Versatile Fidelity Estimation with Confidence

Akshay Seshadri,[1] Martin Ringbauer,[2] Jacob Spainhour,[3] Rainer Blatt,[2,4,5] Thomas Monz,[2,5] and Stephen Becker[3]

[1]Department of Physics, University of Colorado Boulder, Boulder 0309-0390, USA
[2]Universität Innsbruck, Institut für Experimentalphysik, Technikerstrasse 25, 6020 Innsbruck, Austria
[3]Department of Applied Mathematics, University of Colorado Boulder, Boulder 0309-0526, USA
[4]Institut für Quantenoptik und Quanteninformation, Österreichische Akademie der Wissenschaften, Otto-Hittmair-Platz 1, 6020 Innsbruck, Austria
[5]Alpine Quantum Technologies GmbH, 6020 Innsbruck, Austria

(Received 18 February 2024; accepted 2 May 2024; published 10 July 2024)

As quantum devices become more complex and the requirements on these devices become more demanding, it is crucial to be able to verify the performance of such devices in a scalable and reliable fashion. A cornerstone task in this challenge is quantifying how close an experimentally prepared quantum state is to the desired one. Here we present a method to construct an estimator for the quantum state fidelity that is compatible with any measurement protocol. Our method provides a confidence interval on this estimator that is guaranteed to be nearly minimax optimal for the specified measurement protocol. For a well-chosen measurement scheme, our method is competitive in the number of measurement outcomes required for estimation. We demonstrate our method using simulations and experimental data from a trapped-ion quantum computer and compare the results to state-of-the-art techniques. Our method can be easily extended to estimate the expectation value of any observable, such as entanglement witnesses.

DOI: 10.1103/PhysRevLett.133.020402

With quantum devices starting to push the limit of their classical counterparts, it is becoming increasingly challenging to verify and validate that these devices operate as intended. The gold standard for characterizing quantum states and operations is quantum tomography [1,2]. However, tomography, even its more efficient incarnations [3–7], quickly becomes impractical as the system size grows. Moreover, depending on the kind of reconstruction used, providing reliable uncertainty bounds to the fidelity obtained from a reconstructed quantum state has proven very challenging [8,9].

Fortunately, it is often sufficient to ascertain that the quantum state (or process) in the experiment is close to the desired one, rather than fully reconstructing it. A commonly used measure for this "closeness" is the fidelity between the two states. For certain classes of states, the fidelity can be estimated from only a few Pauli measurements using a technique called direct fidelity estimation (DFE) [10,11]. For a given target state, DFE specifies a measurement scheme that is to be sampled on the experimental state. However, depending on experimental constraints, the prescribed measurement scheme might not always be the one that is easiest to implement or achieves the highest accuracy. Moreover, the question of how to compute tighter confidence regions for such estimates remains.

Here we propose a method that not only provides fidelity estimates for arbitrary measurement schemes but also guarantees a nearly minimax optimal (symmetric) confidence interval for the chosen measurement scheme and target state. This method, which we denote the *minimax* method, is based on results in statistics by Juditsky and Nemirovski, who describe the estimation of linear functionals under general conditions [12–14]. In the minimax method, a linear estimator can be constructed for a specified measurement scheme and target state, which can subsequently provide estimates with guaranteed confidence regions from experimental data. Furthermore, in some settings, the performance of the estimator can be precisely characterized in terms of *sample complexity*, which is the minimum number of measurement outcomes necessary to find an estimate within a desired error and confidence level for the chosen measurement setting. For example, we show in the companion paper [15] that for a confidence level $1-\delta$ our method can produce a fidelity estimate within an additive error, or *risk* (defining a symmetric confidence interval) of ε_*, with as few as $\sim \ln(2/\delta)/(2\varepsilon_*^2)$ measurement outcomes for any target state. This sample complexity is achieved when measuring in a basis defined by the target state. In contrast, tomography with Pauli measurements can require exponentially more measurement outcomes [10]. While measuring in the aforementioned basis is often impractical, our method achieves a comparable sample complexity for stabilizer states, like those used in quantum error correction, by

randomly sampling elements from the stabilizer group. For more general states, the sample complexity is comparable to that of DFE, when using the measurement scheme described in Sec. II.E. of Ref. [15], and thus remains practical for typical states.

In the following, we consider an experimenter aiming to characterize an unknown quantum state $\sigma \in \mathcal{X}$ using L different measurement settings, where $\mathcal{X}$ denotes the set of density matrices. Each measurement setting is described by a positive operator-valued measure (POVM) $\{E_1^{(l)}, \ldots, E_{N_l}^{(l)}\}$ such that $\sum_i E_i^{(l)} = \mathbb{I}$ for $l = 1, \ldots, L$. The lth measurement is repeated R_l times (i.e., experimental shots) and the set of possible outcomes $\Omega^{(l)} = \{1, \ldots, N_l\}$ is labeled by the index of the POVM elements. We denote by $p_\sigma^{(l)}$ the distribution of outcomes from the lth measurement setting, given by the Born rule as $p_\sigma^{(l)} = (\mathrm{Tr}(E_1^{(l)}\sigma), \ldots, \mathrm{Tr}(E_{N_l}^{(l)}\sigma))$ [16]. However, since the state σ is unknown, so are the distributions $p_\sigma^{(l)}$. Hence, following Ref. [12], we start with a family of distributions $p_\chi^{(l)}$ over the outcomes $\Omega^{(l)}$, parametrized by density matrices $\chi \in \mathcal{X}$.

We now seek to construct an estimator $\hat{F}_*$ that takes measurement outcomes for the specified measurement scheme as an input and returns an estimate for the fidelity between the state σ and a pure target state ρ with reliable confidence regions. Formally, we define for a given confidence level $1 - \delta \in (0, 1)$ and estimator $\hat{F}$ a δ risk $\mathcal{R}(\hat{F}; \delta)$ as the smallest symmetric interval around $\hat{F}$ such that the probability that the true value is outside that interval is smaller than δ:

$$\mathcal{R}(\hat{F}; \delta) = \inf\left\{ \varepsilon \middle| \sup_{\chi \in \mathcal{X}} \Pr_{\text{outcomes} \sim p_\chi}[|\hat{F}(\text{outcomes}) - \mathrm{Tr}(\rho\chi)| > \varepsilon] < \delta \right\}. \tag{1}$$

Here, "outcomes $\sim p_\chi$" means that the outcomes for lth measurement are distributed according to $p_\chi^{(l)}$ for a given state χ, for all $l = 1, \ldots, L$. Notably, this risk is a property of the measurement strategy, the target state, the chosen estimator, and the desired confidence interval, but *not* of the actual state that was prepared in the lab. Using this definition, we denote the *minimax optimal* δ risk as $\mathcal{R}_*(\delta) = \inf_{\hat{F}} \mathcal{R}(\hat{F}; \delta)$, where the infimum is taken over all real-valued functions over the set of outcomes.

In practice, considering all possible functions as candidate estimators makes the problem computationally unmanageable. To address this challenge we follow Ref. [12] and consider a restricted set $\mathcal{F}$ consisting of estimators ϕ of the form $\phi = \sum_{l=1}^{L} \phi^{(l)}$, where $\phi^{(l)} \in \mathcal{F}^{(l)}$ takes an outcome of the lth POVM as input and returns a number. Here, $\mathcal{F}^{(l)}$ is a finite-dimensional vector space consisting of all estimators for the lth POVM. The risk achieved when we restrict our estimators to $\mathcal{F}$ is called "affine" risk. As shown in Ref. [12], the restriction to $\mathcal{F}$ is unproblematic, as one can still construct an estimator $\hat{F}_* \in \mathcal{F}$ which achieves a risk ε_* that is almost minimax optimal for $\delta \in (0, 1/4)$ in the sense that

$$\varepsilon_* \leq \vartheta(\delta) \mathcal{R}_*(\delta) \tag{2}$$

$$\vartheta(\delta) = 2 + \frac{\ln(64)}{\ln(0.25/\delta)}. \tag{3}$$

For confidence levels greater than 90%, $\vartheta(\delta)$ is of the order of 1, and decreases as the confidence level increases. This estimator $\hat{F}_*$ is constructed as follows: (1) Find the saddle-point value of the concave-convex function

$$\begin{aligned}\Phi(\chi_1, \chi_2; \phi, \alpha) &= \mathrm{Tr}(\rho\chi_1) - \mathrm{Tr}(\rho\chi_2) + 2\alpha \ln(2/\delta) \\ &\quad + \alpha \sum_{l=1}^{L} R_l \left[\ln\left(\sum_{k=1}^{N_l} \exp\left(-\phi_k^{(l)}/\alpha\right) \mathrm{Tr}\left(E_k^{(l)} \chi_1\right) \right) \right. \\ &\quad \left. + \ln\left(\sum_{k=1}^{N_l} \exp\left(\phi_k^{(l)}/\alpha\right) \mathrm{Tr}\left(E_k^{(l)} \chi_2\right) \right) \right]\end{aligned} \tag{4}$$

to a given precision, by maximizing over the density matrices $\chi_1, \chi_2 \in \mathcal{X}$ and minimizing over the candidate estimator $\phi \in \mathcal{F}$ and the positive number $\alpha > 0$; see Ref. [15] for details. Since each $\Omega^{(l)}$ is a finite set, each $\phi^{(l)} \in \mathcal{F}^{(l)}$ can be represented as an N_l-dimensional real vector with its kth component denoted by $\phi_k^{(l)}$. (2) Denote the saddle-point value of Φ by $2\varepsilon_*$, i.e., $2\varepsilon_* = \inf_{\alpha > 0, \phi \in \mathcal{F}} \max_{\chi_1, \chi_2 \in \mathcal{X}} \Phi(\chi_1, \chi_2; \phi, \alpha)$. Suppose that the saddle-point value is attained at $\chi_1^*, \chi_2^* \in \mathcal{X}$, $\phi_* \in \mathcal{F}$, and $\alpha_* > 0$ to within the given precision. Recall that by definition of $\mathcal{F}$, we have $\phi_* = \sum_{l=1}^{L} \phi_*^{(l)}$ with $\phi_*^{(l)} \in \mathcal{F}^{(l)}$. (3) Let $\{o_1^{(l)}, \ldots, o_{R_l}^{(l)}\}$ denote R_l independent and identically distributed outcomes obtained for the lth measurement setting. Then the estimator $\hat{F}_* \in \mathcal{F}$ is given as

$$\hat{F}_*(\text{outcomes}) = \sum_{l=1}^{L} \sum_{k=1}^{R_l} \phi_*^{(l)}(o_k^{(l)}) + c \tag{5}$$

with the constant $c = \frac{1}{2}(\mathrm{Tr}(\rho\chi_1^*) + \mathrm{Tr}(\rho\chi_2^*))$. The risk associated with this estimator satisfies $\mathcal{R}(\hat{F}_*; \delta) \leq \varepsilon_*$, so the final output will be $\hat{F}_*(\text{outcomes}) \pm \varepsilon_*$. An intuitive way to understand the estimator of Eq. (5) is as a simple weighting of the relative frequencies observed in the experiments (see Ref. [15] for details). If we denote $\boldsymbol{a}^{(l)} = (\phi_*^{(l)}(1) \ldots \phi_*^{(l)}(N_l))$ as a "coefficient vector"

obtained using the saddle point, and $\boldsymbol{f}^{(l)}$ as the vector with kth component representing the relative frequency of the outcome k observed in the experiments for the lth measurement setting ($l = 1, \ldots, L$), then the estimator can be written as

$$\hat{F}_*(\boldsymbol{f}^{(1)}, \ldots, \boldsymbol{f}^{(L)}) = \sum_{l=1}^{L} R_l \left\langle \boldsymbol{a}^{(l)}, \boldsymbol{f}^{(l)} \right\rangle + c. \tag{6}$$

The technical details associated with the method as well as details on how we perform the optimization are given in the accompanying Ref. [15]. Importantly, since this method can be used to estimate any linear functional, we can obtain an estimator for the expectation value of any observable $\mathcal{O}$ by replacing ρ with $\mathcal{O}$ in the above equations. Moreover, any so constructed estimator is experimentally robust in the sense that it does not amplify noise, as we show in the accompanying Ref. [15].

Examples.—The above definitions of the estimator and risk are very abstract. To get a better feeling for these quantities, we now consider two thought experiments. First, consider a single qubit with target state $\rho = |1\rangle\langle 1|$, where $\{|0\rangle, |1\rangle\}$ are the eigenstates of the Pauli-Z operator. We simulate 100 repetitions of a Z measurement, with outcomes denoted $o_i \in \{0, 1\}$. For a desired confidence level of 95%, the above method then gives the following estimator (see Sec. II.B of Ref. [15] for details)

$$\hat{F}_*(\{o_i\}_{i=1}^{100}) = \frac{0.952}{100} \sum_{i=1}^{100} o_i + 0.024. \tag{7}$$

Hence, $\hat{F}_*$ turns out to be nothing other than the sample-mean estimator for the Bernoulli parameter, albeit with a small bias to account for finite sample size. Indeed, the problem we have considered is essentially classical, equivalent to estimating the probabilities in a coin flip. As we increase the number of repetitions R, we find that the bias goes to zero, while the coefficient multiplying the outcomes approaches $1/R$ as one would expect.

Next, we consider a random 4-qubit target state and simulated laboratory state that is obtained by applying 10% depolarizing noise to the target state. We simulate 100 repetitions of the first $L = 192 = 0.75 \times 4^4$ Pauli measurements (excluding identity) used in DFE [10], chosen in decreasing order of weights (where 0.75 is an arbitrary cutoff). Figure 1 shows a histogram of the estimates given by the estimator constructed in Eq. (5), obtained using 10^4 sets of simulated outcomes. In order to assess the risk associated with this estimator, we first compute the smallest asymmetric interval around the true fidelity that captures 95% of the estimates [17]. We find that the confidence interval obtained from the minimax method is only 1.74 times larger than this optimal asymmetric interval [18]: a small price to pay for a rigorous confidence interval. Note,

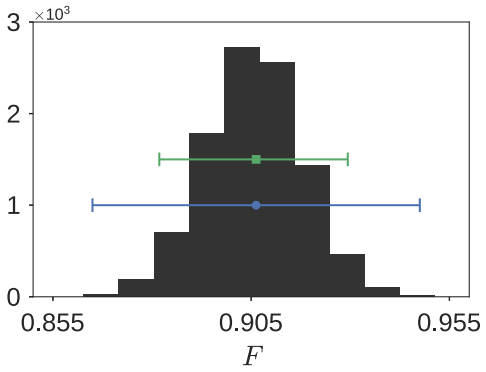

FIG. 1. Histogram of 10^4 estimates obtained using the estimator constructed in Eq. (5) for a random 4-qubit target state using outcomes from three-fourths the total number of Pauli measurements. The true fidelity is marked by a green square and a blue dot. The green bar passing through the square indicates the estimated optimal (asymmetric) confidence interval corresponding to a confidence level of 95%, while the blue bar passing through the circle indicates the risk associated with the estimator of Eq. (5).

that the asymmetric interval is always expected to be tighter than the optimal minimax interval because we use the true fidelity to compute it, which would be inaccessible in practice. Nevertheless, we find that the above ratio is much better than the guaranteed upper bound of $\vartheta(0.05) = 4.58$ in Eq. (3).

Experimental data.—We now proceed to testing the estimator of Eq. (5) on experimental data from a trapped-ion quantum processor [19]. We consider data for three different 4-qubit target states: a GHZ state, a W state, and a locally rotated linear cluster state. Each dataset consists of 81 Pauli measurements with 100 shots each (i.e., $R_l = 100$ for all l). For comparison, we also perform state-of-the-art maximum likelihood estimation (MLE) [2,20]. Table I shows that both methods produce almost identical estimates, yet the risk for the minimax method is larger than the corresponding risk for the MLE estimate, which is *estimated* using Monte Carlo (MC) resampling of the data. Although smaller, the MC risk for the MLE estimator is

TABLE I. Fidelity estimates and risk for a 4-qubit GHZ state, W state, and a cluster state from experimental data. Estimates are calculated using the minimax method as well as MLE. The risk for MLE is obtained from Monte Carlo sampling, which is a heuristic and not guaranteed to be correct. The risk for either method corresponds to a 95% confidence level.

	Minimax method		MLE	
	F Estimate	Risk	F Estimate	MC risk
GHZ	0.84	0.053	0.84	0.023
W	0.89	0.049	0.88	0.019
Cluster	0.79	0.048	0.79	0.020

not computed from raw data, but requires reconstructing the state which is prone to errors and bias [21–25]. Furthermore, in Ref. [15], we give an example where the MC-confidence region for MLE is overconfident when provided with incomplete data. In contrast, the more conservative minimax risk, although not as tight, *is guaranteed to be correct*, irrespective of the underlying actual state and the performed measurements. As this is experimental data, we do not know whether the MC risk estimate is correct.

Stabilizer states.—The experimental data in the previous example used the same set of Pauli measurements for all three target states. However, it is typically desirable to implement measurement schemes that are tailored to a certain target state. A prime example are stabilizer states, which are commonly used in quantum error correction [26,27]. An n-qubit stabilizer state is fully characterized by n stabilizer generators. However, measuring any subset of $n-1$ stabilizer generators is insufficient to determine the state, as that information is still compatible with two orthogonal candidate states (and superpositions thereof). In contrast to MC methods, our method automatically incorporates this fact by returning a risk of 0.5 when using such a measurement scheme (see Ref. [15]). On the other hand, it is known that measuring the elements of the stabilizer group (excluding the identity) uniformly at random is a minimax optimal measurement strategy (when only Pauli measurements are allowed) [28,29]. We show in Ref. [15] that, with this measurement strategy, our method gives a sample complexity better than or equal to $\approx 2[(d-1)/d]^2[\ln(2/\delta)/\varepsilon_*^2]$, where $d=2^n$ is the dimension of the system. In other words, for any given stabilizer state, one needs to measure at most a constant number of Pauli operators to estimate the fidelity with a risk of ε_* and a confidence level of $1-\delta$. This sample complexity rivals that of DFE, showing that our method is highly practical when the measurement scheme is suitably chosen for the target state.

Comparison to direct fidelity estimation.—The above measurement scheme can be generalized to arbitrary pure states, resulting in a similar sample complexity and computational speed as DFE. However, under equal conditions with the same measurements and sampling, we find that the our method typically achieves significantly tighter confidence regions than DFE. One reason for this is that DFE uses only the observed eigenvalues (± 1). In contrast, the minimax method can utilize the additional information about which eigenvectors of the multiqubit Pauli measurements were observed [15]. This information is typically available in the experiments.

Moreover, the minimax method is not constrained to using Pauli measurements like DFE specifies. In fact, we show in Ref. [15] that for random target states we can achieve very similar performance regardless of whether we use the DFE measurements or just random POVMs for the same number of shots. This opens the interesting possibility to design measurement schemes that are tailored to a certain target state in the same way that Pauli measurements are tailored to stabilizer states. Depending on the experimental constraints and the complexity of the target state, this can allow for much more efficient estimation.

We note that our method also has similarities with quantum state verification (QSV) [28] and quantum state certification [30], which tests whether the fidelity is above a fixed threshold. We show that using the measurement protocol introduced in QSV, we can optimally estimate the fidelity with any 2-qubit state. We also compare with classical shadows [31], and show that our method can match their sample complexity in principle. While computing our estimator for global Clifford measurements can be computationally challenging, we show that we can get an exponential advantage over classical shadows using appropriate Pauli measurements. Details of these comparisons are included in our companion paper [15].

Entanglement witnesses.—The minimax method introduced above is not limited to estimating fidelity, but can be applied to arbitrary observables. For example, it is often desirable to estimate the expectation value of an entanglement witness to determine if a (pure or mixed) state is entangled. Consider a Werner state $\rho_{\text{Werner}} = p[(\mathbb{I}+\text{SWAP})/d(d+1)] + (1-p)[(\mathbb{I}-\text{SWAP})/d(d-1)]$, which is a bipartite state, characterized by a parameter $p \in [0,1]$ and is entangled for $p < 0.5$ [32]. The SWAP operator is an entanglement witness for Werner states [32], but often not easy to measure in practice. For the sake of demonstration, we consider a 2-qubit Werner state with $p = 0.25$ as the target state, and apply 10% depolarizing noise to obtain the actual state, resulting in a true expectation value of $\text{Tr}(\text{SWAP}\cdot\sigma) = -0.4$. We perform 3 Pauli measurements, with 100 repetitions each, and obtain an estimate of -0.28 with a risk of 0.23. Note that while this risk is quite large, it is guaranteed to be correct. Since $-0.28 + 0.23 < 0$, we can certify with 95% confidence that the state is entangled.

Computational efficiency.—One of the main challenges of the minimax method is computing the estimator efficiently in large dimensions for arbitrary target states and POVMs. Because we need to perform optimization in such situations, quantifying the computational resources required becomes important. We summarize the time and memory consumed to build the estimator for a random target state and a GHZ state in Table II. Changing the number of repetitions results in negligible overhead. Although the computational time grows exponentially with qubit number (as does the size of the system and number of measurements), structured states give significant improvements. Note that the estimator needs to be constructed only once for a specified measurement scheme and target state. Once the estimator is built, we can give fidelity estimates from raw data almost instantaneously. This is in contrast to

TABLE II. Time taken to construct the estimator for a random n-qubit target state and an n-qubit GHZ state with each of our three implementations (average of 3 simulations). We use $L = 0.75 \times 4^n$ Pauli measurements for the random state, while $L = 2^n$ Pauli measurements for the GHZ state. The computations were performed on a 3.0 GHz CPU without parallelization. Entries marked with * require > 1 GB of memory, entries marked with ** require > 32 GB, and the `cvxpy` algorithm run on a random 5-qubit state requires > 256 GB.

State	n	1	2	3	4	5
Random	L	3	12	48	192	768
	Nesterov	49.8 s	1.12 min	5.16 min	27.7 min	2.86 hr*
	`cvxpy`	0.088 s	0.49 s	4.53 s	1.05 min*	⋯
	`cvxpy-mem`	0.081 s	0.69 s	8.37 s	1.85 min*	39.8 min**
GHZ	L	2	4	8	16	32
	Nesterov	2.77 s	31.6 s	3.21 min	8.75 min	13.9 min
	`cvxpy`	0.065 s	0.17 s	0.68 s	4.05 s	30.6 s*
	`cvxpy-mem`	0.063 s	0.23 s	1.27 s	8.86 s	1.30 min*

methods such as MLE, which incur high computational cost for every evaluation on the data.

We note that it is possible to simplify the optimization problem depending on the target state and the measurement scheme. For example, we reduce the problem of constructing an estimator for stabilizer states (when measuring random elements of the stabilizer group) to a two-dimensional optimization problem, irrespective of the stabilizer state or the dimension. The resulting algorithm is both time and memory efficient for any number of qubits. The same efficient algorithm can be used for an arbitrary target state with a randomized Pauli measurement scheme that we introduce in Sec. II.E of Ref. [15]. However, as in DFE, computing the weights of the different measurements can be cumbersome for complicated states.

Outlook.—We have introduced a method that provides nearly minimax-optimal risk for any specified measurement scheme and target state. The sample complexity of our method is comparable to DFE, when using the same kind of Pauli measurements. However, in contrast to DFE and other methods, we are not imposing any restrictions on the measurement scheme. Hence, depending on the experimental constraints, tailored measurement schemes can be much more efficient than rigid Pauli schemes. In particular, it would be interesting to study the performance of our method for random measurements used in Ref. [31].

The risk is calculated based solely on the measurement scheme, target state, and number of repetitions, so we can use our method as a tool to benchmark experimental protocols. By providing a guarantee on the achievable error for a given protocol, this method can be useful for guiding the experimental design and finding the best protocol for a given target state. Another important avenue for future research is extending the method to estimating nonlinear functions of the state like purity or entanglement measures like negativity. Furthermore, by exploiting the Choi-Jamiołkowski isomorphism, our method can be extended to quantum channels. Since quantum process tomography is even more demanding than state tomography, direct estimation of channel fidelities or other channel properties is extremely useful for characterizing multiqubit channels.

All PYTHON code required to use our method is available at [33].

The authors thank Emanuel Knill, Scott Glancy, and Yanbao Zhang for helpful discussions on the manuscript. This material is based upon work supported by the National Science Foundation under Grants No. 1819251 and No. 2112901. This work utilized the Summit supercomputer, which is supported by the National Science Foundation (Grants No. ACI-1532235 and No. ACI-1532236), the University of Colorado Boulder, and Colorado State University. The Summit supercomputer is a joint effort of the University of Colorado Boulder and Colorado State University. We gratefully acknowledge support by the Austrian Science Fund (FWF Grant-DOI 10.55776/F71) (SFB BeyondC) and the Institut für Quanteninformation GmbH. We also acknowledge funding from the EU H2020-FETFLAG-2018-03 under Grant Agreement No. 820495, by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via U.S. Army Research Office (ARO) Grant No. W911NF-16-1-0070 and W911NF-21-1-0007, and the U.S. Air Force Office of Scientific Research (AFOSR) via IOE Grant No. FA9550-19-1-7044 LASCEM. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 840450.

[1] K. Vogel and H. Risken, Determination of quasiprobability distributions in terms of probability distributions for the rotated quadrature phase, Phys. Rev. A **40**, 2847 (1989).

[2] Z. Hradil, Quantum-state estimation, Phys. Rev. A **55**, R1561 (1997).
[3] S. T. Flammia, D. Gross, Y.-K. Liu, and J. Eisert, Quantum tomography via compressed sensing: Error bounds, sample complexity and efficient estimators, New J. Phys. **14**, 095022 (2012).
[4] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, Quantum state tomography via compressed sensing, Phys. Rev. Lett. **105**, 150401 (2010).
[5] M. Cramer, M. B. Plenio, S. T. Flammia, R. Somma, D. Gross, S. D. Bartlett, O. Landon-Cardinal, D. Poulin, and Y.-K. Liu, Efficient quantum state tomography, Nat. Commun. **1**, 149 (2010).
[6] C. A. Riofrío, D. Gross, S. T. Flammia, T. Monz, D. Nigg, R. Blatt, and J. Eisert, Experimental quantum compressed sensing for a seven-qubit system, Nat. Commun. **8**, 15305 (2017).
[7] B. P. Lanyon, C. Maier, M. Holzäpfel, T. Baumgratz, C. Hempel, P. Jurcevic, I. Dhand, A. S. Buyskikh, A. J. Daley, M. Cramer, M. B. Plenio, R. Blatt, and C. F. Roos, Efficient tomography of a quantum many-body system, Nat. Phys. **13**, 1158 (2017).
[8] R. Blume-Kohout, Robust error bars for quantum tomography, arxiv:1202.5270.
[9] D. Suess, Ł. Rudnicki, T. O. Maciel, and D. Gross, Error regions in quantum state tomography: computational complexity caused by geometry of quantum states, New J. Phys. **19**, 093013 (2017).
[10] S. T. Flammia and Y.-K. Liu, Direct fidelity estimation from few Pauli measurements, Phys. Rev. Lett. **106**, 230501 (2011).
[11] M. P. da Silva, O. Landon-Cardinal, and D. Poulin, Practical characterization of quantum devices without tomography, Phys. Rev. Lett. **107**, 210404 (2011).
[12] A. B. Juditsky and A. S. Nemirovski, Nonparametric estimation by convex programming, Ann. Stat. **37**, 2278 (2009).
[13] A. Goldenshluger, A. Juditsky, and A. Nemirovski, Hypothesis testing by convex optimization, Electron. J. Stat. **9**, 1645 (2015).
[14] A. Juditsky and A. Nemirovski, Near-optimal recovery of linear and N-convex functions on unions of convex sets, IMA Inf. Inference **9**, 423 (2019).
[15] A. Seshadri, M. Ringbauer, J. Spainhour, T. Monz, and S. Becker, companion paper, Theory of versatile fidelity estimation with confidence, Phys. Rev. A **110**, 012431 (2024).
[16] Technically, our approach requires nonzero probabilities for each outcome, which we omit here for brevity and discuss in detail in Ref. [15].
[17] This is not technically a confidence interval, since the true fidelity is unknown in practice. The confidence interval would thus be centered around the estimate, and contain the true value with the specified confidence level.
[18] This interval is mostly symmetric for $F \approx 0.9$, but it is expected to be more asymmetric as we approach a fidelity close to 0 or 1.
[19] P. Schindler, D. Nigg, T. Monz, J. T. Barreiro, E. Martinez, S. X. Wang, S. Quint, M. F. Brandl, V. Nebendahl, C. F. Roos, M. Chwalla, M. Hennrich, and R. Blatt, A quantum information processor with trapped ions, New J. Phys. **15**, 123012 (2013).
[20] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, Measurement of qubits, Phys. Rev. A **64**, 052312 (2001).
[21] N. K. Langford, Errors in quantum tomography: Diagnosing systematic versus statistical errors, New J. Phys. **15**, 035003 (2013).
[22] T. L. Scholten and R. Blume-Kohout, Behavior of the maximum likelihood in quantum state tomography, New J. Phys. **20**, 023050 (2018).
[23] C. Schwemmer, L. Knips, D. Richart, H. Weinfurter, T. Moroder, M. Kleinmann, and O. Gühne, Systematic errors in current quantum state tomography tools, Phys. Rev. Lett. **114**, 080403 (2015).
[24] C. Ferrie and R. Blume-Kohout, Maximum likelihood quantum state tomography is inadmissible, arXiv:1808.01072.
[25] R. Blume-Kohout, Optimal, reliable estimation of quantum states, New J. Phys. **12**, 043034 (2010).
[26] D. Gottesman, Stabilizer codes and quantum error correction, Ph.D. thesis, California Institute of Technology, Pasadena, CA, 1997.
[27] A. Kitaev, Fault-tolerant quantum computation by anyons, Ann. Phys. (N.Y.) **303**, 2 (2003).
[28] S. Pallister, N. Linden, and A. Montanaro, Optimal verification of entangled states with local measurements, Phys. Rev. Lett. **120**, 170502 (2018).
[29] M. Kliesch, *Lecture Notes: Characterization, Certification, and Validation of Quantum Systems* (Heinrich-Heine-Universität Düsseldorf, 2020).
[30] C. Bădescu, R. O'Donnell, and J. Wright, Quantum state certification, in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing* (Association for Computing Machinery, New York, 2019), pp. 503–514.
[31] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, Nat. Phys. **16**, 1050 (2020).
[32] D. Chruściński and G. Sarbicki, Entanglement witnesses: Construction, analysis and classification, J. Phys. A **47**, 483001 (2014).
[33] A. Seshadri, M. Ringbauer, J. Spainhour, R. Blatt, T. Monz, and S. Becker, minimax-fidelity-estimation(2024), https://github.com/akshayseshadri/minimax-fidelity-estimation.