



PAPER • OPEN ACCESS

## Quantum model averaging

To cite this article: Christopher Ferrie 2014 *New J. Phys.* **16** 093035

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

### Related content

- [Accelerated randomized benchmarking](#)  
Christopher Granade, Christopher Ferrie and D G Cory
- [Bayesian group analysis of plasma-enhanced chemical vapour deposition data](#)  
R Fischer
- [Practical Bayesian tomography](#)  
Christopher Granade, Joshua Combes and D G Cory

### Recent citations

- [QInfer: Statistical inference software for quantum applications](#)  
Christopher Granade *et al*
- [Pseudo-Bayesian quantum tomography with rank-adaptation](#)  
The Tien Mai and Pierre Alquier
- [Structured filtering](#)  
Christopher Granade and Nathan Wiebe

## Quantum model averaging

**Christopher Ferrie**

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

E-mail: [csferrie@gmail.com](mailto:csferrie@gmail.com)

Received 3 June 2014, revised 18 July 2014

Accepted for publication 25 July 2014

Published 24 September 2014

*New Journal of Physics* **16** (2014) 093035

doi:[10.1088/1367-2630/16/9/093035](https://doi.org/10.1088/1367-2630/16/9/093035)

### Abstract

Standard tomographic analyses ignore model uncertainty. It is assumed that a given model generated the data and the task is to estimate the quantum state, or a subset of parameters within that model. Here we apply a model averaging technique to mitigate the risk of overconfident estimates of model parameters in two examples: (1) selecting the rank of the state in tomography and (2) selecting the model for the fidelity decay curve in randomized benchmarking.

Keywords: quantum tomography, randomized benchmarking, model selection, Bayesian, quantum information, sequential Monte Carlo

### 1. Introduction

Parameter estimation is an integral part of physics. Accurate estimates of physical parameters in quantum mechanical models allow for precision quantum control [1, 2] which enables practical goals such as quantum computation [3] and quantum metrology [4] which in turn can provide probes of fundamental physics such as gravity wave detection [5].

Quantum parameter estimation shares many similarities with its classical counterpart. But there are many subtle and peculiar differences. Even in single parameter estimation, quantum metrology [4, 7] shows that we can obtain advantages from quantum resources such as squeezed states of light [9] and entanglement [10]. With such rich structure, many new subtleties [11], considerations [12–14] and generalities [15–17] can arise, including entirely new approaches to estimation [18–21] and verification [22–24].



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

At the other end of the parameter spectrum the quantum state of a physical system is our most complete description of it. Thus estimation of quantum states [6] might be seen as the most ambitious form of estimation. There are many approaches to the general problem [25–28], some which specialize for computational efficiency [29–32] and those which go beyond to estimation of regions [33–38].

All of the above mentioned results assume a model. That is, it was taken as given that a particular parametric distribution generated the data. But what if this assumption is not correct? The effect of systematic measurement errors, for example [39], can comprise the security of quantum key distribution protocols [40]. Schemes which guard against measurement errors go by the name self-consistent tomography [41–46]. A recently investigated alternative is the application of classical approaches to model selection, which have been used in a variety of experimental [51, 52] and theoretical [47–50] works. Here we supplement these results with a new approach to parameter estimation: model averaging. The technique shares many similarities with model selection—in fact, model selection is a crucial component of model averaging, but goes beyond it. Although model selection adds an additional layer of security to overconfident estimation, selecting models can be itself a red herring, for the most probable model might only be slightly more probable than others.

The approach considered here combines Bayesian parameter estimation with Bayesian model selection, such that the final estimate of the parameters is the best value of the parameters within each model, averaged over the probability assigned to each model. We will show that such an approach can reduce the error incurred by first selecting a model—which has some probability of being incorrect—then selecting parameters within that model. In fact, the numerical experiments presented here show that model average estimation always does better than the estimates from incorrect models and in some scenarios can perform better than even estimates from the correct model. This is due to the additional hedging afforded by considering multiple models, all of which carry some information.

The paper is organized as follows. In section 2 we outline the problem and review the model selection techniques used so far. In section 3, we present the full Bayesian approach to model selection and define the model average estimate (MAE) of parameters. Section 4 presents two distinct examples where the MAE provides an advantage for parameter estimation. We conclude with a discussion in section 5.

## 2. General problem and common methods

In section 2.1 we overview the problem. In sections 2.2 and 2.3 we review the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) which are by far the most commonly used approaches to model selection (and the only used thus far for quantum estimation). These sections are included for reference and completeness.

### 2.1. Problem setup

Let us begin with the base problem of parameter estimation. A physical model prescribes the probabilities for the outcomes of experiments:  $\Pr(D \mid \mathbf{x}; C, M)$ . Here  $D$  is some hypothetical or

observed data,  $\mathbf{x}$  is a set of real numbers in  $\mathbb{R}^d$ , where  $d$  is the dimension of the model,  $C$  is the experimental context<sup>1</sup> and  $M$  is our model. For example, we could have the model  $M$  be that of a qubit which is parameterized by a Bloch vector  $\mathbf{x} = (x, y, z)$ . The experimental context could be a measurement basis, say that of  $\sigma_x$ . Then, the quantum mechanical model prescribes  $\Pr(\pm|\mathbf{x}; C, M) = (1 \pm x)/2$ . In quantum mechanics this is called the Born rule and in statistics, the likelihood function.

Going from parameters to the probability of data is a deductive process—the model gives us numerical values of  $\Pr(D | \mathbf{x}; C, M)$ . The experiment, on the other hand, gives us a particular data set  $D$ —but what we really want is  $\mathbf{x}$ . This is an example of an inverse problem. What the Bayesian solution provides is  $\Pr(\mathbf{x} | D; C, M)$ , the distribution of  $\mathbf{x}$  given the data. Although we lack certainty about  $\mathbf{x}$ , we can accurately (read: quantitatively, mathematically) describe our state of knowledge of  $\mathbf{x}$  given we have seen the data  $D$ .

The formal path forward is through Bayes' rule

$$\Pr(\mathbf{x} | D; C, M) = \frac{\Pr(D | \mathbf{x}; C, M)\Pr(\mathbf{x}|C, M)}{\Pr(D|C, M)}. \quad (1)$$

From a chronological point of view relative to  $D$ , we begin with the *prior*  $\Pr(\mathbf{x} | C, M)$ , which encodes the information we have about  $\mathbf{x}$  prior to learning about  $D$ . We weight the prior by the likelihood function and normalize by the marginal likelihood<sup>2</sup>

$$\Pr(D | C, M) = \mathbb{E}_{\mathbf{x}|C,M}[\Pr(D | \mathbf{x}; C, M)]. \quad (2)$$

The distribution produced as result of Bayes' rule,  $\Pr(\mathbf{x} | D; C, M)$ , is called the *posterior* and represents our knowledge of  $\mathbf{x}$  after the data has been observed.

At this point we could call the problem solved. This, however, assumes the model is correct. If the model is suspect, then we have the meta-problem of determining the best model.

## 2.2. AIC

The most commonly used model selection<sup>3</sup> technique is the AIC, which arises as follows. First, we suppose there is some true model,  $M = T$ , giving a distribution  $\Pr(D' | T; C)$ . We quantify the discrepancy between this model and our candidate model via the Kullback–Leibler divergence

$$\begin{aligned} \text{KL}(T \parallel M) = & \mathbb{E}_{D'|T;C} \left[ \log \frac{\Pr(D' | T, C)}{\Pr(D' | \mathbf{x}; M, C)} \right] = \mathbb{E}_{D'|T;C} [\log \Pr(D' | T, C)] \\ & - \mathbb{E}_{D'|T;C} [\log \Pr(D' | \mathbf{x}; M, C)], \end{aligned} \quad (3)$$

for some appropriate choice of parameters  $\mathbf{x}$ . The ‘best’ model, then, is the one that minimizes  $\text{KL}(T \parallel M)$ , which is equivalent to maximizing the second term in (3) since the first term only depends on the true model. However, we do not know the true model and we do not know the best set of parameters within our candidate model. The latter problem is naturally addressed by collecting data  $D$  and producing an estimate of the parameters  $\hat{\mathbf{x}}(D)$ , then averaging over

<sup>1</sup> Experimental context,  $C$ , here means any additional information necessary to provide a well defined function  $\Pr(D | \mathbf{x}; C, M)$ .

<sup>2</sup> The notation  $\mathbb{E}_{\mathbf{x}}[f(\mathbf{x})]$  means take the expectation of  $f(\mathbf{x})$  with respect to the distribution of  $\mathbf{x}$  (the distribution itself is implicit).

<sup>3</sup> See [54] for an overview of model selection and the various techniques mentioned here.

possible data sets such that the quantity of interest becomes

$$\mathbb{E}_{D|T;C} \left[ \mathbb{E}_{D'|T;C} \left[ \log \Pr(D' | \hat{\mathbf{x}}(D); M, C) \right] \right]. \quad (4)$$

Akaike showed that, independent of the true model (and under some regularity conditions), an unbiased estimator of this quantity is

$$\text{AIC}(M) = \max_x \log \Pr(D | \mathbf{x}; C, M) - d, \quad (5)$$

where  $d$  is the number of dimensions of the model (the number of free parameters). The preferred model is the one with largest value of  $\text{AIC}(M)$ . The simple linear penalization with dimension makes it is clear how models with more parameters are penalized.

### 2.3. BIC

The Bayesian approach is more general. Being so, it is less obvious how it might penalize complex models. Here we show how an asymptotic approximation leads to a form similar to the AIC. First, we write the marginal likelihood (2) as the integral expectation

$$\Pr(D | M; C) = \int d\mathbf{x} \Pr(\mathbf{x} | M; C) \Pr(D | \mathbf{x}; C, M). \quad (6)$$

We will approximate this integral using Laplace's method. To this end, consider the Taylor expansion of the log of the likelihood function about its peak (note that for brevity, we have dropped the context  $C$  and model  $M$  from the conditionals)

$$l(\mathbf{x}) := \log \Pr(D | \mathbf{x}) \approx l(\mathbf{x}_0) + \nabla^T l(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^T \nabla \nabla^T l(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0). \quad (7)$$

If we assume the number of measurements  $N \rightarrow \infty$ , the law of large numbers gives us

$$\nabla \nabla^T l(\mathbf{x}_0) = \sum_{j=1}^N \nabla \nabla^T l_j(\mathbf{x}_0), \quad (8)$$

$$\rightarrow \sum_{j=1}^N \mathbb{E} \left[ \nabla \nabla^T l_j(\mathbf{x}_0) \right], \quad (9)$$

$$= - \sum_{j=1}^N I_j(\mathbf{x}_0), \quad (10)$$

$$= -NI(\mathbf{x}_0), \quad (11)$$

where  $I_j$  is the Fisher information of the  $j$ th measurement and  $I$  is the arithmetic average of these values. Then, the integral (6) becomes

$$\Pr(D | M; C) = \int d\mathbf{x} \exp(l(\mathbf{x})) \Pr(\mathbf{x}), \quad (12)$$

$$= \exp(l(\mathbf{x}_0)) \Pr(\mathbf{x}_0) \left( \frac{2\pi}{N} \right)^{\frac{d}{2}} |I(\mathbf{x}_0)|^{-\frac{1}{2}} \left( 1 + O\left(\frac{1}{N}\right) \right). \quad (13)$$

Now we take the logarithm to obtain

$$\log \Pr(D | M; C) \approx l(\mathbf{x}_0) - \frac{d}{2} \log N + \frac{d}{2} \log 2\pi + \log \Pr(\mathbf{x}_0) + \log |\mathbf{I}(\mathbf{x}_0)|^{-\frac{1}{2}}. \quad (14)$$

If we ignore the terms not changing with  $N$ , we have a new quantity

$$\text{BIC}(M) = \max_x \log \Pr(D | \mathbf{x}; C, M) - \frac{d}{2} \log N, \quad (15)$$

which is the well-known BIC. Notice the striking similarity to the AIC (5). Being nearly equivalent, the BIC is often considered in addition to the AIC. Next, we will consider the full solution, which will allow us to obtain more accurate estimates of parameters averaged over the competing models.

Recent proposals have used the AIC/BIC on both simulated [47–50] and experimental data [51, 52]. In [50], however, the authors caution its unattended use. The argument against AIC for quantum states, for example, is simple. The AIC is derived from a metric which measures their closeness of model in their *predictive* probability—a certainly well-motivated measure. However, such a measure is only useful if all future measurements will be the same as those used to perform the data analysis. That is, one can measure copies of a quantum system in some fixed set of bases, estimate the state, then use that estimate to predict the outcome of a measurement in a new basis. Thus, in quantum theory, one ought to consider a measure on predictive distributions as maximized over all possible future measurements. As suggested in [50], such a measure might well be the quantum relative entropy, for example. Here we avoid these problems by considering the full Bayesian solution, which we describe next.

### 3. Bayesian model selection and averaging

#### 3.1. MAE

Within the Bayesian framework, the model selection approach is no different than for parameter estimation. Rather than focus on the distribution  $\Pr(\mathbf{x} | D; C, M)$ , we first consider  $\Pr(M | D; C)$ . Using Bayes rule we have

$$\Pr(M | D; C) = \frac{\Pr(D | M; C) \Pr(M | C)}{\Pr(D | C)}. \quad (16)$$

Often, practitioners of Bayesian methods go one step further and compare two models—say,  $M_1$  and  $M_2$ —by taking the ratio of these posteriors

$$\frac{\Pr(M_1 | D; C)}{\Pr(M_2 | D; C)} = \frac{\Pr(D | M_1; C) \Pr(M_1 | C)}{\Pr(D | M_2; C) \Pr(M_2 | C)}, \quad (17)$$

noticing the normalization factor cancels. This quantity is called the posterior odds ratio and first considered by Jeffreys [55]. Clearly, if the posterior odds ratio is larger than 1, we favor  $M_1$ . The last fraction is called the prior odds ratio and the unbiased choice favoring neither model is set this term equal to 1. This leaves us with

$$\frac{\Pr(M_1 | D; C)}{\Pr(M_2 | D; C)} = \frac{\Pr(D | M_1; C)}{\Pr(D | M_2; C)}, \quad (18)$$

which is called the Bayes factor [56]. Each quantity in the ratio is marginal likelihood (2) of its respective model.

For a discrete number of hypothetical models  $\{M_k\}$ , and assuming one model must be chosen, the optimal strategy is to compute the marginal likelihood of each model and select the one with the highest value. Model selection of this type has been used in quantum mechanical problems of Hamiltonian finding [57] and estimating error channels from syndrome measurements [58].

Here we will investigate the idea of using a meta-model, which is an average over those in  $\{M_k\}$ . Assume that we are interested in some subset of parameters  $\mathbf{y}$  common to all models. Given we have taken data  $D$  and computed each marginal likelihood  $\Pr(M_k | D; C)$ , we define the MAE as

$$\hat{\mathbf{y}}_{\text{MAE}}(D; C) = \mathbb{E}_{M_k|D;C} \left[ \mathbb{E}_{\mathbf{y}|D;C,M_k} [\mathbf{y}] \right]. \quad (19)$$

In words, this is the average (over models) of the average (over parameters within models). Variants of this approach are referred to as Bayesian model averaging [53].

Before moving to our examples, a few comments are in order. First, it is not necessary that the models are ‘nested’ in the sense that we can order them into supersets. The only requirement is that the parameters of interest are included in each model. The other parameters in each model we might call nuisance parameters. Part of the appeal of the Bayesian approach is that these parameters are automatically dealt with and we can focus on those parameters which are of immediate interest. Of course, the nuisance parameters can be inferred as well.

Let us wax philosophical for a moment. What is being proposed is to select a meta-model, an average over many different physical models. This may seem awkward for physical theories—after all, there is only one true model, right? Not in our view. Models are human constructs, those Platonic ideals which describe another world, a world we were clever enough to find through our mastery of mathematics and abstraction. Here, we have dropped the idea that the point is to find the capital-T-truth. Rather, we measure our understanding of nature through our ability to predict and control its behavior. By averaging physical models, we can show that this idea has merit.

### 3.2. Sequential Monte Carlo (SMC)

In practice, the Bayesian update rule and the expectations required in the equations above are analytically and computationally intractable since they involve complicated integrals over multidimensional parameter spaces which may include solutions to equations of motions which are themselves intractable. To perform the calculations we turn to Monte Carlo techniques. Our numerical algorithm fits within the subclass of Monte Carlo methods called SMC or particle filtering [59].

The SMC procedure prescribes that we approximate the probability distribution by a weighted sum of Dirac delta-functions

$$\Pr(\mathbf{x}) \approx \sum_{j=1}^n w_j \delta(\mathbf{x} - \mathbf{x}_j), \quad (20)$$

where the weights at each step are iteratively calculated from the previous step via

$$w_j \mapsto \Pr(D | \mathbf{x}_j) w_j, \quad (21)$$

followed by a normalization step. The elements of the set  $\{\mathbf{x}_j\}_{j=1}^n$  are called particles. Here,  $n = |\{\mathbf{x}_i\}|$  is the number of particles and controls the accuracy of the approximation. Like all Monte Carlo algorithms, the SMC algorithm approximates expectation values, such that

$$\mathbb{E}_{\mathbf{x}}[f(\mathbf{x})] \approx \sum_{j=1}^n w_j f(\mathbf{x}_j). \quad (22)$$

In other words, SMC allows us to efficiently compute multidimensional integrals with respect to the measure defined by the probability distribution.

The resultant posterior probability provides a full specification of our knowledge. However, in most applications, it is sufficient—and certainly more efficient—to summarize this distribution. In our context, the optimal single parameter vector to report is the mean of the posterior distribution

$$\hat{\mathbf{x}} = \mathbb{E}_{\mathbf{x}}[\mathbf{x}] = \sum_{j=1}^n w_j \mathbf{x}_j. \quad (23)$$

The SMC approximation can also provide efficient calculation and description of *regions* [38, 60]. For our purpose, we also require the SMC approximation to give an accurate and efficient estimate of the marginal likelihood equation (6), which we need to calculate Bayes' rule at the level of models equation (16). Via the SMC approximation, the integral expectation in the definition of the marginal likelihood, equation (6), is

$$\Pr(D) \approx \sum_{j=1}^n w_j \Pr(D | \mathbf{x}_j). \quad (24)$$

It is not immediately obvious but is easy to see in hindsight that this is exactly the normalization that must be computed already in the SMC algorithm after the weight update, equation (21), is applied. By storing this value, we can apply Bayes rule at the meta-level of models equation (16).

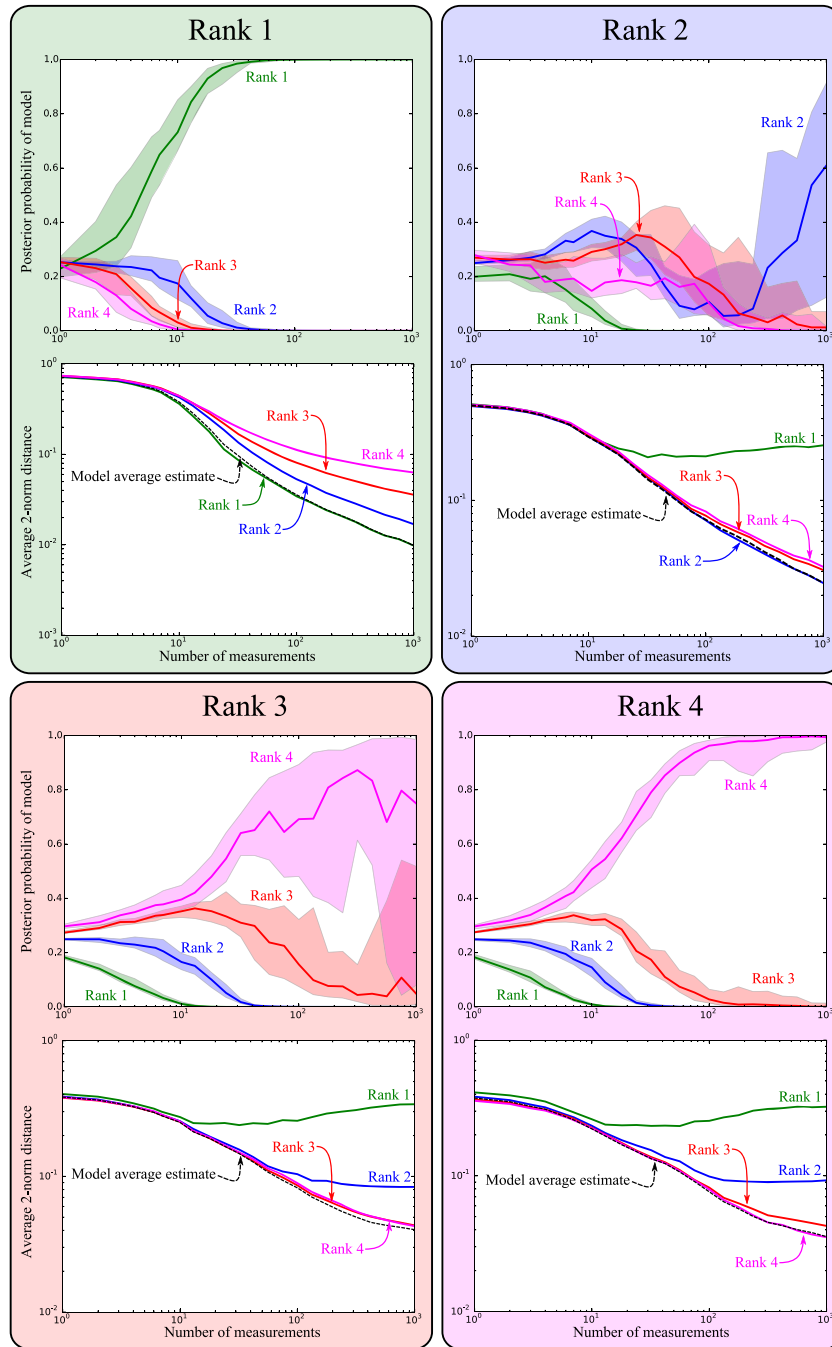
An iterative numerical algorithm such as SMC requires care to ensure stability. Conditions for stability of the algorithm and the specifications of an implementation have been detailed elsewhere [60, 61]. The SMC algorithm has now been used in many quantum mechanical parameter estimation problems [38, 57, 60–63] and a software implementation (the one used here) is available as a Python package [64, 65].

## 4. Examples

### 4.1. Rank selection

We consider first an example similar to Guta, Kypraios and Dryden [52]:  $n$  qubits subjected to random Pauli measurements where the models under consideration are those of differing rank. Each model will be denoted  $M_r$ , where  $r$  is the rank of the unknown quantum state so that the dimension of model  $M_r$  is  $d = 2^{n+1}r$ . We generate unknown quantum states with fixed rank  $r$  as follows [66]. Begin with a matrix  $X \in \mathbb{C}^{2^n \times r}$  where each component  $x_{ij}$  is chosen independently





**Figure 1.** The performance of Bayesian model selection and the model average estimate for two qubits. Each box represents the data for its labeled rank as the ‘true’ model. The lines represent the median of the data and, where present, the shaded areas are the interquartile ranges. Each ‘measurement’ on the horizontal axis corresponds to 100 experiments of a randomly chosen Pauli measurement. In the SMC algorithm,  $10^4$  particles were used in each model. For each true rank, 100 simulated states were generated and measured. The most important things to note here are that the model average estimator does as well as the true model and much better than models whose rank is further way from the true model. Also, we notice that rank one and full rank states are much easier to distinguish than states with ranks two and three. This is explored in more detail for three qubits in figure 2.

according to  $\Re(x_{ij}) \sim \mathcal{N}(0, 1)$  and  $\Im(x_{ij}) \sim \mathcal{N}(0, 1)$ —standard Normal distributions. Then define the rank  $r$  density operator

$$\rho = \frac{XX^\dagger}{\text{Tr}(XX^\dagger)}. \quad (25)$$

If  $r = 2^n$ , this construction is equivalent to a Hilbert-Schmidt random density matrix. The model parameters will be the vectorization of the matrix  $X$ :  $\mathbf{x} = \text{vec}(X)$ .

We label the single qubit Pauli operators  $\{\sigma_0, \sigma_1, \sigma_2, \sigma_3\}$  and the multi-qubit Paulis by

$$\sigma_k = \sigma_{k_1} \otimes \sigma_{k_2} \otimes \cdots \otimes \sigma_{k_n}, \quad (26)$$

where  $k = k_1 + 4k_2 + 4^2k_3 + \cdots + 4^{n-1}k_n$ . Since each Pauli is idempotent,  $\sigma_k^2 = \mathbb{1}$ , each individual measurement has two possible outcomes which we label  $d \in \{+1, -1\}$  for the  $+1$  and  $-1$  eigenvalues. Then the likelihood function of a single measurement can be related to the expectation value via:  $\Pr(\pm 1 | \sigma_k) = (1 \pm \langle \sigma_k \rangle)/2$ . Using the properties of the vec operation, we can write this as an explicit function of  $\mathbf{x}$

$$\Pr(\pm 1 | X, \sigma_k) = \frac{1}{2} \left( 1 \pm \frac{\text{Tr}(XX^\dagger \sigma_k)}{\text{Tr}(XX^\dagger)} \right) = \frac{1}{2} \left( 1 \pm \frac{\mathbf{x} \cdot (\sigma_k \otimes \mathbb{1}) \mathbf{x}}{\mathbf{x} \cdot \mathbf{x}} \right) = \Pr(\pm 1 | \mathbf{x}, \sigma_k). \quad (27)$$

We label the parameter vector within the rank  $r$  model  $M_r$  by  $\mathbf{x}_r$  and the associated density matrix given by equation (25)  $\rho_r$ . Within each model, there is not a one-to-one correspondence between  $\mathbf{x}_r$  and  $\rho_r$ —different vectors will yield the same density matrix. Since we will be interested in obtaining accurate estimates of density matrices as quantified by some norm on the space of density matrices, we will average the models over their  $\rho_r$ 's rather than their  $\mathbf{x}_r$ 's.

Within each model, then, we have the mean density matrix (recall  $D$  is data,  $C$  is the context—that is, which Paulis were measured)

$$\hat{\rho}_r(D; C) = \mathbb{E}_{\mathbf{x}_r | D; C} \left[ \frac{X_r X_r^\dagger}{\text{Tr}(X_r X_r^\dagger)} \right]. \quad (28)$$

Explicitly, the MAE in equation (19) is

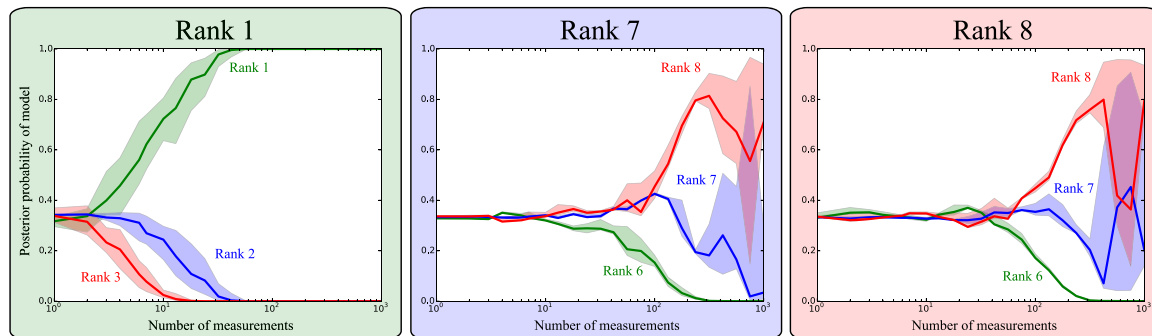
$$\rho_{\text{MAE}}(D; C) = \sum_r \hat{\rho}_r(D; C) \Pr(r | D; C). \quad (29)$$

We assume there is a true density matrix  $\rho_t$  and we judge each estimate of the true state  $\hat{\rho}$  by its spectral distance to  $\rho_t$

$$|\rho_t - \hat{\rho}| = \sigma_{\max}(\rho_t - \hat{\rho}), \quad (30)$$

where  $\sigma_{\max}$  denotes the largest singular value. This is the norm induced by the usual Euclidean norm on vectors and, without giving an operational meaning to the states identified in our toy example, it seems the most convenient.

The data for two qubits is presented in figure 1. The important take-away is that the MAE does as well, or slightly better than the true model in every case. Also, we see that it is quite easy to identify rank 1 models (pure states) as well as rank 4 models (full rank states) while it seems difficult to identify states of non-extreme rank. Notice that it is extremely difficult to distinguish the rank 3 model from the full rank model when the former defines the underlying



**Figure 2.** The performance of Bayesian model selection and the model average estimate for three qubits. Each box represents the data for its labeled rank as the ‘true’ model. The lines represent the median of the data and, where present, the shaded areas are the interquartile ranges. Each ‘measurement’ on the horizontal axis corresponds to 100 experiments of a randomly chosen Pauli measurement. In the SMC algorithm,  $10^5$  particles were used in each model. For each true rank, ten simulated states were generated and measured. Note again that rank one states are easily distinguished. Whereas, it appears that—on average—much more data will be necessary to rule out full rank (that is, rank eight) states when the true rank is seven. Some speculation on why this might be is given in the discussion section 5.

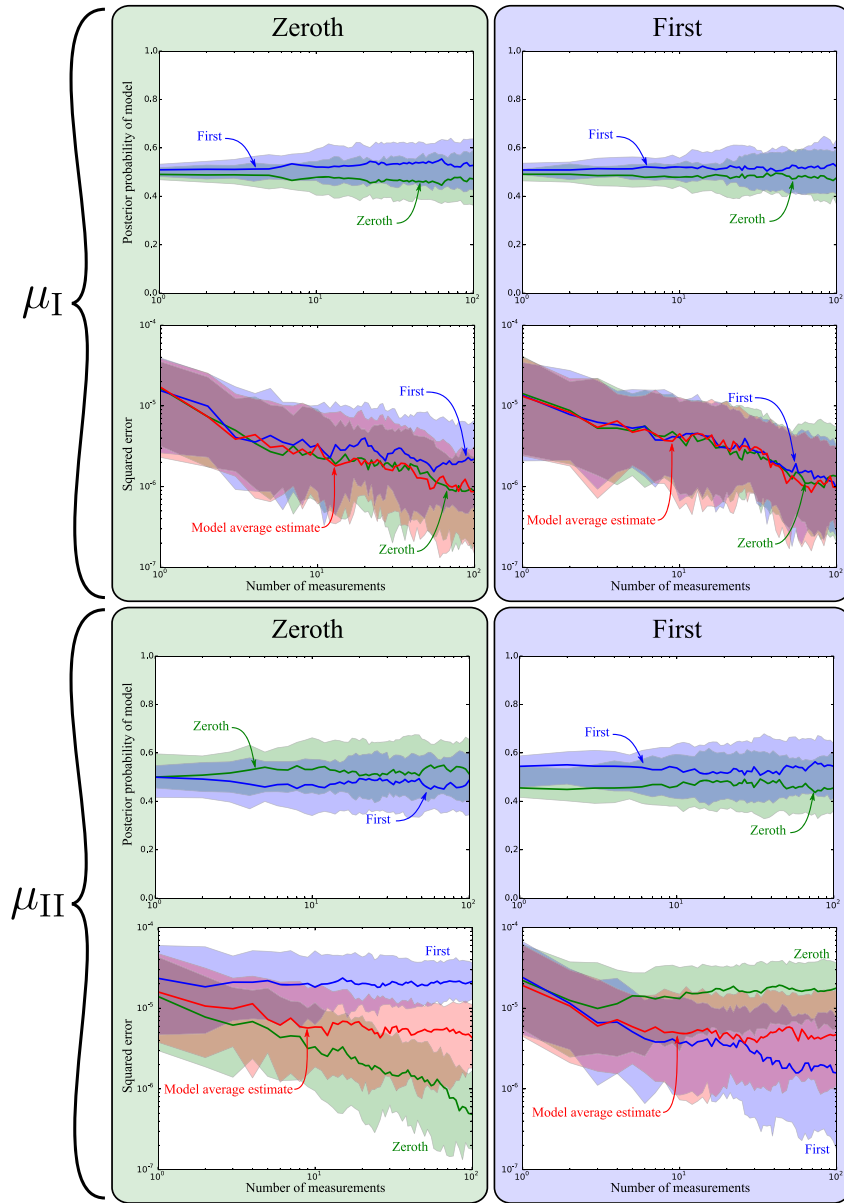
truth. However, both models perform well with respect to the error in the estimated parameters and the MAE does best, on average.

To further illustrate the difficulty in differentiating high rank states, the probabilities assigned to three qubit models is shown in figure 2. Again, we see that pure states and full rank states are correctly identified, yet it is difficult to correctly distinguish between rank 7 and rank 8 states, in the same way as it was difficult to distinguish rank 3 and rank 4 states for two qubits. We conclude that for the models considered here, it is easiest to correctly identify low rank and full rank states, while it is difficult to correctly identify nearly high rank states.

Notice also that models which are far away—in the sense that the ranks differ by relatively large amounts—are quickly ruled out. In these cases, since the SMC algorithm can be run online (in parallel with the experiment), simulating such models can be stopped to mitigate the computational difficulty in simultaneously simulating many quantum models. Still, tomography is at one extreme in the spectrum of methods for estimating quantum mechanical parameters—it is the most complete description of the physical system. At the other extreme is summarizing information from experiments into a single number, such as fidelity.

#### 4.2. Randomized benchmarking

In the second example, we consider the experimental protocol of randomized benchmarking [67] which has been demonstrated in a variety of experimental settings [69–72] to efficiently characterize noise and quantum channels. The protocol consists of stringing together potentially long sequences of gates which is then undone to determine if the initial state has survived. In [67] the approach was shown to give, in expectation, an exponentially decay  $Ap^m + B$ , where  $A$  and  $B$  encode the errors in preparation and measurement,  $p$  is the bare survival probability and  $m$  is the length of the sequence. In the models we consider,  $p$  can be related to the average fidelity over a group of gates, but more specialized protocols exist [24, 73].



**Figure 3.** The performance of Bayesian model selection and the model average estimate for the survival probability in randomized benchmarking experiments. Each box represents the data for its label as the ‘true’ model. The lines represent the median of the data and, where present, the shaded areas are the interquartile ranges. Each ‘measurement’ on the horizontal axis corresponds to a randomized benchmarking experiment with sequence lengths  $\{10, 30, 50, \dots, 200\}$  and 1000 repetitions per sequence length. In the SMC algorithm,  $10^3$  particles were used in for each model. The top row is labeled  $\mu_I$  and the bottom row labeled  $\mu_{II}$  which refer to different prior probabilities over the first-order model parameters, as described in the main text. The prior over the zeroth order model parameters remains fixed. Both  $\mu_I$  and  $\mu_{II}$  represent priors which are ‘close’ to the prior over the zeroth order model parameters, which is why the models are hard to distinguish. Moreover, the models  $\mu_I$  and  $\mu_{II}$  are themselves extremely close. The plots illustrate then one subtle and important fact: our intuitive notion of closeness of models does not translate to closeness in performance of estimates drawn from these models. The model average estimate can hedge the risk of selecting the wrong model.

Typically,  $p$  is the only parameter of interest since it is directly related to the average fidelity of the device which is then compared to some threshold. The other parameters are often considered nuisance parameters. In [74] it was shown that the decay can be interpreted as a probabilistic model where the binary outcome of each measurement sequence of length  $m$  has a probability of survival (labeled 0)

$$\Pr(0 \mid x_0; m, M_0) = A_0 p^m + B_0, \quad (31)$$

where the subscripts refer to this as the zeroth order model. In [67], a hierarchy of models was introduced because the zeroth order model assumes the errors in the gates within the sequence are independent. By dropping this assumption, a richer set of noise models can be studied [68]. Here we will study the zeroth order model and the first order model

$$\Pr(0 \mid x_1; m, M_1) = A_1 p^m + B_1 + C_1(m-1)(q_1 - p^2)p^{m-2}, \quad (32)$$

where  $A_1$  and  $B_1$  again encode the preparation and measurement errors,  $C_1$  encodes the error on the final gate in the sequence and  $q_1 - p^2$  is a measure of the gate dependence in the errors.

For the zeroth order model, we take the prior to be a normal distribution with a mean vector  $(p, A_0, B_0) = (0.95, 0.3, 0.5)$  and equal diagonal covariances given by a deviation of  $\sigma = 0.01$ . Note that the first order model is equal to the zeroth order model when either  $C_1 = 0$  or  $q_1 = p^2$ . In order to not make the model so different that it would be trivial to distinguish them, we look at two priors for the first order model which are close to the zeroth order model. The first is a normal distribution with a mean vector  $(p, A_1, B_1, C_1, q_1) = (0.95, 0.3, 0.5, 0.03, 0.95) =: \mu_I$  and equal diagonal covariances given by a deviation of  $\sigma = 0.01$  and the second slightly closer with the same covariance matrix but mean vector  $(p, A_1, B_1, C_1, q_1) = (0.95, 0.3, 0.5, 0.02, 0.92) =: \mu_{II}$ . Note that the difference between these two distributions in the relative entropy divergence is only 0.050, and so we might expect them to behave the same. Since the both models are close to the zeroth order model, we expect it to be difficult to distinguish them.

In figure 3 we simulate the models noting again that, via the priors, they are very close. This intuition is quantified by the fact that the models are hard to distinguish, regardless of which is true. On the top in figure 3, we see that parameters in the first order model are so close to those in the zeroth order model that it is irrelevant which is chosen, for the purpose of estimating  $p$ . However, what is ‘close’ can be deceiving as we see in the bottom of figure 3. Recall that relative entropy from  $\mu_I$  to  $\mu_{II}$  is only 0.050 (which explains why they are so difficult to distinguish). In this case, the accuracy of the estimates of the parameter  $p$  depend crucially on which model is actually correct. In such cases, the MAE can be seen as providing a more conservative estimate of the average gate fidelity by hedging what is at best a 50/50 guess on which model is correct.

## 5. Conclusion and discussion

We have introduced a Bayesian model averaging approach to estimating parameters in quantum mechanical models describing data. In the examples considered, the MAE performance as well as the unknown true model in most cases. In situations where models are difficult to distinguish, the MAE can slightly *outperform* the true model.

For the quantum state estimation example (section 4.1) we considered models of differing rank of the density matrix. Ranks which differ by large amounts from the true rank are rapidly

ruled out—that is, the probability assigned to them quickly approaches zero. Thus, they contribute nothing to the MAE. On the other hand, ranks which are close to the true rank—especially when the true rank is high—are not so easily distinguished. This means that first selecting a rank and then performing estimation within that model can lead to overconfident estimates of the state. By averaging, we can allow those estimates to only contribute with the relative probability that we deem them to be true.

The mechanism for why higher rank states are hard to distinguish is yet unclear. Although the SMC algorithm and the implementation used here have been extensively studied for a wide range of problems, it is possible that state tomography is an outlier. That is, some major modification to modeling or the algorithm itself may be required to obtain the best performance. This resolution would be less interesting than a physical explanation, such as Pauli measurement tomography (which we have used in the example here) is not the optimal scheme to distinguish rank. These questions are left for future work.

In the example of randomized benchmarking (section 4.2), we explored a situation where the presence of higher order perturbations were difficult to detect. The Bayesian model selection approach accurately predicts this by assigning roughly 50/50 probability assignments to the models. Surprisingly, the ‘closeness’ of the models as measured by our ability to distinguish them does not translate into our ability to accurately estimate parameters common to each. In some cases we can do no more than guess which is correct, yet guessing the wrong model may have disastrous consequences in our ability to accurately infer the parameters of interest. Again, the MAE mitigates the risk of improperly ‘guessing’ which model is correct.

We have noted in the introduction the numerous approaches to estimation within quantum theory. This should urge one to ask, are all of these distinct approaches necessary—is there not some unified approach? Yes! The Bayesian framework outlined here is remarkably powerful in its generality. We note that Bayesian ideas have already been put to good use in quantum information theoretic and foundational problems [77–85] as well as for tomographic and parameter estimation problems [26, 86–93] and experimental design [94, 95]. Importantly, the Bayesian algorithm can provide solutions to these problems *online*, while the experiment is running, with the same software tools [64].

## Acknowledgments

The author thanks C Granade and R Blume-Kohout for helpful discussions. This work was supported in part by National Science Foundation Grant No. PHY-1212445 and by the Canadian Government through the NSERC PDF program.

## References

- [1] Khaneja N, Reiss T, Kehlet C, Schulte-Herbrüggen T and Glaser S J 2005 Optimal control of coupled spin dynamics: design of NMR pulse sequences by gradient ascent algorithms *J. Magn. Reson.* **172** 296
- [2] Brif C, Chakrabarti R and Rabitz H 2010 Control of quantum phenomena: past, present and future *New J. Phys.* **12** 075008
- [3] Nielsen M A and Chuang I L 2010 *Quantum Computation and Quantum Information* (Cambridge: Cambridge University Press)
- [4] Caves C M 1981 Quantum-mechanical noise in an interferometer *Phys. Rev. D* **23** 1693



- [5] Aasi J *et al* 2013 Enhanced sensitivity of the LIGO gravitational wave detector by using squeezed states of light *Nat. Photonics* **7** 613
- [6] Paris M and Rehacek J 2004 *Quantum State Estimation* (New York: Springer)
- [7] Helstrom C W 1976 *Quantum Detection and Estimation Theory* (New York: Academic)
- [8] Hayashi M 2011 Comparison between the Cramer–Rao and the mini-max approaches in quantum channel estimation *Commun. Math. Phys.* **304** 689
- [9] Lang M D and Caves C M 2013 Optimal quantum-enhanced interferometry using a laser power source *Phys. Rev. Lett.* **111** 173601
- [10] Giovannetti V, Lloyd S and Maccone L 2006 Quantum metrology *Phys. Rev. Lett.* **96** 010401
- [11] Barndorff-Nielsen O and Gill R 2000 Fisher information in quantum statistics *J. Phys. A: Math. Gen.* **33** 4481
- [12] Fujiwara A and Imai H 2008 A fibre bundle over manifolds of quantum channels and its application to quantum statistics *J. Phys. A: Math. Theor.* **41** 255304
- [13] Knysh S, Smelyanskiy V N and Durkin G A 2011 Scaling laws for precision in quantum interferometry and the bifurcation landscape of the optimal state *Phys. Rev. A* **83** 021804
- [14] Escher B M, de Matos Filho R L and Davidovich L 2011 General framework for estimating the ultimate precision limit in noisy quantum-enhanced metrology *Nat. Phys.* **7** 406
- [15] Boixo S, Flammia S T, Caves C M and Geremia J M 2007 Generalized limits for single-parameter quantum estimation *Phys. Rev. Lett.* **98** 090401
- [16] Tsang M, Wiseman H M and Caves C M 2011 Fundamental quantum limit to waveform estimation *Phys. Rev. Lett.* **106** 090401
- [17] Ferrie C 2014 The best Fisher is upstream: data processing inequalities for quantum metrology *Phys. Rev. A* **90** 014101
- [18] Higgins B L, Berry D W, Bartlett S D, Wiseman H M and Pryde G J 2007 Entanglement-free Heisenberg-limited phase estimation *Nature* **450** 393
- [19] Hentschel A and Sanders B C 2011 Efficient algorithm for optimizing adaptive quantum metrology processes *Phys. Rev. Lett.* **107** 233601
- [20] Mahler D, Rozema L A, Darabi A, Ferrie C, Blume-Kohout R and Steinberg A 2013 Adaptive quantum state tomography improves accuracy quadratically *Phys. Rev. Lett.* **111** 183601
- [21] Ferrie C and Granade C E 2014 Likelihood-free methods for quantum parameter estimation *Phys. Rev. Lett.* **112** 130402
- [22] Flammia S T and Liu Y-K 2011 Direct fidelity estimation from few Pauli measurements *Phys. Rev. Lett.* **106** 230501
- [23] da Silva M P, Landon-Cardinal O and Poulin D 2011 Practical characterization of quantum devices without tomography *Phys. Rev. Lett.* **107** 210404
- [24] Magesan E *et al* 2012 Efficient measurement of quantum gate error by interleaved randomized benchmarking *Phys. Rev. Lett.* **109** 080505
- [25] Hradil Z 1997 Quantum-state estimation *Phys. Rev. A* **55** R1561
- [26] Blume-Kohout R 2010 Optimal, reliable estimation of quantum states *New J. Phys.* **12** 043034
- [27] Blume-Kohout R 2010 Hedged maximum likelihood quantum state estimation *Phys. Rev. Lett.* **105** 200504
- [28] Gross D, Liu Y-K, Flammia S T, Becker S and Eisert J 2010 Quantum state tomography via compressed sensing *Phys. Rev. Lett.* **105** 150401
- [29] Gottesman D 2008 Identifying Stabilizer States *Quantum Estimation: Theory Practice* Perimeter Institute for Theoretical Physics <http://pirsa.org/08080052/>
- [30] Cramer M, Plenio M B, Flammia S T, Somma R, Gross D, Bartlett S D, Landon-Cardinal O, Poulin D and Liu Y-K 2010 Efficient quantum state tomography *Nat. Commun.* **1** 149
- [31] Tóth G, Wieczorek W, Gross D, Krischek R, Schwemmer C and Weinfurter H 2010 Permutationally invariant quantum tomography *Phys. Rev. Lett.* **105** 250403
- [32] Landon-Cardinal O and Poulin D 2012 Practical learning method for multi-scale entangled states *New J. Phys.* **14** 085004

- [33] Řeháček J, Mogilevtsev D and Hradil Z 2008 Tomography for quantum diagnostics *New J. Phys.* **10** 043022
- [34] Christandl M and Renner R 2012 Reliable quantum state tomography *Phys. Rev. Lett.* **109** 120403
- [35] Arrazola J M, Gittsovich O, Donohue J M, Lavoie J, Resch K J and Lütkenhaus N 2013 Reliable entanglement verification *Phys. Rev. A* **87** 062331
- [36] Blume-Kohout R 2012 Robust error bars for quantum tomography [arXiv:1202.5270](#)
- [37] Shang J, Ng H K, Sehwat A, Li X and Englert B-G 2013 Optimal error regions for quantum state estimation *New J. Phys.* **15** 123026
- [38] Ferrie C 2014 High posterior density ellipsoids of quantum states *New J. Phys.* **16** 023006
- [39] Rosset D, Ferretti-Schbitz R, Bancal J-D I, Gisin N and Liang Y-C 2012 Imperfect measurement settings: implications for quantum state tomography and entanglement witnesses *Phys. Rev. A* **86** 062325
- [40] Gerhardt I, Liu Q, Lamas-Linares A, Skaar J, Scarani V, Makarov V and Kurtsiefer C 2011 Experimentally faking the violation of bells inequalities *Phys. Rev. Lett.* **107** 170404
- [41] Braczyk A M, Mahler D H, Rozema L A, Darabi A, Steinberg A M and James D F V 2012 Self-calibrating quantum state tomography *New J. Phys.* **14** 085003
- [42] Medford J, Beil J, Taylor J M, Bartlett S D, Doherty A C, Rashba E I, DiVincenzo D P, Lu H, Gossard A C and Marcus C M 2013 Self-consistent measurement and state tomography of an exchange-only spin qubit *Nat. Nanotechnology* **8** 654
- [43] Merkel S T, Gambetta J M, Smolin J A, Poletto S, Crcoles A D, Johnson B R, Ryan C A and Steffen M 2013 Self-consistent quantum process tomography *Phys. Rev. A* **87** 062119
- [44] Takahashi M, Bartlett S D and Doherty A C 2013 Tomography of a spin qubit in a double quantum dot *Phys. Rev. A* **88** 022120
- [45] Blume-Kohout R, King Gamble J, Nielsen E, Mizrahi J, Sterk J D and Maunz P 2013 Robust, self-consistent, closed-form tomography of quantum logic gates on a trapped ion qubit [arXiv:1310.4492](#)
- [46] Stark C 2014 Self-consistent tomography of the state-measurement Gram matrix *Phys. Rev. A* **89** 052109
- [47] Lougovski P and van Enk S J 2009 Characterizing entanglement sources *Phys. Rev. A* **80** 052324
- [48] Yin J O S and van Enk S J 2011 Information criteria for efficient quantum state estimation *Phys. Rev. A* **83** 062110
- [49] Schwarz L and van Enk S J 2013 Error models in quantum computation: an application of model selection *Phys. Rev. A* **88** 032318
- [50] van Enk S J and Blume-Kohout R 2013 When quantum tomography goes wrong: drift of quantum sources and other errors *New J. Phys.* **15** 025024
- [51] Usami K, Nambu Y, Tsuda Y, Matsumoto K and Nakamura K 2003 Accuracy of quantum-state estimation utilizing Akaike's information criterion *Phys. Rev. A* **68** 022314
- [52] Guta M, Kypraios T and Dryden I 2012 Rank-based model selection for multiple ions quantum tomography *New J. Phys.* **14** 105002
- [53] Hoeting J A, Madigan D, Raftery A E and Volinsky C T 1999 Bayesian model averaging: a tutorial *Stat. Sci.* **14** 382
- [54] Burnham K P and Anderson D R 2002 *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (Berlin: Springer)
- [55] Jeffreys H 1998 *The Theory of Probability* (Oxford: Oxford University Press)
- [56] Kass R E and Raftery A E 1995 Bayes factors *J. Am. Stat. Assoc.* **90** 773
- [57] Wiebe N, Granade C, Ferrie C and Cory D G 2014 Quantum Hamiltonian learning using imperfect quantum resources *Phys. Rev. A* **89** 042314
- [58] Combes J, Ferrie C, Cesare C, Tiersch M, Milburn G J, Briegel H J and Caves C M 2014 *In situ* characterization of quantum devices with error correction [arXiv:1405.5656](#)
- [59] Doucet A, Godsill S and Andrieu C 2000 On sequential monte carlo sampling methods for Bayesian filtering *Stat. Comput.* **10** 197
- [60] Granade C E, Ferrie C, Wiebe N and Cory D G 2012 Robust online Hamiltonian learning *New J. Phys.* **14** 103013



- [61] Wiebe N, Granade C, Ferrie C and Cory D G 2013 Hamiltonian learning and certification using quantum resources *Phys. Rev. Lett.* **112** 190501
- [62] Chase B A and Geremia J M 2009 Single-shot parameter estimation via continuous quantum measurement *Phys. Rev. A* **79** 022314
- [63] Huszar F and Houlisby N M T 2012 Adaptive Bayesian quantum tomography *Phys. Rev. A* **85** 052120
- [64] Granade C *et al* 2012 *QInfer: Library for Statistical Inference in Quantum Information* (<https://github.com/csferrie/python-qinfer>)
- [65] Jones E *et al* 2001 *SciPy: Open Source Scientific Tools For Python* ([www.scipy.org](http://www.scipy.org))
- [66] Zyczkowski K and Sommers H-J 2001 Induced measures in the space of mixed quantum states *J. Phys. A: Math. Gen.* **34** 7111
- [67] Magesan E, Gambetta J M and Emerson J 2011 Scalable and robust randomized benchmarking of quantum processes *Phys. Rev. Lett.* **106** 180504
- [68] Magesan E, Gambetta J M and Emerson J 2012 Characterizing quantum gates via randomized benchmarking *Phys. Rev. A* **85** 042311
- [69] Ryan C A, Laforest M and Laflamme R 2009 Randomized benchmarking of single-and multi-qubit control in liquid-state NMR quantum information processing *New J. Phys.* **11** 013034
- [70] Chow J M, Gambetta J M, Tornberg L, Koch J, Bishop L S, Houck A A, Johnson B R, Frunzio L, Girvin S M and Schoelkopf R J 2009 Randomized benchmarking and process tomography for gate errors in a solid-state qubit *Phys. Rev. Lett.* **102** 090502
- [71] Olmschenk S, Chicreanu R, Nelson K D and Porto J V 2010 Randomized benchmarking of atomic qubits in an optical lattice *New J. Phys.* **12** 113007
- [72] Brown K R, Wilson A C, Colombe Y, Ospelkaus C, Meier A M, Knill E, Leibfried D and Wineland D J 2011 Single-qubit-gate error below  $10^{-4}$  in a trapped ion *Phys. Rev. A* **84** 030303
- [73] Moussa O, da Silva M P, Ryan C A and Laflamme R 2012 Practical experimental certification of computational quantum gates using a twirling procedure *Phys. Rev. Lett.* **109** 070504
- [74] Granade C, Ferrie C and Cory D G 2014 Accelerated randomized benchmarking arXiv:1404.5275
- [75] Kimmel S, da Silva M P, Ryan C A, Johnson B R and Ohki T 2014 Robust extraction of tomographic information via randomized benchmarking *Phys. Rev. X* **4** 011050
- [76] Epstein J M, Cross A W, Magesan E and Gambetta J M 2014 Investigating the limits of randomized benchmarking protocols *Phys. Rev. A* **89** 062321
- [77] Schack R, Brun T A and Caves C M 2001 Quantum Bayes rule *Phys. Rev. A* **64** 014305
- [78] Caves C M, Fuchs C A and Schack R 2002 Unknown quantum states: The quantum de Finetti representation *J. Math. Phys.* **43** 4537
- [79] Caves C M, Fuchs C A and Schack R 2002 Quantum probabilities as Bayesian probabilities *Phys. Rev. A* **65** 022305
- [80] Fuchs C A and Schack R 2009 Priors in quantum bayesian inference *AIP Conf. Proc.* **1101** 255
- [81] Fuchs C A and Schack R 2009 Quantum-Bayesian coherence *Rev. Mod. Phys.* **85** 1693
- [82] Fuchs C A 2010 QBism, the perimeter of quantum Bayesianism arXiv:1003.5209
- [83] Fuchs C A and Schack R 2011 Bayesian conditioning, the reflection principle, and quantum decoherence *Probability in Physics* (Berlin: Springer)
- [84] Leifer M S and Spekkens R W 2011 A Bayesian approach to compatibility, improvement, and pooling of quantum states *J. Phys. A: Math. Theor.* **47** 275301
- [85] Leifer M S and Spekkens R W 2013 Towards a formulation of quantum theory as a causally neutral theory of Bayesian inference *Phys. Rev. A* **88** 052130
- [86] Jones K 1991 Principles of quantum inference *Ann. Phys., NY* **207** 140
- [87] Derka R, Buzek V, Adam G and Knight P L 1996 From quantum Bayesian inference to quantum tomography *J. Fine Mech. Opt.* **11-12** 341
- [88] Bužek V 1998 Reconstruction of Liouvillian superoperators *Phys. Rev. A* **58** 1723

- [89] Teklu B, Olivares S and Paris M G A 2009 Bayesian estimation of one-parameter qubit gates, *Journal of Physics B: Atomic J. Phys. B: At. Mol. Opt. Phys.* **42** 035502
- [90] Gill R D 2008 Conciliation of bayes and pointwise quantum State estimation: asymptotic information bounds in quantum statistics *Quantum Stochastics and Information: Statistics, Filtering and Control* ed V P Belavkin and M Guta (Singapore: World Scientific) pp 239–61
- [91] Sergeevich A, Chandran A, Combes J, Bartlett S and Wiseman H 2011 Characterization of a qubit Hamiltonian using adaptive measurements in a fixed basis *Phys. Rev. A* **84** 052315
- [92] Brivio D, Cialdi S, Vezzoli S, Gebrehiwot B T, Genoni M G, Olivares S and Paris M G A 2010 Experimental estimation of one-parameter qubit gates in the presence of phase diffusion *Phys. Rev. A* **81** 012305
- [93] Oi D and Schirmer S 2012 Quantum system characterization with limited resources *Phil. Trans. R. Soc. A* **370** 5386
- [94] Ferrie C, Granade C E and Cory D G 2012 Adaptive Hamiltonian estimation using Bayesian experimental design *AIP Conf. Proc.* **1443** 165
- [95] Ferrie C, Granade C E and Cory D G 2013 How to best sample a periodic probability distribution, or on the accuracy of Hamiltonian finding strategies *Quantum Inf. Process.* **12** 611