Systematic Errors in Current Quantum State Tomography Tools

Christian Schwemmer,[1] Lukas Knips,[1] Daniel Richart,[1] Harald Weinfurter,[1] Tobias Moroder,[2] Matthias Kleinmann,[2,3,4] and Otfried Gühne[2]

[1]*Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Straße 1, D-85748 Garching, Germany and Department für Physik, Ludwig-Maximilians-Universität, D-80797 München, Germany*

[2]*Naturwissenschaftlich-Technische Fakultät, Universität Siegen, Walter-Flex-Straße 3, D-57068 Siegen, Germany*

[3]*Departamento de Matemtica, Universidade Federal de Minas Gerais, Caixa Postal 702, Belo Horizonte, Minas Gerais 31270-901, Brazil*

[4]*Department of Theoretical Physics, University of the Basque Country UPV/EHU, P.O. Box 644, E-48080 Bilbao, Spain*

(Received 31 October 2013; published 26 February 2015)

Common tools for obtaining physical density matrices in experimental quantum state tomography are shown here to cause systematic errors. For example, using maximum likelihood or least squares optimization to obtain physical estimates for the quantum state, we observe a systematic underestimation of the fidelity and an overestimation of entanglement. Such strongly biased estimates can be avoided using linear evaluation of the data or by linearizing measurement operators yielding reliable and computational simple error bounds.

DOI: 10.1103/PhysRevLett.114.080403 PACS numbers: 03.65.Wj, 03.65.Ud, 06.20.Dk

Introduction.—Quantum state tomography (QST) [1] enables us to fully determine the state of a quantum system and, thereby, to deduce all its properties. As such, QST and the closely related quantum process tomography (QPT) are widely used to characterize and to evaluate numerous experimentally implemented qubit states or their dynamics, e.g., in ion trap experiments [2,3], photonic systems [4,5], superconducting circuits [6], or nuclear magnetic resonance systems [7,8]. The increasing complexity of today's multiqubit or multiqudit quantum systems brought new challenges but, also, progress. Now, highly efficient methods allow an even scalable analysis for important subclasses of states [9,10]. The calculation of errors of QST was significantly improved, although the errors remain numerically expensive to evaluate for larger systems [11]. Moreover, QST and QPT were adopted to detect systematic errors in the alignment of an experiment itself [12].

A central step in QST is to establish the state from the acquired experimental data. A direct, linear evaluation of the data returns, almost for sure, an unphysical density matrix with negative eigenvalues [13]. Thus, several schemes have been developed to obtain a physical state which resembles the observed data as closely as possible [4,14,15]. From classical statistics, it is known that a constraint, such as the physicality of a state, can lead to systematic deviations, called bias, in parameter estimation for finite statistics [16,17]. However, in quantum tomography experiments, this effect has hardly ever been considered.

In this Letter, we test whether the naive expectation is met that QST delivers meaningful estimates for physical quantities. We test this for the two most commonly used reconstruction schemes—maximum likelihood (ML) [15] and least squares (LS) [4]—using Monte Carlo simulations. In detail, we investigate whether or not a possibly occurring bias of these reconstruction schemes is relevant at all on the example of some of the most prominent multiqubit quantum states. We find that, due to the constraint of physicality, both ML and LS return states which deviate systematically from the true state. Foremost, for small sample sizes, as they are typical in multiqubit experiments, it leads to significantly differing estimates for physical quantities as illustrated for the fidelity with respect to the Greenberger-Horne-Zeilinger (GHZ) state in Fig. 1 [18]. These deviations depend on the experimental and statistical noise but are typically larger than commonly deduced errors [19].

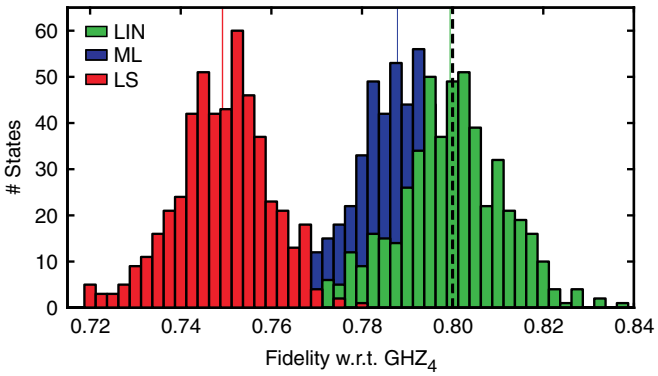

FIG. 1 (color online). Histogram of the fidelity estimates of 500 independent simulations of QST of a noisy four-party GHZ state for three different reconstruction schemes. The values obtained via ML (blue) or LS (red) fluctuate around a value that is lower than the initial fidelity of 80% (dashed line). For comparison, we also show the result using LIN (green), which does not have this systematic error called bias.

We demonstrate that, for linear quantities, one can provide meaningful confidence regions directly from the raw data and that it is also possible to compute meaningful lower (upper) bounds on convex (concave) functions like the bipartite negativity.

Standard state tomography tools.—The aim of QST is to find an estimate together with some confidence region for the initially unknown state ϱ_0 of a system via appropriate measurements on multiple preparations of this state. For an n-qubit system, the so-called Pauli tomography scheme consists of measuring in the eigenbases of all 3^n possible combinations of local Pauli operators, each yielding 2^n possible results [4]. In more general terms, in a tomography protocol, one repeats, for each measurement setting s, the experiment a certain number of times N_s and obtains c_r^s times the result r. These numbers then yield the frequencies $f_r^s = c_r^s/N_s$. The probability to observe the outcome r for setting s is given by $P_{\varrho_0}^s(r) = tr(\varrho_0 M_r^s)$. Here, M_r^s labels the measurement operator corresponding to the result r when measuring setting s. The probabilities $P_{\varrho_0}^s(r)$ will uniquely identify the unknown state ϱ_0, if the set of operators M_r^s spans the space of traceless Hermitian operators.

Provided the data f, i.e., the experimentally determined frequencies f_r^s, one requires a method to determine the estimate $\hat{\varrho} \equiv \hat{\varrho}(f)$ of the unknown state ϱ_0. Simply inverting the relations for $P_{\varrho_0}^s(r)$ we obtain

$$\hat{\varrho}_{\mathrm{LIN}} = \sum_{r,s} A_r^s f_r^s, \tag{1}$$

where A_r^s are determined from the measurement operators M_r^s [8,20]. Note that there is a canonical construction of A_r^s even for the case of an overcomplete set of M_r^s, see SM 1 in the Supplemental Material (SM) [21]. This reconstruction of $\hat{\varrho}_{\mathrm{LIN}}$ is computationally simple and has become known as linear inversion (LIN) [23]. In principle, Gaussian error propagation could also be used here to determine confidence regions.

Yet, due to unavoidable statistical fluctuations, the estimate $\hat{\varrho}_{\mathrm{LIN}}$ is not a physical density operator for typical experimental situations; i.e., generally some eigenvalues are negative. Apart from causing issues related to a physical interpretation of such a "state", negative eigenvalues impedes the evaluation of interesting functions like the von Neumann entropy, the quantum Fisher information, or an entanglement measure like the negativity, as these functions are defined, or meaningful, only for valid, i.e., positive semidefinite, quantum states.

For this reason, different methods have been introduced that mostly follow the paradigm that the reconstructed state $\hat{\varrho} = \arg\max_{\varrho \geq 0} T(\varrho|f)$ maximizes a target function $T(\varrho|f)$ within the set of valid density operators. This target function, thereby, measures how well a density operator ϱ agrees with the observed data f. Two common choices are ML [15] where $T_{\mathrm{ML}} = \sum_{r,s} f_r^s \log[P_\varrho^s(r)]$, and LS [4] where $T_{\mathrm{LS}} = -\sum_{r,s} [f_r^s - P_\varrho^s(r)]^2 / P_\varrho^s(r)$. We denote the respective optima by $\hat{\varrho}_{\mathrm{ML}}$ and $\hat{\varrho}_{\mathrm{LS}}$. From these estimates, one then easily computes any physical quantity of the observed state, e.g., the fidelities $\hat{F}_{\mathrm{ML}} = \langle\psi|\hat{\varrho}_{\mathrm{ML}}|\psi\rangle$ and $\hat{F}_{\mathrm{LS}} = \langle\psi|\hat{\varrho}_{\mathrm{LS}}|\psi\rangle$ with respect to the target state $|\psi\rangle$.

Numerical simulations.—To enable detailed analysis of the particular features of the respective state reconstruction algorithm and to exclude influence of systematic experimental errors, we perform Monte Carlo simulations. For a chosen state ϱ_0, the following procedure is used: (i) Compute the single event probabilities $P_{\varrho_0}^s(r)$, (ii) toss coins to get frequencies distributed according to the multinomial distribution determined by $P_{\varrho_0}^s(r)$ and N_s, (iii) reconstruct the state with either reconstruction method and compute the functions of interest, (iv) carry out steps (ii) and (iii) 500 times. Note that the optimality of the maximizations for ML and LS in step (iii) is certified by convex optimization [10,24].

First, we consider the four-qubit GHZ state $|\mathrm{GHZ}_4\rangle = (|0000\rangle + |1111\rangle)/\sqrt{2}$ mixed with white noise, i.e., $\varrho_0 = p|\mathrm{GHZ}_4\rangle\langle\mathrm{GHZ}_4| + (1-p)\mathbb{1}/16$ where p is chosen such that the fidelity is $\langle\mathrm{GHZ}_4|\varrho_0|\mathrm{GHZ}_4\rangle = 0.8$. This state is used to simulate the Pauli tomography scheme. Figure 1 shows a typical histogram of the resulting fidelities for $N_s = 100$ measurement repetitions, which is a typical value used for various multiqubit experiments. The fidelities obtained via LIN reconstruction fluctuate around the initial value ($\bar{F}_{\mathrm{LIN}} = 0.799 \pm 0.012$). (The values given there are the mean and the standard deviation obtained from the 500 reconstructed states). In stark contrast, both ML ($\bar{F}_{\mathrm{ML}} = 0.788 \pm 0.010$) and even more LS ($\bar{F}_{\mathrm{LS}} = 0.749 \pm 0.010$) systematically underestimate the fidelity, i.e., are strongly biased. Evidently, the fidelities of the reconstructed states differ by more than 1 standard deviation for ML and even more than 5 standard deviations for LS. The question of how these systematic errors depend on the parameters of the simulation arises. Let us start by investigating the dependence on the number of repetitions N_s. Figure 2(a) shows the mean and the standard deviations of histograms like the one shown in Fig. 1 for different N_s. As expected, the systematic errors are more profound for low numbers of repetitions N_s per setting s and decrease with increasing N_s. Yet, even for $N_s = 500$, a number hardly used in multiqubit experiments, $\bar{F}_{\mathrm{LS}}$ still deviates by 1 standard deviation from the value for the initial state. The effect is also, by no means, special for the GHZ state but was equally observed for other prominent four-party states, here, also, chosen with a fidelity of 80%, see Figs. 2(b)–2(d) and the SM [21].

The systematic deviations also vary with the number of qubits or the purity of the initial state. Figure 3(a) shows the respective dependencies of the fidelity for n-qubit states $\varrho_0 = p|\mathrm{GHZ}_n\rangle\langle\mathrm{GHZ}_n| + (1-p)\mathbb{1}/2^n$ (for $N_s = 100$). Here, a significant increase of the bias with the number of qubits is observed especially for LS. Also, when varying the purity or fidelity with the GHZ state, respectively,

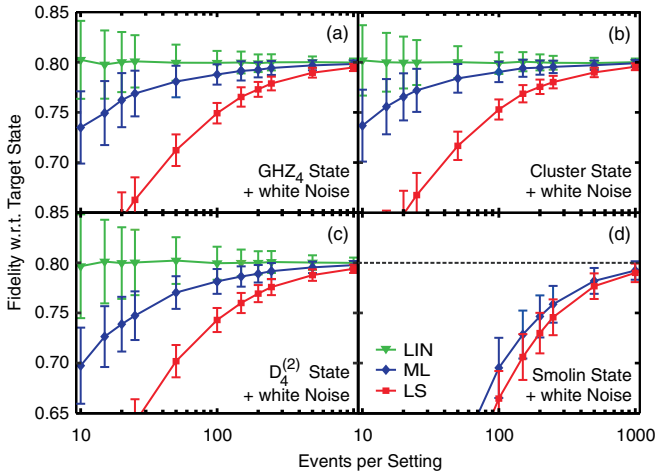

FIG. 2 (color online). The performance of ML, LS, and LIN methods depending on the number of events N_s per setting and for four different noisy initial states ϱ_0. Note that the fidelity can only be calculated linearly if the reference state is pure which is not the case for the Smolin state [25]. Therefore, only the curves for ML and LS are plotted for the Smolin state.

we observe a large deviation for ML and LS estimators [Fig. 3(b)]. If the initial fidelity is very low, the effect is negligible, but large fidelity values suffer from stronger deviations, especially for LS.

The reliability of the estimates $\hat{\varrho}$ or of physical quantities deduced thereof are quantified by the size of confidence regions which commonly are deduced by bootstrapping methods [19]. Starting either from the estimate $\hat{\varrho}_{\mathrm{EST}}$ (EST $\in \{\mathrm{ML}, \mathrm{LS}\}$) or the observed data set f, this error is typically accessed by Monte Carlo sampling: One repeatedly simulates data $f^{(i)}$ according to the state ϱ_{EST} or f together with a reasonable noise model for the respective experiment and reconstructs the state $\hat{\varrho}(f^{(i)})$. From the resulting empirical distribution, one then reports the standard deviation (or a region including, say, 68% of the simulated states) for the matrix elements or for quantities of interest [19], see also SM 3. However, the problem with such error regions is that they are typically too small since they reflect only statistical fluctuations of the biased estimate, which can easily be smaller than the systematic error [26].

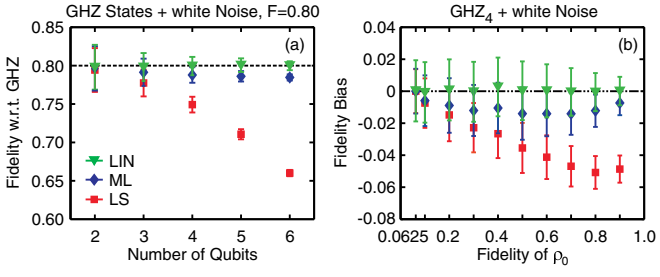

FIG. 3 (color online). The behavior of ML, LS, and LIN depending on the number of qubits n (left) and the fidelity of ϱ_0 (right).

In summary, we observe systematic errors, which depend on the state reconstruction method and the strength of the statistical fluctuations of the count rates. Any manual correction of the bias and the statistical fluctuations [17] seems out of reach, since the effect depends on the unknown initial state which cannot be calculated from the observed data [16]. Let us emphasize that in most cases the initial value differs by more than the statistical error determined via bootstrapping (cf. SM 3 [21]).

Biased and unbiased estimators.—The systematic offset discussed above is well known in the theory of point estimates [26]. Expressed for QST, an estimator $\hat{\varrho}$ is called unbiased if its fluctuations are centered around the true mean, such that, for its expectation value,

$$\mathbb{E}_{\varrho_0}(\hat{\varrho}) \equiv \sum_f P_{\varrho_0}(f)\hat{\varrho}(f) = \varrho_0 \tag{2}$$

holds for all possible states ϱ_0 with $P_{\varrho_0}(f)$ the probability to observe the data f. An estimator that violates Eq. (2) is called biased. Similar definitions hold, for instance, for fidelity estimators, $\mathbb{E}_{\varrho_0}(\hat{F}) = \langle\psi|\varrho_0|\psi\rangle \equiv F_0$. This terminology is motivated by the form of the mean squared error, which decomposes, for example, for the fidelity into

$$\mathbb{E}_{\varrho_0}[(\hat{F} - F_0)^2] = \mathbb{V}_{\varrho_0}(\hat{F}) + [\mathbb{E}_{\varrho_0}(\hat{F}) - F_0]^2, \tag{3}$$

where $\mathbb{V}(\hat{F}) \equiv \mathbb{E}(\hat{F}^2) - \mathbb{E}(\hat{F})^2$ denotes the variance. Equation (3) consists of two conceptually different parts. The first is a statistical term quantifying the fluctuations of the estimator $\hat{F}$ itself. The second, purely systematic term, is called bias and vanishes for unbiased estimators [27]. Note that, since the expectation values of the frequencies are the probabilities, $\mathbb{E}_{\varrho_0}(f_r^s) = P_{\varrho_0}^s(r)$, and because $\hat{\varrho}_{\mathrm{LIN}}$ as given by Eq. (1) is linear in f_r^s, the determination of a quantum state using LIN is unbiased. However, as shown below, for QST, the bias is inherent to estimators constrained to giving only physical answers.

Proposition.—A reconstruction scheme for QST that always yields valid density operators is biased.

Proof.—For a tomography experiment on the state $|\psi_i\rangle$ with finite measurement time, there is a set of possible data $\mathcal{S}_i = \{f_i | P_{|\psi_i\rangle}(f_i) > 0\}$, with $P_{|\psi_i\rangle}(f_i)$ the probability to obtain data f_i when observing state $|\psi_i\rangle$.

Consider two pure nonorthogonal states $|\psi_1\rangle$ and $|\psi_2\rangle$ ($\langle\psi_1|\psi_2\rangle \neq 0$). For these two states, there exists a nonempty set of data $\mathcal{S}_{12} = \{f' | P_{|\psi_1\rangle}(f') \cdot P_{|\psi_2\rangle}(f') > 0\} = \mathcal{S}_1 \cap \mathcal{S}_2$, which can occur for both states.

Now, let us assume that a reconstruction scheme $\hat{\varrho}$ provides a valid quantum state $\hat{\varrho}(f)$ for all possible outcomes f and that Eq. (2) is satisfied for $|\psi_1\rangle$, i.e., $\sum_{\mathcal{S}_1} P_{|\psi_1\rangle}(f_1)\hat{\varrho}(f_1) = |\psi_1\rangle\langle\psi_1|$. This incoherent sum over all $\hat{\varrho}(f_1)$ can be equal to the pure state $|\psi_1\rangle\langle\psi_1|$ only for the

(already pathological) case that $\hat{\varrho}(f_1) = |\psi_1\rangle\langle\psi_1|$ for all $f_1 \in \mathcal{S}_1$. This means that the outcome of the reconstruction is fixed for all f_1 including all data $f' \in \mathcal{S}_{12}$. As these data also occur for state $|\psi_2\rangle$, there exist $f_2 \in \mathcal{S}_{12}$ with $\hat{\varrho}(f_2) = |\psi_1\rangle\langle\psi_1| \neq |\psi_2\rangle\langle\psi_2|$. Thus, in Eq. (2), the sum over all reconstructed states now is an incoherent mixture of at least two pure states, and the condition $\sum_{\mathcal{S}_2} P_{|\psi_2\rangle}(f_2)\hat{\varrho}(f_2) = |\psi_2\rangle\langle\psi_2|$ is violated for $|\psi_2\rangle$. Hence, $\hat{\varrho}$ does not obey Eq. (2) for $|\psi_2\rangle$ and is, therefore, biased [28].

This leaves us with the tradeoff: Should one necessarily use an algorithm like ML or LS to obtain a valid quantum state but suffer from a bias, or should one use LIN which is unbiased but typically delivers an unphysical result? In the following, we propose a scheme using linearized operators to provide a valid, lower or upper bound and an easily computable confidence region for many quantities of interest.

Parameter estimation by linear evaluation.—Many relevant functions are either convex, like most entanglement measures or the quantum Fisher information, or concave, like the von Neumann entropy. Thus, these operators can be linearized around some properly chosen state in order to obtain a reliable lower (upper) bound. Note that, typically, a lower bound on an entanglement measure is often suited for evaluating experimental states, whereas an upper bound does not give much additional information.

Recall that a differentiable function $g(x)$ is convex if $g(x) \geq g(x') + \nabla g(x')^T(x - x')$ holds for all x, x'. In our case, we are interested in a function $g(x) = g[\varrho(x)]$ where x is a variable to parametrize a quantum state ϱ in a linear way. From convexity, it follows that it is possible to find an operator L, such that

$$\mathrm{tr}(\varrho_0 L) \leq g(\varrho_0) \tag{4}$$

holds for all ϱ_0 (similarly an upper bound is obtained for concave functions). This operator L can be determined from the derivatives of $g(x)$ with respect to x at a suitable point x', from the Legendre transformation [29], or directly inferred from the definition of the function $g(x)$ [30]. A detailed discussion is given in SM 5 [21].

For this bound, a confidence region, i.e., the error region for the frequentistic approach, can be calculated. For example, a one-sided confidence region of level γ can be described by a function $\hat{C}$ on the data f such that $\mathrm{Prob}_{\varrho_0}[\hat{C} \leq g(\varrho_0)] \geq \gamma$ holds for all ϱ_0 [26]. According to Hoeffding's tail inequality [31] and a given decomposition of $L = \sum l_r^s M_r^s$ into the measurement operators M_r^s, a confidence region, then, is

$$\hat{C} = \mathrm{tr}(\hat{\varrho}_{\mathrm{LIN}} L) - \sqrt{\frac{h^2 |\log(1-\gamma)|}{2N_s}}, \tag{5}$$

where h^2 is given by $h^2 = \sum_s (l^s_{\max} - l^s_{\min})^2$, and $l^s_{\max/\min}$ denotes the respective extrema of l_r^s over r for each setting s. Although not optimal, such error regions are easy to evaluate and valid without extra assumptions. Since we directly compute a confidence region on $g(x)$, we obtain, generally, a tighter result than what would be obtained from a "smallest" confidence region on density operators which tend to drastically overestimate the error (see SM 4 [21] for an example).

In the following, we show how to use a linearized operator on the example of the bipartite negativity [30]. (For the quantum Fisher information [32] and additional discussion, see SM 5 [21].) A lower bound on the negativity $N(\varrho_{AB})$ of a bipartite state ϱ_{AB} is given by $N(\varrho_{AB}) \geq \mathrm{tr}(\varrho_{AB} L)$ for any L satisfying $\mathbb{1} \geq L^{T_A} \geq 0$, where the superscript T_A denotes partial transposition [33] with respect to party A. This bound is tight if L is the projector on the negative eigenspace of $\varrho_{AB}^{T_A}$. Using this linear expression, one can directly compute the lower bound on the negativity and, by using Eq. (5), the one-sided confidence region. Any choice of L is, in principle, valid, but, for a good performance, L should be chosen according to the experimental situation. We assume, however, no prior knowledge and rather estimate L independent of the tomographic data by the projection on the negative eigenspace of $\hat{\varrho}_{\mathrm{ML}}^{T_A}$ deduced from an additional tomography, again with $N_s = 100$ counts per setting. One can, of course, also start with an educated guess of L motivated by the target state one wants to prepare.

Figure 4 shows the distributions of the negativity between qubits $A = \{1, 2\}$ and $B = \{3, 4\}$ for the four-qubit GHZ state and for the separable four-qubit state $|\psi_{\mathrm{sep}}\rangle \propto (|0\rangle + |+\rangle)^{\otimes 4}$, with $|+\rangle \propto (|0\rangle + |1\rangle)$, each mixed with white noise such that the fidelity with the respective pure state is 80%. In both cases, we observe that ML and LS overestimate the amount of entanglement. Even if no entanglement is present, ML and LS clearly indicate entanglement. In contrast, the lower bound of the negativity does not indicate spurious entanglement. As negativity

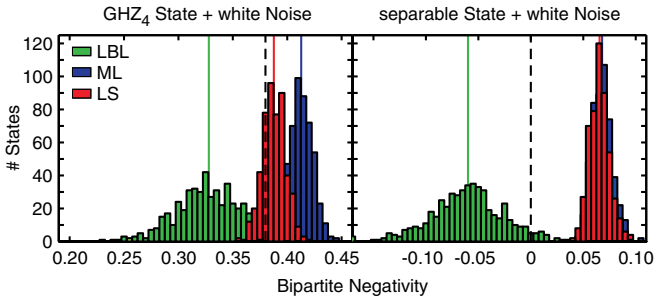

FIG. 4 (color online). Lower bound obtained by linearizing bipartite negativity (LBL) for the GHZ (left) and a four-qubit product state (right) both mixed with white noise resulting in 80% fidelity. The ML and LS reconstruction leads to a systematic overestimation of the negativity, while the lower bound yields a valid estimate.

gives lower bounds of other entanglement measures, those would overestimate entanglement of a state, too [34].

Conclusion.—Any state reconstruction algorithm enforcing physicality of the result suffers from systematic deviations. We have shown that, for the commonly used methods, this bias is significant for data sizes typical in current experiments. Quantities that are computed from such a point estimate can easily inherit this bias and lead to erroneous conclusions, as shown here on the examples of the fidelity, the negativity, and the Fisher information. Equivalent statements can be inferred for process tomography.

Recently, methods have been used to obtain confidence regions via the likelihood function. However, these are notoriously difficult to compute. The linearization method developed here yields a well defined confidence region for interesting quantities. This quantity is easily calculable, yet pessimistic. The quest is, thus, open for finding tighter, but still computationally accessible, confidence regions.

We would like to thank P. Drummond, B. Englert, D. Gross, H. K. Ng, M. Raymer, and J. Shang for stimulating discussions. This work has been supported by the EU (Marie Curie CIG 293993/ENFOQI, ERC Starting Grant GEDENTQOPT, FET Project QWAD, ERC QOLAPS), the BMBF (Chist-Era Project QUASAR), the Brazilian agency CAPES, through the program Science without Borders, and the Elite Network of Bavaria (PhD programs QCCC and ExQM).

[1] *Quantum State Estimation*, edited by M. G. A. Paris and J. Řeháček, (Springer, Berlin, 2004).

[2] H. Häffner, W. Hänsel, C. F. Roos, J. Benhelm, D. Chek-al-kar, M. Chwalla, T. Körber, U. D. Rapol, M. Riebe, P. O. Schmidt *et al.*, Nature (London) **438**, 643 (2005).

[3] J. P. Home, D. Hanneke, J. D. Jost, J. M. Amini, D. Leibfried, and D. J. Wineland, Science **325**, 1227 (2009); J. P. Gaebler, A. M. Meier, T. R. Tan, R. Bowler, Y. Lin, D. Hanneke, J. D. Jost, J. P. Home, E. Knill, D. Leibfried, and D. J. Wineland, Phys. Rev. Lett. **108**, 260503 (2012).

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

[5] K. J. Resch, P. Walther, and A. Zeilinger, Phys. Rev. Lett., **94**, 070402 (2005); N. Kiesel, C. Schmid, G. Tóth, E. Solano, and H. Weinfurter, Phys. Rev. Lett. **98**, 063604 (2007); M. W. Mitchell, C. W. Ellenor, S. Schneider, and A. M. Steinberg, Phys. Rev. Lett. **91**, 120402 (2003).

[6] L. DiCarlo, J. M. Chow, J. M. Gambetta, L. S. Bishop, B. R. Johnson, D. I. Schuster, A. Majer, J. Blais, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, Nature (London) **460**, 240 (2009); M. Neeley, R. C. Bialczak, M. Lenander, E. Lucero, M. Mariantoni, A. D. O'Connell, D. Sank, H. Wang, M. Weides, J. Wenner *et al.*, Nature (London) **467**, 570 (2010); A. Fedorov, L. Steffen, M. Baur, M. da Silva, and A. Wallraff, Nature (London) **481**, 170 (2012); J. M. Chow, J. M. Gambetta, A. D. Córcoles, S. T. Merkel, J. A. Smolin, C. Rigetti, S. Poletto, G. A. Keefe, M. B. Rothwell, J. R. Rozen, M. B. Ketchen, and M. Steffen, Phys. Rev. Lett. **109**, 060501 (2012).

[7] O. Mangold, A. Heidebrecht, and M. Mehring, Phys. Rev. A **70**, 042307 (2004); M. A. Nielsen, E. Knill, and R. Laflamme, Nature (London) **396**, 52 (1998).

[8] I. L. Chuang and M. A. Nielsen, J. Mod. Opt. **44**, 2455 (1997).

[9] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, Phys. Rev. Lett. **105**, 150401 (2010); G. Tóth, W. Wieczorek, D. Gross, R. Krischek, C. Schwemmer, and H. Weinfurter, Phys. Rev. Lett. **105**, 250403 (2010); M. Cramer, M. B. Plenio, S. T. Flammia, D. Gross, S. D. Bartlett, R. Somma, O. Landon-Cardinal, Y.-K. Liu, and D. Poulin, Nat. Commun. **1**, 149 (2010); O. Landon-Cardinal and D. Poulin, New J. Phys. **14**, 085004 (2012).

[10] T. Moroder, P. Hyllus, G. Tóth, C. Schwemmer, A. Niggebaum, S. Gaile, O. Gühne, and H. Weinfurter, New J. Phys. **14**, 105001 (2012).

[11] M. Christandl and R. Renner, Phys. Rev. Lett. **109**, 120403 (2012); T. Sugiyama, P. S. Turner, and M. Murao, Phys. Rev. Lett. **111**, 160406 (2013); R. Blume-Kohout, arXiv:1202.5270; J. Shang, H. K. Ng, A. Sehrawat, X. Li, and B.-G. Englert, New J. Phys. **15**, 123026 (2013).

[12] T. Moroder, M. Kleinmann, P. Schindler, T. Monz, O. Gühne, and R. Blatt, Phys. Rev. Lett. **110**, 180401 (2013); D. Rosset, R. Ferretti-Schöbitz, J.-D. Bancal, N. Gisin, and Y.-C. Liang, Phys. Rev. A **86**, 062325 (2012); C. Stark, Phys. Rev. A **89**, 052109 (2014); N. K. Langford, New J. Phys. **15**, 035003 (2013); S. J. van Enk and R. Blume-Kohout, New J. Phys. **15**, 025024 (2013); S. T. Merkel, J. M. Gambetta, J. A. Smolin, S. Poletto, A. D. Córcoles, B. R. Johnson, C. A. Ryan, and M. Steffen, Phys. Rev. A **87**, 062119 (2013).

[13] D. Smithey, M. Beck, M. G. Raymer, and A. Faridani, Phys. Rev. Lett. **70**, 1244 (1993); G. M. D'Ariano, C. Macchiavello, and N. Sterpi, Quantum Semiclass. Opt. **9**, 929 (1997); U. Leonhardt, M. Munroe, T. Kiss, T. Richter, and M. Raymer, Opt. Commun. **127**, 144 (1996).

[14] R. Blume-Kohout, New J. Phys. **12**, 043034 (2010); Phys. Rev. Lett. **105**, 200504 (2010).

[15] Z. Hradil, Phys. Rev. A **55**, R1561 (1997).

[16] T. Sugiyama, P. S. Turner, and M. Murao, New J. Phys. **14**, 085005 (2012).

[17] C. W. Helstrom, Phys. Lett., **25A**, 101 (1967); S. L. Braunstein and C. M. Caves, Phys. Rev. Lett. **72**, 3439 (1994).

[18] We stress that ML methods can, in principle, be used for directly estimating the fidelity without reconstructing the quantum state, but most experiments using ML first reconstruct the state and then calculate the fidelity. This procedure is the subject of our criticism.

[19] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap* (Chapman & Hall, London, 1994).

[20] J. F. Poyatos, J. I. Cirac, and P. Zoller, Phys. Rev. Lett. **78**, 390 (1997); N. Kiesel, Ph.D. thesis, Ludwig-Maximilians-Universität München, 2007; T. Kiesel, Phys. Rev. A **85**, 052114 (2012).

[21] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.114.080403 for details on LIN, additional examples of states with a strong bias, examples for bootstrapping, the difference between multi-parameter and scalar confidence regions, and further details

on the proposed linearization method for reliable confidence regions, which includes Ref. [22].

[22] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis* (Cambridge University Press, Cambridge, England, 1991).

[23] Clearly, LIN is only reasonable if tomographically complete information about the state is available.

[24] S. Boyd and S. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, England, 2004).

[25] J. A. Smolin, Phys. Rev. A **63**, 032306 (2001).

[26] A. F. Mood, *Introduction to the Theory of Statistics* (McGraw-Hill Inc., New York, 1974).

[27] It is well known that estimators can be biased for a finite number of samples or for particular additional conditions and constraints. An example is the estimate of the variance σ^2 of a Gaussian distribution with known mean μ. The ML estimate from N samples x_i gives $\hat{\sigma}^2_{\mathrm{ML}}(x_i) = 1/N\sum_i (x_i - \mu)^2$, while an unbiased estimator is $\hat{\sigma}^2(x_i) = (1/N-1)\sum_i (x_i - \mu)^2$.

[28] This proof, given here for pure states, can be generalized for mixed states ϱ_1 and ϱ_2 with the ranges of the two states not sharing any vector.

[29] O. Gühne, M. Reimpell, and R. F. Werner, Phys. Rev. Lett. **98**, 110502 (2007); J. Eisert, F. G. S. L. Brandao, and K. M. R. Audenaert, New J. Phys. **9**, 46 (2007).

[30] G. Vidal and R. F. Werner, Phys. Rev. A **65**, 032314 (2002).

[31] M. Tomamichel, C. C. W. Lim, N. Gisin, and R. Renner, Nat. Commun. **3**, 634 (2012); S. T. Flammia and Y.-K. Liu, Phys. Rev. Lett. **106**, 230501 (2011); M. P. da Silva, O. Landon-Cardinal, and D. Poulin, Phys. Rev. Lett. **107**, 210404 (2011).

[32] D. Petz and C. Ghinea, QP-PQ: Quantum Probab. White Noise Anal. **27**, 261 (2011).

[33] A. Peres, Phys. Rev. Lett. **77**, 1413 (1996).

[34] K. Chen, S. Albeverio, and S.-M. Fei, Phys. Rev. Lett. **95**, 040504 (2005); **95**, 210501 (2005).