

PAPER • OPEN ACCESS

Demonstration of machine-learning-enhanced Bayesian quantum state estimation

To cite this article: Sanjaya Lohani *et al* 2023 *New J. Phys.* **25** 083009

View the [article online](#) for updates and enhancements.

You may also like

- [Practical Bayesian tomography](#)
Christopher Granade, Joshua Combes and D G Cory
- [Simple informative prior distributions for Type A uncertainty evaluation in metrology](#)
Anthony O'Hagan and Maurice Cox
- [Bayesian error regions in quantum estimation I: analytical reasonings](#)
Yong Siah Teo, Changhun Oh and Hyunseok Jeong



PAPER

Demonstration of machine-learning-enhanced Bayesian quantum state estimation

OPEN ACCESS

RECEIVED

13 March 2023

REVISED

28 June 2023

ACCEPTED FOR PUBLICATION

12 July 2023

PUBLISHED

2 August 2023

Original Content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.



Sanjaya Lohani^{1,*} , Joseph M Lukens^{2,3} , Atiyya A Davis¹, Amirali Khannejad¹, Sangita Regmi¹ ,
Daniel E Jones⁴ , Ryan T Glasser⁵, Thomas A Searles^{1,*}  and Brian T Kirby^{4,5,*} 

¹ Department of Electrical & Computer Engineering University of Illinois Chicago, Chicago, IL 60607, United States of America

² Research Technology Office and Quantum Collaborative, Arizona State University, Tempe, AZ 85287, United States of America

³ Quantum Information Science Section, Oak Ridge National Laboratory, Oak Ridge, TN 37831, United States of America

⁴ DEVCOM Army Research Laboratory, Adelphi, MD 20783, United States of America

⁵ Tulane University, New Orleans, LA 70118, United States of America

* Authors to whom any correspondence should be addressed.

E-mail: slohan3@uic.edu, tsearles@uic.edu and brian.t.kirby4.civ@army.mil

Keywords: quantum tomography, Bayesian techniques, machine learning

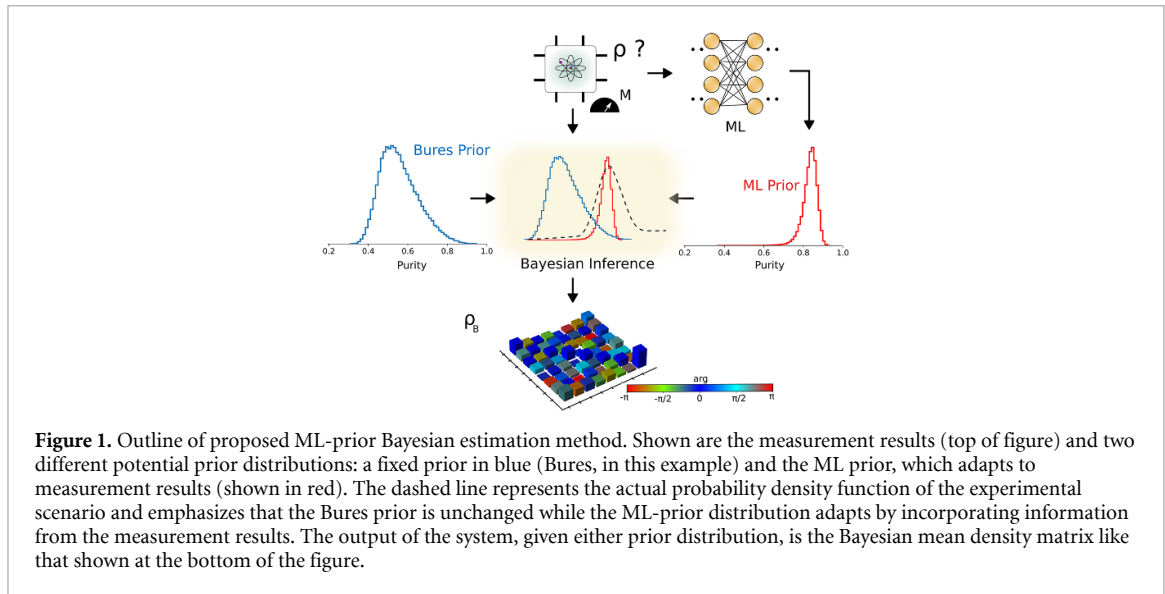
Abstract

Machine learning (ML) has found broad applicability in quantum information science in topics as diverse as experimental design, state classification, and even studies on quantum foundations. Here, we experimentally realize an approach for defining custom prior distributions that are automatically tuned using ML for Bayesian quantum state estimation methods that generally better conform to the physical properties of the underlying system than standard fixed prior distributions. Previously, researchers have looked to Bayesian quantum state tomography for advantages like uncertainty quantification, the return of reliable estimates under any measurement condition, and minimal mean-squared error. However, practical challenges related to long computation times and conceptual issues concerning how to incorporate prior knowledge most suitably can overshadow these benefits. Using both simulated and experimental measurement results, we demonstrate that ML-defined prior distributions reduce net convergence times and provide a natural way to incorporate both implicit and explicit information directly into the prior distribution. These results constitute a promising path toward practical implementations of Bayesian quantum state tomography.

1. Introduction

Quantum state tomography (QST) is a process for characterizing the state of a quantum system through a sequence of measurements. In general, QST consists of two distinct steps: the repeated measurement of identically prepared systems and the entirely classical post-processing step of determining the density matrix most consistent with the measurement results. However, given the substantial experimental challenges associated with collecting measurements of high-dimensional quantum systems, the computational resources involved with state reconstruction have only relatively recently become a significant bottleneck in their own right [1, 2]. Various approaches for performing state estimation given a set of measurement results have appeared in the literature, including maximum likelihood estimation (MLE) [2–6], compressive sensing [7, 8], machine learning (ML) [9–17], and Bayesian inference [18–24]. The existing inference methods define a trade-space between different characteristics such as general applicability and assumptions about the system, computation time, and accuracy guarantees.

Among the available state estimation approaches, Bayesian estimation enjoys such features as automatic uncertainty quantification, the return of reliable estimates under any measurement condition, and minimal mean-squared error [25]. Unfortunately, these advantages come with the practical computational challenges related to the calculation of high-dimensional integrals, although the estimation of these integrals through Monte Carlo methods has proven useful [18–24]. Further, efforts to include prior information in the



Bayesian estimation framework are often ad hoc, relying on custom likelihood functions or the direct manipulation of prior distributions. While the former has shown some promise in terms of improving efficiency [22, 23], it is conceptually unsatisfying, as in principle all prior information should enter through the prior distribution. On the other hand, current approaches for including prior information through direct engineering of prior distributions are coarse and rely on assumptions about a system that may change over time, such as the expected rank of output states, while also requiring manual tuning [16, 26].

Here we propose and experimentally realize a method, pictured in figure 1, for mitigating Bayesian state reconstruction's practical and conceptual challenges through ML-defined prior distributions that automatically adapt to input data and, in general, better conform to the physical properties of the underlying system than standard prior distributions. In particular, we use a pre-trained neural network reconstruction system to define a custom prior distribution for any arbitrary input dataset by biasing a general distribution of random density matrices toward a rapid initial reconstruction of the state. We find a net reduction in the time required for Bayesian inference to reach a given reconstruction fidelity, even when including the additional time required by the pre-trained neural network to define the custom prior distribution. We first demonstrate the efficacy of our approach on simulated measurement results for systems of up to four qubits and then experimentally with a fiber-based source of polarization-entangled photon pairs. In all cases, our ML-enhanced Bayesian inference approach reaches a lower-error estimate with reduced computational resources than a traditional prior distribution.

Our approach is sufficiently general to encompass several previous methods for manually tuned prior distributions and allows for either direct or automatic tuning of prior distribution parameters. It can incorporate any resource-efficient estimation of a density matrix—not only those coming from ML, but also other common methods such as linear inversion and compressive sensing—into a prior distribution to jumpstart Bayesian calculations closer to the converged result and reduce computational resource requirements. Consequently, the broad framework provided by our approach should offer a variety of pathways for computational and conceptual enhancements in Bayesian QST.

2. Bayesian quantum state reconstruction

Intuitively, reconstruction of quantum states using Bayesian inference takes some prior distribution (an initial set of beliefs regarding what states might describe the system, possibly based on prior information or intentionally chosen to be uniform) and, through Bayes' rule, incorporates measurements to update the prior and generate a posterior distribution. Hence, unlike other approaches to quantum state reconstruction, such as MLE [3, 4], Bayesian inference produces a complete distribution of possible states whose average minimizes mean squared error [25]. In order to perform this Bayesian updating, we first require a parameterized description of a density matrix, given by $\rho(\mathbf{w})$, where all possible values of the vector of parameters \mathbf{w} result in a valid physical density matrix—i.e. unit-trace, Hermitian, and positive-semidefinite. Further, we must define a prior distribution $\pi_0(\mathbf{w})$ over these parameters, possibly containing descriptive information.

Then, following an experiment, Bayes' rule defines the posterior distribution for \mathbf{w} as

$$\pi(\mathbf{w}) = \frac{1}{\mathcal{Z}} L_{\mathcal{D}}(\mathbf{w}) \pi_0(\mathbf{w}), \quad (1)$$

where the likelihood $L_{\mathcal{D}}(\mathbf{w}) \propto P(\mathcal{D}|\mathbf{w})$ (the probability of observing dataset \mathcal{D} given parameters \mathbf{w}) and \mathcal{Z} is a normalizing constant (which in practice often does not need to be determined explicitly). In the context of quantum state reconstruction the likelihood can be expressed as $L_{\mathcal{D}}(\mathbf{w}) = \prod_{m=1}^M \text{Tr} \rho(\mathbf{w}) \Lambda_m$ where each of M measurements, described by an operator Λ_m , is performed on identically prepared input states. As QST in practical settings is often based on the simpler class of projective measurements, we can further express the likelihood as

$$L_{\mathcal{D}}(\mathbf{w}) = \prod_{m=1}^M \langle \psi_m | \rho(\mathbf{w}) | \psi_m \rangle, \quad (2)$$

with $|\psi_m\rangle$ the eigenstate observed in measurement m . The Bayesian mean estimate of the reconstructed quantum density matrix ρ follows as

$$\rho_B = \int d\mathbf{w} \pi(\mathbf{w}) \rho(\mathbf{w}). \quad (3)$$

The computation of this integral is frequently a severe roadblock to implementation. While several methods for estimating this integral exist, we utilize preconditioned Crank–Nicolson Markov chain Monte Carlo (MCMC) in this investigation [23, 27].

Regardless of the specific prior distribution, as long as it provides nonzero support to the full Hilbert space and sufficient measurement results are available, Bayes' rule will converge to a sharply peaked single answer. However, in several important contexts, the selection of a prior distribution can significantly impact the performance of a Bayesian reconstruction system, for example whenever experimental data are severely limited or incomplete. In addition, even when unlimited data are available, typically the further the prior distribution deviates from the ground truth, the longer it will take computational methods to sample the posterior.

Previous work considering the selection of prior distributions for Bayesian reconstruction has often defaulted to standard distributions based on fair sampling according to the Bures [28] or Hilbert–Schmidt [29] measures. Unfortunately, these distributions are, on average, significantly more mixed than many relevant experimental scenarios and hence have motivated the use of alternatives such as the Mai–Alquier (MA) distribution [22], which can be tuned to favor lower-rank states than those obtained on average from Bures or Hilbert–Schmidt metrics [16, 26]. Recent results show that manually tuning the MA distribution with even very coarse knowledge, such as the mean state purity expected in an experimental scenario, can improve the performance of Bayesian reconstruction methods [26]. Unfortunately, these approaches require manually updating the prior distribution if experimental conditions change.

3. Bayesian inference with an ML-defined prior

Significant recent research effort has focused on leveraging ML to perform quantum state reconstruction. While different results reveal that ML-based reconstruction has potential advantages over conventional approaches when reconstructing data that include experimental errors, missing measurements, or high statistical noise, the primary motivator has been to reduce computation times [16]. In particular, ML methods that utilize pre-trained networks effectively ‘frontload’ expensive computations into training the network so that, once trained, it can perform inference indefinitely without further computationally intensive training [30]. However, despite these potential advantages, ML-based reconstruction utilizing pre-training does not provide absolute performance guarantees, error quantification, or a natural way to incorporate prior information without further training.

Here we propose leveraging the main advantage of ML-based quantum state reconstruction—rapid inference—to automatically define prior distributions for use with Bayesian inference. In this way, we can translate some of the reduced computation time from pre-trained networks to lower the net time required for Bayesian inference. In particular, we propose using a pre-trained ML-based reconstruction system to first supply a rapid estimate ρ_{ML} based on input measurement data, which—assuming a sufficiently high fidelity—serves to bias the prior distribution such that it better reflects the physical properties of the underlying system than would a standard prior distribution. Then, similar to the MA distribution [22], we generate the complete prior distribution through a Dirichlet-weighted convex sum of ρ_{ML} and Haar-distributed random pure states.

The Dirichlet distribution is defined for vectors $\mathbf{x} = (x_1, \dots, x_K)$ whose elements belong to the open $K - 1$ simplex. The probability density function of the Dirichlet distribution is defined by

$$\text{Dir}(\mathbf{x}|\boldsymbol{\alpha}) = \frac{\Gamma\left(\sum_{i=1}^K \alpha_i\right)}{\prod_{i=1}^K \Gamma(\alpha_i)} \prod_{i=1}^K x_i^{\alpha_i-1}, \tag{4}$$

where $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K)$ with all $\alpha_i \geq 0$ defines the concentration parameters and $\Gamma(\cdot)$ denotes the standard gamma function; $\boldsymbol{\alpha}$ are free parameters that determine the moments of the distribution.

One special case is $\alpha_1 = \alpha_a$ and $\alpha_{2 \leq i \leq K} = \alpha_b$. Then the sum of the α_i becomes $\alpha_0 = \alpha_a + (K - 1)\alpha_b$, and the first-order moments are

$$\begin{aligned} E(X_1) &= \frac{\alpha_a}{\alpha_0} \\ E(X_{2 \leq i \leq K}) &= \frac{\alpha_b}{\alpha_0}. \end{aligned} \tag{5}$$

We can subsequently define a family of distributions of D -dimensional quantum states consisting of the convex sum of ρ_{ML} and $K - 1$ Haar-random pure states $|\psi_i\rangle$ as

$$\rho = x_1 \rho_{\text{ML}} + \sum_{i=2}^K x_i |\psi_i\rangle \langle \psi_i|. \tag{6}$$

The degrees of freedom of this family of distributions are the system dimension D , the number of terms in the convex sum K , and the two concentration parameters α_a and α_b . For simplicity, and motivated by previous related work using similar distributions of quantum states [26], we set $K = D + 1$. This choice means the sum of Haar-random pure states in ρ has D terms and hence, regardless of ρ_{ML} , ρ is full-rank with high probability. Full-rank estimates are desirable in this context to construct prior distributions that are capable of representing any state of a D -dimensional Hilbert space.

As a particularly convenient reparametrization, we consider the sum α_0 above and define the ratio μ as

$$\mu = \frac{\alpha_a}{\alpha_b} = \frac{E(X_1)}{E(X_{2 \leq i \leq K})}, \tag{7}$$

so that the expectation values become

$$\begin{aligned} E(X_1) &= \frac{\mu}{K + \mu - 1} \\ E(X_{2 \leq i \leq K}) &= \frac{1}{K + \mu - 1} \end{aligned} \tag{8}$$

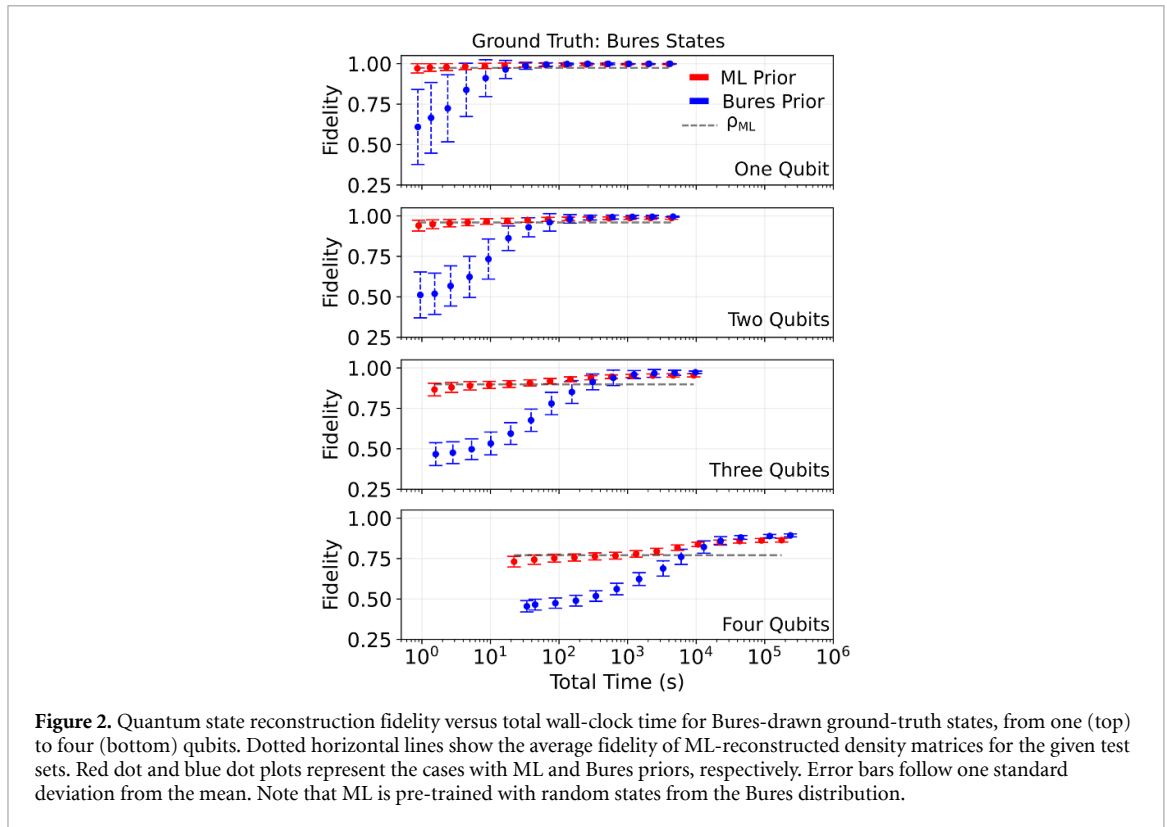
and variances

$$\begin{aligned} \text{var}(X_1) &= \frac{\mu(K - 1)}{(\alpha_0 + 1)(\mu + K - 1)^2} \\ \text{var}(X_{2 \leq i \leq K}) &= \frac{\mu + K - 2}{(\alpha_0 + 1)(\mu + K - 1)^2}. \end{aligned} \tag{9}$$

Thus, the mean values are determined only by μ , while both μ and α_0 contribute to the variances.

In particular, α_0 determines the average sparsity of the Dirichlet vector \mathbf{x} while μ determines the bias of the distribution toward ρ_{ML} . The sparsity effects of α_0 are most readily apparent when $\mu = 1$, meaning the Dirichlet parameters are unbiased. Here a flat distribution of the entire K -simplex is obtained when $\alpha_0 = K$; a tendency for all components of \mathbf{x} to equal $1/K$ occurs when $\alpha_0 \gg K$; and finally, the distribution becomes sparse as $\alpha_0 \rightarrow 0$, with one component equal to unity (at random) and all others zero. Alternatively, to see the impact of altering μ we can express the mean of the ML-prior distribution as $E(\rho) \propto \mu \rho_{\text{ML}} + \frac{(K-1)}{D} I_D$, with I_D being the $D \times D$ identity and where we have used the expectation values from equation (8). Unlike a uniform prior that averages to the maximally mixed state $\frac{1}{D} I_D$, this state is biased toward ρ_{ML} with a weight determined by μ . For example, for a fixed α_0 , as $\mu \rightarrow \infty$ the distribution approaches a sharp peak around ρ_{ML} ; the opposite limit $\mu \rightarrow 0$ returns a uniform prior with $E(\rho) = \frac{1}{D} I_D$.

Interestingly, the state construction defined by equation (6) with Dirichlet-distributed coefficients (equation (4)) is capable of incorporating both *implicit* and *explicit* prior information, where for our purposes ‘implicit’ denotes information integrated into the prior without any deliberate or obvious specification of beliefs, and ‘explicit’ refers to prior information encoded quantitatively through e.g. tunable



parameters. The information contained within ρ_{ML} can therefore be viewed as *implicit* since ρ_{ML} is automatically generated based on the measurement inputs and hence does not necessarily rely on prior information about the experimental situation. Alternatively, we can manually tune μ and α_0 to reflect *explicit* prior information. As described above these two parameters allow us to control both the sparsity (related to the mean purity of the distribution) and bias toward ρ_{ML} (related to our confidence in the estimate ρ_{ML}).

While explicitly tunable priors have been considered in previous Bayesian QST studies [20, 22, 23, 26], implicit information has to our knowledge remained unexplored. From a mathematical and computational viewpoint, both types of information sources are treated on equal footing by Bayes' theorem; they represent prior knowledge that influences the final result. On the other hand, from an experimental and practical standpoint, they are fundamentally distinct, as explicit information is directly connected to beliefs specified by the experimentalist, while implicit information is supplied by the ML system acting on the experimental results, without any human input—apart from, of course, any guidance provided during the training and optimization of the neural network itself.

To analyze the interplay of both types of information in Bayesian state reconstruction, we first test the performance of an ML-defined-prior using simulated measurement results for systems of one to four qubits. We consider ground-truth states sampled uniformly according to the Bures measure and reconstruct the states using either our ML-defined-prior or a more traditional Bures prior. For each ground truth state (200 sampled at each qubit number), we simulate the outcomes for 16 000 randomly chosen Pauli strings and use these results to perform reconstruction. In the case of the ML prior, we first train and save an ML model as described in appendix A with quantum states generated using a fixed distribution (Bures in this section, MA in the next [22]). Note that the underlying neural network used to generate ρ_{ML} is pre-trained and remains unchanged regardless of experimental inputs; hence any adaptability of the ML prior is a reflection of the inherent flexibility of the ML approach and not a case-by-case incorporation of information.

The reconstruction fidelities with ML-based (with $\mu = 25$ and $\alpha_0 = 11.6$) and Bures priors are shown in figure 2 for successively doubled MCMC chain lengths from 2^5 to 2^{18} , where the x -axis includes the time required to compute ρ_{ML} . All times are based on a desktop computer with an 11th Generation Intel i7-11 700 K processor operating at 3.60 GHz with 16 threads and 32 GB of RAM. A GeForce RTX 3080 graphical processing unit is additionally leveraged by the neural network producing ρ_{ML} . For a fair comparison, the ML-prior times do omit the neural network training period, since this represents a one-time cost which is not repeated. The averages shown in figure 2, and the accompanying error bars, are calculated from the results for all 200 randomly sampled Bures ground-truth states. In all cases the ML-defined and Bures priors each eventually reach similar reconstruction fidelities, before the Bures prior marginally

outperforms the ML one, consistent with the known mean-squared optimality of a prior which precisely matches the ground-truth draws [25]. However, the ML-defined prior significantly outperforms Bures at shorter times: in all four scenarios in figure 2 the net time to reach a low-error estimate is reduced using the ML-defined prior compared to the ideal situation of prior and ground-truth states being sampled from the same distribution.

Intuitively, the ML-defined prior biases the Bayesian reconstruction process to start nearer to the final reconstruction, as highlighted by the horizontal lines in figure 2 that mark the mean fidelity of ρ_{ML} , the neural-network-produced guess based on the measurement results; the more accurate this initial guess, the better the head start. Ultimately, Bayesian inference with either the ML-based or traditional Bures prior succeeds in surpassing this original ML estimate, but the former is able to do so more rapidly than the latter which begins further from the final reconstruction. However, as we will see in the next section, the benefit of rapid convergence from biasing the distribution toward the initial guess can come at the cost of ultimately limiting the peak reconstruction fidelity. Hence, a tradeoff between improving the short-computation-time reconstruction fidelities and maximizing the long-computation-time fidelities will emerge.

4. Performance in optical experiment

To demonstrate the efficacy of our ML-defined prior approach in a real-world scenario, we apply it to states generated by a commercially available polarization-entangled photon pair source [31]. The experimental setup includes an entangled photon source (EPS) connected to two detector stations (DSs) with telecom optical fibers. A schematic of the setup is shown in figure 3(a), along with the real part of an example density matrix in figure 3(b), calculated from experimental data by the ML-prior Bayesian estimation method with ($\mu = 25, \alpha_0 = 11.6$) (details below). The EPS outputs photon pairs via four-wave mixing in a dispersion-shifted fiber (DSF) [32] pumped by a 50 MHz pulsed fiber laser centered at 1552.52 nm. The DSF is cooled to -86°C in a laboratory-grade freezer to reduce Raman-scattered noise photons. For the experiment performed here, the average number of photon pairs generated per pulse is set to ~ 0.1 . The signal and idler photons are entangled in polarization by arranging the DSF in a Sagnac loop with a polarizing beam splitter and are then spectrally demultiplexed into 100 GHz-spaced ITU outputs [33].

The DSs consist of 50 MHz-gated single-photon detectors (SPDs) preceded by polarization analyzers. The SPDs have detection efficiencies of $\sim 20\%$ and dark count probabilities of $\sim 4 \times 10^{-5}$ per gate. Full polarization state tomography is performed by collecting measurements in 36 different projections. Maximum coincidence rates of up to several thousand per measurement are achieved by performing each of the 36 measurements over 10 million detector gates. Tomography is performed 1000 consecutive times, resulting in a total experiment duration of ~ 4 hr. The density matrix is calculated for all 1000 runs using Bayesian inference with ML-based and Bures priors. As in the previous section, the red curve in figure 4(a) is the reconstruction using the ML prior with $\mu = 25$ and $\alpha_0 = 11.6$, but now utilizing a neural network trained on MA-distributed states with $\alpha = 0.4$ to infer ρ_{ML} , motivated by previous work that found this distribution balances reconstruction fidelity across the purity spectrum of arbitrary two-qubit states [16, 26]. Even though we *test* inference on experimental data in this section, we still *train* our neural network using simulated data in order to avoid the inclusion of false correlations from experimental error and keep our system more general; simulated data are often significantly more accessible than comparable quantities of accurately labeled experimental data. In addition, and again analogous to section 3, in blue we include the curve showing the reconstruction fidelity assuming a Bures prior. For all fidelity calculations in figure 4, we compare the inferred results against the state obtained from MLE as a proxy for the (unknown) ground truth.

As before, the ML prior significantly outperforms the Bures prior for short timescales but with the Bures prior ultimately reaching a higher peak fidelity at long computation times. In some sense, it is surprising that the Bures prior outperforms the ML prior here, considering that—unlike in the previous section—neither distribution matches that of the experimental ground truth. Hence, we would intuitively assume the early lead by the ML prior to persist. One likely explanation for the observed outcome is that the ML prior is in effect *overly* biased toward the initial guess. To test this, we also include the green curve ($\mu = 0.25, \alpha_0 = 1.7$), for which the ML prior is defined with a significantly reduced bias. This reduced affinity toward ρ_{ML} lowers the initial reconstruction fidelity but, in the limit of long computation times, ultimately reaches a noticeably higher fidelity than the previous ML prior (red) and lower uncertainty than the Bures prior (blue). This interplay highlights the flexibility afforded by our approach via explicit tuning of μ and α_0 , as well as the tradeoff between optimizing short-computation-time fidelity ($\mu = 25$ case) and long-computation-time fidelity ($\mu = 0.25$ case). Incidentally, optimal balancing of this tradeoff is highly application-dependent, based on the relative importance of speed versus maximum-attainable accuracy; e.g. one could envision

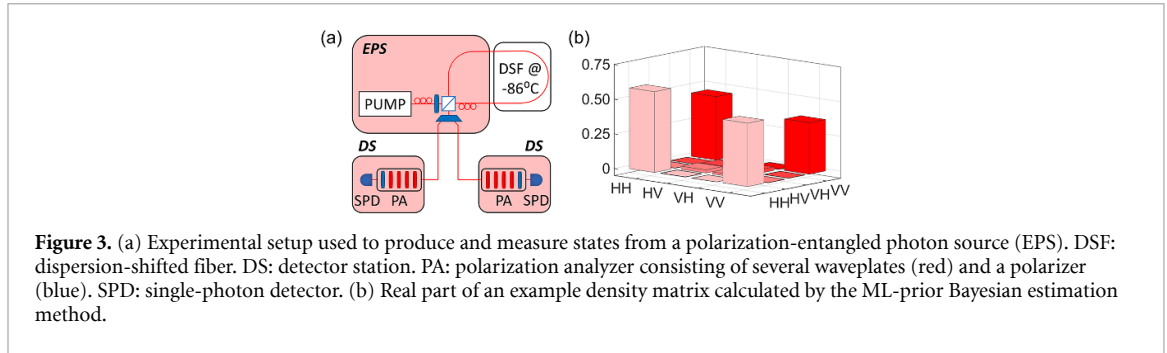


Figure 3. (a) Experimental setup used to produce and measure states from a polarization-entangled photon source (EPS). DSF: dispersion-shifted fiber. DS: detector station. PA: polarization analyzer consisting of several waveplates (red) and a polarizer (blue). SPD: single-photon detector. (b) Real part of an example density matrix calculated by the ML-prior Bayesian estimation method.

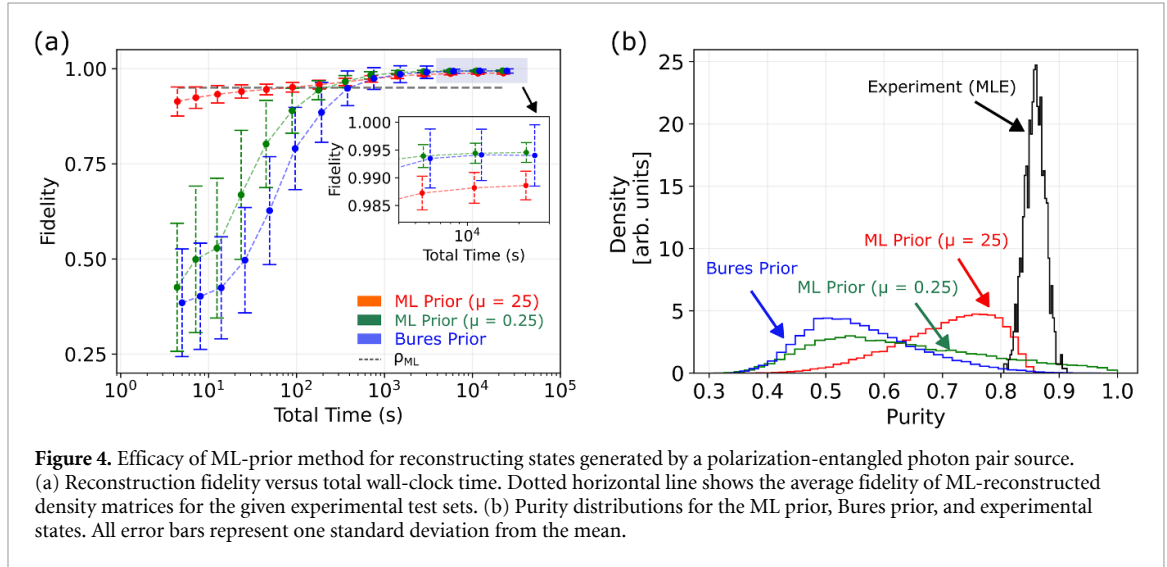


Figure 4. Efficacy of ML-prior method for reconstructing states generated by a polarization-entangled photon pair source. (a) Reconstruction fidelity versus total wall-clock time. Dotted horizontal line shows the average fidelity of ML-reconstructed density matrices for the given experimental test sets. (b) Purity distributions for the ML prior, Bures prior, and experimental states. All error bars represent one standard deviation from the mean.

speed proving critical in a real-time measurement-based feedback system, with accuracy instead assuming prominence in post-experiment analyses. Ultimately, this tradeoff could be quantified through a cost function encompassing both effects, tailored to the needs of a specific application.

To further understand the behavior in figure 4(a), we also include in figure 4(b) numerically generated probability density functions of purity for the Bures prior (blue), an exemplar of each ML-prior distribution given a randomly chosen set of experimental data (red and green), and the density matrices obtained from MLE for all 1000 collected experimental datasets (black). Since the Bures prior is fixed regardless of the experimental scenario we see very little overlap in the probability density functions, as compared with the ML-prior distributions which successfully adapt based on the rapid reconstruction of ρ_{ML} . Interestingly, the lower-bias ML prior ($\mu = 0.25$) covers the experimental distribution more uniformly than the high-bias case ($\mu = 25$), which may contribute to its higher final fidelity in figure 4(a).

Considering such priors more broadly, whenever the actual test states align with a fiducial distribution (e.g. Bures in figure 2), no ML tuning can surpass the long-computation-time accuracy obtainable with the standard prior, although modifications can lead to computational speedups. However, in most *experimental* systems, such standard prior distributions do not reflect the actual variations in the states produced, unlocking the possibility for enhancements in both short- and long-time accuracy with tailored priors. In particular, the physical system of figure 3 is not pure—and hence using Haar-distributed pure states would be inappropriate—but it also is not nearly as mixed as states from the Bures distribution (with similar results for the Hilbert–Schmidt distribution [16, 26]). Therefore, from a physical perspective, our results confirm the interpretation from [26] that even very general pieces of prior information can improve the performance of state reconstruction systems, but with the additional insight that such prior information can be incorporated in an *implicit* manner as well.

5. Discussion

In Bayesian inference, the prior distribution ideally encapsulates all prior knowledge about a system. In practice, when performing QST using Bayesian methods, one often selects a prior based on broad experience with a given system and not on a case-by-case basis. This paper has considered a more nuanced approach where a prior distribution is selected after measurement results are collected—and an ML estimate ρ_{ML} obtained—but before Bayesian reconstruction. Ostensibly, such a data-informed prior would seem to complicate the typical Bayesian paradigm that clearly separates initial knowledge (in the prior) from observations (in the likelihood). Nevertheless, our procedure poses no problems mathematically: from the perspective of Bayes' rule, ρ_{ML} is simply a high-dimensional hyperparameter whose origin has no impact on the method, even if it is in fact obtained anachronistically *after* the experiment in question.

Moreover, our approach for defining biased prior distributions based on an initial estimate can be straightforwardly applied regardless of the method of generating the proposal. We chose to train our network using the Bures distribution in section 3 and MA distribution in section 4; however, the approach described is equally applicable regardless of how the states used to train the network are obtained, with the only difference potentially being the quality of the ρ_{ML} reconstruction.

Indeed, our method does not even require ML but can utilize any approach for generating proposed initial reconstructed density matrices. ML-based strategies are attractive due to their resource-efficiency after the initial training period when compared to other common techniques (e.g. MLE). However, different resource-efficient approaches, such as linear inversion or compressive sensing, could be readily substituted. The primary considerations when enlisting an alternative method for an initial point estimate are the resources required and the fidelity of the reconstruction. While linear inversion has similar resource requirements to ML-based approaches, prior work indicates that ML methods outperform linear inversion in terms of reconstruction fidelity [14]—hence our use of ML here.

It is important to reiterate that the implicit prior defined by equation (6) does not circumvent known optimality conditions for more standard priors. For example, in the tests of figure 2(a) on Bures-drawn states, Bayesian inference with a Bures prior ultimately converges to higher average fidelities with respect to the ground truth than inference with an ML prior. Such results are consistent with the mean-squared-error optimality of the Bayesian mean estimator when tested on ground-truth states drawn from the prior [25]; in other words, irrespective of the process used to tune the prior with ML, the fundamental limit set by a ground-truth-matched prior cannot be surpassed. Of course, in many practical settings the prior is simply a convenient choice and may differ widely from the distribution of states tested, so that adhering strictly to a fiducial distribution like Bures may not be particularly meaningful. In any case, the primary advantages of the ML prior lie in reducing the total computational cost to reach a given inference fidelity, which can be attributed to its initial weighting of the MCMC chain closer to the final converged answer.

Other approaches to Bayesian quantum state reconstruction have also aimed to reduce required computational resources by including dataset-specific information through a pseudo-likelihood based on a simple loss function [22, 23], rather than a physics-inspired likelihood like Born's rule in equation (2). While the pseudo-likelihood also leverages low-resolution state reconstructions to improve computational efficiency, the ML prior proposed here is arguably preferable conceptually in that it introduces this information as prior knowledge instead of through modifying the likelihood. Yet regardless of any theoretical distinctions, both approaches offer complementary procedures to utilize some initial estimate and improve thereon with Bayesian inference, either through the prior or the likelihood.

In summary, we have leveraged ML to tune prior distributions in Bayesian quantum state estimation. The low computational overhead of our neural network allows us to outperform standard priors in total time to convergence, even when including the computation time required to define the custom prior. Aside from demonstrating this for systems of up to four qubits in simulation, we also apply our approach to two-qubit optical systems and find the advantages to persist. By offering a general framework for computational speedups, our work should provide a practical path toward the more widespread adoption of Bayesian inference techniques for quantum state reconstruction.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

The code and a web application to generate all datasets are openly available at the following URLs: <https://github.com/slohani-ai/machine-learning-for-physical-sciences> and <https://mlphys.streamlit.app/>.

Acknowledgments

Work by UIC was supported in part by the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Co-design Center for Quantum Advantage (C2QA) under Contract Number DE-SC0012704. A portion of this work was performed at Oak Ridge National Laboratory, operated by UT-Battelle for the U.S. Department of Energy under Contract No. DE-AC05-00OR22725. J M L acknowledges funding by the U.S. Department of Energy, Office of Science, Advanced Scientific Computing Research, through the Early Career Research Program (ERKJ353), and the Quantum Collaborative, led by Arizona State University. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein. This material is also based upon work supported by, or in part by, the Army Research Laboratory and the Army Research Office under Contract/Grant Numbers W911NF-19-2-0087 and W911NF-20-2-0168.

Author Contributions

S L, J M L, and B T K conceived, and all authors contributed to, the design of the research and the analysis of results. S L, S R implemented the neural networks and ran computational experiments. S L and A K wrote the code using python. A A D and D E J performed the experiments. S L, B T K, and J M L drafted the manuscript, with input from all authors. T A S and B T K supervised all aspects of the work.

Conflicts of interest

The authors declare no conflicts of interest.

Appendix A. ML details

A.1. Model

We design a convolutional neural network (CNN) with a convolutional unit of kernel size (2, 2), strides of 1, ReLU as an activation function, and 25 filters. The output of the CNN is fed into a layer that performs pooling with a pool size (2, 2), followed by a second convolutional unit of the same configuration. Then, we combine two dense layers, followed by a dropout layer with a rate of 0.5, which is then attached to an output layer predicting τ -vectors (Cholesky coefficients of the density matrix [34, 35]). We keep the depth (number of hidden layers) of the network fixed for all qubit cases while increasing the number of neurons in the dense layers as shown in table 1.

A.2. Training the model

To generate the training sets, we sample 40 000 states (for each qubit number) from either the Bures distribution (for section 3) or the MA distribution with $\alpha = 0.4$ (for section 4; see appendix B), and take 35 000 for training and 5000 for validation. For each state we compute the ground-truth τ vector and simulate 16 000 random Pauli measurements [16] (one measurement per each random basis selection), feeding these measurements into the network. The mean-square loss between the target and predicted τ is evaluated and fed back to optimize the network's trainable parameters using the adaptive gradient (Adagrad) optimizer. We use a learning rate of 0.01 and batch size of 100 for up to 75 epochs to train the network. Additional details related to the impact and ultimate selection of neural network hyperparameters in quantum information tasks can be found in appendix figures 1 and 5 of [16], where we show how both the learning rate and the number of network parameters alter the reconstruction fidelity. At the output layer is attached a pipeline that rearranges the predicted τ -vectors into density matrices. The pipeline is built into the same graph of the network for the purposes of computing the average fidelity per epoch for cross-validation and outputting the density matrix directly to avoid postprocessing. As an example, in the two-qubit case, the predicted τ -vectors are rearranged to lower triangular matrices T expressed as

$$T = \begin{bmatrix} \tau_0 & 0 & 0 & 0 \\ \tau_4 + i\tau_5 & \tau_1 & 0 & 0 \\ \tau_{10} + i\tau_{11} & \tau_6 + i\tau_7 & \tau_2 & 0 \\ \tau_{14} + i\tau_{15} & \tau_{12} + i\tau_{13} & \tau_8 + i\tau_9 & \tau_3 \end{bmatrix}. \quad (\text{A1})$$

Table 1. Number of neurons in fully connected layers as a function of the number of qubits in the state. The $[x, y]$ notation indicates the width (x) and length (y) of the two-dimensional array of input data after reshaping.

Qubit number	Input shape	Dense neurons	Dense neurons	Output neurons
1	[2, 3]	250	150	4
2	[6, 6]	750	450	16
3	[6, 36]	2500	1000	64
4	[36, 36]	4500	2500	256

The density matrices follow as $\tilde{\rho} = \frac{TT^\dagger}{\text{Tr}(TT^\dagger)}$. Note that the physicality of $\tilde{\rho}$ is guaranteed through the Cholesky decomposition, which ensures positive semidefiniteness [4, 35]. Finally, at the end of the network, the fidelity F between the predicted density matrix $\tilde{\rho}$ and the target ρ is evaluated as $F = \left| \text{Tr} \sqrt{\sqrt{\tilde{\rho}} \rho \sqrt{\tilde{\rho}}} \right|^2$. An in-depth description of the network architecture is given in [30].

Appendix B. MA distribution

This distribution was originally introduced as a prior for Bayesian QST [22], has been applied in several subsequent studies [23, 36–38], and was recently utilized in [26] to generate training sets for ML-based state reconstruction methods. The MA distribution is defined as a mixture of Haar-random pure states with coefficients drawn from the Dirichlet distribution:

$$\rho = \sum_{i=1}^K x_i |\psi_i\rangle \langle \psi_i|, \quad (\text{B1})$$

where the vector \mathbf{x} is a random variable distributed according to $\text{Dir}(\mathbf{x}|\alpha)$. This construction matches that of our ML-based prior in equation (6) of the main text, but now with the $i=1$ term a random pure state rather than the fixed ρ_{ML} . If we specialize to the symmetric case $\alpha = (\alpha, \dots, \alpha)$ and assume a Hilbert space dimension of D , the expectation value of the purity is

$$\text{E}_{\text{MA}} [\text{Tr}(\rho^2)] = \frac{D + \alpha(D + K - 1)}{D(1 + \alpha K)}. \quad (\text{B2})$$

Finally, we note that in [26] strong evidence was presented that the MA distribution reduces to the HS distribution for $K = \alpha = D$.

Appendix C. Parameterizing the ML prior

For the ML prior, we define $\mathbf{w} = (\mathbf{y}, \mathbf{z}_2, \dots, \mathbf{z}_K)$, where $\mathbf{y} = (y_1, \dots, y_K)$ is a real K -dimensional vector of positive scalars, and each \mathbf{z}_j is a D -dimensional complex vector. The prior follows $\pi_0(\mathbf{w}) \propto y_1^{\alpha_1-1} e^{-y_1} \prod_{i=2}^K y_i^{\alpha_i-1} e^{-y_i} e^{-\frac{1}{2} \mathbf{z}_i^\dagger \mathbf{z}_i}$, and we define $\rho(\mathbf{w})$ according to

$$\rho(\mathbf{w}) = \frac{y_1}{\sum_j y_j} \rho_{\text{ML}} + \sum_{i=2}^K \left(\frac{y_i}{\sum_j y_j} \right) \frac{\mathbf{z}_i \mathbf{z}_i^\dagger}{|\mathbf{z}_i|^2}. \quad (\text{C1})$$

In the main text we take $K = D + 1$. For convenience we utilize independent and unnormalized parameters; our combined gamma and complex-normal prior ensures that the normalized entities $\mathbf{y}/\sum_j y_j$ and $\mathbf{z}_i/|\mathbf{z}_i|$ are Dirichlet- and Haar-distributed, respectively, as required [22].

Appendix D. Bures distribution

A random quantum state ρ from the Bures ensemble can be generated according to

$$\rho = \frac{(\mathbb{1} + U) G G^\dagger (\mathbb{1} + U^\dagger)}{\text{Tr}[(\mathbb{1} + U) G G^\dagger (\mathbb{1} + U^\dagger)]}, \quad (\text{D1})$$

where G is a random matrix from the Ginibre ensemble and U is a Haar-distributed random unitary from $U(D)$ [39]. Both matrices can be represented as functions of a parameter vector $\mathbf{w} = (w_1, \dots, w_{2D^2})$, with each element independently distributed according to a complex standard normal distribution $w_k \sim \mathcal{CN}(0, 1)$. We assign D^2 of the components to populate the $D \times D$ matrix G ; the remaining D^2 elements comprise a second, independent Ginibre matrix which is then fed into the algorithm of [40] to produce the Haar-random unitary U . Thus the Bures prior can be written as $\pi_0(\mathbf{w}) \propto \prod_{k=1}^{2D^2} e^{-\frac{1}{2} |w_k|^2}$.

ORCID iDs

Sanjaya Lohani  <https://orcid.org/0000-0003-0699-0669>
Joseph M Lukens  <https://orcid.org/0000-0001-9650-4462>
Sangita Regmi  <https://orcid.org/0000-0002-7857-4253>
Daniel E Jones  <https://orcid.org/0000-0002-9854-5767>
Thomas A Searles  <https://orcid.org/0000-0002-0532-7884>
Brian T Kirby  <https://orcid.org/0000-0002-2698-9887>

References

- [1] Hou Z *et al* 2016 *New J. Phys.* **18** 083036
- [2] Smolin J A, Gambetta J M and Smith G 2012 *Phys. Rev. Lett.* **108** 070502
- [3] Hradil Z 1997 *Phys. Rev. A* **55** R1561
- [4] James D F V, Kwiat P G, Munro W J and White A G 2001 *Phys. Rev. A* **64** 052312
- [5] James D F, Kwiat P G, Munro W J and White A G 2005 *Asymptotic Theory of Quantum Statistical Inference: Selected Papers* (World Scientific) pp 509–38
- [6] Teo Y S, Zhu H, Englert B-G, Řeháček J and Hradil Z 2011 *Phys. Rev. Lett.* **107** 020404
- [7] Gross D, Liu Y-K, Flammia S T, Becker S and Eisert J 2010 *Phys. Rev. Lett.* **105** 150401
- [8] Liu W-T, Zhang T, Liu J-Y, Chen P-X and Yuan J-M 2012 *Phys. Rev. Lett.* **108** 170403
- [9] Carrasquilla J, Torlai G, Melko R G and Aolita L 2019 *Nat. Mach. Intell.* **1** 155
- [10] Cha P, Ginsparg P, Wu F, Carrasquilla J, McMahon P L and Kim E-A 2021 *Mach. Learn.: Sci. Technol.* **3** 01LT01
- [11] Tiunov E S, Tiunova V, Ulanov A E, Lvovsky A and Fedorov A 2020 *Optica* **7** 448
- [12] Torlai G *et al* 2019 *Phys. Rev. Lett.* **123** 230504
- [13] Neugebauer M, Fischer L, Jäger A, Czischek S, Jochim S, Weidemüller M and Gärttner M 2020 *Phys. Rev. A* **102** 042604
- [14] Lohani S, Kirby B T, Brodsky M, Danaci O and Glasser R T 2020 *Mach. Learn. Sci. Technol.* **1** 035007
- [15] Xu Q and Xu S 2018 Neural network state estimation for full quantum state tomography (arXiv:1811.06654)
- [16] Lohani S, Lukens J M, Glasser R T, Searles T A and Kirby B T 2022 *Mach. Learn. Sci. Technol.* **3** 04LT01
- [17] Ahmed S, Muñoz C S, Nori F and Kockum A F 2021 *Phys. Rev. Lett.* **127** 140502
- [18] Blume-Kohout R 2010 *New J. Phys.* **12** 043034
- [19] Seah Y-L, Shang J, Ng H K, Nott D J and Englert B-G 2015 *New J. Phys.* **17** 043018
- [20] Granade C, Combes J and Cory D G 2016 *New J. Phys.* **18** 033024
- [21] Williams B P and Lougovski P 2017 *New J. Phys.* **19** 043003
- [22] Mai T T and Alquier P 2017 *J. Stat. Plan. Inference* **184** 62
- [23] Lukens J M, Law K J H, Jasra A and Lougovski P 2020 *New J. Phys.* **22** 063038
- [24] Lukens J M, Law K J and Bennink R S 2021 *npj Quantum Inf.* **7** 113
- [25] Robert C P and Casella G 1999 *Monte Carlo Statistical Methods* (Springer)
- [26] Lohani S, Lukens J M, Jones D E, Searles T A, Glasser R T and Kirby B T 2021 *Phys. Rev. Res.* **3** 043145
- [27] Cotter S L, Roberts G O, Stuart A M and White D 2013 *Statist. Sci.* **28** 424
- [28] Sommers H-J and Życzkowski K 2003 *J. Phys. A: Math. Gen.* **36** 10083
- [29] Życzkowski K and Sommers H-J 2003 *J. Phys. A: Math. Gen.* **36** 10115
- [30] Lohani S, Searles T A, Kirby B T and Glasser R 2021 *IEEE Trans. Quantum Eng.* **2** 2103410
- [31] NuCrypt 2022 Quantum optical instrumentation (available at: <http://nucrypt.net/quantum-optical-instrumentation.html>)
- [32] Fiorentino M, Voss P L, Sharping J E and Kumar P 2002 *IEEE Photonics Technol. Lett.* **14** 983
- [33] Wang S X and Kanter G S 2009 *IEEE J. Sel. Top. Quantum Electron.* **15** 1733
- [34] Banaszek K, D'Ariano G M, Paris M G A and Sacchi M F 1999 *Phys. Rev. A* **61** 010304
- [35] Altepeter J B, Jeffrey E R and Kwiat P G 2005 *Adv. At. Mol. Opt. Phys.* **52** 105
- [36] Lu H-H, Simmerman E M, Lougovski P, Weiner A M and Lukens J M 2020 *Phys. Rev. Lett.* **125** 120503
- [37] Lingaraju N B, Lu H-H, Seshadri S, Leaird D E, Weiner A M and Lukens J M 2021 *Optica* **8** 329
- [38] Alshowkan M *et al* 2021 *PRX Quantum* **2** 040304
- [39] Al Osipov V, Sommers H-J and Życzkowski K 2010 *J. Phys. A: Math. Theor.* **43** 055302
- [40] Mezzadri F 2007 *Not. Am. Math. Soc.* **54** 592 (arXiv:math-ph/0609050)