Increasing the statistical significance of entanglement detection in experiments

Bastian Jungnitsch,[1] Sönke Niekamp,[1] Matthias Kleinmann,[1] Otfried Gühne,[1,2]
He Lu,[3] Wei-Bo Gao,[3] Yu-Ao Chen,[3,4] Zeng-Bing Chen,[3] and Jian-Wei Pan[3,4]

[1] *Institut für Quantenoptik und Quanteninformation, Österreichische Akademie der Wissenschaften, Technikerstraße 21A, A-6020 Innsbruck, Austria*
[2] *Institut für Theoretische Physik, Universität Innsbruck, Technikerstraße 25, A-6020 Innsbruck, Austria*
[3] *Hefei National Laboratory for Physical Sciences at Microscale and Department of Modern Physics, University of Science and Technology of China, Hefei, Anhui 230026, China*
[4] *Physikalisches Institut, Ruprecht-Karls-Universität Heidelberg, Philosophenweg 12, 69120 Heidelberg, Germany*

(Dated: December 3, 2009)

Entanglement is often verified by a violation of an inequality like a Bell inequality or an entanglement witness. Considerable effort has been devoted to the optimization of such inequalities in order to obtain a high violation. We demonstrate theoretically and experimentally that such an optimization does not necessarily lead to a better entanglement test, if the statistical error is taken into account. Theoretically, we show for different error models that reducing the violation of an inequality can improve the significance. Experimentally, we observe this phenomenon in a four-photon experiment, testing the Mermin and Ardehali inequality for different levels of noise.

Introduction — Quantum theory is a statistical theory, predicting in general only probabilities for experimental results. Consequently, in most experiments observing quantum effects, several copies of a quantum state are generated and individually measured, to determine the desired probabilities. As only a finite number of states can be generated, this leads to an unavoidable statistical error.

Many of today's experiments aim at the generation of entanglement, which is considered as a central resource in quantum information processing [1, 2]. So far, entanglement of up to ten qubits has been achieved using trapped ions or photons [3, 4]. For the experimental verification of entanglement, often inequalities for the correlations — such as Bell inequalities or entanglement witnesses — are used [2], for which a violation indicates entanglement. The maximization of this violation has been investigated in detail, cf. Refs. [2, 5]. In fact, making such inequalities more sensitive is a crucial step in order to allow advanced experiments with more particles.

In this paper we demonstrate theoretically and experimentally that such an optimization does not necessarily lead to a better entanglement test, if the statistical nature of quantum theory is taken into account. It was already noted [6] that, when aiming at ruling out local realism, highly entangled states do not necessarily deliver a stronger test than weakly entangled states, but this does not answer the question which inequality to use for a given state and it remains unclear how to apply it to actual error models used in experiments. In Ref. [7] different entanglement detection methods for two qubits have been compared, but most of these methods cannot be applied to multiparticle systems.

From the theoretical side, we show for different error models that decreasing the violation of an inequality can improve the significance. Experimentally, we demonstrate this phenomenon in a four-photon experiment, in which we measure the Mermin and Ardehali inequality and find that the former inequality leads to a higher significance than the latter, despite of a lower violation. Finally, we discuss consequences for future experiments and possible applications to other problems.

Statement of the problem — A witness $\mathcal{W}$ is an observable which has a non-negative expectation value on all separable states (that is, states which can be written as a mixture of product states, $\varrho = \sum_k p_k |a_k, b_k\rangle\langle a_k, b_k|$ with some probabilities p_k). Hence, a negative expectation value of a witness signals the presence of entanglement. Similarly, a Bell inequality is an inequality $\langle\mathcal{B}\rangle \leq C_{\text{lhv}}$ where $\mathcal{B}$ is a sum of certain correlation terms and which holds if the measurement outcomes originate from a local hidden variable (LHV) model. As separable states allow a description by LHV models, a violation of a Bell inequality implies the presence of entanglement.

In both cases, we define $\mathcal{V}$ as the violation of the corresponding inequality. That is, for a witness we have $\mathcal{V}(\mathcal{W}) = -\langle\mathcal{W}\rangle$ while for a Bell inequality $\mathcal{V}(\mathcal{B}) = \langle\mathcal{B}\rangle - C_{\text{lhv}}$. Then, the significance of an entanglement test can be defined as

$$\mathcal{S} = \frac{\mathcal{V}}{\mathcal{E}} \tag{1}$$

where $\mathcal{E}$ is the statistical error for the experiment. Clearly, $\mathcal{E}$ depends on the particular experimental implementation and on the error model used. Nevertheless, in any experiment $\mathcal{S}$ is a well characterized quantity and its notion is widely used in the literature, when the violation is expressed in terms of "standard deviations".

Previously, much effort has been devoted to improving entanglement tests in order to achieve a higher violation. For instance, for entanglement witnesses a mature theory how to optimize witnesses has been developed [5]. Here, for a given witness $\mathcal{W}$ one tries to find a positive operator P which one can subtract from the witness, such that

$\mathcal{W}' = \mathcal{W} - P$ is still a witness. In order to have a more significant result, however, one can either increase $\mathcal{V}$ in Eq. (1) or decrease $\mathcal{E}$. It is a central result of this paper that decreasing $\mathcal{E}$ is often superior.

Variance as the error — Let us first consider a simple model, where we take the square root of the variance as the error of a witness,

$$\mathcal{E}(\mathcal{W}) = \Delta(\mathcal{W}) = \sqrt{\langle \mathcal{W}^2 \rangle - \langle \mathcal{W} \rangle^2}. \tag{2}$$

An experimentally relevant model will be discussed below. This simple model already allows us to conclude that the standard optimization of witnesses is, for a large class of situations, not the appropriate approach to increase the significance. We can formulate:

Observation. Let $\varrho = |\psi\rangle\langle\psi|$ be a pure state detected by the witness $\mathcal{W}$. Then, one can always increase the significance of $\mathcal{W}$ at the expense of optimality (i.e., by *adding* a positive operator). With this method one can make the significance arbitrarily large.

The idea to prove this is to find a positive observable P, such that $|\psi\rangle$ is an eigenstate of $\mathcal{W}' = \mathcal{W} + P$; then the error vanishes. Indeed, such a P can be found, details are given in the Appendix.

Multi-photon experiments — Let us now consider a realistic situation, in which other and more specific error models are used. As our later implementation uses multi-photon entanglement, we concentrate on this type of experiments but our ideas can also be applied to other implementations, such as trapped ions [8].

The basic experimental quantities are the numbers of detection events n_i from the different detectors i. From these data, all other quantities such as correlations or mean values of observables are derived.

In the standard error model for photonic experiments [4, 9], it is assumed that the counts are distributed according to a Poissonian distribution. Its free parameter is fixed by taking the observed value as mean value. That is, for a certain measurement outcome i one sets the mean value as $\langle n_i \rangle = n_i$ and the error as $\mathcal{E}(n_i) = \sqrt{n_i}$, where the square root comes in as the standard deviation of a Poissonian distribution. In general, for a quantity $f = f(n_i)$ that is a function of several counts, Gaussian error propagation has to be applied to obtain the error (see also below).

To give an example, consider a two-qubit correlation

$$\mathcal{M} = \alpha Z_1 Z_2 + \beta Z_1 \mathbb{1}_2 + \gamma \mathbb{1}_1 Z_2. \tag{3}$$

Here and in the following, Z_k (or $\mathbb{1}_k$) denotes the Pauli matrix σ_z (or the identity matrix) acting on the kth qubit, and tensor product symbols are omitted. $\langle \mathcal{M} \rangle$ can be determined by measuring in the common eigenbasis of all three terms in $\mathcal{M}$, i.e., by projecting onto $|00\rangle, |01\rangle, |10\rangle$ and $|11\rangle$. Repeating this with many copies of the state will lead to count numbers n_{kl} with $k, l = 0$ or 1 and to count rates $p_{kl} = n_{kl}/n_{\text{tot}}$, where $n_{\text{tot}} = n_{00} + n_{01} + n_{10} + n_{11}$ is the total number of events. The mean value $\langle \mathcal{M} \rangle$ can be written as a linear combination of p_{kl}, namely $\langle \mathcal{M} \rangle = \lambda_{00} p_{00} + \lambda_{01} p_{01} + \lambda_{10} p_{10} + \lambda_{11} p_{11}$ with $\lambda_{00} = \alpha + \beta + \gamma, \lambda_{01} = -\alpha + \beta - \gamma, \lambda_{10} = -\alpha - \beta + \gamma$, and $\lambda_{11} = \alpha - \beta - \gamma$. Then, according to Gaussian error propagation, the squared error is given by [10]

$$\mathcal{E}(\mathcal{M})^2 = \sum_{k,l} \left[\frac{\partial \langle \mathcal{M} \rangle}{\partial n_{kl}} \right]^2 \mathcal{E}(n_{kl})^2 = \sum_{k,l} \left[\frac{\lambda_{kl}}{n_{\text{tot}}} - \frac{\langle \mathcal{M} \rangle}{n_{\text{tot}}} \right]^2 n_{kl}. \tag{4}$$

Let us finally discuss the underlying assumptions of this error model. The first main assumption is that the n_{kl} are Poisson distributed and their errors are uncorrelated. This is well motivated by the experimental observations. Moreover, Gaussian error propagation stems from a Taylor expansion of the function f. Finally, if one interprets the standard deviation as a confidence interval, one tacitly assumes that the distribution is Gaussian, as for other distributions the connection is not so direct. If the number of events for all detectors is sufficiently large (e.g. $n_{kl} \gtrsim 10$), however, the Poissonian distribution is well approximated by a Gaussian distribution.

Bell inequalities for four particles — Let us now discuss the Mermin and Ardehali inequality as experimentally relevant examples. For the Mermin inequality [11] we consider

$$\mathcal{B}_{\text{M}} = X_1 X_2 X_3 X_4 - [X_1 X_2 Y_3 Y_4 + \text{ perm.}] + Y_1 Y_2 Y_3 Y_4, \tag{5}$$

where the bracket $[\ldots]$ is meant as a sum over all permutations of the term $X_1 X_2 Y_3 Y_4$ that yield distinct operators. For states allowing an LHV description, $\langle \mathcal{B}_{\text{M}} \rangle \leq 4$ holds. We have written $\mathcal{B}_{\text{M}}$ with the Pauli matrices as observables, since they are used later, however, one might replace them by arbitrary dichotomic measurements.

Second, we consider the Ardehali inequality $\langle \mathcal{B}_{\text{A}} \rangle \leq 2\sqrt{2}$ [12], where

$$\begin{aligned} \mathcal{B}_{\text{A}} = \Big[& X_1 X_2 X_3 A_4 + X_1 X_2 X_3 B_4 \\ & - [X_1 Y_2 Y_3 A_4 + \text{ perm.}] - [X_1 Y_2 Y_3 B_4 + \text{ perm.}] \\ & - [X_1 X_2 Y_3 A_4 + \text{ perm.}] + [X_1 X_2 Y_3 B_4 + \text{ perm.}] \\ & + Y_1 Y_2 Y_3 A_4 - Y_1 Y_2 Y_3 B_4 \Big] / \sqrt{2}. \end{aligned} \tag{6}$$

The sums in brackets include all distinct permutations on the first three qubits. We will later set $A_4 = (X_4 + Y_4)/\sqrt{2}$ and $B_4 = (X_4 - Y_4)/\sqrt{2}$, but again, in order to test local realism, the observables can remain arbitrary [13].

The Mermin and Ardehali inequality are designed to reveal the non-local correlations of the four-qubit GHZ state,

$$|GHZ_4\rangle = \frac{1}{\sqrt{2}} (|0000\rangle + |1111\rangle). \tag{7}$$

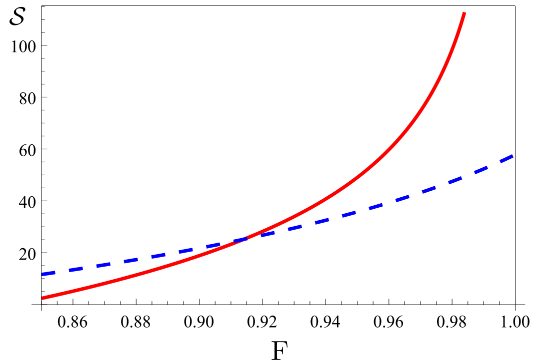

FIG. 1: Significance $\mathcal{S}$ for the Mermin (red, solid) and the Ardehali inequality (blue, dashed) for bit flip noise. On the horizontal axis, we show the fidelity with respect to a perfect GHZ state. We assumed that the experimenter prepares 8000 instances of a GHZ state, and chooses either to measure the eight terms of the Mermin inequality (each term with 1000 realizations of the state) or the 16 terms of the Ardehali inequality with 500 states per correlation term. See text for further details.

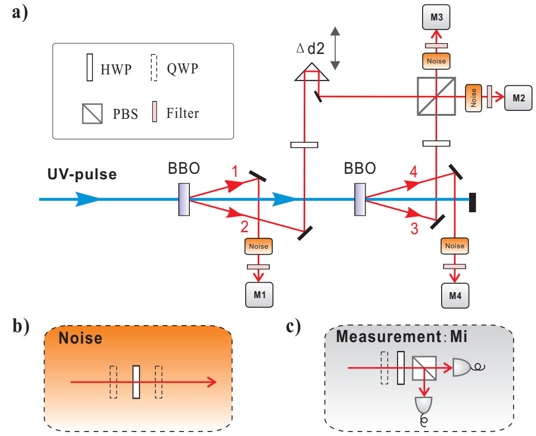

FIG. 2: Scheme of the experimental setup. **a**. The setup to generate the required four-photon GHZ state. Femtosecond laser pulses (≈ 200 fs, 76 MHz, 788 nm) are converted to ultraviolet pulses through a frequency doubler LiB_3O_5 (LBO) crystal (not shown). The pulses go through two main β-barium borate (BBO) crystals (2 mm), generating two pairs of photons. The observed two-fold coincidence count rates are about 1.6×10^4/s with a visibility of 96% (94%) in the H/V $(+/-)$ basis. **b**. Setup for engineering the bit-flip noise. **c**. The measurement setup.

For this state we have $\langle \mathcal{B}_{\mathrm{M}} \rangle = \langle \mathcal{B}_{\mathrm{A}} \rangle = 8$. As the bound for LHV models for the Ardehali inequality is smaller, the violation $\mathcal{V}$ is larger. This may lead to the opinion that the Ardehali inequality is "better" than the Mermin inequality for the state $|GHZ_4\rangle$.

However, this belief is easily shattered, if the significance $\mathcal{S}$ is considered as the relevant figure of merit. This can be seen directly from the expression of the error in Eq. (4). The GHZ state is an eigenstate for each of the correlation measurements in the Mermin inequality (they are so-called stabilizing operators of the GHZ state). Hence, if the Mermin inequality for a perfect GHZ state is measured, we have in the last term of Eq. (4) for each case k, l either $\lambda_{kl} = \langle \mathcal{M} \rangle$ (since the mean value is an eigenvalue) or $n_{kl} = 0$, hence the error $\mathcal{E}(\mathcal{M})$ vanishes. The Ardehali inequality, however, does not contain stabilizer terms, and the error remains finite.

For an experimental application it is important that the Mermin inequality still leads to a higher significance than the Ardehali inequality, even if noise is introduced [14]. In order to see this, we considered bit flip noise, which can easily be simulated in experiment. Therefore, we used a GHZ state whose qubits are locally affected by the bit flip operation f with probability p, i.e. $f(\varrho_i) = (1-p)\varrho_i + pX_i\varrho_i X_i$ for each qubit i. In Fig. 1, we plotted the significance $\mathcal{S}$ versus the fidelity F of the prepared state with respect to a perfect GHZ state, i.e. $F = \langle GHZ_4 | \varrho_{exp} | GHZ_4 \rangle$. The results for white noise or the dephasing noise model are qualitatively the same.

Experimental setup — Let us proceed with the experimental implementation. Spontaneous down conversion has been used to produce the desired 4-photon state [see Fig. 2(a)]. With the help of polarizing beam splitters (PBSs), half-wave plates (HWPs), and conventional photon detectors, we prepare a 4-qubit GHZ state [see Eq. (7)], where $|0\rangle = |H\rangle$ ($|1\rangle = |V\rangle$) denotes horizontal (vertical) polarization. In the experiment, as a proof-of-principle, we have chosen the bit-flip noise channel to demonstrate the theory introduced in this paper. As shown in Fig. 2(b), the noisy quantum channels are engineered by one HWP sandwiched with two quarter-wave plates (QWPs) [15]. The HWP is switched randomly between angles $+\theta$ and $-\theta$ and the QWPs are set at $0°$ with respect to the vertical direction. In this way, the noisy quantum channel can be engineered with a bit-flip probability $p = \sin^2(2\theta)$. The Pauli matrix measurements required in the Bell inequality test can be implemented by a combination of HWP, QWP and PBS [see Fig. 2(c)]. The fidelity of the prepared GHZ state can be calculated from the identity $F = \frac{1}{2}(\langle |0000\rangle\langle 0000| + |1111\rangle\langle 1111| \rangle) + \frac{1}{16}\langle \mathcal{B}_M \rangle$. In the experiment without added noise its value is $F = 0.84 \pm 0.01$.

Experimental results — With different noise levels, the experimental results of the violation, the statistical error and the significance are shown in Table I. The first observation is that, when there is no engineered noise, the violation of the Mermin inequality is smaller than the violation of the Ardehali inequality. Its significance, however, is larger than that of the Ardehali inequality; this proves that testing the Mermin inequality is a better choice to characterize the entanglement in this case. The second observation is that when the noise level increases, the significance in the Mermin inequality decreases more quickly. When $\theta = \pm 6°, \pm 8°$, the significance for the Ardehali inequality is already larger than that for the

θ	p	$\mathcal{V}(M)$	$\mathcal{E}(M)$	$\mathcal{S}(M)$	$\mathcal{V}(A)$	$\mathcal{E}(A)$	$\mathcal{S}(A)$
$\pm 0°$	0	2.37	0.05	44.3	5.16	0.15	35.0
$\pm 2°$	0.005	2.00	0.06	33.4	4.45	0.15	29.2
$\pm 4°$	0.019	1.57	0.07	23.7	3.51	0.16	21.8
$\pm 6°$	0.043	1.13	0.07	16.2	2.90	0.16	17.8
$\pm 8°$	0.076	0.67	0.08	8.8	2.31	0.17	13.7

TABLE I: Experimental values of the violation, the statistical error and the significance for different values of θ (and the corresponding p). $\mathcal{V}(M)$, $\mathcal{E}(M)$, $\mathcal{S}(M)$ represent the values of $\mathcal{V}$, $\mathcal{E}$ and $\mathcal{S}$ in testing the Mermin inequality; $\mathcal{V}(A)$, $\mathcal{E}(A)$, $\mathcal{S}(A)$ represent the corresponding values for the Ardehali inequality. Each setting $X_1X_2X_3X_4$ etc. in the Mermin inequality is measured for 800 s, while each setting $X_1X_2X_3A_4$ etc. in the Ardehali inequality is measured for 400 s. The average total count number for each inequality is about 7500.

Mermin inequality. This confirms the theoretical predictions in Fig. 1.

Discussion — We have proved that in order to justify a claim of entanglement in an experiment with a high statistical significance, it can be favorable to use an entanglement witness or a Bell inequality that results in a lower violation. We confirmed this experimentally using four-photon GHZ states. Our results show that the usual way of optimizing witnesses will not necessarily lead to more powerful tools for the analysis of many-particle experiments. It is important to note that when the number of photons in multi-photon experiments is increased the count rates decrease; consequently, the statistical error becomes more and more relevant.

Moreover, our results also provide a direction to improve entanglement tests for low count rates. For the case of the Ardehali and Mermin inequality, the observed effect relied on the fact that in the Mermin inequality only stabilizer measurements were made. There are already powerful approaches available to construct witnesses from stabilizer observables [16] and also other Mermin-like or Ardehali-like Bell inequalities have been explored [17]. Consequently, these approaches are promising candidates for developing sensitive analysis tools. Further, inequalities similar to witnesses have also been proposed and used to characterize quantum gate fidelities [18], which is another application of our theory. Finally, we believe that a careful analysis of the statistical error can also help in designing loophole-free Bell tests.

We thank A. Cabello, R. Gill, O. Gittsovich, D. James, J.-Å. Larsson, and C. Roos for discussions. This work has been supported by the FWF (START Prize), the EU (SCALA, OLAQUI QICS, Marie Curie Excellence Grant), the NNSFC, the NFRP (2006CB921900), the CAS, the ICP at HFNL, and the A. v. Humboldt Foundation.

Appendix — To prove the Observation, we use as an ansatz for the improved witness $\mathcal{W}' = \mathcal{W} + \gamma P$, where γ is positive and P is a positive observable with unit trace. For small γ, we expand

$$-\frac{\langle \mathcal{W}' \rangle}{\Delta(\mathcal{W}')} = -\frac{\langle \mathcal{W} \rangle}{\Delta(\mathcal{W})} + \gamma \frac{\langle \mathcal{W} \rangle}{2\Delta^3(\mathcal{W})} \Big[\langle \mathcal{W}P + P\mathcal{W} \rangle - 2\frac{\langle \mathcal{W}^2 \rangle}{\langle \mathcal{W} \rangle} \langle P \rangle \Big] + O(\gamma^2) \ . \quad (8)$$

Maximizing this expression over all positive P with $\mathrm{Tr}(P) = 1$ is equivalent to minimizing $\mathrm{Tr}(QP)$, where $Q = \varrho\mathcal{W} + \mathcal{W}\varrho - 2\langle \mathcal{W}^2 \rangle / \langle \mathcal{W} \rangle \varrho$. Hence the optimal P is a one-dimensional projector $P = |\varphi\rangle\langle\varphi|$, where $|\varphi\rangle$ is an eigenvector corresponding to the minimal eigenvalue of Q. We still have to show that this minimal eigenvalue is negative [19]. To this end, we make the ansatz $|\varphi\rangle = \alpha|\psi\rangle + \beta|\psi^\perp\rangle$, where $|\psi^\perp\rangle$ is orthogonal to $|\psi\rangle$. We then have to minimize

$$\mathrm{Tr}(QP) = 2\,\mathrm{Re}(\alpha^*\beta\langle\psi|\mathcal{W}|\psi^\perp\rangle) - 2|\alpha|^2 \frac{\Delta^2_\psi(\mathcal{W})}{\langle\psi|\mathcal{W}|\psi\rangle} \ . \quad (9)$$

We can always choose the phases of α and β such that $\mathrm{Re}(\ldots)$ is negative. Therefore the optimal $|\psi^\perp\rangle$ is the vector orthogonal to $|\psi\rangle$ which maximizes $|\langle\psi|W|\psi^\perp\rangle|$, that is, $|\psi^\perp_{\rm opt}\rangle = [\mathbf{1} - |\psi\rangle\langle\psi|]\mathcal{W}|\psi\rangle/\Delta_\psi(\mathcal{W})$. Furthermore, we can always choose the moduli of α and β such that the negative term $2\,\mathrm{Re}(\ldots)$ dominates the positive second term. This shows that the minimal eigenvalue of Q is negative.

For finite γ we can iterate this procedure. We will always find the same $|\psi^\perp_{\rm opt}\rangle$ (though α and β will be different in each iteration step). We therefore make the following ansatz for the final result of the iteration:

$$\gamma P = a|\psi\rangle\langle\psi| + b|\psi^\perp_{\rm opt}\rangle\langle\psi^\perp_{\rm opt}| + c|\psi\rangle\langle\psi^\perp_{\rm opt}| + \mathrm{h.\,c.} \quad (10)$$

If we choose $c = -\Delta_\psi(\mathcal{W})$, $ab \geq |c|^2$, and $a, b > 0$, then γP is positive, $|\psi\rangle$ is an eigenstate of $\mathcal{W}'$, and the standard deviation $\Delta_\psi(\mathcal{W}')$ is zero, so $\mathcal{S}$ diverges. $\square$

[1] R. Horodecki *et al.*, Rev. Mod. Phys. **81**, 865 (2009); M. Plenio and S. Virmani, Quant. Inf. Comp. **7**, 1 (2007).
[2] O. Gühne and G. Tóth, Phys. Reports **474**, 1 (2009).
[3] H. Häffner *et al.*, Nature **438**, 643 (2005); W.B. Gao *et al.*, arXiv:0809.4277.
[4] C.-Y. Lu *et al.*, Nature Physics **3**, 91 (2007).
[5] M. Lewenstein *et al.*, Phys. Rev. A **62**, 052310 (2000); P. Hyllus and J. Eisert, New J. Phys. **8** 51 (2006); G. Tóth *et al.*, New J. Phys. **11**, 083002 (2009).
[6] W. van Dam, R.D. Gill, and P. Grünwald, IEEE-Transactions on Information Theory **51**, 2812 (2005); A. Acín, R.D. Gill, and N. Gisin, Phys. Rev. Lett. **95**, 210402 (2005); R.D. Gill, IMS Lecture Notes Monograph Series **55**, 135 (2007), math/0610115.
[7] J.B. Altepeter *et al.*, Phys. Rev. Lett. **95**, 033601 (2005).
[8] H. Häffner, C.F. Roos, and R. Blatt, Phys. Rep. **469**, 155 (2008).

[9] D.F.V. James *et al.*, Phys. Rev. A **64**, 052312 (2001).

[10] In Ref. [9] the second term, $-\langle\mathcal{M}\rangle/n_{\rm tot}$, is omitted, leading to a systematic overestimation of the standard deviation. Such an overestimation is problematic, if the claim is that certain state properties (e.g. having a negative partial transpose) are not significant. This can occur, e.g. in the analysis of bound entanglement, see A. Elias and M. Bourennane, Nature Phys. **5**, 748 (2009).

[11] N. Mermin, Phys. Rev. Lett. **65**, 1838 (1990).

[12] M. Ardehali, Phys. Rev. A **46**, 5375 (1992).

[13] Indeed, with the given choice of observables the operators $\mathcal{B}_{\rm M}$ and $\mathcal{B}_{\rm A}$ are identical within quantum mechanics. However, the use of A and B within the Ardehali inequality results in a more restrictive test for LHV models.

[14] This is also important, as for nearly perfect GHZ states, some count numbers n_{kl} will be close to zero. Then, the interpretation of the statistical error as a confidence interval may be questioned.

[15] Y.-A. Chen *et al.*, Phys. Rev. Lett. **96**, 220504 (2006).

[16] G. Tóth and O. Gühne, Phys. Rev. Lett. **94**, 060501 (2005).

[17] V. Scarani *et al.*, Phys. Rev. A **71**, 042325 (2005); A. Cabello, O. Gühne, and D. Rodriguez, *ibid.* **77** 062106 (2008); O. Gühne and A. Cabello, *ibid.* **77**, 032108 (2008).

[18] H.F. Hofmann, Phys. Rev. Lett. **94**, 160504 (2005); W.-B. Gao, *et al.*, arXiv:0905.2103.

[19] Up to this point we did not use that the state ϱ is pure. Our method also works for such mixed states for which Q has a negative eigenvalue, but then one cannot make $\mathcal{S}$ arbitrarily large.