Confidence Polytopes in Quantum State Tomography

Jinzhao Wang, Volkher B. Scholz, and Renato Renner
Institute for Theoretical Physics, ETH Zurich, 8093 Zürich, Switzerland

(Received 29 October 2018; published 15 May 2019)

Quantum state tomography is the task of inferring the state of a quantum system from measurement data. A reliable tomography scheme should not only report an estimate for that state, but also well-justified error bars. These may be specified in terms of confidence regions, i.e., subsets of the state space which contain the system's state with high probability. Here, building upon a quantum generalization of Clopper-Pearson confidence intervals—a notion known from classical statistics—we present a simple and reliable scheme for generating confidence regions. These have the shape of a polytope and can be computed efficiently. We provide several examples to demonstrate the practical usability of the scheme in experiments.

DOI: 10.1103/PhysRevLett.122.190401

Quantum state tomography (QST) may be regarded as the quantum variant of statistical estimation theory. Given data obtained from measuring a quantum system, the goal is to estimate the system's state. QST has become an increasingly important tool in experimental physics, especially in the area of quantum information technology. Accordingly, a lot of work has been put into the development of techniques to increase its efficiency. Among them are methods to reduce the number of different measurements needed and to keep the (generally unfavorable) scaling of the amount of required data in the dimension of the system under control [1–9].

Nonetheless, only relatively little attention has been paid to the problem of statistical errors in QST. Statistical errors are due to unavoidable fluctuations, resulting from the fact that the collected data always represent a finite sample. In other words, they are those errors that remain even if the experiment is implemented perfectly and shielded from any environmental noise.

In experimental sciences, statistical errors are generally reported in terms of error bars, which are obtained by standard methods from classical statistics. In the context of quantum information, techniques to determine error estimates have been developed for specific tasks, such as entanglement verification and quantum metrology [10–16]. These are however not universal enough to be applicable to QST. In fact, an agreed-upon scheme for reporting the accuracy of estimates in QST does not seem to exist. Experimental results in QST are therefore often stated without error bars, or with error bars that do not have a well-defined operational meaning. A widespread approach is to use point estimators for the system's state, such as maximum likelihood estimation (MLE) [17,18] (for examples, see Refs. [19–22]), and take the width of the likelihood function as a measure for their accuracy [23]. Another common, heuristic, method to determine the accuracy is numerical bootstrapping, or resampling [24,25]. The resulting error bars then correspond to the variance of the point estimators. But since these are generally highly biased, they do not correctly reflect the uncertainty in the state estimate (see Ref. [26] for a discussion). A notable exception is Ref. [27], where a point estimator has been proposed whose distance to the true state is provably below a given bound with high probability.

The problems described above can be avoided with methods that, rather than giving point estimates, yield regions in state space. The idea is that these regions contain, with high probability, the (unknown) state ρ, i.e., the state in which the system was prepared. Depending on what is meant by "high probability," one talks about credibility regions or confidence regions.

Credibility regions are motivated by the Bayesian approach to probability theory, where probabilities are interpreted as measures for personal belief or knowledge [28–31]. To use this approach in QST, it is necessary to specify a prior, i.e., a probability distribution over the possible states ρ, that reflects one's personal belief before considering the measurement data. The corresponding credibility region obtained from QST then has the property that it contains ρ with high probability according to the posterior belief, i.e., the updated belief one would have after taking into account the measurement data. The reported credibility region thus has a well-defined operational meaning—but only for those who agree with the prior. Unfortunately, there is no unique natural choice for the latter; even when demanding certain symmetries, the class of possible priors is usually infinitely large.

Confidence regions avoid this prior dependence. While they are generally larger than the credibility regions of the Bayesian approach, they contain the unknown state ρ with high probability—independently of what the prior was. Currently, there exist two approaches to obtain confidence regions. One of them, due to Blume-Kohout [26] and

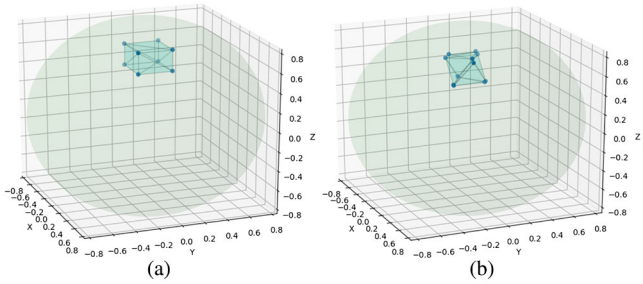

FIG. 1. Confidence polytopes for QST on a qubit. The plots show confidence polytopes with confidence level 0.999, obtained from data of simulated measurements on a single qubit. The polytopes lie within the Bloch sphere, which represents the entire state space of the qubit. In (a), the measurement is defined by six Pauli projectors, along the X, Y, and Z directions, and the polytope is a rectangular box whose normal directions are given by the three Pauli operators. In (b), the measurement operators are chosen such that they form a symmetrical informationally complete (SIC) POVM [36,37]. In this case, the resulting confidence polytope is the intersection of two tetrahedrons whose normals are given by the measurement directions.

Glancy *et al.* [32], uses a construction based on the computation of likelihood ratios. Although supported by heuristic arguments, it has, however, to the best of our knowledge, not been established rigorously that the constructed regions are valid confidence regions. In the other approach, due to Christandl and Renner [33] and Faist and Renner [34], confidence regions are constructed by extending credibility regions for a particular symmetric prior. While the validity of these regions has been proved rigorously, their size is far from optimal (see the discussion below).

In this Letter, we propose an alternative method to determine confidence regions in QST. It is based upon a generalization of a notion from classical statistics, known as Clopper-Pearson confidence intervals [35]. Given data from any informationally complete measurement, the corresponding confidence regions have the shape of a polytope (see Fig. 1), with facets that can be computed efficiently. As we shall demonstrate, this simple structure can also be exploited to optimize the choice of tomographic measurements for more accuracy.

In classical estimation theory, one of the most basic problems is to determine the bias P of a biased coin from a given sample of tosses. The Clopper-Pearson interval solves this problem "exactly," i.e., without involving approximations. In particular, the interval represents a reliable confidence region for P, even in extreme cases, e.g., when $P \approx 0$ or $P \approx 1$, in which other schemes may fail. This feature turns out to be crucial for QST, where the measurement statistics often contains such extreme cases, especially when the unknown state is close to the boundary of the state space.

In QST, one usually considers the following scenario (see Ref. [33] for a more general treatment, which does not assume identically repeated preparations). A d-dimensional quantum system is repeatedly prepared in the same unknown state ρ, i.e., an element of the set $\mathcal{S}(\mathcal{H}_d)$ of density operators on a d-dimensional Hilbert space. After each preparation, a measurement, described by a positive operator valued measure (POVM) on $\mathcal{H}_d$ with elements E_i, for $i = 1, \ldots, k$, is applied. In the common case of projective measurements, these elements are just the projectors belonging to the different possible measurement outcomes i. The results presented here are, however, valid for arbitrary (not necessarily projective) POVM elements E_i, which is useful, for instance, to model noisy measurements [37]. After n preparation-and-measurement rounds, the data can be brought into the form of a k-tuple $\mathbf{n} \coloneqq (n_1, \ldots, n_k)$, where n_i denotes the number of occurrences of an outcome corresponding to E_i.

We are interested in constructing a QST procedure that computes, from the measurement outcome $\mathbf{n}$, a confidence region, denoted by $\Gamma(\mathbf{n})$, for any desired confidence level $1 - \varepsilon$, where $\varepsilon > 0$. This means that, except with probability ε, the unknown state ρ is contained in $\Gamma(\mathbf{n})$; i.e.,

$$\Pr[\rho \in \Gamma(\mathbf{n})] \geq 1 - \varepsilon.$$

Crucially, this bound is supposed to hold for any arbitrary ρ (which would not be the case for a credibility region, where ρ must be sampled from a given prior), whereas $\Pr[\cdot]$ should be understood as the probability taken over all possible outcomes $\mathbf{n}$ [40].

We break down the problem of constructing confidence regions into determining a confidence half-space for each POVM element E_i, depending on the measured frequency n_i/n of the corresponding outcome. The intersection of all such half-spaces, for all the measurement POVM elements, then forms a confidence region, as asserted by the following theorem. For its formulation, we use the binary relative entropy, which is defined as $D(x\|y) = x\log(x/y) + (1-x)\log(1-x/1-y)$. We also introduce a function $\delta_n(m, \nu)$ that returns the positive root δ of $D[(m/n)\|(m/n) + \delta] = -(1/n)\log(\nu)$.

Theorem 1.—Consider a QST setup as described above, with unknown state $\rho \in \mathcal{S}(\mathcal{H}_d)$ and measurements defined by a k-elements POVM $\{E_i\}$. Let $1 - \varepsilon$ be the desired confidence level. For any possible measurement data $\mathbf{n} = (n_1, \ldots, n_k)$ and for any i, define

$$\Gamma_i(n_i) = \left\{\sigma \in \mathcal{S}(\mathcal{H}_d) : \mathrm{tr}(E_i\sigma) \leq \frac{n_i}{n} + \delta_n\left(n_i, \frac{\varepsilon}{k}\right)\right\}.$$

Then their intersection $\Gamma(\mathbf{n}) \coloneqq \bigcap_i \Gamma_i(n_i)$ is a confidence region with confidence level $1 - \varepsilon$.

Proof.—The proof consists of two steps. We first show that the conditions for the confidence region can be reduced to the Clopper-Pearson construction. We then

use known results about this construction to conclude the argument [39]. ∎

Given a family $\boldsymbol{\lambda} = \{\lambda_\alpha\}$ of $d^2 - 1$ generalized Pauli matrices satisfying the orthogonality relation $\mathrm{tr}\lambda_\alpha\lambda_\beta = 2\delta_{\alpha\beta}$, we can embed the space $\mathcal{S}(\mathcal{H}_d)$ of density operators ρ into the Euclidean space $\mathbb{R}^{d^2-1}$ of vectors $\mathbf{r}$ via the relation [41]

$$\rho = \frac{1}{d}\left(\mathbf{1} + \sqrt{\frac{d(d-1)}{2}}\mathbf{r}\cdot\boldsymbol{\lambda}\right).$$

The Euclidean metric on $\mathbb{R}^{d^2-1}$ then corresponds to the Hilbert-Schmidt metric for $\mathcal{S}(\mathcal{H}_d)$ [42]. Similarly, we can represent each POVM element E_i by a vector $\boldsymbol{\eta}_i$ in $\mathbb{R}^{d^2-1}$; i.e.,

$$E_i = \frac{1}{m_i}\left(\mathbf{1} + \sqrt{\frac{d(d-1)}{2}}\boldsymbol{\eta}_i\cdot\boldsymbol{\lambda}\right),$$

with m_i such that $\sum_i (1/m_i) = 1$. Theorem 1 may now be rephrased in terms of these representations in $\mathbb{R}^{d^2-1}$.

Corollary 1.—Consider a QST setup as in Theorem 1, with an unknown state ρ parametrized by $\mathbf{r} \in \mathbb{R}^{d^2-1}$ and POVM $\{E_i\}$ parametrized by $\boldsymbol{\eta}_i \in \mathbb{R}^{d^2-1}$. Then the intersection of the embedding of the state space $\mathcal{S}(\mathcal{H}_d)$ in $\mathbb{R}^{d^2-1}$ with the half-spaces of all $\mathbf{r}$ that satisfy

$$1 + (d-1)\mathbf{r}\cdot\boldsymbol{\eta}_i \le m_i\left[\frac{n_i}{n} + \delta_n\left(n_i, \frac{\varepsilon}{k}\right)\right] \tag{1}$$

represents a confidence region with confidence level $1 - \varepsilon$.

The results above may be generalized by replacing the argument ϵ/k to δ_n by any partition ε_i of unity [39]. One may then optimize ε_i for the tightest confidence region. However, even without this optimization (which may be hard to carry out), the confidence region is rather tight, as discussed below.

If the measurement POVM is informationally complete, the inequalities Eq. (1) define the facets of a polytope, the confidence polytope. From this polytope one can estimate any desired figure of merit (such as the fidelity to a reference state; see Table I for other examples) and obtain error bars for it. The latter are given in terms of $(1 - \varepsilon)$ confidence intervals, which one can obtain via the following procedure.

(i) Choose a convenient basis $\boldsymbol{\lambda}$, e.g., the basis corresponding to the measurement axis, and compute the representation $\boldsymbol{\eta}_i$, m_i for all k elements E_i of the POVM.

(ii) Compute $\delta_n[n_i, (\varepsilon/k)]$ for any of the k measurement outcomes i and collect the corresponding inequalities Eq. (1).

(iii) Sample states from the polytope defined by the k inequalities and compute the figure of merit for each of them [37].

TABLE I. QST of simulated noisy Bell state. A confidence polytope with confidence level 0.999 was generated for data from simulated SIC POVM measurements on 10^4 copies of a noisy Bell state $\rho = 0.9\Phi^+ + 0.1\frac{1}{4}$. The confidence intervals are shown, which are derived from the confidence polytope, for various figures of merit, such as the fidelity to and the distance from particular reference states (MLE denotes the state obtained by maximum likelihood estimation), or the negativity, which is a measure for entanglement [43].

Reference	Fidelity	Trace distance	Negativity
MLE state	>0.973	<0.0902	
Perfect Bell state Φ^+	(0.944, 0.980)	(0.0546, 0.133)	(0.393, 0.459)

(iv) The confidence interval is approximated by the maximum and minimum of the figure of merit among all sampled states.

For illustration we provide examples of simple QST scenarios. The first is QST on a single qubit, where the state space is three dimensional, so that the confidence polytopes can be depicted easily (Fig. 1). We also demonstrate QST on a noisy Bell state with simulated measurement data (Table I) and on s-qubit GHZ states [45] for $s = 2, 3, 4$ with data from IBM's Q Experience [37,44] (Table II). For all our examples we chose a confidence level defined by $\varepsilon = 0.001$.

The shape of the confidence polytopes provides information about the distribution of the statistical errors. This, in turn, enables the choice of particular additional measurements to improve the precision of QST. We demonstrate this with QST on a single qubit. (Its low dimensionality allows us to illustrate the idea by intuitive plots in the Bloch sphere picture, but a generalization to higher-dimensional spaces is straightforward.) We start with a biased informationally complete POVM, which may be regarded as a skewed version of a symmetrical informationally complete (SIC) POVM [see Fig. 2(a)]. The polytope obtained after 5000 measurements is much more extended in the X and the Y direction than in the Z direction

TABLE II. QST of GHZ states on IBM's Q Experience. GHZ states of 2, 3, and 4 qubits were prepared on IBM's 5-qubit device "ibmqx2" and then measured with respect to the Pauli basis on each qubit. The sample size is given by the number of different measurement directions times the shot count (each measurement is repeated 1024 times). The third and fourth columns show the deviation from perfect GHZ states. The confidence level is set at 0.999.

	Data size n	Fidelity	Trace distance	Negativity
GHZ2	9×1024	(0.903, 0.940)	(0.131, 0.208)	(0.318, 0.386)
GHZ3	27×1024	(0.837, 0.869)	(0.313, 0.371)	not applicable
GHZ4	81×1024	(0.944, 0.980)	(0.0546, 0.133)	not applicable

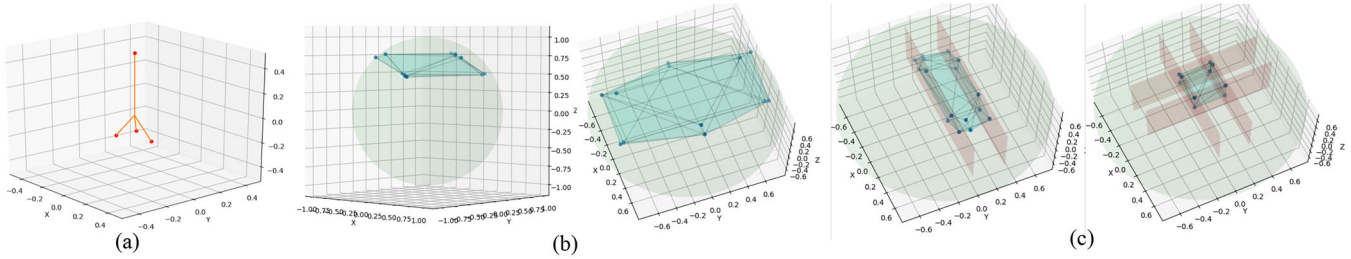

FIG. 2. Optimizing the information content of measurements. The red dots in (a) represent the elements of a skewed SIC POVM, which has more distinguishing power in the Z direction than in the X and Y directions. (b) Confidence polytope that is obtained from 5000 measurements defined by this POVM. Panel (c) depicts the effect of 1000 additional projective measurements in both the X and the Y direction. The red planes represent the new facets introduced by the extra measurements.

[Fig. 2(b)]. Therefore, 1000 extra measurements along both the X and the Y direction help to "refine" the polytope, yielding a smaller confidence region [Fig. 2(c)].

In higher dimensions, extracting the relevant geometrical information can be computationally expensive. One may, however, simplify this task by considering a bounding box in the representation space $\mathbb{R}^{d^2-1}$, defined as the minimum enclosing hyperrectangle with faces perpendicular to the axes given by the basis $\boldsymbol{\lambda}$. Ideally, this basis should be chosen such that it contains experimentally accessible observables (e.g., tensor powers of Pauli matrices). Since the orientation of the bounding box is fixed by the basis, the corners of the box can be determined via simple linear programs. If a particular edge of the bounding box is long, it implies that the confidence polytope is more extended in that direction, and further measurements along the corresponding axis would be effective in reducing its size.

Confidence regions have the advantage over Bayesian credible regions that they do not rely on any prior knowledge. Conversely, credibility regions are generally smaller than confidence regions, thus giving tighter state estimates [47]. Clearly, if the prior is already highly peaked around the actual (unknown) state of the system, the credibility regions obtained by QST can be arbitrarily small. However, numerical results indicate that, in the case of relatively flat priors, the resulting credibility regions are comparable in size to the confidence polytopes introduced here.

Specifically, we take priors defined by the Hilbert-Schmidt measure $d\rho$ [46]. A region $\Gamma(\mathbf{n})$ of the state space has credibility $1-\varepsilon_b$ with respect to this prior if the condition

$$\int_{\Gamma(\mathbf{n})} \mu_{\mathbf{n}}(\rho) d\rho \geq 1-\varepsilon_b \tag{2}$$

holds, where $\mu_{\mathbf{n}}$ is the posterior conditioned on the collected data $\mathbf{n}$ [37]. For our comparison, we take $\Gamma(\mathbf{n})$ to be a $(1-\varepsilon)$ confidence polytope as in Theorem 1 and determine its credibility level ε_b by Eq. (2). We then plot the ratio $\varepsilon/\varepsilon_b$ for randomly chosen states. As shown in Fig. 3, the numerics indicate that this ratio does not scale with the dimension of the measured quantum system nor with the data size. Confidence polytopes therefore provide rather tight estimates for the unknown state. In particular, they outperform the earlier construction proposed by Christandl and Renner [33]. In the latter, $(1-\varepsilon)$ confidence regions are obtained from particular $(1-\varepsilon_b)$ credibility regions, but ε is larger than ε_b by a factor polynomial in the dimension of the measured system [48].

The method we presented here is based on Clopper-Pearson confidence intervals. In classical statistics, there exist several alternative methods to determine confidence intervals, many of which rely however on approximations [48,50–54]. Some of these methods yield confidence intervals that are smaller than Clopper-Pearson intervals and thus seem to have more prediction power [52,54]. Conversely, Clopper-Pearson confidence intervals are a safe choice, in the sense that they never result in an overestimation of the confidence level. Furthermore, for sample sizes n of the order 10^5, which is common in QST, the actual coverage probability of the Clopper-Pearson intervals is very close to the claimed confidence level [52].

Confidence polytopes as proposed here may also be combined with methods for dimension-scalable QST.

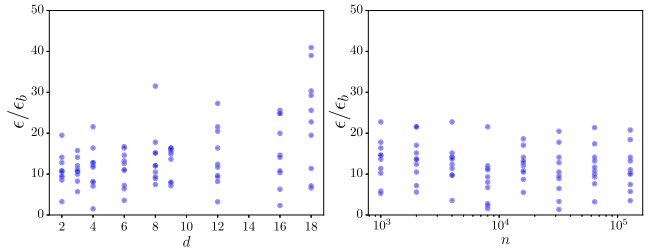

FIG. 3. Confidence versus credibility regions. The plots show the ratio $\varepsilon/\varepsilon_b$ between the confidence and credibility levels of a polytope constructed according to the prescription of Theorem 1, interpreted as a confidence region and as a credibility region, respectively. Each dot was obtained by QST on n copies of a simulated state chosen at random from a d-dimensional state space, evaluated with Ref. [49]. Although there are fluctuations due to the different choices of states and measurement outcomes, no scaling in the data size n or the dimension d is observed.

These are based on additional assumptions about the unknown state, e.g., that it has bounded rank [55], that it is permutation invariant [6], or that it has a matrix product state structure [1]. These assumptions generally restrict the relevant state space. Accordingly, it is sufficient to construct confidence polytopes within this restricted space.

As shown above, rather than reporting the full confidence polytope as the outcome of a QST experiment, it is often sensible to characterize it with one (or a few) parameters. One could treat the state obtained from any point estimation scheme, such as MLE or constrained least square, as a reference and report its maximum distance to the polytope boundary as the error bar. In this sense, our methods, rather than replacing current state estimation schemes, endow them with error bars that characterize the statistical (un)certainty of the estimates.

We thank Philippe Faist for providing the Tomographer software, and Rotem Arnon-Friedman, Kenny Choo, Johannes Heinsoo, Raban Iten, Joseph Renes, Ernest Tan, Yuxiang Yang, and Chi Zhang for useful discussions. This work received support from the Swiss National Science Foundation via the National Center for Competence in Research "QSIT" and from the Air Force Office of Scientific Research (AFOSR) via Grant No. FA9550-16-1-0245.

[1] M. Cramer, M. B. Plenio, S. Flammia, D. Gross, S. Bartlett, R. Somma, O. Landon-Cardinal, Y.-K. Liu, and D. Poulin, Nat. Commun. **1**, 149 (2010).
[2] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, Phys. Rev. Lett. **105**, 150401 (2010).
[3] S. Aaronson, Proc. R. Soc. A **463**, 3089 (2007).
[4] C. Ferrie, Phys. Rev. Lett. **113**, 190404 (2014).
[5] F. Huszár and N. M. T. Houlsby, Phys. Rev. A **85**, 052120 (2012).
[6] G. Tóth, W. Wieczorek, D. Gross, R. Krischek, C. Schwemmer, and H. Weinfurter, Phys. Rev. Lett. **105**, 250403 (2010).
[7] D. Goyeneche, G. Cañas, S. Etcheverry, E. S. Gómez, G. B. Xavier, G. Lima, and A. Delgado, Phys. Rev. Lett. **115**, 090401 (2015).
[8] G. I. Struchalin, I. A. Pogorelov, S. S. Straupe, K. S. Kravtsov, I. V. Radchenko, and S. P. Kulik, Phys. Rev. A **93**, 012103 (2016).
[9] R. J. Chapman, C. Ferrie, and A. Peruzzo, Phys. Rev. Lett. **117**, 040402 (2016).
[10] M. Walter and J. M. Renes, IEEE Trans. Inf. Theory **60**, 8007 (2014).
[11] T. Sugiyama, Phys. Rev. A **91**, 042126 (2015).
[12] M. Baur, A. Fedorov, L. Steffen, S. Filipp, M. P. da Silva, and A. Wallraff, Phys. Rev. Lett. **108**, 040502 (2012).
[13] D. Rosset, R. Ferretti-Schöbitz, J.-D. Bancal, N. Gisin, and Y.-C. Liang, Phys. Rev. A **86**, 062325 (2012).
[14] R. Blume-Kohout, J. O. S. Yin, and S. J. van Enk, Phys. Rev. Lett. **105**, 170501 (2010).
[15] J. Řeháček, D. Mogilevtsev, and Z. Hradil, New J. Phys. **10**, 043022 (2008).
[16] B. Jungnitsch, S. Niekamp, M. Kleinmann, O. Gühne, H. Lu, W.-B. Gao, Y.-A. Chen, Z.-B. Chen, and J.-W. Pan, Phys. Rev. Lett. **104**, 210401 (2010).
[17] Z. Hradil, Phys. Rev. A **55**, R1561 (1997).
[18] Z. Hradil, J. Řeháček, J. Fiurášek, and M. Ježek, in *Quantum State Estimation*, edited by M. Paris and J. Řeháček (Springer-Verlag, Berlin Heidelberg, 2004), p. 59.
[19] C. F. Roos, G. P. T. Lancaster, M. Riebe, H. Häffner, W. Hänsel, S. Gulde, C. Becher, J. Eschner, F. Schmidt-Kaler, and R. Blatt, Phys. Rev. Lett. **92**, 220402 (2004).
[20] K. J. Resch, P. Walther, and A. Zeilinger, Phys. Rev. Lett. **94**, 070402 (2005).
[21] R. Blatt and D. Wineland, Nature (London) **453**, 1008 (2008).
[22] S. Filipp, P. Maurer, P. J. Leek, M. Baur, R. Bianchetti, J. M. Fink, M. Göppl, L. Steffen, J. M. Gambetta, A. Blais *et al.*, Phys. Rev. Lett. **102**, 200402 (2009).
[23] T. Sugiyama, P. S. Turner, and M. Murao, Phys. Rev. A **83**, 012105 (2011).
[24] J. P. Home, D. Hanneke, J. D. Jost, J. M. Amini, D. Leibfried, and D. J. Wineland, Science **325**, 1227 (2009).
[25] R. J. Tibshirani and B. Efron, *An Introduction to the Bootstrap* (CRC Press, New York, 1993), p. 168.
[26] R. Blume-Kohout, arXiv:1202.5270.
[27] T. Sugiyama, P. S. Turner, and M. Murao, Phys. Rev. Lett. **111**, 160406 (2013).
[28] R. Blume-Kohout, New J. Phys. **12**, 043034 (2010).
[29] J. Shang, H. K. Ng, A. Sehrawat, X. Li, and B. G. Englert, New J. Phys. **15**, 123026 (2013).
[30] C. Ferrie, New J. Phys. **16**, 023006 (2014).
[31] C. Granade, J. Combes, and D. G. Cory, New J. Phys. **18**, 033024 (2016).
[32] S. Glancy, E. Knill, and M. Girard, New J. Phys. **14**, 095017 (2012).
[33] M. Christandl and R. Renner, Phys. Rev. Lett. **109**, 120403 (2012).
[34] P. Faist and R. Renner, Phys. Rev. Lett. **117**, 010404 (2016).
[35] C. J. Clopper and E. S. Pearson, Biometrika **26**, 404 (1934).
[36] J. M. Renes, R. Blume-Kohout, A. J. Scott, and C. M. Caves, J. Math. Phys. (N.Y.) **45**, 2171 (2004).
[37] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.122.190401 for measurements modeled by arbitrary POVMs which are discussed in Sec. D, Sec. A for a general formulation and for the proof of Theorem 1, which includes Refs. [38,39], Sec. F for a description of the method we used for the sampling of states from a confidence polytope, Sec. B for the details of IBM's Q Experience demonstration, and also refer Sec. C.
[38] W. Hoeffding, J. Am. Stat. Assoc. **58**, 13 (1963).
[39] H. Chernoff, Ann. Math. Stat. **23**, 493 (1952).
[40] The bound does not necessarily hold conditioned on the event that a particular outcome $\mathbf{n}$ occurred, i.e., the conditional probability that $\rho \in \Gamma(\mathbf{n})$ may be larger or smaller than $1-\varepsilon$; see Ref. [33] for a discussion.
[41] R. A Bertlmann and P. Krammer, J. Phys. A **41**, 235303 (2008).
[42] I. Bengtsson and K. Życzkowski, *Geometry of Quantum States: An Introduction to Quantum Entanglement*

(Cambridge University Press, Cambridge, England, 2017), p. 310.
[43] G. Vidal and R. F. Werner, Phys. Rev. A **65**, 032314 (2002).
[44] IBM Quantum Experience, https://quantumexperience.ng.bluemix.net/qx.
[45] An s-qubit GHZ state is defined as $|\psi\rangle = (|0\rangle^{\otimes s} + |1\rangle^{\otimes s})/\sqrt{2}$.
[46] Given the Haar measure $d\phi$ over the purification space $\mathcal{H} \otimes \mathcal{K}$, with $\mathcal{K}$ isomorphic to $\mathcal{H}$, the Hilbert-Schmidt measure over $\mathcal{S}(\mathcal{H})$, $d\rho$, is induced by tracing out $\mathcal{K}$.
[47] E. T. Jaynes, in *E. T. Jaynes: Papers on Probability, Statistics and Statistical Physics*, edited by R. D. Rosenkrantz, Lecture Notes in Computer Science Vol. 158 (Springer Science, Netherlands, 1989), p. 149.
[48] C. R. Blyth and H. A. Still, J. Am. Stat. Assoc. **78**, 108 (1983).
[49] P. Faist, The Tomographer Project, https://github.com/Tomographer/tomographer.
[50] I. Sakakibara, E. Haramo, A. Muto, I. Miyajima, and Y. Kawasaki, Am. J. Biostat. **4**, 11 (2014).
[51] A. Agresti and B. A. Coull, Am. Stat. **52**, 119 (1998).
[52] M. Thulin, J. Stat. **8**, 817 (2014).
[53] E. B Wilson, J. Am. Stat. Assoc. **22**, 209 (1927).
[54] L. D. Brown, T. T. Cai, and A. DasGupta, Stat. Sci. **16**, 101 (2001).
[55] C. H. Baldwin, I. H. Deutsch, and A. Kalev, Phys. Rev. A **93**, 052105 (2016).