

Minimax Quantum Tomography: Estimators and Relative Entropy Bounds

Christopher Ferrie^{1,2} and Robin Blume-Kohout³

¹Center for Quantum Information and Control, University of New Mexico, Albuquerque, New Mexico 87131-0001, USA

²Centre for Engineered Quantum Systems, School of Physics, The University of Sydney, Sydney, New South Wales 2006, Australia

³Center for Computing Research (Sandia National Laboratories), Albuquerque, New Mexico 87185, USA

(Received 13 May 2015; published 4 March 2016)

A *minimax* estimator has the minimum possible error (“risk”) in the worst case. We construct the first minimax estimators for quantum state tomography with relative entropy risk. The minimax risk of nonadaptive tomography scales as $O(1/\sqrt{N})$ —in contrast to that of classical probability estimation, which is $O(1/N)$ —where N is the number of copies of the quantum state used. We trace this deficiency to *sampling mismatch*: future observations that determine risk may come from a different sample space than the past data that determine the estimate. This makes minimax estimators very biased, and we propose a computationally tractable alternative with similar behavior in the worst case, but superior accuracy on most states.

DOI: 10.1103/PhysRevLett.116.090407

Quantum information processing relies on physical systems that store and process quantum information, usually in the form of qubits. Testing and characterizing qubit devices is the business of quantum tomography [1], and quantum state tomography in particular is used to estimate the quantum state (density matrix) ρ produced by an initialization procedure. Tomography comprises two steps: (1) data gathering, accomplished by measuring a quorum of different observables on N samples of ρ and (2) an estimator that maps the data to a final estimate $\hat{\rho}$. The goal, of course, is an accurate estimate—we want a $\hat{\rho}$ close to the true state ρ , minimizing some error metric $d(\rho:\hat{\rho})$.

We define an optimal estimator to be one which achieves the highest accuracy in the worst case. One might expect tomographers to choose an estimator that is optimal (or at least near optimal). Surprisingly, this is not done. Although several estimators are known and used (linear inversion [2], maximum likelihood [3], Bayesian mean [4], hedged maximum likelihood [5], L_1 regularization [6], best linear unbiased estimator [7]), none of them is known to have optimal pointwise accuracy [8] for a finite N . Until now, it has not even been possible to evaluate whether any of these estimators is good enough, because the bounds on achievable pointwise accuracy have not been known either.

We address this situation in the present Letter by constructing minimax estimators [depicted in Fig. 1; see the detailed explanation after Eq. (7)] with absolutely optimal performance. These estimators are unwieldy, but (i) their performance yields tight upper bounds on accuracy, effectively delineating what “good enough” means, and (ii) their construction provides quite a lot of insight into the structure of the problem. Armed with these results, we show that hedged maximum likelihood (HML) is remarkably close to optimal and outperforms minimax for most states (though, of course, its worst-case risk is higher). We also identify the value for the hedging parameter β that appears in HML, which leads to the minimax solution within that class.

Prerequisites.—Defining “accuracy” requires making several choices. For example, an optimal estimator for one error metric $d(\rho:\hat{\rho})$ is generally not optimal for a different metric $d'(\rho:\hat{\rho})$. Here [14], we quantify inaccuracy by the quantum relative entropy,

$$d(\rho:\hat{\rho}) = \text{Tr}[\rho(\log \rho - \log \hat{\rho})]. \quad (1)$$

Relative entropy [19] is a well-motivated measure of predictive (and information-theoretic) inaccuracy [4]. It is a uniquely well-motivated error metric [9]; critically, it is Fisher adjusted (i.e., it agrees locally with the unique metric of statistical distinguishability [20]). Non-Fisher-adjusted metrics are ill motivated and yield arbitrary results. Analysis of a different Fisher-adjusted metric (e.g., infidelity) would produce results qualitatively similar to those we derive here.

An estimator’s pointwise risk is a function of the true state ρ and is given by the average of $d(\rho:\hat{\rho})$ over all possible data sets D of finite size N :

$$\bar{d}(\rho) = \sum_D \text{Pr}(D|\rho) d(\rho:\hat{\rho}(D)). \quad (2)$$

In the minimax paradigm, we quantify an estimator’s accuracy by its worst-case risk, $\bar{d}_{\max} = \max_{\rho} \bar{d}(\rho)$. The minimax risk of the estimation problem is the minimum achievable risk (minimized over all possible estimators), and a minimax estimator is one that achieves this bound.

In most inference problems, the sample space of possible observations (data) is fixed by the problem. Not so in quantum tomography. Quantum systems can be measured in many different and incomparable ways. This is the single most significant difference between quantum and classical estimation. This freedom is often removed in quantum problems by choosing the best or worst possible measurement (e.g., as in the definition of quantum relative entropy as the classical relative entropy of the most difficult-to-predict measurement). This is usually not done in tomography, because the measurements which have the lowest expected

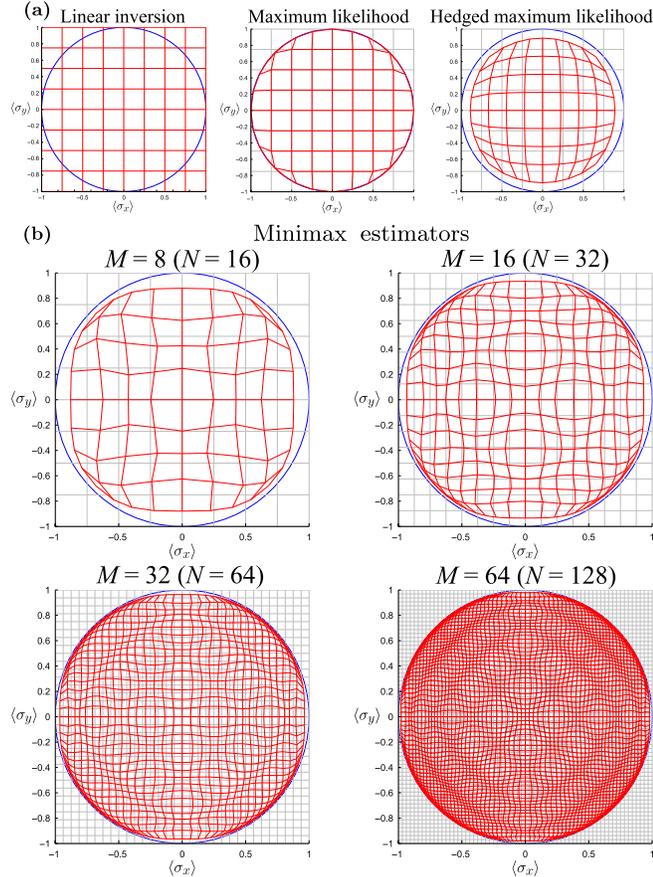


FIG. 1. Estimators for Pauli measurements on a rebit, depicted as distortions of the linear inversion grid [see the text after Eq. (7)]. (a) Three standard estimators for $M = 8$ measurements of X and Y . Vertices of the red grid correspond to estimated states. Linear inversion estimates extend outside the Bloch disk of physical states. MLE's estimates are non-negative; HML's are strictly positive. (b) Minimax estimators for $M = 8, 16, 32, 64$ measurements of X and Y on a rebit. Ripples indicate local bias toward support points of the least favorable prior [9].

risk are far too difficult. In this Letter, we follow the majority of experiments and analyze tomography based on Pauli measurements on a single qubit. However, we also prove analytic lower bounds on minimax risk that apply to *any* nonadaptive measurement and any d -dimensional quantum system. In some parts of our analysis, we use a rebit—a quantum system with a two-dimensional real Hilbert space, whose state space corresponds to the equatorial plane of the Bloch sphere—as an easier-to-analyze proxy for a qubit.

Minimax risk.—The first main result of this Letter is a lower bound on the asymptotic ($N \rightarrow \infty$) minimax relative entropy risk of Pauli tomography on qubits and rebits,

$$\bar{d}_{\max} \geq \frac{e^{-(1/2)} \sqrt{\mathfrak{D} - 1}}{4 \sqrt{N}}, \quad (3)$$

where $\mathfrak{D} = 2$ for rebits and $\mathfrak{D} = 3$ for qubits. Its $O(1/\sqrt{N})$ scaling contrasts sharply with the minimax risk of estimating a classical bit, which is almost exactly $0.5/N$ [22,23].

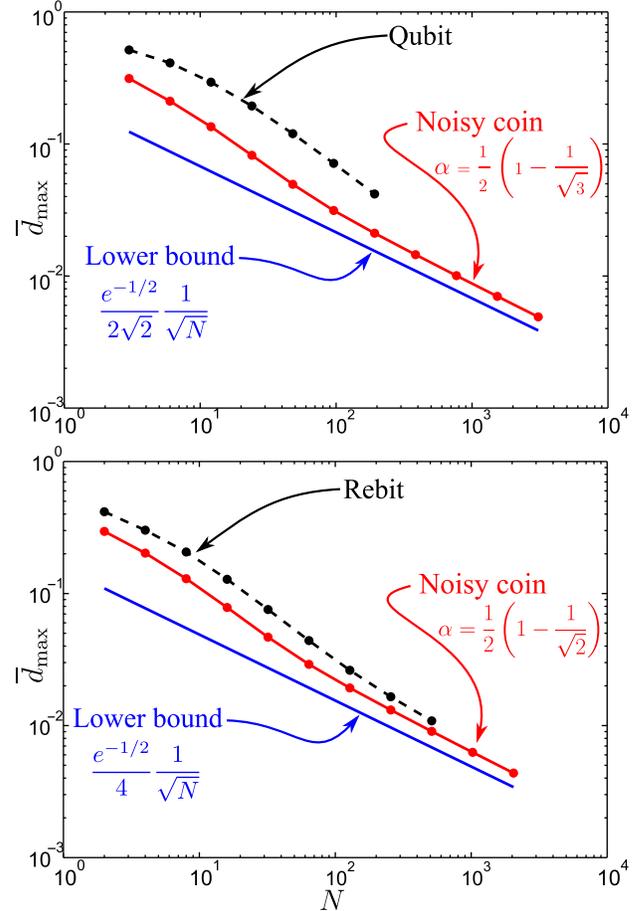


FIG. 2. Numerical minimax risk for qubits, rebits, and noisy coins. Black curves show the risk of numerically constructed minimax estimators for a qubit and a rebit, as a function of the number of samples (N), up to the maximum that was numerically feasible. Red curves illustrate the numerically computed risk of noisy coin systems whose noise levels are chosen to match the effective noise of the qubit and the rebit, respectively. Blue lines show the lower bound given in Eq. (6).

We derive this bound below by mapping the minimax risk of qubit and rebit state tomography to a classical noisy coin model. In Fig. 2, we compare these bounds to numerical calculations of the minimax risk, for small N 's, of qubits, rebits, and noisy coins.

A d -dimensional quantum state is analogous in many ways to a classical d -outcome probability distribution. However, its minimax risk scales differently because of a phenomenon intrinsic to quantum tomography (though not uniquely quantum) that we call *sampling mismatch*: the sample space for the observed events is neither unique nor isomorphic to the underlying state space. For example, the possible statistics for the three two-outcome Pauli measurements on a qubit naturally define a cube, whereas the possible quantum states form a sphere (the Bloch ball).

Sampling mismatch can be reproduced in a simple classical model called the noisy coin [24]. It is a classical system with a two-outcome sample space (i.e., a coin flip) where each observation is erroneous, with the known

probability α . Sampling mismatch arises when we attempt to assign probabilities to future noiseless observations using data from noisy measurements. The noisy coin's minimax risk is $O(1/\sqrt{N})$, because nearly pure states are hard to estimate accurately from noisy statistics. The corresponding minimax estimators are strongly biased toward nearly pure states (see Ref. [24] for details). We are going to use a variant of the noisy coin model to bound the risk of tomography.

We define “tomography” thusly: N samples (copies) of a single-qubit state ρ will be prepared; each sample will be measured independently (not jointly together with other samples) in a predefined fashion (not adaptively). The k th sample is measured in an arbitrary basis, and this measurement can be described by a positive operator-valued measure (POVM) $\mathcal{M}_k = \{\Pi_k, \mathbb{1} - \Pi_k\}$, whose outcomes have probabilities $\{q, 1 - q\}$, with $q = \text{Tr}\Pi_k\rho$. Based on the N measurement results, we report a state $\hat{\rho}$ and seek to minimize the relative entropy cost.

Now, suppose that before analyzing the data (but after choosing the measurements) we are told the eigenbasis of ρ . This helps us (only ρ 's spectrum must be estimated), so the risk of spectrum estimation is a strict lower bound on the risk of full tomography [25].

We define $\{|0\rangle, |1\rangle\}$ to be the eigenstates of ρ and write

$$\rho = p|0\rangle\langle 0| + (1 - p)|1\rangle\langle 1|. \quad (4)$$

Now, we need only estimate $p \in [0, 1]$. This parameter manifold is identical to that of a coin. Furthermore, the quantum relative entropy between two diagonal density matrices is identical to the classical relative entropy between the corresponding distributions. So, since ρ 's eigenbasis is known, estimating ρ is identical to estimating the bias of a coin. However, unless the eigenbases of ρ and the Π_k happen to coincide, the measurement data obtained from the N samples of ρ are not noiseless. Even if $p = 0$ (i.e., ρ is pure), the data remain somewhat random. The probability of observing Π_k is not p , but

$$\begin{aligned} q &= p\langle 0|\Pi_k|0\rangle + (1 - p)\langle 1|\Pi_k|1\rangle \\ &= p(1 - 2\alpha_k) + \alpha_k, \end{aligned}$$

where the effective noise in sample k is

$$\alpha_k = \langle 1|\Pi_k|1\rangle. \quad (5)$$

We can model this situation perfectly by a noisy coin (as in Ref. [24]) where each observation fails with a different error probability. The error probability for the k th sample is α_k . In Ref. [9], we bound this estimation problem's minimax risk by

$$\bar{d}_{\max} \geq \frac{e^{-(1/2)}}{2\sqrt{\bar{\beta}}} \frac{1}{\sqrt{N}}, \quad (6)$$

where $\bar{\beta}$ is the average resolution provided by the N noisy samples:

$$\bar{\beta} = \frac{1}{N} \sum_{k=1}^N \beta_k = \frac{1}{N} \sum_{k=1}^N \frac{(1 - 2\alpha_k)^2}{\alpha_k(1 - \alpha_k)}. \quad (7)$$

For any fixed measurement strategy—e.g., the standard one where $N/3$ samples are measured in the X , Y , Z bases—the maximum risk occurs when we choose the eigenbasis of ρ to maximize $\bar{\beta}$ in Eq. (7). This least favorable basis is the one that lies as far as possible from all measured bases. For a rebit, it lies halfway between the X and Y bases, and $\alpha_k = \frac{1}{2}(1 - 1/\sqrt{2})$. For a qubit, it is the geometric mean of the X , Y , and Z bases, and $\alpha_k = \frac{1}{2}(1 - 1/\sqrt{3})$. Inserting these values for α_k yields the final bound given in Eq. (3).

This argument applies (qualitatively) to tomography on any finite-dimensional system with any discrete POVM. As long as no samples are measured in a basis that diagonalizes ρ , the minimax risk scales as $O(1/\sqrt{N})$ (although the prefactor will vary). However, if any nonvanishing fraction of the N samples is measured in a basis that diagonalizes ρ , then Eq. (6) no longer applies. Thus, continuous POVMs such as the unitarily invariant Haar-uniform rank-1 POVM (also known as the uniform POVM) require a slightly different argument. In Ref. [9], we prove that, even in this case, the minimax risk is lower bounded by $O((N \log N)^{-1/2})$.

Estimators.—To confirm the bound given by Eq. (3) and to explore the minimax risk at small N , we use numerics to find minimax estimators. An estimator is a map from the set of all possible data sets into the set of density matrices. The possible outcomes of the measurement(s) performed are represented by a set of positive operators $\{E_k\}$, and the data themselves by a set of raw counts $D = \{n_k\}$. For qubit Pauli tomography, the data comprise $M = N/3$ samples of each of the σ_x , σ_y , and σ_z measurements; for rebits, they comprise $M = N/2$ samples of each of the σ_x and σ_y measurements [26].

We used numerical optimization (over the set of possible estimators) to find minimax estimators. The algorithms are described in Ref. [9]. In Fig. 1, we depict the resulting estimators and compare them to three canonical estimators: 1) Linear inversion ($\hat{\rho}_{\text{LI}}$): The first tomographic estimator, it is obtained by equating each probability $\text{Pr}(k|\hat{\rho}_{\text{LI}}) = \text{Tr}E_k\hat{\rho}_{\text{LI}}$ to its observed frequency n_k/M . 2) Maximum likelihood ($\hat{\rho}_{\text{ML}}$): MLE assigns the density matrix that maximizes the probability of the observed data (the likelihood), $\mathcal{L}(\rho) = \text{Pr}(D|\rho) = \prod_k [\text{Tr}(E_k\rho)^{n_k}]$. 3) Hedged maximum likelihood ($\hat{\rho}_{\text{HML},\beta}$): The HML estimator maximizes the product of $\mathcal{L}(\rho)$ and a hedging function $h(\rho) = \det(\rho)^\beta$. This function is strictly convex and vanishes for rank-deficient states, so the HML estimate is always full rank. To simplify visualization, we depict rebit estimators, which are qualitatively similar to qubit estimators and easier to depict. A rebit estimator is a map from data sets to Bloch vectors, as $\hat{\rho}: \{0, \dots, M\}^2 \rightarrow \mathbb{R}^2$. We use the linear inversion estimator as a reference. As a linear map from the two-dimensional space of data sets ($\{0 \dots M\}^2$) and the two-dimensional space of rebit states (the unit disk in \mathbb{R}^2), the linear inversion estimator is

represented by a uniform grid on the Bloch square [see Fig. 1(a)]. Every *other* estimator is represented as a distortion of this grid. The vertices of the grid are estimates $\hat{\rho}$, and the position of such a vertex within the grid indicates what data set it came from.

Minimax estimators for $N = 16, 32, 64,$ and 128 (total) Pauli measurements on a rebit are shown in Fig. 1(b). The most striking feature of these estimators is a pronounced ripple phenomenon. This is not a numerical artifact. Instead, it represents a consistent bias toward certain discrete points within the state space (support points of the least favorable prior—see Fig. 1 in Ref. [9]), which can be identified in Fig. 1 as regions where the grid lines cluster together. The minimax estimator demonstrates this bias because these points are, in a particular sense, the most difficult to estimate accurately.

Improving on minimax.—The minimax criterion is an elegant concept, but a dangerous one. In its single-minded quest to improve the maximum risk, it has no concern for the pointwise risk at states that are easier to estimate. In such regions, it may incur extreme bias and inaccuracy, for the sole purpose of achieving a tiny reduction in the maximum risk. For quantum tomography, this effect becomes extreme. While $O(1/N)$ risk can be achieved on all full-rank states, the risk is unavoidably $O(1/\sqrt{N})$ near the boundary. Our numerical experiments confirm that the minimax estimator’s pointwise risk is $O(1/\sqrt{N})$ everywhere, whereas other estimators easily achieve $O(1/N)$ risk in the interior of the Bloch sphere [see Fig. 3(b)]. If ρ really were selected adversarially, then minimax would be a wise strategy. But in realistic cases, we would prefer an estimator that achieved $O(1/N)$ scaling where possible, even at the cost of slightly worse worst-case behavior.

A good estimator should achieve $O(1/N)$ risk in the interior, while coming as close as possible to minimax performance near the boundary. The maximum likelihood estimator (MLE) is disqualified because its pointwise expected risk is uniformly infinite [it has a nonzero probability of returning a rank-deficient estimate for every ρ , so $\bar{d}(\rho) = \infty$]. However, HML does not display this behavior. Introduced in Ref. [5] as a full-rank alternative to MLE, HML generalizes classical add- β estimators. Like them, it never assigns zero probabilities, and it has an adjustable parameter β that governs how much it avoids zero eigenvalues. Classical add- β estimators are very nearly minimax (for $\beta \approx 1/2$), which suggests that HML estimators might have similar near-optimality properties.

All HML estimators have good behavior [$O(1/N)$ pointwise risk] in the interior, so we choose β to be the one which is minimax among HML estimators. As illustrated in Fig. 3(b), a HML estimator’s pointwise risk has local maxima at the boundary (pure states) and/or at a slightly depolarized state (with purity $\sim 1 - 1/\sqrt{N}$). To minimize its maximum, we choose β to equalize the risk at these two local maxima. The asymptotically minimax β for the noisy coin model was shown in Ref. [24] to be

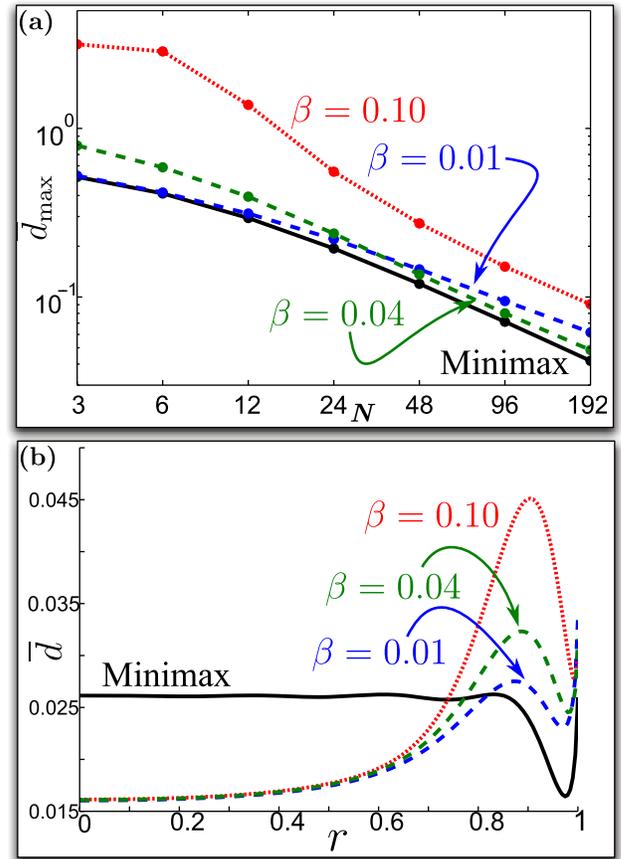


FIG. 3. Maximum and pointwise risk of minimax and HML estimators. (a) The maximum risk, for qubit tomography, of the minimax estimator and three different HML estimators ($\beta = 0.01, 0.04, 0.10$) for $N \leq 192$ samples, distributed equally among the three Pauli bases. (b) The pointwise risk, along the axis oriented at 45° to both X and Y , of the same estimators for $N = 128$ samples for a rebit [this minimax estimator is depicted in Fig. 1(b)]. The two local maxima of $\bar{d}(\rho)$ are at $r = 1$ and $r \approx 1 - 1/\sqrt{N}$. Choosing $\beta \approx 0.04$ balances these risks and is therefore minimax among HML estimators. This HML estimator comes very close to matching the worst-case performance of the minimax estimator and outperforms it dramatically in the interior of the state space.

$\beta_{\text{optimal}} \approx 0.0389$, and our numerics confirm that $\beta \approx 0.04$ is minimax to within the available numerical precision for rebit tomography as well [see Fig. 3(b); qubit results for smaller N 's are not shown, but they confirm that $\beta \approx 0.04$ has nearly minimax performance].

For this value of β , HML compares favorably with minimax estimators. Its worst-case risk is very close to the minimax risk [see Fig. 3(a)], and it dramatically outperforms minimax in the interior of the state space [see Fig. 3(b)]. So, while hedging estimators do not offer strictly optimal performance by the global minimax criterion, they are (i) easy to specify and calculate, (ii) close to minimax, and (iii) more accurate than minimax estimators for almost all states ρ . We do not know why the minimax β is so different for noiseless coins (≈ 0.5) and for qubits, rebits, or noisy

coins (≈ 0.04), but it suggests fundamental differences between noiselessly sampled systems and those (like qubits and noisy coins) where sampling mismatch is important.

C. F. was supported by National Science Foundation Grant No. PHY-1212445; by the Canadian Government through the NSERC PDF program, the IARPA MQCO program, and the ARC, via EQU S Project No. CE11001013; and by U.S. Army Research Office Grants No. W911NF-14-1-0098 and No. W911NF-14-1-0103. Sandia National Laboratories is a multiprogram laboratory operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration, under Contract No. DE-AC04-94AL85000.

-
- [1] M. Paris and J. Rehacek, *Quantum State Estimation* (Springer, New York, 2004).
- [2] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, England, 2010).
- [3] Z. Hradil, Quantum-state estimation, *Phys. Rev. A* **55**, R1561 (1997).
- [4] R. Blume-Kohout, Optimal, reliable estimation of quantum states, *New J. Phys.* **12**, 043034 (2010).
- [5] R. Blume-Kohout, Hedged Maximum Likelihood Quantum State Estimation, *Phys. Rev. Lett.* **105**, 200504 (2010).
- [6] 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).
- [7] H. Zhu, Quantum state estimation with informationally over-complete measurements, *Phys. Rev. A* **90**, 012115 (2014).
- [8] Bayesian mean is known to provide optimal average accuracy for certain important error metrics $d(\rho; \hat{\rho})$.
- [9] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.116.090407>, which includes Refs. [10–13], for a discussion of error metrics and estimators, a proof of the bound in Eq. (3), and the numerical algorithms used.
- [10] K. M. R. Audenaert, J. Calsamiglia, R. Muñoz-Tapia, E. Bagan, Ll. Masanes, A. Acín, and F. Verstraete, Discriminating States: The Quantum Chernoff Bound, *Phys. Rev. Lett.* **98**, 160501 (2007).
- [11] C. A. Fuchs, Ph.D. thesis, University of New Mexico, 1996.
- [12] E. L. Lehmann and G. Casella, *Theory of Point Estimation* (Springer, New York, 1998).
- [13] P. J. Kempthorne, Numerical specification of discrete least favorable prior distributions, *SIAM J. Sci. Stat. Comput.* **8**, 171 (1987).
- [14] Without a doubt, there are other reasonable definitions that would lead to different answers. Examples are given by Refs. [15–18], which each obtain different conclusions by choosing different loss functions. We expect future work to consider other variations, and we hope that our results and discussion will help to inform and guide such future work.
- [15] H. K. Ng and B.-G. Englert, A simple minimax estimator for quantum states, *Int. J. Quantum. Inform.* **10**, 1250038 (2012).
- [16] N. K. Langford, Errors in quantum tomography: Diagnosing systematic versus statistical errors, *New J. Phys.* **15**, 035003 (2013).
- [17] R. Schmied, Quantum state tomography of a single qubit: Comparison of methods, [arXiv:1407.4759](https://arxiv.org/abs/1407.4759).
- [18] S. T. Flammia, D. Gross, Y.-K. Liu, and Jens Eisert, Quantum tomography via compressed sensing: Error bounds, sample complexity and efficient estimators, *New J. Phys.* **14**, 095022 (2012).
- [19] V. Vedral, The role of relative entropy in quantum information theory, *Rev. Mod. Phys.* **74**, 197 (2002).
- [20] To be precise, there is a unique classical metric of statistical distinguishability (the Fisher metric), and for quantum states there exists a limited but nontrivial family of such metrics [21]. However, all of the quantum Fisher metrics necessarily coincide for spectral (eigenvalue) deviations, which dominate the pointwise risk. Thus, our statement in the text (while technically flawed) is fully accurate in practice; all quantum Fisher-adjusted error metrics agree asymptotically on the relevant errors.
- [21] D. Petz and C. Sudar, Geometries of quantum states, *J. Math. Phys. (N.Y.)* **37**, 2662 (1996).
- [22] A. Rukhin, Minimax estimation of the binomial parameter under entropy loss, *Stat Decision* **13**, 69 (1993).
- [23] R. E. Krichevskiy, Laplace's law of succession and universal encoding, *IEEE Trans. Inf. Theory* **44**, 296 (1998).
- [24] C. Ferrie and R. Blume-Kohout, Estimating the bias of a noisy coin, *AIP Conf. Proc.* **1443**, 14 (2012).
- [25] The risk is a sum of terms due to (i) spectral (eigenvalue) error and (ii) eigenbasis error. The minimax risk is dominated by spectral error because $O(\epsilon)$ spectral variations (changes in eigenvalues) can induce $O(\epsilon)$ changes in probabilities p that are themselves $O(\epsilon)$ and thus change the risk by $O(\epsilon)$ —whereas an $O(\epsilon)$ unitary variation changes p by, at most, $O(\epsilon p)$, and it changes the risk by, at most, $O(\epsilon^2)$. Thus, giving up the eigenbasis error does not lower the risk substantially.
- [26] The assumption of perfectly known measurements can be relaxed [27,28].
- [27] A. M. Branczyk, D. H. Mahler, L. A. Rozema, A. Darabi, A. M. Steinberg, and D. F. V. James, Self-calibrating quantum state tomography, *New J. Phys.* **14**, 085003 (2012).
- [28] D. Greenbaum, Introduction to quantum gate set tomography, [arXiv:1509.02921](https://arxiv.org/abs/1509.02921).