Adaptive Quantum State Tomography Improves Accuracy Quadratically

D. H. Mahler,[1,2,*] Lee A. Rozema,[1,2] Ardavan Darabi,[1,2] Christopher Ferrie,[3]
Robin Blume-Kohout,[4] and A. M. Steinberg[1,2]

[1]*Department of Physics, Centre for Quantum Information and Quantum Control and Institute for Optical Sciences, University of Toronto, 60 St. George Street, Toronto, Ontario M5S 1A7, Canada*
[2]*Canadian Institute for Advanced Research, Toronto, Ontario M5G1Z8, Canada*
[3]*Center for Quantum Information and Control, University of New Mexico, Albuquerque, New Mexico 87131-0001, USA*
[4]*Sandia National Laboratories, Advanced Device Technologies (01425), Albuquerque, New Mexico 87185, USA*

(Received 2 March 2013; revised manuscript received 22 August 2013; published 29 October 2013)

We introduce a simple protocol for adaptive quantum state tomography, which reduces the worst-case infidelity $[1 - F(\hat{\rho}, \rho)]$ between the estimate and the true state from $O(1/\sqrt{N})$ to $O(1/N)$. It uses a single adaptation step and just one extra measurement setting. In a linear optical qubit experiment, we demonstrate a full order of magnitude reduction in infidelity (from 0.1% to 0.01%) for a modest number of samples ($N \approx 3 \times 10^4$).

DOI: 10.1103/PhysRevLett.111.183601 PACS numbers: 42.50.Dv, 03.65.Wj, 42.50.Ex, 42.50.Xa

Quantum information processing requires reliable, repeatable preparation and transformation of quantum states. Quantum state tomography is used to identify the density matrix ρ that was prepared by such a process. No finite ensemble of N samples is sufficient to uniquely identify ρ, so we estimate it, reporting either a single state $\hat{\rho}$ that is "close" to ρ with high probability [1–5], or a confidence region of nonzero radius that contains ρ with high probability [6,7]. Both approaches must accept some inaccuracy (the discrepancy between $\hat{\rho}$ and ρ) or imprecision (the diameter of the confidence region). The universal goal of state tomography is to minimize this discrepancy, which has been quantified with various metrics (e.g., trace norm, fidelity, relative entropy, etc.). In this Letter, we focus on the particularly well-motivated quantum infidelity,

$$1 - F(\hat{\rho}, \rho) = 1 - \mathrm{Tr}\left(\sqrt{\sqrt{\rho}\hat{\rho}\sqrt{\rho}}\right)^2, \tag{1}$$

and show that as $N \to \infty$, adaptive tomography reduces expected infidelity from $O(1/\sqrt{N})$ to $O(1/N)$.

Unlike alternative metrics, $1 - F(\hat{\rho}, \rho)$ quantifies an important operational quantity: how many copies are required to reliably distinguish $\hat{\rho}$ from ρ? Without doing justice to the rich body of research behind this simple statement (e.g., [8–13]...), we summarize as follows. The discrepancy between $\hat{\rho}$ and ρ given a single sample is well described by the trace distance, $|\hat{\rho} - \rho|_1$. But tomography (i) requires $N \gg 1$ samples, (ii) is used to predict experiments on $N \gg 1$ samples, and (iii) yields errors that cannot be detected without $N \gg 1$ samples. So the operationally relevant quantity is $|\hat{\rho}^{\otimes N} - \rho^{\otimes N}|_1$, which for $N \gg 1$ behaves as $1 - e^{-D(\hat{\rho},\rho)N}$. The exponent D is the quantum Chernoff bound [13], and $N \approx D\log(1/\epsilon)$ samples are necessary and sufficient to distinguish ρ from $\hat{\rho}$ with confidence $1 - \epsilon$. D is tightly bounded by the logarithm of the fidelity (see [12], Eq. 28); when $1 - F(\hat{\rho}, \rho) \ll 1$ (which should always be true in tomography), $-\log(F) \approx 1 - F$ and

$$\frac{1 - F}{2} \leq D \leq 1 - F. \tag{2}$$

Thus, $1 - F$ really does (almost uniquely) quantify tomographic inaccuracy; $N \approx [1 - F(\hat{\rho}, \rho)]^{-1}$ samples are (up to a factor of 2) necessary and sufficient [14] to falsify $\hat{\rho}$. In contrast, Hilbert-Schmidt- and trace-distance have no such N-sample meaning, and give wildly misleading metrics of tomographic error.

We show that standard tomography with static measurements can't beat $1 - F = O(1/\sqrt{N})$ as $N \to \infty$ for a large and important class of states, then introduce and explain a simple adaptive protocol that achieves $1 - F = O(1/N)$ for every state. Finally, we demonstrate this effect in a linear optical experiment, achieving a tenfold improvement in infidelity (from 0.1% to 0.01% with $N = 3 \times 10^4$ measurements) over standard tomography. We believe this protocol will have wide application, particularly in situations where the rate of data collection is small, such as postselected optical systems (e.g., [15], where data were collected at approximately 9 measurements per hour).

Adaptivity has been proposed in various contexts. Single-step adaptive tomography was first analyzed by [16], then refined in [17–19]. A scheme similar to ours (and its efficacy for *pure* states) was analyzed in [20]. Reference [21] recently treated state estimation as parameter estimation, obtaining results complementary, but largely orthogonal, to those reported here. Here, we present both an experimental demonstration and simple, self-contained derivation of (i) why quantum fidelity is significant, (ii) why adaptive tomography achieves far better infidelity, and (iii) how the adaptation should be done. We optimize worst-case infidelity over all states, not just pure states [20] or specific ensembles of mixed states (e.g., Ref. [18] achieved high average fidelity, but low fidelity on nearly pure states).

Adaptive tomography.—Static tomography uses data from a fixed set of measurements. Different measurements yield subtly different tomographic accuracy [22], but to leading order, "good" protocols for single-qubit tomography provide equal information [23] about every component of the unknown density matrix ρ,

$$\rho = \frac{1}{2}(11 + \langle\sigma_x\rangle\sigma_x + \langle\sigma_y\rangle\sigma_y + \langle\sigma_z\rangle\sigma_z). \quad (3)$$

The canonical example involves measuring the three Pauli operators $(\sigma_x, \sigma_y, \sigma_z)$. This minimizes the variance of the estimator $\hat{\rho}$—but not the expected infidelity, for two reasons.

First, the variance of the estimate $\hat{\rho}$ depends also on ρ itself. Consider the linear inversion estimator $\hat{\rho}_{\text{lin}}$, defined by estimating $\langle\sigma_z\rangle = (n_\uparrow - n_\downarrow)/(n_\uparrow + n_\downarrow)$ (and similarly for $\langle\sigma_x\rangle$ and $\langle\sigma_y\rangle$), and substituting into Eq. (3). Each measurement behaves like $N/3$ flips of a coin with bias $p_k = (1/2)(1 + \langle\sigma_k\rangle)$, and yields

$$\hat{p}_k = p_k \pm \sqrt{\frac{3}{N}}\sqrt{p_k(1-p_k)} \quad (4)$$

$$\Rightarrow \langle\sigma_k\rangle_{\text{estimated}} = \langle\sigma_k\rangle_{\text{true}} \pm \sqrt{\frac{3}{2N}}\sqrt{1 - \langle\sigma_k\rangle^2}. \quad (5)$$

When $\langle\sigma_k\rangle \approx 0$, its estimate has a large variance—but when $\langle\sigma_k\rangle \approx \pm 1$, the variance is very small. As a result, the variance of $\hat{\rho}$ around ρ is anisotropic and ρ dependent [see Fig. 1(a)].

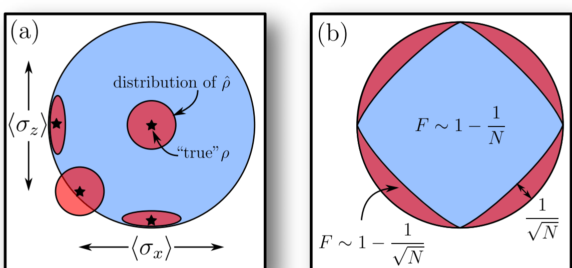

FIG. 1 (color online). Two features of qubit tomography with Pauli measurements (shown for an equatorial cross section of the Bloch sphere): (a) The distribution or "scatter" of any unbiased estimator $\hat{\rho}$ (depicted by dull red ellipses) varies with the true state ρ (black stars at the center of ellipses). (b) The expected infidelity between $\hat{\rho}$ and ρ as a function of ρ. Within the Bloch sphere, the expected infidelity is $O(1/N)$. But in a thin shell of nearly pure states (of thickness $O(1/\sqrt{N})$), it scales as $O(1/\sqrt{N})$—except when ρ is aligned with a measurement axis (Pauli X, Y, or Z).

Second, the dependence of infidelity on the error, $\Delta = \hat{\rho} - \rho$, also varies with ρ. Infidelity is hypersensitive to misestimation of small eigenvalues. A Taylor expansion of $1 - F(\hat{\rho}, \rho)$ yields (in terms of ρ's eigenbasis $\{|i\rangle\}$),

$$1 - F(\rho, \rho + \epsilon\Delta) = \frac{1}{4}\sum_{i,j}\frac{\langle i|\Delta|j\rangle^2}{\langle i|\rho|i\rangle + \langle j|\rho|j\rangle} + O(\Delta^3). \quad (6)$$

Infidelity is quadratic in Δ—except that as an eigenvalue $\langle i|\rho|i\rangle$ approaches 0, its sensitivity to $\langle i|\Delta|i\rangle$ diverges; $1 - F$ becomes linear [24] in Δ:

$$1 - F(\rho, \rho + \epsilon\Delta) = \epsilon \sum_{i:\ \langle i|\rho|i\rangle = 0} \langle i|\Delta|i\rangle + O(\Delta^2). \quad (7)$$

To minimize infidelity, we must accurately estimate the small eigenvalues of ρ, particularly those that are (or appear to be) zero. For states deep within the Bloch sphere, static tomography achieves infidelity of $O(1/N)$ [16,25]. Typical errors scale as $|\Delta| = O(1/\sqrt{N})$ (Eq. (5)), and infidelity scales as $1 - F = O(|\Delta|^2)$. But for states with eigenvalues less than $O(1/\sqrt{N})$, infidelity scales as $O(1/\sqrt{N})$. Quantum information processing relies on nearly pure states, so this poor scaling is significant.

To achieve better performance, we observe that if ρ is diagonal in one of the measured bases (e.g., σ_z), then infidelity *always* scales as $O(1/N)$. The increased sensitivity of $1 - F$ to error in small eigenvalues [Eq. (6)] is precisely cancelled by the reduced inaccuracy that accompanies a highly biased measurement-outcome distribution [Eq. (5)]. This suggests an obvious (if naïve) solution: we should simply ensure that we measure the diagonal basis of ρ!

This is unreasonable—knowing ρ would render tomography pointless. But we can perform standard tomography on $N_0 < N$ samples, get a preliminary estimate $\hat{\rho}_0$, and measure the remaining $N - N_0$ samples so that one basis diagonalizes $\hat{\rho}_0$. This measurement will not diagonalize ρ exactly, but if $N_0 \gg 1$ it will be fairly close. The angle θ between the eigenbases of ρ and $\hat{\rho}_0$ is $O(|\Delta|) = O(1/\sqrt{N_0})$. This implies that if ρ has an eigenvector $|\psi_k\rangle$ with eigenvalue $\lambda_k = 0$, then corresponding measurement outcome $|\phi_k\rangle\langle\phi_k|$ will have probability at most $p_k = \sin^2\theta \approx \theta^2 = O(1/N_0)$. Since we make this measurement on $O(N - N_0)$ copies [26], the final error in the estimated $\hat{p}_k$ (and, therefore, in the eigenvalue λ_k) is $O(1/\sqrt{N_0(N - N_0)})$. So using a constant fraction $N_0 = \alpha N$ of the available samples for the preliminary estimation should yield $O(1/N)$ infidelity for all states.

A similar protocol was suggested in Ref. [18], but that analysis concluded that $N_0 \propto N^p$ for $p \geq 2/3$ would be sufficient. This works for average infidelity over a particular ensemble, but yields $1 - F = O(N^{-5/6})$ for almost all nearly pure states.

Simulation results.—We performed numerical simulations of single-qubit tomography using four different protocols: (1) standard fixed-measurement tomography; (2) adaptive tomography with $N_0 = N^{2/3}$, as proposed in [18]; (3) adaptive tomography with $N_0 = \alpha N$ (for a range of α); and (4) ''known basis'' tomography, wherein we cheat by aligning our measurement frame with ρ's eigenbasis (for all N samples). We simulated many true states ρ, but present a representative case: a pure state with $(\langle\sigma_x\rangle, \langle\sigma_y\rangle, \langle\sigma_z\rangle) = (0.5, 1/\sqrt{2}, 0.5)$

$$|\nearrow\rangle = \frac{1}{2}\begin{pmatrix} \sqrt{3} \\ \frac{1}{\sqrt{3}} - \frac{2i}{\sqrt{6}} \end{pmatrix} \tag{8}$$

Our results are not particularly sensitive to the exact estimator used; we used maximum-likelihood estimation (MLE) with a quadratic approximation to the negative loglikelihood function:

$$l(\rho) = -\log \mathcal{L}(\rho) \approx \sum_{k=1}^{3} \frac{N_k(\mathrm{Tr}[\rho E_k] - f_k)^2}{f_k(1 - f_k)}, \tag{9}$$

where $f_k = n_k/N_k$ are the observed frequencies of the $+1$ eigenvectors of the three Pauli operators σ_k, E_k is the corresponding projector, and N_k is the number of samples on which σ_k was measured. Convex optimization (in MATLAB [27]) was used to find $\hat{\rho}_{\mathrm{MLE}}$. Results were averaged over many (typically 150) randomly generated measurement records.

Figure 2 shows average infidelity versus N. We fit these simulated data to power laws of the form $1 - F = \beta N^p$, and found $p = -0.513 \pm 0.006$ (for static tomography), $p = -0.868 \pm 0.008$ (for adaptive tomography with $N_0 = N^{2/3}$), $p = -0.980 \pm 0.006$ (for adaptive tomography with $N_0 = 0.5N$), and $p = -0.993 \pm 0.09$ (for known-basis tomography). These results are not significantly different [28] from predictions of the simple theory ($p = -(1/2)$, $-(5/6)$, 1, and 1, respectively). The borderline-significant discrepancy is, we believe, due to boundary effects ($\hat{\rho}_{\mathrm{MLE}}$ is constrained to be positive). We also varied $\alpha = N_0/N$ (Fig. 2, inset) and found that $\alpha = 1/2$ optimizes the prefactor (β).

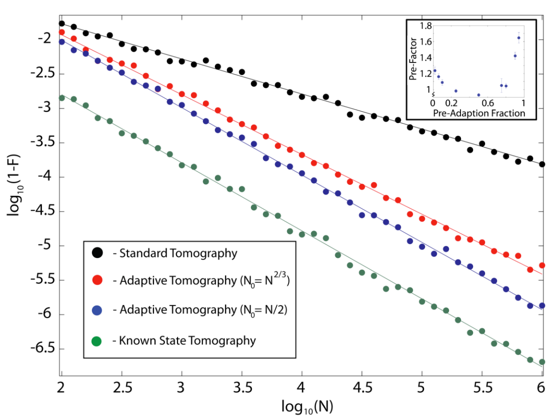

FIG. 2 (color online). Average infidelity $1 - F(\hat{\rho}, \rho)$ vs sample size N for Monte Carlo simulations of four different tomographic procotocols: standard tomography (black), the procedure proposed in [18] using $N_0 = N^{2/3}$ (red), our procedure using $N_0 = N/2$ (blue), and known basis tomography (green). Both adaptive procedures clearly outperform static tomography, but our procedure clearly outperforms the $N_0 = N^{2/3}$ approach, and matches the asymptotic scaling of known-basis tomography. The inset shows the dependence of the prefactor (β) on $\alpha = N_0/N$.

Experimental results.—We implemented our protocol experimentally in linear optics (Fig. 3). Using type-1 spontaneous parametric down-conversion in a nonlinear crystal, photon pairs were created. One of these photons was sent immediately to a single photon counting module (SPCM) to act as a trigger. The second photon was sent through a Glan-Thomson polarizer to prepare it in a state of very pure linear polarization. Computer-controlled wave plates were first used to prepare the polarization state of the photon, and subsequently used in tandem with a polarization beam splitter to project onto any state on the Bloch sphere.

We compared static and adaptive tomography protocols on a measured state given (in the H/V basis) by

$$\rho = \begin{pmatrix} 0.7711 & 0.2010 + 0.3624i \\ 0.2010 - 0.3624i & 0.2289 \end{pmatrix}, \tag{10}$$

which has purity $\mathrm{Tr}(\rho^2) = 0.991$ and fidelity $F = 0.992$ with $|\nearrow\rangle$ [see Eq. (8)]. We identified ρ to within an uncertainty which is at most $O(1/\sqrt{\tilde{N}})$ using one very long ($\tilde{N} = 10^7$) static tomography experiment, whose overwhelming size ensures accuracy sufficient to calibrate the other experiments, all of which involve $N \leq 3 \times 10^4$ photons.

Our ''standard'' (static) protocol involved repeatedly preparing our target state, collecting $N/3$ photons at each of the three measurement settings corresponding to σ_x, σ_y,

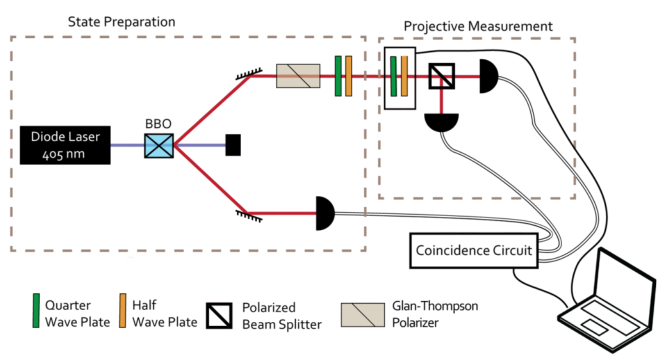

FIG. 3 (color online). Spontaneous parametric down-conversion is performed by pumping a nonlinear BBO crystal with linearly polarized light. One photon is sent directly to a detector as a trigger. A rotation using a quarter-half wave plate combination prepares the other photon in any desired polarization state. Finally, a projective measurement onto any axis of the Bloch sphere is performed by a quarter-half wave plate combination followed by a polarizing beam splitter. The measurement wave plates are connected to a computer to enable adaptation.

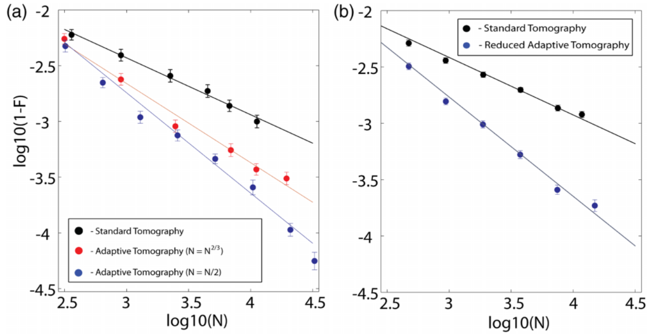

FIG. 4 (color online). Experimental data: (a) The average infidelity $1 - F(\hat{\rho}, \rho)$ for the three tomographic protocols shown in Fig. 2 vs the number of samples N. Each average is over 150 different realizations of the experiment. (b) Average infidelity $1 - F(\hat{\rho}, \rho)$ for standard tomography (black) and reduced adaptive tomography (blue) is plotted vs N. Each average is over 200 different realizations of the experiment; error bars are the standard deviation of the mean of these samples.

and σ_z, and computing $\hat{\rho}_{\mathrm{MLE}}$ as outlined in [29]. Each data point in Fig. 4(a) represents an average over many (~ 150) repetitions.

To do adaptive tomography, we measured $N_0 = N/2$ photons, used the data to generate an ML estimate $\hat{\rho}_0$, then rotated the measurement bases so that one diagonalized $\hat{\rho}_0$. So, if the preliminary estimate is

$$\hat{\rho}_0 = \lambda_1 |\psi_1\rangle\langle\psi_1| + \lambda_2 |\psi_2\rangle\langle\psi_2|,$$

we define $|\psi_{3/4}\rangle = (1/2)(|\psi_1\rangle \pm |\psi_2\rangle)$ and $|\psi_{5/6}\rangle = (1/2)(|\psi_1\rangle \pm i|\psi_2\rangle)$, and then measure the bases $\{\{|\psi_1\rangle, |\psi_2\rangle\}, \{|\psi_3\rangle, |\psi_4\rangle\}, \{|\psi_5\rangle, |\psi_6\rangle\}\}$. We measured the remaining $N - N_0$ photons in these new bases and constructed a final ML estimate using the data from both phases.

We fit a power law ($1 - F = \beta N^p$) to the average infidelity of each protocol [Fig. 4(a)], and found $p = -0.51 \pm 0.02$ for standard tomography, $p = -0.71 \pm 0.04$ for the procedure of Ref. [18], and $p = -0.90 \pm 0.04$ for our adaptive procedure.

Our data generally match the theory; adaptive tomography outperforms standard tomography by an order of magnitude even for modest ($\sim 10^4$) N. Experiments that achieve very low infidelities ($\sim 10^{-4}$) show small but statistically significant deviations from theory, which we believe can be explained by wave plate misalignment. Fluctuations on the order of 10^{-3} radians reproduce the observed deviations in simulations. For a detailed discussion of systematic error and how it affects our results please see [30].

There is an even simpler adaptive procedure. After obtaining a preliminary estimate $\hat{\rho}_0$, we measured *all* of the remaining $N/2$ samples in the diagonal basis of $\hat{\rho}_0$, neglecting the second and third bases presented in the previous section's protocol. This reduced adaptive tomography procedure requires just one extra measurement setting (full adaptive tomography requires three), but achieves the same $O(1/N)$ infidelity [Fig. 4(b)]. The best fits to the exponent p in $1 - F = \beta N^p$ are $p = -0.51 \pm 0.02$ for standard tomography and $p = -0.88 \pm 0.05$ for reduced adaptive tomography [not significantly different from the results shown in Fig. 4(a)]. In higher dimensional systems, reduced adaptive tomography should provide even greater efficiency advantages.

Discussion.—We demonstrated two easily implemented adaptive tomography procedures that achieve $1 - F(\hat{\rho}, \rho) = O(1/N)$ for *every* qubit state. In contrast, any static tomography protocol will yield infidelity $O(1/\sqrt{N})$ for most nearly pure states. Our simplest procedure requires only one additional measurement setting than standard tomography. We see almost no reason not to use reduced adaptive tomography in future experiments.

Previous work [18] optimized average fidelity over Bures measure, a very respectable choice [31–33]. Unfortunately, the "hard-to-estimate" states lie in a thin shell at the surface of the Bloch sphere, whose Bures measure vanishes as $N \to \infty$. So although the scheme with $N_0 \propto N^{2/3}$ proposed in [18] achieves Bures-average infidelity $O(1/N)$, it achieves only $O(1/N^{5/6})$ infidelity for nearly all of the (important) nearly pure states [34].

The $O(1/N)$ infidelity scaling achieved by our scheme is optimal, but the constant can surely be improved—i.e., if our scheme has asymptotic error α/N, a more sophisticated scheme can achieve α'/N with $\alpha' < \alpha$. The absolutely optimal protocol requires joint measurements on all N samples [35], and will outperform any local measurement. There is undoubtedly some marginal benefit to adapting more than once, but we have shown that a single adaptation is sufficient to achieve $O(1/N)$ scaling.

D. H. M., L. A. R., A. D., and A. M. S. thank NSERC and CIFAR for support, and Alan Stummer for designing the coincidence circuit. C. F. was supported in part by NSF Grants No. PHY-1212445 and No. PHY-1005540 and an NSERC PDF. R. B. K. was supported by the LDRD program at Sandia National Laboratories, 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.

*dmahler@physics.utoronto.ca

[1] Z. Hradil, Phys. Rev. A **55**, R1561 (1997).
[2] M. Paris and J. Rehacek, *Quantum State Estimation* (Springer, New York, 2004), Vol. 649.
[3] R. Blume-Kohout, New J. Phys. **12**, 043034 (2010).
[4] R. Blume-Kohout, Phys. Rev. Lett. **105**, 200504 (2010).
[5] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, Phys. Rev. Lett. **105**, 150401 (2010).

[6] M. Christandl and R. Renner, Phys. Rev. Lett. **109**, 120403 (2012).
[7] R. Blume-Kohout, arXiv:1202.5270.
[8] W. K. Wootters, Phys. Rev. D **23**, 357 (1981).
[9] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976), Vol. 84.
[10] C. A. Fuchs, arXiv:quant-ph/9601020.
[11] C. Fuchs and J. van de Graaf, IEEE Trans. Inf. Theory **45**, 1216 (1999).
[12] J. Calsamiglia, R. Muñoz-Tapia, L. Masanes, A. Acin, and E. Bagan, Phys. Rev. A **77**, 032311 (2008).
[13] K. M. R. Audenaert, J. Calsamiglia, R. Muñoz-Tapia, E. Bagan, L. Masanes, A. Acin, and F. Verstraete, Phys. Rev. Lett. **98**, 160501 (2007).
[14] Remarkably, for large N, local measurements can discriminate almost as well as joint measurements on all N samples. If D_Q and D_C are the optimal error exponents for joint and local measurements, respectively, then $(1-F)/2 \leq D_C \leq D_Q \leq 1-F$ [12].
[15] X.-C. Yao, T.-X. Wang, P. Xu, H. Lu, G.-S. Pan, X.-H. Bao, C.-Z. Peng, C.-Y. Lu, Y.-A. Chen, and J.-W. Pan, Nat. Photonics **6**, 225 (2012).
[16] R. D. Gill and S. Massar, Phys. Rev. A **61**, 042312 (2000).
[17] E. Bagan, M. A. Ballester, R. D. Gill, A. Monras, and R. Munoz-Tapia, Phys. Rev. A **73**, 032301 (2006).
[18] E. Bagan, M. A. Ballester, R. D. Gill, R. Munoz-Tapia, and O. Romero-Isart, Phys. Rev. Lett. **97**, 130501 (2006).
[19] F. Huszár and N. M. T. Houlsby, Phys. Rev. A **85**, 052120 (2012).
[20] J. Řeháček, B.-G. Englert, and D. Kaszlikowski, Phys. Rev. A **70**, 052321 (2004).
[21] R. Okamoto, M. Iefuji, S. Oyama, K. Yamagata, H. Imai, A. Fujiwara, and S. Takeuchi, Phys. Rev. Lett. **109**, 130404 (2012).
[22] M. D. de Burgh, N. K. Langford, A. C. Doherty, and A. Gilchrist, Phys. Rev. A **78**, 052122 (2008).
[23] A. J. Scott, J. Phys. A **39**, 13 507 (2006).
[24] Because ρ lies on the state set's boundary, the gradient of F need not vanish in order for $\hat{\rho} = \rho$ to be a local maximum.
[25] T. Sugiyama, P. S. Turner, and M. Murao, New J. Phys. **14**, 085005 (2012).
[26] The "O" notation is necessary here because some of the remaining $N - N_0$ copies may be measured in other bases that make up a complete measurement frame.
[27] J. Löfberg, in *Proceedings of the CACSD Conference, Taipei, 2004.*
[28] All quoted uncertainties herein are 1σ, or 68% confidence intervals. Therefore, we don't expect the "true" value to lie within the error bars more than 68% of the time. Most of the results given here agree with theoretical predictions to within 2σ (95% confidence intervals), a common criterion for consistency between data and theory.
[29] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, Phys. Rev. A **64**, 052312 (2001).
[30] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.111.183601 for a detailed discussion of systematic error and how it affects our results.
[31] M. Hübner, Phys. Lett. A **163**, 239 (1992).
[32] D. Petz and C. Sudar, J. Math. Phys. (N.Y.) **37**, 2662 (1996).
[33] K. Życzkowski and H.-J. Sommers, Phys. Rev. A **71**, 032313 (2005).
[34] Ironically, restricting the problem to pure states falsely trivializes it—the average *and* worst-case infidelity is $O(1/N)$ even for static tomography. The difficulty is not in estimating which pure state we have, but in distinguishing between small eigenvalues ($\lambda = 0$ vs $\lambda = 1/\sqrt{N}$).
[35] S. Massar and S. Popescu, Phys. Rev. Lett. **74**, 1259 (1995).