

Experimental adaptive Bayesian tomography

K. S. Kravtsov,¹ S. S. Straupe,^{2,*} I. V. Radchenko,¹ N. M. T. Houlby,³ F. Huszár,³ and S. P. Kulik²

¹*A. M. Prokhorov General Physics Institute RAS, Moscow, Russia*

²*Faculty of Physics, M. V. Lomonosov Moscow State University, Moscow, Russia*

³*Computational and Biological Learning Laboratory, Department of Engineering, University of Cambridge, Cambridge, United Kingdom*

(Received 26 March 2013; published 28 June 2013)

We report an experimental realization of an adaptive quantum state tomography protocol. Our method takes advantage of a Bayesian approach to statistical inference and is naturally tailored for adaptive strategies. For pure states, we observe close to N^{-1} scaling of infidelity with overall number of registered events, while the best nonadaptive protocols allow for $N^{-1/2}$ scaling only. Experiments are performed for polarization qubits, but the approach is readily adapted to any dimension.

DOI: [10.1103/PhysRevA.87.062122](https://doi.org/10.1103/PhysRevA.87.062122)

PACS number(s): 03.65.Wj, 03.67.–a, 02.50.Ng, 42.50.Dv

Introduction. The main goal of quantum state tomography is to provide an estimate $\hat{\rho}$ for an unknown quantum state ρ based on the data collected in a series of measurements [1]. The estimator is supposed to be close to the real state in some reasonable sense, therefore various notions of statistical distance between quantum states are used [2,3]. One of the most widely used measures of statistical distance is *infidelity* [4], defined as $1 - F(\rho, \hat{\rho}) = 1 - \text{Tr}(\sqrt{\sqrt{\rho}\hat{\rho}\sqrt{\rho}})^2$. The ultimate goal of any tomographic protocol is to minimize this quantity for a fixed overall number N of measurements made. Usually a protocol makes use of some fixed number of measurement settings determined before the actual experiment. For such a protocol, the infidelity scales as $1 - F \sim N^{-1/2}$ for the set of almost-pure states, which is the most interesting case for applications. Although one can significantly alter the prefactor by a clever choice of measurements [5–7], the scaling law for large N is unaffected. A natural question is whether it is possible to beat this limit. The answer turns out to be positive if one allows for adaptivity—the measurement performed at some step of the protocol should be dependent upon the data obtained in the previous ones [8,9].

Here we report an experimental approach to adaptive quantum state tomography based on a recently proposed adaptive Bayesian estimation algorithm [10]. We achieve almost $1/N$ scaling of infidelity for pure states of polarization qubits and demonstrate a clear advantage over ordinary nonadaptive protocols. Our approach is completely different from that of another recent experimental realization [11], where adaptive measurements were used to estimate a single unknown parameter of a quantum state.

Bayesian tomography. Let us start with describing a general framework for quantum state estimation and, in particular, the Bayesian approach. A tomographic protocol is a set of positive operator valued measures (POVMs), $\mathcal{M} = \{\mathbb{M}_\alpha\}$, with index $\alpha \in \mathcal{A}$ numbering the different configurations of the experimental apparatus. In a given configuration, the probabilistic outcome of each measurement γ being observed is determined according to the Born's rule,

$$p(\gamma|\rho, \alpha) = \text{Tr}[M_{\alpha\gamma}\rho], \quad (1)$$

where $M_{\alpha\gamma}$ are POVM elements, obeying $\sum_{\gamma=0}^{\Gamma-1} M_{\alpha\gamma} = I$, and ρ is the density matrix of the state to be determined. The set \mathcal{D} of all outcomes observed in an experiment forms the data set used to estimate density matrix elements. The Bayesian approach to statistical inference dictates the following rules:

(i) A *prior* distribution over the space of density matrices $p(\rho)$ is specified.

(ii) The collected data are used to obtain the posterior distribution $p(\rho|\mathcal{D}) \propto \mathcal{L}(\rho; \mathcal{D})p(\rho)$, where $\mathcal{L}(\rho; \mathcal{D}) = \prod_{n=1}^N p(\gamma_n|\rho, \alpha_n)$ is the *likelihood* function, and it contains our statistical model that encodes probabilistic mapping from the state to the observed data.

(iii) Quantities of interest are estimated using expected values under the posterior distribution: for example, we may obtain the *Bayesian mean estimate* of the state as $\hat{\rho} = \mathbb{E}_{p(\rho|\mathcal{D})}[\rho]$. Variance, infidelity, or any other statistical quantity of interest may be obtained similarly.

The Bayesian approach has many advantages over the maximum-likelihood estimation (MLE) [12], which is more standard in the quantum information community. It offers, in a natural way, a distribution over the space of density matrices, which provides the most complete description of our knowledge about the quantum state, inferred from data \mathcal{D} [13].

Adaptive approach. Bayesian tomography is a natural framework for construction of adaptive estimation protocols. Indeed, the posterior distribution may be updated as soon as one observes some data—in the extreme case, after each measurement—and the new knowledge about the state may be used to select the next measurement setting α in the most optimal way. Choosing the criterion for “optimality” is a task of *optimal experiment design* and may be solved in various ways. In the Bayesian framework, a natural strategy is to choose a measurement that reduces the Shannon entropy of the posterior maximally, $\mathbb{H}[p(\rho|\mathcal{D})]$, which means that our knowledge about the state, obtained after such measurement, is maximized [10]. This may be formulated as choosing a measurement configuration α as a solution to the following optimization procedure:

$$\alpha = \arg \max_{\alpha \in \mathcal{A}} \{\mathbb{H}[p(\rho|\mathcal{D})] - \mathbb{E}_{p(\gamma|\alpha, \mathcal{D})}[\mathbb{H}[p(\rho|\gamma, \alpha, \mathcal{D})]]\}, \quad (2)$$

with $p(\rho|\gamma, \alpha, \mathcal{D})$ being the new posterior after the outcome γ is observed. Note that because we do not know which outcome γ

*straups@yandex.ru

will be observed, we use the *expected* information gain (under the posterior) as our objective.

Approximate inference. One of the principal reasons that Bayesian methods enjoy less popularity in quantum tomography than MLE is the fact that posterior normalization requires computing an (in general, high-dimensional) integral of likelihood function, which is computationally hard. Usually, when faced with intractable Bayesian inference, the posterior is approximated via sampling [13] or by approximating the posterior with a simpler distribution, such as a Gaussian [14].

This computational difficulty is further compounded when one is performing adaptive quantum tomography, in which one must keep track of the current posterior after making n measurements in order to calculate the optimal $(n + 1)$ 'st measurement. To perform inference based on all of the observed data is at best an $O(n)$ operation, which becomes increasingly problematic as the experiment progresses. Fortunately, fast algorithms for solving online Bayesian inference problems exist; they update the posterior after inclusion of each new data point without revisiting all previous data. We briefly review the core idea behind this approach, and refer the reader to [15] for details.

The algorithm is a variant of the sequential importance sampling (SIS) algorithm with resampling. The idea is to construct a particle filter, approximating the posterior with a set of weighted samples, i.e., $p(\rho|\mathcal{D}_n) \approx \sum_{s=1}^S w_s^{(n)} \delta(\rho - \rho_s)$. After each observation, one updates the weights $w_s^{(n)}$; this can be done incrementally, using the current set of particles and weights, and the likelihood corresponding to the new observation. Integration of the likelihood has $O(1)$ cost, which means that it can be applied online at every step of the adaptive protocol, irrespective of the current amount of data collected. If too many particles get almost zero weight, then the particles are resampled and the weights are equalized.

We now return to computation of the objective function for adaptive tomography (2); although this objective is theoretically attractive, it is computationally expensive. It is highly beneficial to work with the following equivalent formulation:

$$\alpha = \arg \max_{\alpha \in \mathcal{A}} \{ \mathbb{H}[p(\gamma|\alpha, \mathcal{D})] - \mathbb{E}_{p(\rho|\mathcal{D})} \mathbb{H}[p(\gamma|\rho, \alpha)] \}. \quad (3)$$

Here, $p(\gamma|\alpha, \mathcal{D}) = \sum_s w_s^{(n)} p(\gamma|\rho_s, \alpha)$ is the *average* predictive probability of outcome γ . In (3), only predictive entropies are required rather than state-space entropies, which is much easier because output space typically has a much lower dimensionality. Furthermore, only the *current* posterior $p(\rho|\mathcal{D})$ is needed.

Simulations. We performed numerical simulations to empirically evaluate the performance of Bayesian adaptive tomography. For our performance metric, we use the mean infidelity as measured against the true state, $\bar{\rho}: 1 - \hat{F}(\rho, \bar{\rho}) = \mathbb{E}_{p(\rho|\mathcal{D}_n)} [1 - F(\rho, \bar{\rho})]$. Note that Bayesian mean $1 - \hat{F}(\rho, \bar{\rho})$ is a “fairer” score than the fidelity of a point estimate, e.g., the posterior mean (i.e., $[1 - F(\mathbb{E}_{p(\rho|\mathcal{D}_n)}[\rho], \bar{\rho})]$), because the posterior mean can happen to be correct even if we have no knowledge about the system, e.g., for a completely mixed state. The Bayesian estimator rewards posterior distributions that are both centered in the correct location *and* have low variance.

To achieve statistically significant results, we perform multiple runs within each simulation, each with a different

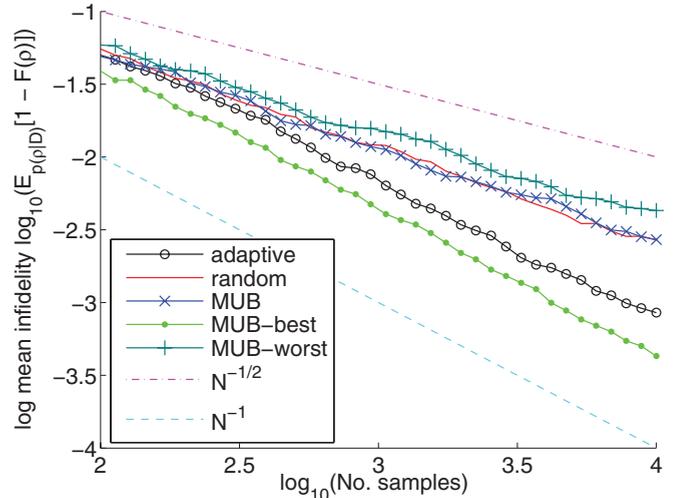


FIG. 1. (Color online) Simulated tomography using three measurement selection methods: randomly sampled (red continuous line), MUBs (blue \times), and fully adaptive Bayesian tomography (black \circ). For these methods, the true state is random and pure, and the results presented here are the average of 20 independent runs. Overlaid dashed lines indicate the power-law fit. The functions $1 - F = N^{-1/2}$ (magenta, dash-dotted line) and $1 - F = N^{-1}$ (cyan, dashed line) are shown for comparison. To account for the state dependence of MUB tomography, we also present its performance for the “worst” and “best” (see text) true states (dark green $+$ and light green \bullet , respectively).

random pure “true state.” We compare this to random uniform measurements and mutually unbiased bases (MUBs), which are the optimal fixed measurements (in terms of information gain) [16]. The random and adaptive schemes are independent of the true state (the algorithms are agnostic to rotations of the Stokes coordinates), but fixed MUBs are not. Drawing intuition from the fact that the adaptive scheme predominantly selects measurements that align with the true state, we find that the “best case” for MUB tomography is when the state is aligned with one of the MUB measurements, and the worst case is when the state is equally biased to all measurements, i.e., $\{\sigma_x = \sigma_y = \sigma_z = 1/\sqrt{3}\}$.

The results are presented in Fig. 1. We fit a power law, $1 - \hat{F} \propto N^a$, to the data. Random tomography yields $a = -0.66 \pm 0.03$, which is in reasonable correspondence with the expected asymptotic scaling $N^{-1/2}$. However, adaptive tomography performs close to the N^{-1} level with average $a = -0.90 \pm 0.03$. In its most favorable scenario, MUBs also perform close to the N^{-1} rate (with a small multiplicative constant improvement over the adaptive scheme). In practice, the optimal MUBs are unknown *a priori*. In the case of arbitrarily chosen MUBs, we observe that on average the rate is near $N^{-1/2}$: $a = -0.64 \pm 0.05$.

Experimental imperfections. In practice, quantum tomography is subject to experimental noise. This noise is not modeled in the likelihood function given by Born’s rule (1). In our experiment, we have identified two major additional sources of noise: detector dark counts with detector-specific rates and attenuation in both channels due to detector inefficiency and losses or reflections at the optical elements.

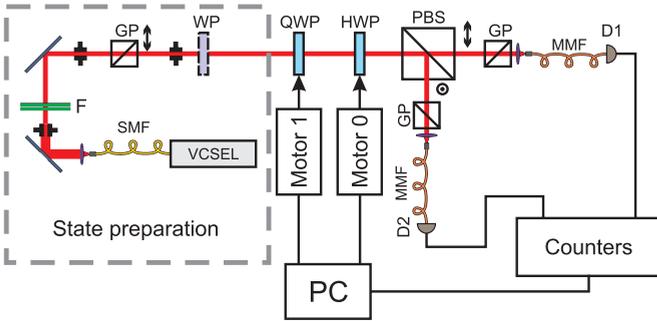


FIG. 2. (Color online) Experimental setup. An attenuated laser is used as a source. The polarization state is prepared by a custom wave plate and analyzed by a sequence of quarter- and half-wave plates, followed by a polarizing beam splitter and two single-photon counters. Wave plates are rotated by electronically controlled step-motor drivers to allow for adaptivity.

A popular approach to modeling dark counts is to model the observed state as a linear mixture of the true state and the maximally mixed state [17]. Although with this assumption one can describe certain simple noise sources, such as equal dark counts arriving with equal rates at each detector, we address the problem more directly. We assume that the production of photons by the laser source and the generation of dark counts by the detectors can be modeled using independent homogeneous Poisson processes with rate parameters λ_s for the source and λ_d^γ for each detector. These rates are estimated *a priori* in a preliminary experiment. From these assumptions, one can derive the new likelihood function,

$$\mathbb{P}(\gamma|\rho, \alpha, \lambda_s, \lambda_d^1, \dots, \lambda_d^\gamma) = \frac{\text{Tr}[M_{\alpha\gamma}\rho]\lambda_s + \lambda_d^\gamma}{\lambda_s + \sum_\gamma \lambda_d^\gamma}. \quad (4)$$

To deal with channel inefficiency, we assume a fixed probability of a photon being “lost” from each channel, denoted by $1 - \eta_\gamma$. These probabilities are also estimated *a priori*. The likelihood now becomes

$$\mathbb{P}(\gamma|\rho, \alpha, \eta_1, \dots, \eta_\gamma) = \frac{\text{Tr}[M_{\alpha\gamma}\rho]\eta_\gamma}{\sum_\gamma \text{Tr}[M_{\alpha\gamma}\rho]\eta_\gamma}. \quad (5)$$

Note that in both cases, both the numerator and denominator contain only linear terms in the additional parameters (λ, η). Therefore, one only requires estimates of the ratio of the dark count rates to the source rate $\lambda_d^\gamma/\lambda_s$, and, for single-qubit tomography, the ratio of the efficiencies of the two channels η_1/η_2 .

Experiment. Figure 2 depicts our experimental setup. We use a CW 850 nm vertical-cavity surface-emitting laser (VCSEL) coupled to a single-mode fiber as a laser source. The radiation is attenuated to the single-photon level by a set of neutral density filters (F) and additionally spatially filtered with small iris apertures. The input polarization state is defined by a Glan-Taylor prism (GP) with high extinction ratio (more than 6000:1); the prism transmits horizontally polarized light, which may be transformed to some arbitrary state with a proper choice of a quartz wave plate (WP).

The measurement scheme consists of an effective zero-order quarter-wave plate (QWP) and a half-wave plate (HWP). The plates are rotated by step-motor-driven stages, with a

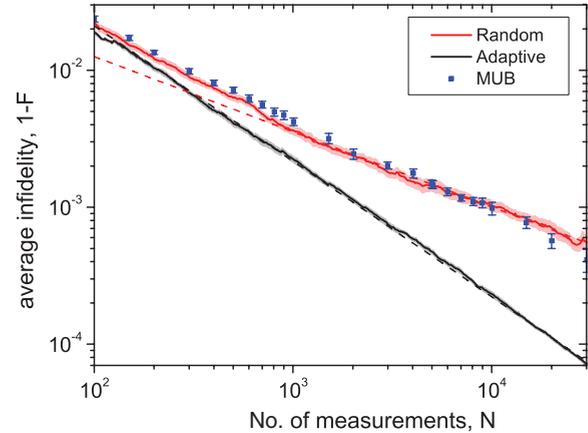


FIG. 3. (Color online) Experimental results: mean infidelity $1 - \hat{F}(\rho, \hat{\rho})$ with current Bayesian mean estimate $\hat{\rho}$ for random measurements (red, upper line), adaptive measurements (black, lower line), and measurements in MUBs [blue (dark gray) points]. Data points are averages over 10 experimental runs; shaded areas and error bars show the standard deviation of the mean. Dashed straight lines are power-law fits.

minimal angular step of 0.1° . The zero position is controlled by a Hall sensor providing uncertainty of 0.2° . We clean up the polarization states in the output channels of a polarization beam-splitting (PBS) cube with two additional Glan-Taylor prisms to ensure a high extinction ratio. Effectively, this is equivalent to introducing some losses in the *ideal* PBS cube without altering the output polarization states. In each channel, photons are coupled to multimode fibers and detected by single-photon counting modules (SPCMs), D1 and D2 (Perkin-Elmer). Electronic pulses from SPCMs are sent to in-house built counters, which may operate in two regimes: count for a fixed period of time or count until the specified number of counts is reached.

We show the advantage of adaptive tomography using a direct comparison in the Bayesian framework. In the adaptive estimation scheme, we used two strategies, namely, adaptation after every single detection event and adapting after blocks of measurements increasing in size with the amount of data collected as $\lceil N/100 \rceil$, and found no statistically significant difference. Figure 3 shows the dependence of mean infidelity, $1 - \hat{F}(\rho, \hat{\rho}) = \mathbb{E}_{p(\rho|D_n)} [1 - F(\rho, \hat{\rho})]$, with current estimate $\hat{\rho}$ (for which we used the Bayesian posterior mean) on the number of measurements performed. Note that we intentionally *did not average* over many realizations at each step of the algorithm; the data points in Fig. 3 are averaged over several *full runs* of the experiment. The convergence rate behaves regularly from run to run. The power-law fits are $a = -0.60 \pm 0.05$ and $a = -0.62 \pm 0.02$ for the random and MUB protocols, respectively, and $a = -0.98 \pm 0.01$ for the adaptive protocol.

In a real world application of tomography, the “true” state is unknown, and the Bayesian estimate given above is the only figure of merit at hand. However, in our experiments, we can estimate the prepared state by averaging over many runs of adaptive protocol and comparing the convergence with our simulations. Scaling of the posterior mean infidelity to the (estimated) true state with N is depicted in Fig. 4. Power-law

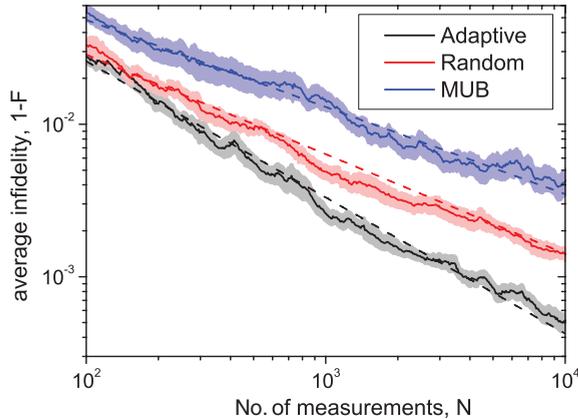


FIG. 4. (Color online) Experimental results: mean infidelity $1 - \hat{F}(\rho, \bar{\rho})$ with “true” state $\bar{\rho}$ for random measurements (red, middle line), adaptive measurements (black, lower line), and measurements in MUBs [blue, upper line]. Data points are averages over 10 experimental runs; shaded areas show the standard deviation of the mean. Dashed straight lines indicate the power-law fits.

fits give $a = -0.64 \pm 0.02$ and $a = -0.60 \pm 0.05$ for the random and MUB protocols, respectively, while the adaptive strategy yields $a = -0.92 \pm 0.03$.

Within the errors bands, the experimentally obtained scaling laws agree with simulations. We observed much more regular behavior with the adaptive protocol, i.e., trajectories of individual runs are close to each other, while for random and MUB measurements, the behavior resembles more of a random walk with a much larger variance. Our model does not take into account systematic errors caused, for example, by inaccuracies in wave-plate rotation. However, for the reached values of infidelities on the order of 10^{-4} – 10^{-3} , we did not observe any deviations from the expected behavior and we were not able to identify the influence of systematic errors. Further investigation with much larger statistics is required to address this issue.

Finally, let us note that using an attenuated laser source is equivalent to a true single-photon source for the purposes of this particular single-qubit experiment.

Conclusion. Our experiments clearly demonstrate the advantages of adaptive strategies in quantum state tomography. Besides the aforementioned favorable properties, the Bayesian approach is convenient from a purely practical point of view. It does not require any additional precomputation, and since posterior updates may be easily carried out after a single detection event, we expect that this approach will be particularly useful in the case of extremely weak signals. The N^{-1} scaling of infidelity in the adaptive case is the theoretical limit for any tomographic protocol, and further improvement may only affect prefactors in this power law. Simulation results show that our strategy of choosing adaptively between general measurements outperforms any nonadaptive protocol, especially for the most interesting case of nearly-pure states. Our experimental results taken for MUB and random measurements show good agreement with simulations, while direct experimental comparison with more sophisticated nonadaptive strategies is underway. Although in this paper we report single-qubit tomography experiments, the developed adaptive protocol is general and may be used for states in Hilbert space of any dimensions. Experimental realizations of adaptive tomography for two-qubit polarization states and higher-dimensional systems (such as spatial modes of a biphoton field) will be reported elsewhere.

Note added. Recently, we became aware of a highly relevant work [18] taking a different approach to adaptive state estimation and achieving similar performance.

Acknowledgments. This work was supported in part by the Federal Program of the Russian Ministry of Education and Science (Grant No. 8393) and state Contract No. 11.519.11.4009, European Union Seventh Framework Programme under Grant Agreement No. 308803 (project BRISQ2), ERA.Net RUS Project NANOQUINT, the NATO project Grant No. EAP.SFPP 984397 “Secure Communication Using Quantum Information Systems,” and by RFBR Grants No. 12-02-31041 and No. 12-02-31792. S.S.S. is grateful to the “Dynasty” foundation for financial support. N.H. is grateful for support from the Google European Doctoral Fellowship programme.

-
- [1] M. G. A. Paris and J. Řeháček, *Quantum State Estimation*, Lecture Notes in Physics, Vol. 649 (Springer-Verlag, Berlin, Heidelberg, 2004).
- [2] S. L. Braunstein and C. M. Caves, *Phys. Rev. Lett.* **72**, 3439 (1994).
- [3] I. Bengtsson and K. Życzkowski, *Geometry of Quantum States* (Cambridge University Press, Cambridge, UK, 2006).
- [4] Sometimes, e.g., in [7], this quantity is called *fidelity loss*.
- [5] J. Řeháček, B.-G. Englert, and D. Kaszlikowski, *Phys. Rev. A* **70**, 052321 (2004).
- [6] M. D. de Burgh, N. K. Langford, A. C. Doherty, and A. Gilchrist, *Phys. Rev. A* **78**, 052122 (2008).
- [7] Y. I. Bogdanov, G. Brida, I. D. Bukeev, M. Genovese, K. S. Kravtsov, S. P. Kulik, E. V. Moreva, A. A. Soloviev, and A. P. Shurupov, *Phys. Rev. A* **84**, 042108 (2011).
- [8] D. G. Fischer, S. H. Kienle, and M. Freyberger, *Phys. Rev. A* **61**, 032306 (2000).
- [9] T. Hannemann, D. Reiss, C. Balzer, W. Neuhauser, P. E. Toschek, and C. Wunderlich, *Phys. Rev. A* **65**, 050303 (2002).
- [10] F. Huszár and N. M. T. Houlby, *Phys. Rev. A* **85**, 052120 (2012).
- [11] R. Okamoto, M. Iefuji, S. Oyama, K. Yamagata, H. Imai, A. Fujiwara, and S. Takeuchi, *Phys. Rev. Lett.* **109**, 130404 (2012).
- [12] Z. Hradil, *Phys. Rev. A* **55**, R1561 (1997).
- [13] Robin Blume-Kohout, *New J. Phys.* **12**, 043034 (2010).
- [14] K. M. Audenaert and S. Scheel, *New J. Phys.* **11**, 023028 (2009).
- [15] A. Doucet, N. De Freitas, N. Gordon *et al.*, *Sequential Monte Carlo Methods in Practice* (Springer, New York, 2001).
- [16] W. Wootters and B. Fields, *Ann. Phys.* **191**, 363 (1989).
- [17] A. I. Lvovsky, H. Hansen, T. Aichele, O. Benson, J. Mlynek, and S. Schiller, *Phys. Rev. Lett.* **87**, 050402 (2001).
- [18] D. H. Mahler, L. A. Rozema, A. Darabi, C. Ferrie, R. Blume-Kohout, and A. M. Steinberg, arXiv:1303.0436v1.