

# Adaptive Bayesian Quantum Tomography

F. Huszár and N.M.T. Housby

*Computational and Biological Learning Lab, Department of Engineering, University of Cambridge, Cambridge, UK*

In this paper we revisit the problem of optimal design of quantum tomographic experiments. In contrast to previous approaches where an optimal set of measurements is decided in advance of the experiment, we allow for measurements to be adaptively and efficiently re-optimised depending on data collected so far. We develop an adaptive statistical framework based on Bayesian inference and Shannon’s information, and demonstrate a ten-fold reduction in the total number of measurements required as compared to non-adaptive methods, including mutually unbiased bases.

Quantum tomography is a valuable tool in quantum information processing, being essential for characterisation of quantum states, gates, and measurement equipment. Quantum state tomography (QST) aims to determine an unknown quantum state from the outcome of measurements performed on an ensemble of identically prepared systems. Measurements in quantum systems are non-deterministic, hence QST is a classical statistical estimation problem. Full tomography is inherently resource-intensive: even in moderately sized systems, the number of measurements required is often prohibitive. There is a need for methods that allow for shorter experiments. Optimal experiment design (OED) aims to achieve this by selecting cleverly which measurements to use during the experiment.

Most existing approaches to OED determine, prior to collecting data, an optimal set of measurements to be used throughout the experiment. In this sense, whenever they exist, mutually unbiased bases (MUBs) are known to be optimal [1, 2]. Research since has focused mainly on proving or disproving existence of, and implement MUBs in various dimensions [2–4]. Other work, [5, 6] considered OED based on the Cramér-Rao bound. Here we argue that these approaches, including MUBs, provide only a partial solution to the problem of optimal experiment design inasmuch as they do not take partial data into account. If we are allowed to revise our choice of measurements during the experiment based on data collected so far, we may be in a better position to reduce redundancy. This strategy is generally known as active learning or adaptive sampling; such adaptive experimental design have been applied in a number of other fields, such as clinical trials [7], cognitive science [8] and computer vision [9]. In physics, this approach has been referred to as self-learning measurements [10, 11]. However, due to the expensive computations that are involved, these methods have been restricted to two dimensional pure quantum states, or very few measurements. Recently advances in Bayesian methods allow us to build a fast, online algorithm that allows self-learning in arbitrary dimensions with many measurements.

Here we propose a new algorithmic framework that we call *Adaptive Bayesian Quantum Tomography* (ABQT), that builds on full Bayesian inference and Shannon information. To achieve adaptivity in practice, we need a fast algorithm for performing Bayesian state reconstruction from partial data after each measurement. Current

sampling methods such as in [12] are inappropriate as their costs increase with the number of measurement configurations tried so far. As a solution, we present a sequential importance sampling scheme [13], that does not suffer from this. We then use the developed algorithm in conjunction with an information theoretic objective to adaptively optimise measurements. We assess the relative performance of our adaptive method in Monte Carlo simulations of qubit systems, and demonstrate a ten-fold reduction in the number of measurements needed for full tomography of two-qubit pure states. We also investigate the trade off between entangling and separable measurements in multipartite systems. Our central finding is that via adaptive tomography one can achieve, and even surpass, the statistical efficiency of MUB tomography using only separable measurements, that require experimental apparatus that is substantially easier to build using current technology.

*Quantum state tomography* involves determining from experimental data the quantum state,  $\rho$ , of a system by performing measurements on several identical copies. For a  $D$ -dimensional system ( $D = 2^m$  for  $m$ -qubit systems),  $\rho$  is an  $D \times D$  complex-valued density matrix.  $\rho$  has to be Hermitian and have unit trace, so  $D^2 - 1$  real degrees of freedom must be estimated. The apparatus for a tomographic experiment may be configured in several different ways; we use  $\alpha \in \mathcal{A}$  to index all accessible configurations. Each measurement configuration  $\alpha$  is characterised by a positive operator-valued measure (POVM). For each configuration, a measurement results in observing one of a finite number,  $\Gamma$ , of distinguishable outcomes. A POVM is defined by a set,  $\mathbb{M}_\alpha$ , of Hermitian operators  $M_{\alpha\gamma}$ , indexed by possible outcomes  $\gamma \in \{1, \dots, \Gamma\}$ , satisfying  $\sum_{\gamma=1}^{\Gamma} M_{\alpha\gamma} = I$ . These POVMs jointly constitute our tomographic model  $\mathcal{M} = \{\mathbb{M}_\alpha : \alpha \in \mathcal{A}\}$  and determine the probability of observing outcome  $\gamma$  in configuration  $\alpha$  when the measured system is in state  $\rho$  via Born’s rule:

$$\mathbb{P}(\gamma|\rho, \alpha; \mathcal{M}) = \text{tr} \{M_{\alpha\gamma}\rho\}$$

State reconstruction has been approached with several methods, the most popular being maximum likelihood estimation (MLE). MLE finds a physically feasible state  $\rho$  that is most likely to have produced the observed data,

$\mathcal{D}$ , by maximising the likelihood:

$$\mathcal{L}(\rho; \mathcal{D}) = \prod_{n=1}^N \mathbb{P}(\gamma_n | \rho, \alpha_n) = \prod_{\alpha \in \mathcal{A}} n_\alpha! \prod_{\gamma=1}^{\Gamma} \frac{\text{tr}\{M_{\alpha\gamma}\rho\}^{n_{\alpha\gamma}}}{n_{\alpha\gamma}!} \quad (1)$$

where  $n_\alpha$  is the number of times configuration  $\alpha$  was used,  $n_{\alpha\gamma}$  is the number of times outcome  $\gamma$  was observed in configuration  $\alpha$ . All probabilities are conditional on  $\mathcal{M}$ , for brevity this is omitted. A well-known drawback of MLE is that it often yields rank-deficient estimates, and thus assigns zero predictive probability to certain observations [12]. This seems an unreasonable conclusion on the basis of a finite sample. Additionally, MLE provides no measure of uncertainty in its point estimate.

More sophisticated methods for quantum tomography use Bayesian inference and suffer from neither of these problems [12, and refs.]. In Bayesian inference a prior probability density,  $p(\rho)$ , over feasible states is specified. This prior is then augmented with the likelihood from Eqn. (1) using Bayes' rule to yield a posterior distribution:

$$p(\rho | \mathcal{D}) \propto \mathcal{L}(\rho; \mathcal{D}) p(\rho) \quad (2)$$

Should we want a point estimate, we may report, say, the Bayesian mean estimate (BME) which is known to maximise expected operational divergences [12, 14]. But importantly, Bayesian inference also provides *error bars*, and more: the posterior captures richly our remaining uncertainty in the true state having seen the data  $\mathcal{D}$ .

For Bayesian inference one has to provide the prior  $p(\rho)$ , which is typically chosen to be non-informative or uniform. Here we adopt the representation and prior introduced in [12], that treats our original system of interest as part of a larger,  $D \times K$  dimensional bipartite system. Our prior over the mixed state  $\rho$  is then defined as the measure induced by the uniform (Haar) measure over pure states in  $D \times K$  dimensions. It is easy to see that, tracing out the  $K$  dimensional ancillary part leaves us with a rank- $K$  mixed state  $\rho$ . Thus, by tuning this parameter we can trade off between computational efficiency and estimation accuracy, in a similar manner to compressed sensing [15].

Unfortunately, normalisation of the posterior distribution (Eqn. (2)) becomes analytically intractable, and therefore we have to approximate it, usually via Markov chain Monte Carlo (MCMC) methods. Several MCMC approaches have been suggested in this context [12, and refs. therein]. These methods require evaluation of the full likelihood (1), which has  $\mathcal{O}(n)$  cost with the number of different configurations used so far. This is undesirable for adaptive tomography, where inference has to be performed after each measurement. To address this problem we developed a fast sequential importance sampling (SIS) algorithm, with  $\mathcal{O}(1)$  likelihood evaluation cost. As we are not aware of this approach being used in the context of QST, we briefly explain the basic version below. The interested reader is referred to [13] for a thorough overview.

In SIS, one keeps track of a number of samples, often called particles,  $\rho_s$ , ( $s = 1 \dots S$ ) and corresponding weights  $w_s$ , ( $\sum_s w_s = 1$ ) which are updated sequentially, every time a new measurement is made. Assume that after  $n$  measurements, having observed data  $\mathcal{D}_n$ , the particles and weights  $w_s^{(n)}$  constitute an approximation to the posterior:

$$p(\rho | \mathcal{D}_n) \approx \sum_{s=1}^S w_s^{(n)} \delta(\rho - \rho_s) \quad (3)$$

Using this approximation, and Bayes' rule, one can derive an approximation to the next posterior, after observing a new outcome  $\gamma_{n+1}$  in configuration  $\alpha_{n+1}$ , as:

$$\begin{aligned} p(\rho | \alpha_{n+1}, \gamma_{n+1}, \mathcal{D}_n) &= \frac{\mathbb{P}(\gamma_{n+1} | \rho, \alpha_{n+1}) p(\rho | \mathcal{D}_n)}{\int \mathbb{P}(\gamma_{n+1} | \rho, \alpha_{n+1}) p(\rho | \mathcal{D}_n) d\rho} \quad (4) \\ &\approx \sum_{s=1}^S \underbrace{\frac{\mathbb{P}(\gamma_{n+1} | \rho_s, \alpha_{n+1}) w_s^{(n)}}{\sum_{r=1}^S \mathbb{P}(\gamma_{n+1} | \rho_r, \alpha_{n+1}) w_r^{(n)}}}_{w_s^{(n+1)}} \delta(\rho - \rho_s) \end{aligned}$$

The new weights  $w_s^{(n+1)}$  are the renormalised product of our current weights  $w_s^{(n)}$  and observation probabilities  $\mathbb{P}(\gamma_{n+1} | \rho_s, \alpha_{n+1})$ . This update is fast, and only requires computing one term of the full likelihood, thus its complexity is independent of how many configurations have been tried before. This computational efficiency comes at a price; as time progresses, several weights decay to almost zero, and thus the quality of our approximation drops. This issue can be detected and handled by monitoring the effective sample size and resampling appropriately [13].

Having discussed our method for estimating the state based on partial data, we now turn to the problem of optimal experiment design. Different state determination schemes have different OED strategies associated with them. Maximum likelihood methods usually use some form of the Cramér-Rao bound [5, 6]. Bayesian experiment design on the other hand is based on Shannon information [1, 16]. The posterior characterises our remaining uncertainty in the parameter, and this uncertainty can be quantified using Shannon's entropy. A sensible aim is to pick an experimental configuration  $\alpha$ , such that after observing the outcome  $\gamma$ , the entropy  $\mathbb{H}$  of the new posterior is reduced the most:

$$\arg\max_{\alpha \in \mathcal{A}} \{ \mathbb{H}[p(\rho | \mathcal{D})] - \mathbb{E}_{p(\gamma | \alpha, \mathcal{D})} [\mathbb{H}[p(\rho | \gamma, \alpha, \mathcal{D})]] \} \quad (5)$$

The expectation with respect to  $\gamma$  is needed as the measurement outcome is unknown *a priori*. This objective naturally allows us to address the question 'Having seen the outcome of the first few measurements, which measurement should we carry out next?' Rather, it was used to determine a single best set of measurements which are then

uniformly sampled throughout the experiment [1, 16]. Under these circumstances mutually unbiased bases (MUBs) are optimal, whenever they exist. We exploit the dependence of Eqn. (5) on past observations, and allow for measurements to be re-optimised adaptively as the experiment progresses.

However, Eqn. (5) is impractical to work with directly, as it involves computing entropies of high-dimensional intractable posterior densities. Recall that we approximate our posterior by samples, with which it is notoriously hard to estimate differential entropies. Furthermore, in Eqn. (5) the posterior has to be re-computed for every possible outcome  $\gamma$ . Therefore, instead of working with Eqn. (5) directly, we propose to use an equivalent reformulation thereof in terms of predictive distributions [16]:

$$\operatorname{argmax}_{\alpha \in \mathcal{A}} \left\{ \mathbb{H}[\mathbb{P}(\gamma|\alpha, \mathcal{D})] - \mathbb{E}_{p(\rho|\mathcal{D})} [\mathbb{H}[\mathbb{P}(\gamma|\alpha, \rho)]] \right\} \quad (6)$$

In previous studies [10] the system is limited to pure single qubit states, calculating the intractable Bayesian normalising constant can be realised with simple numerical integration; this could not be extended easily to higher dimensions. They consider two active learning algorithms: firstly, uncertainty sampling, which uses an approximate version of Eqn. (6), where the second term was ignored. This arguably leads to suboptimal selection behaviour; the experimenter's uncertainty may be confounded with inherent uncertainty of quantum measurements.

Within the Bayesian framework, one could use other measures of uncertainty about the Quantum state; we note that minimizing the Shannon Entropy (Eqn. (5)) is equivalent to minimization of the Bayes risk when one uses the log loss to evaluate probabilistic estimate of the state [17]. One could construct a number of other loss functions - the second algorithm in [10] seeks to minimize the Bayes risk, but using fidelity as the loss function. Although this loss is theoretically attractive, only the log loss allows the particular analytic reformulation to Eqn. (6) that permits efficient online computations. Other loss functions require one full posterior update for every possible outcome for each measurement under consideration, ABQT requires only one update per complete cycle. Therefore, if one does not use log loss online computation is infeasible; in [11] experimental designs for all  $2^N$  possible experimental outcome successions are pre-computed, they are therefore limited to very short experiments ( $< 20$  measurements). Combining Eqn. (6) with our SIS Bayesian update scheme allows for fast online experimental design.

The equivalence between Eqns. (5) and (6) becomes clear realising that they both express the conditional mutual information between  $\rho$  and  $\gamma$ . Eqn. (6) offers computational advantages over Eqn. (5): it only involves computing discrete entropies  $\mathbb{H}[\mathbb{P}(\gamma|\alpha, \rho)]$  and expectations of these under the posterior. This objective function is generally non-convex in  $\alpha$ , but its value - and derivatives with respect to  $\alpha$  - can now be efficiently computed using our weighted posterior samples from Eqn. (3), allowing us to find the most informative  $\alpha$  by direct optimisation.

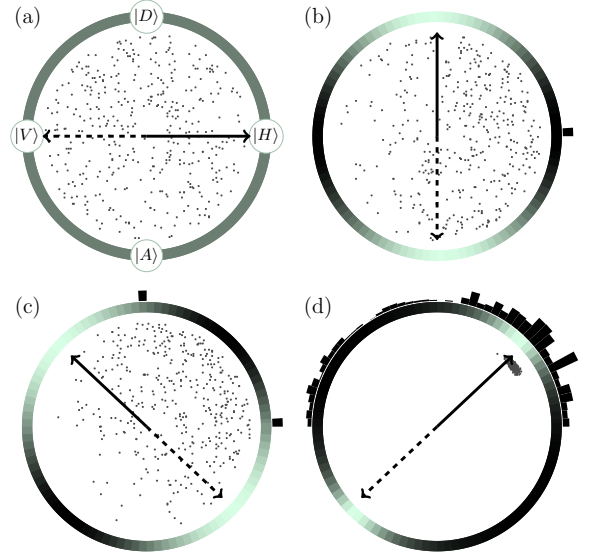


FIG. 1. Adaptive selection of measurements based on partial data. Scatter plots show 400 samples from current posterior. Shaded circles around the ‘Bloch disk’ show relative value of the objective in Eqn. (6) for different measurement directions (lighter is higher). Pairs of arrows show the most informative next measurement. Circular histograms show the number of times measurement directions have been used. (a) Initially, no observations are made, samples shown are from the uniform prior. All measurements are equally informative, we chose to start with  $\{|H\rangle, |V\rangle\}$ . (b) After one measurement, the posterior is updated, the next best measurement is mutually unbiased w.r.t. the first one. It is now  $\{|D\rangle, |A\rangle\}$ . (c) After two observations, the next best measurement is equally biased to the first two bases. (d) Posterior after 1000 observations concentrates around true state. The method tries a range of measurements, with a tendency to point towards the solution.

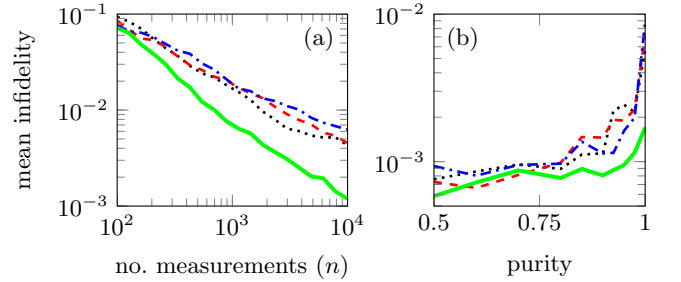


FIG. 2. One qubit tomography using projective measurements. (a) Improvement of mean posterior fidelity as the experiment progresses. Results are shown for uniformly sampled measurements (.....), uniformly sampled Pauli measurements (---), ABQT selecting adaptively amongst the 3 Pauli measurements (-.-.-) and ABQT picking general measurements (—). Adaptive optimisation of measurements allows for an almost  $n^{-1}$  rate of convergence, while other methods are more consistent with a  $n^{-\frac{1}{2}}$  rate. (b) Final value of the mean posterior infidelity after 6000 measurements using the four methods as before, as a function of purity of the state to be estimated. The advantage of ABQT is greatest for purer states.

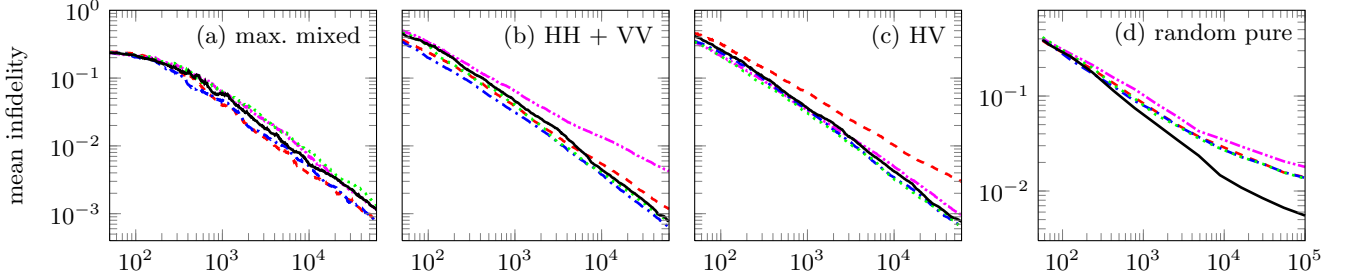


FIG. 3. Two qubit QST with uniformly chosen amongst MUB (---) or SSQT bases (---) and ABQT picking from the same set of MUBs (---), SSQT bases (---) or a more flexible set of 81 separable bases (—). Cases (a)-(c) are the same as those in [2], (d) shows average results over 20 randomly generated entangled pure states. (a) As expected, for the maximally mixed state the choice of measurement strategy has little effect. (b) On the entangled state  $(|HH\rangle + |VV\rangle)/\sqrt{2}$  MUB outperforms SSQT when uniformly sampled, but by allowing for adaptivity we can close the performance gap. (c) SSQT outperforms MUBs on the separable state  $|HV\rangle$ , but again, picking measurements adaptively the two sets perform similarly. (d) For random pure states a large improvement in performance is made when performing ABQT with the flexible set of separable measurements. Using this set, ABQT only needs  $10^4$  measurements to achieve  $\approx 98.7\%$  mean fidelity for which MUB needs  $10^5$  measurements.

It is interesting to note that Eqn. (6) is equivalent to the Jensen-Shannon divergence of the class of predictive distributions, with the posterior being the mixing distribution.

In summary, we propose the following algorithm, called Adaptive Bayesian Quantum Tomography. After each single measurement, ABQT updates its approximate posterior using Eqn. (4), then chooses the next measurement by direct numerical maximisation of the information theoretic objective in Eqn. (6).

*EX 1: single qubit tomography.* In our first simulated experiments we study tomography of single qubits ( $D = 2$ ). Mixed state qubits have three real degrees of freedom,  $\rho$  is represented as a point in a unit ball, called the Bloch sphere. For illustration purposes we first omit the third component, and only infer two remaining parameters, which will lie in a unit (Bloch) disk. This corresponds to e.g. determining linear polarisation of a photon, assuming that the circular polarisation is zero. We allow for arbitrary projective measurements with binary ( $\Gamma = 2$ ) outcomes. These are represented by pairs of antipodal points on the perimeter of the Bloch disk. Now  $\alpha \in [0, \pi)$  codes for the orientation. Fig. 1 shows the progression of measurement bases chosen by ABQT. The first two measurements are mutually unbiased, however, the third measurement is equally biased with respect to both previous bases, demonstrating that using a fixed MUB set is suboptimal in the adaptive framework. Throughout the rest of the experiment the algorithm explores a wide range of measurements.

Fig. 2 shows that the posterior mean fidelity - this time inferring all three coordinates in the full Bloch sphere - improves at a faster rate when measurements are adaptively optimised. We quantify performance as mean posterior fidelity, rather than the fidelity of the Bayesian mean estimate, as the latter gives no indication of the confidence in our estimate. The rate is more consistent with a  $n^{-1}$  law rather than  $n^{-\frac{1}{2}}$  as predicted for non-adaptive methods [2, and refs.]. Fig. 2.b shows a larger advantage for states of high purity (defined as sum of squared eigenvalues).

#### EX 2: Separable vs. MUB tomography of two qubits.

In multipartite systems, such as  $m$ -qubit registers, there are two fundamentally different classes of measurements one can apply: separable or entangling. Separable tomographic experiments are straightforward and cheap to implement, while entangling measurements are statistically more powerful. Notably, entanglement is required for implementing MUBs. These differences are discussed extensively in [2]. To investigate this trade-off in the light of adaptive tomography, we reproduce and extend the experiments in [2]. Results are shown in Fig. 3. Notably, all substantial differences between MUB and standard separable tomography (SSQT) vanish as we allow for adaptivity (Fig. 3.a-c). Furthermore, for random pure states, we are able to realise a ten-fold improvement over MUBs when using flexible separable measurements (Fig. 3.d). The results indicate that allowing for adaptivity with an imperfect, but flexible set of measurements offers greater advantages than using a fixed set of MUBs.

In summary, we have presented a new adaptive optimal experimental design framework and method based on Bayesian inference and Shannon's information. We showed that mutually unbiased bases, widely accepted as *the* optimal measurements, represent only a partial solution and are suboptimal in the adaptive framework. Moreover, the adaptive framework applies regardless of dimensionality, and can be applied to spaces where MUBs do not even exist [3, 16]. This motivates a shift in experimental focus from implementing complex entangling measurements to implementing quickly reconfigurable simpler measurements. In quantum optics, this could be feasibly achieved via mechanically or electronically controlled liquid crystal wave plates.

Although our algorithm demonstrated a substantial leap forward in terms of empirical performance, it is important to keep in mind that it still does not resolve the curse of dimensionality: the size of the parameter space still scales exponentially with the number of qubits in question. Other successful approaches address the question of dimensionality by restricting the search space.



Compressed sensing [15] constrains estimation onto a lower-dimensional manifold of rank-deficient states. It is even possible to carry out quantum homodyne tomography in infinite dimensional spaces, assuming the Wigner function is infinitely differentiable and falls into a particular smoothness class [18]. These simplifying assumptions and smoothness constraints can be incorporated into a Bayesian framework via priors.

We thank our advisors M Lengyel, Z Ghahramani and CE Rasmussen as well as G Csányi, S Lacoste-Julien, E Snelson, J Rau and S Strelchuk for useful comments. We are supported by EPSRC and Trinity College Cambridge (FH) and Google Europe (NMTH).

- 
- [1] W. Wootters and B. Fields, *Ann. Phys.* **191**, 363 (1989).
  - [2] R. B. A. Adamson and A. M. Steinberg, *Phys. Rev. Lett.* **105**, 030406 (2010).
  - [3] P. Raynal, X. Lü, and B.-G. Englert, *Phys. Rev. A* **83**,

- 062303 (2011).
- [4] F. Yan, M. Yang, and Z.-L. Cao, *Phys. Rev. A* **82**, 044102 (2010).
- [5] R. Kosut, I. A. Walmsley, and H. Rabitz, (2004), [arXiv:quant-ph/0411093](#).
- [6] J. Nunn *et al*, *Phys. Rev. A* **81**, 042109 (2010).
- [7] D. Berry, *Nat. Rev. Drug Disc.* **5**, 27 (2006).
- [8] D. Cavagnaro, J. Myung, M. Pitt, and J. Kujala, *Neur. Comp.* **22**, 887 (2010).
- [9] C. Vondrick, U. Irvine, and D. Ramanan, *NIPS* (2011).
- [10] D. G. Fischer, S. H. Kienle, and M. Freyberger, *Phys. Rev. A* **61**, 032306 (2000).
- [11] T. Hannemann *et al*, *Phys. Rev. A* **65**, 050303 (2002).
- [12] R. Blume-Kohout, *New J. Phys.* **12**, 043034.
- [13] A. Doucet, N. de Freitas, and N. Gordon, eds., *Sequential Monte Carlo in Practice* (Springer-Verlag, 2001).
- [14] R. Blume-Kohout and P. Hayden, (2006), [arXiv:quant-ph/0603116](#).
- [15] D. Gross *et al*, *Phys. Rev. Lett.* **105**, 150401 (2010).
- [16] M. K. Patra, *J. Phys. A* **40**, 10887 (2007).
- [17] A. Dawid, *Ann. ISM* **59**, 77 (2007).
- [18] C. Butucea, M. Guță, and L. Artiles, *The Annals of Statistics* **35**, pp. 465 (2007).