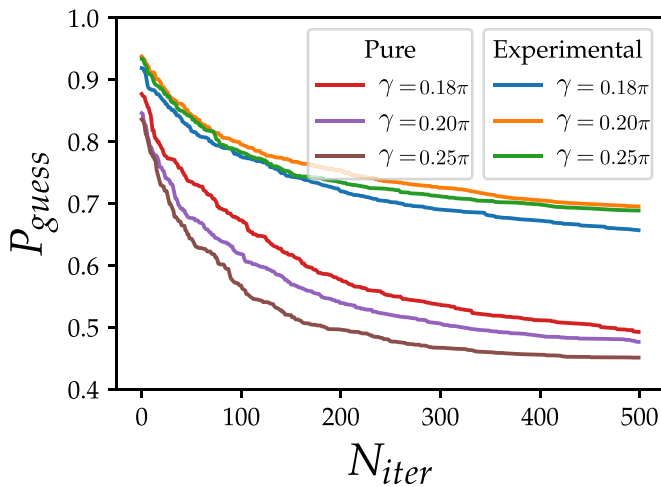

FIG. 8. Computational simulation of extracted randomness. Optimal p_{guess} of an adversary as a function of iterations for selected values of the parameter γ for states of the form $|\psi^\gamma\rangle = \cos\gamma |HV\rangle + \sin\gamma |VH\rangle$. The experimental curves are obtained by considering density matrices of states reconstructed using experimental data. As in the main text, the behaviors used to evaluate Eve's p_{guess} are obtained directly through Born's rule $p(a,b|x,y) = \text{Tr}(A_{a|x} \otimes B_{b|y}\rho)$, where ρ is the reconstructed density matrix. This is the trend of p_{guess} in Eq. (9) as a function of the iterations.

VII. DISCUSSION

The violation of a Bell inequality is often pictured as the paradigmatic example of a device-independent task: From the observed statistic alone, without knowing the internal working of devices, one can conclude the nonclassical nature of it. To obtain such violation, however, one often has to rely on a precise description of the devices. Here, we propose a practical solution to this problem, exploiting an adaptive automated algorithm able to maximize the Bell inequality violation in a fully black-box setting. Employing both simulation and actual experiments, we demonstrated the protocol by optimizing the violation of many Bell inequalities for different unknown photonic bipartite states and measurement responses . Nicely, the optimum values are achieved after a few hundred iterations.

Our approach can also be applied to quantum networks of growing size and complexity that have started to be experimentally and theoretically explored in the recent years [60–72]. Finding Bell inequalities in these cases can be very difficult, but heuristic machine-learning-based approaches
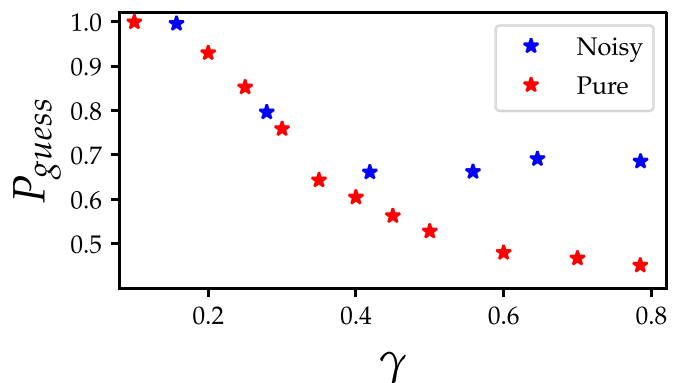

FIG. 9. Simulated randomness certification. Optimal p_{guess}, obtained as an average over 50 runs of the optimal value reached by the employed algorithm after 500 iterations, as a function of the parameter γ for pure states in the form (4) (red stars) and for experimentally reconstructed noisy density matrices (blue stars). The noisy points can certify less randomness than the ideal ones.

have been proposed to find and quantify nonclassicality [16,17,73] opening the possibility of applying our framework also in such scenarios. Moreover, from a more near-term perspective, practical experimental quantum information tasks, where multiple parameters are tuned to optimize a desired cost function, such as quantum communication and state engineering in noisy environments, can also benefit from our approach. To demonstrate that, we have applied our framework for a paradigmatic application of Bell's theorem, showing that even with no knowledge of states or devices, one can directly (i.e., without the need of Bell inequalities) maximize the amount of certified randomness.

APPENDIX A: DETAILS OF THE STOCHASTIC NELDER-MEAD ALGORITHM

The stochastic Nelder-Mead algorithm is composed of three main steps, continuously repeated during the optimization process:

(1) The Bell inequality parameter S for all the points t of the simplex Ω_{k-1} is evaluated, and the point with the maximum value is removed from the simplex, if $\dim(\Omega_{k-1}) > n$. Then the values $S(t)$ are sorted, and three elements are individuated: $t^{\max}$, $t^{2\mathrm{nd\,max}}$, and $t^{\min}$, which are the points for which the CHSH parameter $S(\cdot)$ assumes the maximum, the second maximum, and the minimum values, respectively.

(2) The barycenter t^{bar} of the points in $\Omega_{k-1} \setminus \{t^{\max}\}$ is calculated, and a new point t^{ref} is generated by the following reflection rule: $t^{\mathrm{ref}} = (1 + \delta)t^{\mathrm{bar}} - \delta t^{\max}$, where $\delta > 0$ is the reflection coefficient.

(3) Either step (a), (b), or (c) below is followed.

(a) If $S(t^{\min}) \leqslant S(t^{\mathrm{ref}}) < S(t^{2\mathrm{nd\,max}})$, impose $\Omega_k = \Omega_{k-1} \cup \{t^{\mathrm{ref}}\}$.

(b) If $S(t^{\mathrm{ref}}) < S(t^{\min})$, generate the expansion $t^{\mathrm{exp}} = \gamma t^{\mathrm{ref}} + (1 - \gamma)t^{\mathrm{bar}}$, where $\gamma > 1$ is the expansion coefficient; and if $S(t^{\mathrm{exp}}) < S(t^{\mathrm{ref}})$, impose $\Omega_k = \Omega_{k-1} \cup \{t^{\mathrm{exp}}\}$, otherwise $\Omega_k = \Omega_{k-1} \cup \{t^{\mathrm{ref}}\}$.

(c) If $S(t^{\mathrm{ref}}) \geqslant S(t^{2\mathrm{nd\,max}})$, then either step (i) or step (ii) below is followed.

(i) If $S(t^{2\mathrm{nd\,max}}) \leqslant S(t^{\mathrm{ref}}) < S(t^{\max})$, there will be an *external* contraction given by

$$t^{\mathrm{cont}} = \beta t^{\mathrm{ref}} + (1 - \beta)t^{\mathrm{cent}} \quad 0 \leqslant \beta \leqslant 1.$$

If $S(t^{\mathrm{cont}}) \leqslant S(t^{\mathrm{ref}})$, the contraction is accepted.

(ii) If $S(t^{\mathrm{ref}}) \geqslant S(t^{\max})$, there will be an *internal* contraction given by

$$t^{\mathrm{cont}} = \beta t^{\max} + (1 - \beta)t^{\mathrm{cent}} \quad 0 \leqslant \beta \leqslant 1.$$

If $S(t^{\mathrm{cont}}) \leqslant S(t^{\max})$, the contraction is accepted.

If the contraction is accepted, then $\Omega_{k+1} = \Omega_k \cup \{t^{\mathrm{cont}}\}$, otherwise an *adaptive random search* (ARS) is exploited as described in the main text.

The standard control parameters used in the employed algorithm are $\{\delta, \eta, \gamma\} = 1, 0.5, 2$ [28] and ϵ equal to $\approx 10\%$ of the value of the parameter variation used to perform the ARS global search process.

With respect to Ref. [15], we avoided the use of an adaptive sample size scheme, which would require us to resample every point of the simplex at each iteration of the algorithm. This is motivated by the time-constrained nature of our experimental test bed, since every cost function estimation requires approximately 20 s to be performed, considering both the data acquisition and the response time of the optical apparatus. The computational time is negligible, being ~ 1 ms for each run on a commercial PC.

APPENDIX B: COMPARISON BETWEEN STOCHASTIC NELDER-MEAD ALGORITHM AND NONADAPTIVE ALGORITHMS

In this Appendix we compare our approach with two basic nonadaptive procedures for gradient-free optimization in high-dimensional parameter space.

The first is a brute-force approach where the function is optimized on an equally spaced grid in the parameter space. The exponential increase in the number of points in the grid given by n^d, where d is the number of optimization parameters (for instance, $d = 8$ for the CHSH case) and n is the
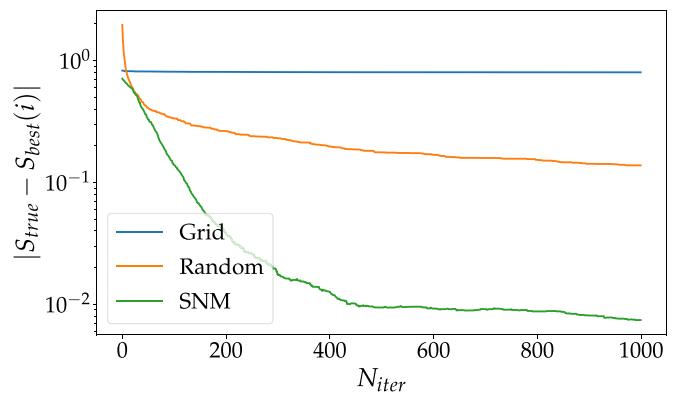

FIG. 10. Comparison between different optimization procedures (simulation). The plot considers the eight-dimensional parameter space of the CHSH scenario. The grid optimization (blue curve) is a brute-force search on an equally spaced grid in this space. The plot shows the average of 100 runs where the starting point of the grid was chosen at random. The random search method (orange curve) instead performs sampling on random points trying to find the optimum. Also in this case the average of 100 runs is shown. As can be seen, the stochastic Nelder-Mead algorithm (green curve) performs much better. All curves were generated by simulating noisy samples following Poissonian statistics corresponding to $N = 10^4$ events.
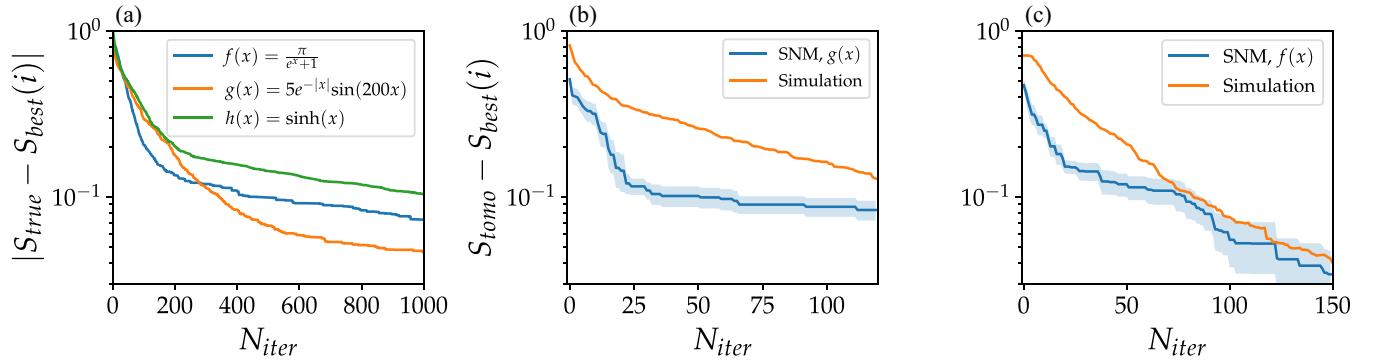
FIG. 11. Optimization with different measurement responses (simulation and experiment). (a) The difference $|S_{\text{true}} - S_{\text{best}}|$, averaged over 100 runs, is plotted as a function of the algorithm's iteration. Here, the optimization of the CHSH inequality is performed over pure singlet states using nonlinear maps to the eight total projector parameters, accounting for Poissonian fluctuations with $N = 10^4$. (b) and (c) Mean difference $S_{\text{tomo}} - S_{\text{best}}$ as a function of the algorithm's iteration considering the $\text{Osc}(\alpha)$ and $\text{Logi}(\alpha)$ nonlinear maps as the measurement response in both Alice's and Bob's stations. In both cases the QWP position has been fixed on its optical axis; the algorithm is controlling only the HWP. For both (b) and (c) the mean rate of events per measurement is $\approx 5 \times 10^4$, and nine experimental runs have been collected. The orange lines indicate the average simulations (repeated 100 times) for a Werner state with $p = 0.90$, employing $N = 5 \times 10^4$ Poissonian events with added Gaussian noise ($\sigma = 0.05$) to simulate the real experimental conditions. In (b) we have obtained $S_{\text{tomo}} = 2.574 \pm 0.015$, and our average over the optimization runs is $S_{\text{best}} = 2.491 \pm 0.016$ with a variance of 0.0023, while in (c) we have $S_{\text{tomo}} = 2.469 \pm 0.017$ and $S_{\text{best}} = 2.435 \pm 0.026$ with a variance of 0.006.

number of samples for a single parameter, severely limits the applicability of this method for high parameter space. As a matter of fact, usually this approach is used in combination with other local optimization methods. To illustrate the method, here we employ $n = 3$ which gives 6561 grid points to sample.

Among the simplest and effective nonadaptive approaches to multiparameter black-box optimization is to perform a random search in the parameter space. Here, at each step, sets of parameters are selected uniformly at random in the allowed set and independently in respect to previous steps.

Employing such approaches and comparing them with our optimization procedure shows a definite advantage of the latter. In particular, the convergence to the optimum is notably faster, a feature of fundamental importance for real applica-

FIG. 12. Simulations for noisy states. Simulated scenario with different mixed states, parametrized by p and λ, with a number of events $N = 10^4$. The difference $|S_{\text{true}} - S_{\text{best}}|$ is plotted as a function of the algorithm's iteration. All the curves are obtained by averaging over 100 simulation runs.

tions where the number of samples one can afford is severely limited.

For comparison we performed both the grid and random optimization simulating noisy measurements with Poissonian statistics corresponding to $N = 10^4$ events. The results are shown in Fig. 10, considering the average of 100 optimization procedures, where in the grid case we randomized the starting point of the grid in the parameter space.

APPENDIX C: CHSH INEQUALITY OPTIMIZATION WITH DIFFERENT MEASUREMENT RESPONSES AND NOISY STATES

In our photonic implementation with qubits, each projective measurement is described by two continuous angles (from the half- and quarter-wave plates). In the black-box scenario, however, we cannot assume how the actual measurement operator depends on such inputs. To illustrate the role that the choice of the response functions might have in the optimization, we considered the following three response functions:

$$f(\alpha) = \begin{cases} 5e^{-|\alpha|}\sin(200\alpha) & \equiv \text{Osc}(\alpha), \\ \frac{\pi}{e^{\alpha}+1} & \equiv \text{Logi}(\alpha), \\ \sinh(\alpha). & \end{cases}$$

As can be seen in Fig. 11(a) all such response functions reach a reasonable accuracy in our simulations. As shown in Figs. 11(b) and 11(c), we compare the numerical simulations with an actual experiment, by choosing the oscillating and logistic response functions. The experimental minimization is done by fixing the QWP positions and optimizing on the four measurement parameters associated with the HWP positions.

Furthermore, we simulate the optimization of CHSH inequality for different states of the form $\rho = p|\psi^{\gamma}\rangle\langle\psi^{\gamma}| + (1-p)(\lambda \frac{|\psi^{-}\rangle\langle\psi^{-}| + |\psi^{+}\rangle\langle\psi^{+}|}{2} + (1-\lambda)\frac{\mathbb{I}}{4})$, with different values of noise parameters p and λ. The results are shown in Fig. 12 and demonstrate how the algorithm can optimize all the tested states with similar performances.
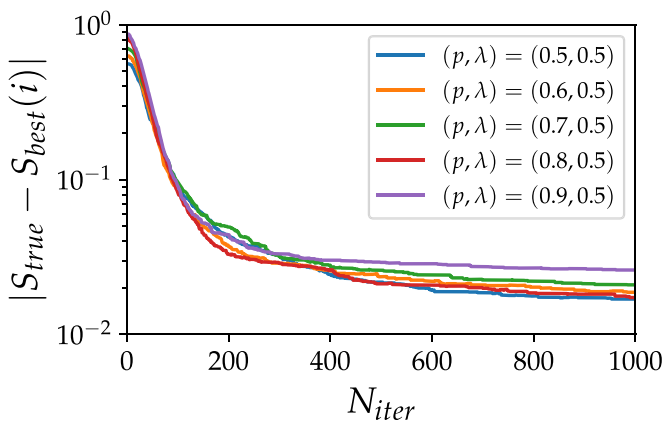
[1] S. Pironio, V. Scarani, and T. Vidick, Focus on device independent quantum information, New J. Phys. **18**, 100202 (2016).

[2] J. S. Bell, On the Einstein Podolsky Rosen paradox, Phys. Phys. Fiz. **1**, 195 (1964).

[3] 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).

[4] A. Acín and L. Masanes, Certified randomness in quantum physics, Nature (London) **540**, 213 (2016).

[5] I. Šupić and J. Bowles, Self-testing of quantum systems: A review, Quantum **4**, 337 (2020).

[6] H. Buhrman, R. Cleve, S. Massar, and R. De Wolf, Nonlocality and communication complexity, Rev. Mod. Phys. **82**, 665 (2010).

[7] J. Ho, G. Moreno, S. Brito, F. Graffitti, C. L. Morrison, R. Nery, A. Pickston, M. Proietti, R. Rabelo, A. Fedrizzi, and R. Chaves, Quantum communication complexity beyond Bell nonlocality, arXiv:2106.06552.

[8] 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 *et al.*, Strong Loophole-Free Test of Local Realism, Phys. Rev. Lett. **115**, 250402 (2015).

[9] M. Giustina, M. A. M. Versteegh, S. Wengerowsky, J. Handsteiner, A. Hochrainer, K. Phelan, F. Steinlechner, J. Kofler, J. A. Larsson, C. Abellan, W. Amaya, V. Pruneri, M. W. Mitchell, J. Beyer, T. Gerrits, A. E. Lita, L. K. Shalm, S. W. Nam, T. Scheidl, R. Ursin *et al.*, Significant-Loophole-Free Test of Bell's Theorem with Entangled Photons, Phys. Rev. Lett. **115**, 250401 (2015).

[10] B. Hensen, H. Bernien, A. E. Dréau, A. Reiserer, N. Kalb, M. S. Blok, J. Ruitenberg, R. F. Vermeulen, R. N. Schouten, C. Abellán, 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).

[11] B. G. Christensen, Y.-C. Liang, N. Brunner, N. Gisin, and P. G. Kwiat, Exploring the Limits of Quantum Nonlocality with Entangled Photons, Phys. Rev. X **5**, 041052 (2015).

[12] C. Marletto and V. Vedral, Gravitationally Induced Entanglement between Two Massive Particles is SufficientEvidence of Quantum Effects in Gravity, Phys. Rev. Lett. **119**, 240402 (2017).

[13] S. Bose, A. Mazumdar, G. W. Morley, H. Ulbricht, M. Toroš, M. Paternostro, A. A. Geraci, P. F. Barker, M. Kim, and G. Milburn, Spin Entanglement Witness for Quantum Gravity, Phys. Rev. Lett. **119**, 240401 (2017).

[14] N. Lambert, Y.-N. Chen, Y.-C. Cheng, C.-M. Li, G.-Y. Chen, and F. Nori, Quantum biology, Nat. Phys. **9**, 10 (2013).

[15] K.-H. Chang, Stochastic Nelder–Mead simplex method–a new globally convergent direct search method for simulation optimization, Eur. J. Oper. Res. **220**, 684 (2012).

[16] A. Canabarro, S. Brito, and R. Chaves, Machine Learning Nonlocal Correlations, Phys. Rev. Lett. **122**, 200401 (2019).

[17] T. Kriváchy, Y. Cai, D. Cavalcanti, A. Tavakoli, N. Gisin, and N. Brunner, A neural network oracle for quantum

nonlocality problems in networks, npj Quantum Inf. **6**, 1 (2020).

[18] K. Bharti, T. Haug, V. Vedral, and L.-C. Kwek, How to teach AI to play Bell non-local games: Reinforcement learning, arXiv:1912.10783.

[19] J. Wallnöfer, A. A. Melnikov, W. Dür, and H. J. Briegel, Machine Learning for Long-Distance Quantum Communication, PRX Quantum **1**, 010301 (2020).

[20] V. Dunjko and H. J. Briegel, Machine learning & artificial intelligence in the quantum domain: A review of recent progress, Rep. Prog. Phys. **81**, 074001 (2018).

[21] D.-L. Deng, Machine Learning Detection of Bell Nonlocality in Quantum Many-Body Systems, Phys. Rev. Lett. **120**, 240402 (2018).

[22] A. A. Melnikov, P. Sekatski, and N. Sangouard, Setting up Experimental Bell Test with Reinforcement Learning, Phys. Rev. Lett. **125**, 160401 (2020).

[23] G. M. D'Ariano, M. G. Paris, and M. F. Sacchi, Quantum tomography, Adv. Imaging Electron Phys. **128**, 206 (2003).

[24] F. Buscemi, All Entangled Quantum States are Nonlocal, Phys. Rev. Lett. **108**, 200401 (2012).

[25] L. Lydersen, C. Wiechers, C. Wittmann, D. Elser, J. Skaar, and V. Makarov, Hacking commercial quantum cryptography systems by tailored bright illumination, Nat. Photonics **4**, 686 (2010).

[26] 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).

[27] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Bell nonlocality, Rev. Mod. Phys. **86**, 419 (2014).

[28] J. A. Nelder and R. Mead, A simplex method for function minimization, Comput. J. **7**, 308 (1965).

[29] A simplex is the n-dimensional generalization of the triangle. Given $n-1$ measurement parameters, a simplex contains n vertices that correspond to the parameter points in the $(n-1)$-dimensional space.

[30] The goal of the ARS is to sample the space of measurement parameters using a local or global search, in order to find a point able to improve the simplex, when the contraction step does not provide a satisfactory point. The process is randomly driven: With a probability $P_s > 0$ the Bell parameter S is evaluated in one point uniformly chosen in the parameter space (global search), while with probability $1 - P_s > 0$ a point of the simplex is randomly chosen and a point t_i is uniformly extracted from the hypersphere with center t_i and radius ϵ (local search). If $S(t^{\mathrm{ARS}}) \leqslant S(t^{\max})$, the search stops; otherwise, the ARS is repeated.

[31] 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).

[32] S. L. Braunstein and C. M. Caves, Wringing out better Bell inequalities, Ann. Phys. (Amsterdam) **202**, 22 (1990).

[33] A. Acín, S. Massar, and S. Pironio, Randomness versus Nonlocality and Entanglement, Phys. Rev. Lett. **108**, 100402 (2012).

[34] L. Masanes, Necessary and sufficient condition for quantum-generated correlations, arXiv:quant-ph/0309137.

[35] R. Horodecki, P. Horodecki, and M. Horodecki, Violating Bell inequality by mixed spin-12 states: Necessary and sufficient condition, Phys. Lett. A **200**, 340 (1995).

[36] T. Kim, M. Fiorentino, and F. N. C. Wong, Phase-stable source of polarization-entangled photons using a polarization Sagnac interferometer, Phys. Rev. A **73**, 012316 (2006).

[37] A. Fedrizzi, T. Herbst, A. Poppe, T. Jennewein, and A. Zeilinger, A wavelength-tunable fiber-coupled source of narrowband entangled photons, Opt. Express **15**, 15377 (2007).

[38] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, Measurement of qubits, Phys. Rev. A **64**, 052312 (2001).

[39] A. Grudka, K. Horodecki, M. Horodecki, P. Horodecki, M. Pawłowski, and R. Ramanathan, Free randomness amplification using bipartite chain correlations, Phys. Rev. A **90**, 032322 (2014).

[40] I. Šupić, R. Augusiak, A. Salavrakos, and A. Acín, Self-testing protocols based on the chained Bell inequalities, New J. Phys. **18**, 035013 (2016).

[41] This value becomes 1.6 bits if we employ the von Neumann entropy instead of the min-entropy to characterize the generated randomness.

[42] K. T. Goh, J. Kaniewski, E. Wolfe, T. Vértesi, X. Wu, Y. Cai, Y.-C. Liang, and V. Scarani, Geometry of the set of quantum correlations, Phys. Rev. A **97**, 022104 (2018).

[43] A. Rai, C. Duarte, S. Brito, and R. Chaves, Geometry of the quantum set on no-signaling faces, Phys. Rev. A **99**, 032106 (2019).

[44] F. Xu, X. Ma, Q. Zhang, H.-K. Lo, and J.-W. Pan, Secure quantum key distribution with realistic devices, Rev. Mod. Phys. **92**, 025002 (2020).

[45] C. A. Miller and Y. Shi, Robust protocols for securely expanding randomness and distributing keys using untrusted quantum devices, J. ACM **63**, 33 (2016).

[46] U. Vazirani and T. Vidick, Certifiable quantum dice: Or, true random number generation secure against quantum adversaries, in *Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing*, STOC '12 (Association for Computing Machinery, New York, 2012), pp. 61–76.

[47] R. Arnon-Friedman, F. Dupuis, O. Fawzi, R. Renner, and T. Vidick, Practical device-independent quantum cryptography via entropy accumulation, Nat. Commun. **9**, 459 (2018).

[48] M. Herrero-Collantes and J. C. Garcia-Escartin, Quantum random number generators, Rev. Mod. Phys. **89**, 015004 (2017).

[49] 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 (London) **464**, 1021 (2010).

[50] R. Colbeck and A. Kent, Private randomness expansion with untrusted devices, J. Phys. A: Math. Theor. **44**, 095305 (2011).

[51] I. Agresti, D. Poderini, L. Guerini, M. Mancusi, G. Carvacho, L. Aolita, D. Cavalcanti, R. Chaves, and F. Sciarrino, Experimental device-independent certified randomness generation with an instrumental causal structure, Commun. Phys. **3**, 110 (2020).

[52] R. Gallego, L. Masanes, G. De La Torre, C. Dhara, L. Aolita, and A. Acín, Full randomness from arbitrarily deterministic events, Nat. Commun. **4**, 2654 (2013).

[53] F. G. S. L. Brandão, R. Ramanathan, A. Grudka, K. Horodecki, M. Horodecki, P. Horodecki, T. Szarek, and H. Wojewódka, Realistic noise-tolerant randomness amplification using finite number of devices, Nat. Commun. **7**, 11345 (2016).

[54] Y. Liu, Q. Zhao, M.-H. Li, J.-Y. Guan, Y. Zhang, B. Bai, W. Zhang, W.-Z. Liu, C. Wu, X. Yuan, H. Li, W. J. Munro, Z. Wang, L. You, J. Zhang, X. Ma, J. Fan, Q. Zhang, and J.-W. Pan, Device-independent quantum random-number generation, Nature (London) **562**, 548 (2018).

[55] O. Nieto-Silleras, C. Bamps, J. Silman, and S. Pironio, Device-independent randomness generation from several Bell estimators, New J. Phys. **20**, 023049 (2018).

[56] O. Nieto-Silleras, S. Pironio, and J. Silman, Using complete measurement statistics for optimal device-independent randomness evaluation, New J. Phys. **16**, 013035 (2014).

[57] S. Pironio, L. Masanes, A. Leverrier, and A. Acín, Security of Device-Independent Quantum Key Distribution in the Bounded-Quantum-Storage Model, Phys. Rev. X **3**, 031007 (2013).

[58] The guessing probability is directly connected to the conditional min-entropy $H_{min} = -\log_2(p_{guess})$ representing the bits of randomness certifiable by Alice and Bob.

[59] M. Navascués, S. Pironio, and A. Acín, Bounding the Set of Quantum Correlations, Phys. Rev. Lett. **98**, 010401 (2007).

[60] C. Branciard, D. Rosset, N. Gisin, and S. Pironio, Bilocal versus nonbilocal correlations in entanglement-swapping experiments, Phys. Rev. A **85**, 032119 (2012).

[61] R. Chaves, C. Majenz, and D. Gross, Information–theoretic implications of quantum causal structures, Nat. Commun. **6**, 5766 (2015).

[62] T. Fritz, Beyond Bell's theorem: Correlation scenarios, New J. Phys. **14**, 103001 (2012).

[63] T. Fritz, Beyond Bell's theorem II: Scenarios with arbitrary causal structure, Commun. Math. Phys. **341**, 391 (2016).

[64] M.-O. Renou, E. Bäumer, S. Boreiri, N. Brunner, N. Gisin, and S. Beigi, Genuine Quantum Nonlocality in the Triangle Network, Phys. Rev. Lett. **123**, 140401 (2019).

[65] N. Gisin, Entanglement 25 years after quantum teleportation: testing joint measurements in quantum networks, Entropy **21**, 325 (2019).

[66] R. Chaves, G. Moreno, E. Polino, D. Poderini, I. Agresti, A. Suprano, M. R. Barros, G. Carvacho, E. Wolfe, A. Canabarro, R. W. Spekkens, and F. Sciarrino, Causal Networks and Freedom of Choice in Bell's Theorem, PRX Quantum **2**, 040323 (2021).

[67] G. Carvacho, F. Andreoli, L. Santodonato, M. Bentivegna, R. Chaves, and F. Sciarrino, Experimental violation of local causality in a quantum network, Nat. Commun. **8**, 14775 (2017).

[68] D. J. Saunders, A. J. Bennet, C. Branciard, and G. J. Pryde, Experimental demonstration of nonbilocal quantum correlations, Sci. Adv. **3**, e1602743 (2017).

[69] Q.-C. Sun, Y.-F. Jiang, B. Bai, W. Zhang, H. Li, X. Jiang, J. Zhang, L. You, X. Chen, Z. Wang, Q. Zhang, J. Fan, and J.-W. Pan, Experimental demonstration of non-bilocality with truly independent sources and strict locality constraints, Nat. Photonics **13**, 687 (2019).

[70] D. Poderini, I. Agresti, G. Marchese, E. Polino, T. Giordani, A. Suprano, M. Valeri, G. Milani, N. Spagnolo, G. Carvacho, R. Chaves, and F. Sciarrino, Experimental violation of *n*-locality in a star quantum network, Nat. Commun. **11**, 2467 (2020).

[71] A. Tavakoli, A. Pozas-Kerstjens, M.-X. Luo, and M.-O. Renou, Bell nonlocality in networks, arXiv:2104.10700.

[72] I. Agresti, B. Polacchi, D. Poderini, E. Polino, A. Suprano, I. Šupić, J. Bowles, G. Carvacho, D. Cavalcanti, and F. Sciarrino, Experimental Robust Self-Testing of the State Gen-

erated by a Quantum Network, PRX Quantum **2**, 020346 (2021).

[73] T. Kriváchy, Y. Cai, J. Bowles, D. Cavalcanti, and N. Brunner, Fast semidefinite programming with feedforward neural networks, arXiv:2011.05785.