

PAPER • OPEN ACCESS

High posterior density ellipsoids of quantum states

To cite this article: Christopher Ferrie 2014 *New J. Phys.* **16** 023006

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

Related content

- [Practical Bayesian tomography](#)
Christopher Granade, Joshua Combes and D G Cory
- [Robust online Hamiltonian learning](#)
Christopher E Granade, Christopher Ferrie, Nathan Wiebe et al.
- [Error regions in quantum state tomography: computational complexity caused by geometry of quantum states](#)
Daniel Suess, ukasz Rudnicki, Thiago O maciel et al.

Recent citations

- [Error regions in quantum state tomography: computational complexity caused by geometry of quantum states](#)
Daniel Suess *et al*
- [Structured filtering](#)
Christopher Granade and Nathan Wiebe
- [Efficient Bayesian Phase Estimation](#)
Nathan Wiebe and Chris Granade

High posterior density ellipsoids of quantum states

Christopher Ferrie

Center for Quantum Information and Control, University of New Mexico, Albuquerque,
NM 87131-0001, USA
E-mail: csferrie@gmail.com

Received 14 October 2013, revised 12 January 2014

Accepted for publication 15 January 2014

Published 4 February 2014

New Journal of Physics **16** (2014) 023006

doi:[10.1088/1367-2630/16/2/023006](https://doi.org/10.1088/1367-2630/16/2/023006)

Abstract

Regions of quantum states generalize the classical notion of error bars. High posterior density (HPD) credible regions are the most powerful of region estimators. However, they are intractably hard to construct in general. This paper reports on a numerical approximation to HPD regions for the purpose of testing a much more computationally and conceptually convenient class of regions: posterior covariance ellipsoids (PCEs). The PCEs are defined via the covariance matrix of the posterior probability distribution of states. Here it is shown that PCEs are near optimal for the example of Pauli measurements on multiple qubits. Moreover, the algorithm is capable of producing accurate PCE regions even when there is uncertainty in the model.

1. Introduction

Characterizing a physical system at the most fundamental level requires specifying its quantum mechanical state. Arriving at such a description from measured data is called quantum tomography and the output of such a process is often a single point in the space of allowed parameters. In theory, by considering an infinite amount of data, a unique state can be identified. In practice, however, only a finite amount of data can be obtained. In such cases, it is impossible for a single reported state to coincide with the true state. In classical data fitting, *error bars* give a measure of the *accuracy* of the estimate. In the quantum state tomography setting, *regions* generalize this concept [1]. A region of quantum states should colloquially be understood to



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.

contain the true state with high probability (with the exact interpretation depending subtly on how the region is constructed).

Although this idea is quite simple, formalizing the concept of region estimation is not straightforward and there exists many competing alternatives, each with its own set of advantages and drawbacks. For example, *bootstrap resampling* is a common technique to produce error bars in tomographic experiments. However, as in [2] for example, bootstrapping has so far been exclusively used to calculate statistics on quantities derived from state estimates, such as fidelity to some target state. Bootstrapping is conceptually simple and easy to implement. However, the errors bootstrapping estimate come with no guarantees and it can grossly underestimate errors for estimators which produce states near the boundary [3]. The Fisher information (matrix) is most often used, via the Cramer–Rao inequality, to lower bound the variance of unbiased estimators. In this sense, it gives the errors one would expect in the asymptotic limit, provided an efficient estimator is used. In terms of regions, the Fisher information matrix is also asymptotically the inverse of the covariance matrix of the posterior distribution of parameters, which in turn defines an error ellipse [4]. For most problems, however, even computing the Fisher information numerically is an intractable problem¹. In some estimation strategies, such as compressed sensing [5] (also [6, 7]), the estimate of the state comes with a *certificate*. That is, an estimated state is provided along with an upper bound on the distance to the true state. This implicitly defines a ball in state space centered on the estimated state. However, the statistical meaning of this ball is not clear nor does a ball provide information about correlated errors. Confidence regions [3, 8] (and comparable constructions [9]) are the most stringently defined regions from the frequentist statistical perspective, although they are difficult to implement computationally (see, however, [10]).

There is one overarching theme to note here: trade-offs. On one end of the spectrum is conceptual simplicity, ease of implementation and computational tractability; on the other is statistical rigor, precision, accuracy and optimality. Here we take the approach of constructing statistically rigorous region estimators via a numerical algorithm which possess tunable parameters to control the trade-off between optimality and computational efficiency. The regions we construct here are approximations to *high posterior density (HPD) credible regions* and are in some sense the Bayesian analogues of confidence regions. To aid in the descriptive simplicity of the regions, we use *ellipsoids* which are well understood and easy to conceptualize geometrical objects. Moreover, ellipsoids provide a useful summary of how the state parameters are correlated.

The numerical algorithm we use is a Monte Carlo approximation to Bayesian inference and has been used to in the tomographic estimation of one and two qubit states [11] as well as in the continuous measurement of a qubit [12]. It has also been recently used to construct point and region estimates of Hamiltonian parameters in [13, 14]. In particular, the region used was the ellipse defined by the posterior covariance matrix. Here we show that the same method can be applied to quantum states and, more importantly, such regions approximate HPD credible regions.

One of the major advantages to this approach is that it naturally accommodates the possibility of unknown errors in modeling. For example, we might assume that the source of quantum states is fixed when it is not; or, we might assume that the measurements are known

¹ On the other hand, the algorithm presented here can be seen as an approximation to the Fisher matrix in the large sample limit.

exactly when they are not. Previous analysis of errors in quantum state estimation have focused on assessing its effect [15]; *detecting* its presence [16, 17]; and, most recently, selecting the best model for it [18, 19]. Here we demonstrate that our approach can estimate and construct regions for both quantum mechanical state and error model parameters *simultaneously*. That is, our algorithm produces a region in the space defined by all parameters.

This work is outlined as follows. In section 2 we overview the precise problem and theoretical solution. In section 3 we give the numerical algorithm which constructs the regions. In section 4 we describe the examples used to test the method and in section 5 the results of the numerical experiments are presented. Finally, in section 6 we conclude discussion.

2. Bayesian learning and credible regions

Each quantum mechanical problem specification produces a probability distribution $\Pr(d_k|\vec{x}; c_k)$, where d_k is the data obtained and c_k are the experimental designs (or *controls*) chosen for measurement k , and where \vec{x} is a vector parameterizing the system of interest.

Suppose we have performed experiments with control settings $C := \{c_1, c_2, \dots, c_N\}$ and obtained data $D := \{d_1, d_2, \dots, d_N\}$. The model specifies the likelihood function

$$\Pr(D|\vec{x}; C) = \prod_{k=1}^N \Pr(d_k|\vec{x}; c_k). \quad (1)$$

However, we are ultimately interested in $\Pr(\vec{x}|D; C)$, the probability distribution of the model parameters \vec{x} given the experimental data. We achieve this using Bayes' rule

$$\Pr(\vec{x}|D; C) = \frac{\Pr(D|\vec{x}; C)\Pr(\vec{x})}{\Pr(D|C)}, \quad (2)$$

where $\Pr(\vec{x})$ is the *prior*, which encodes any *a priori* knowledge of the model parameters. The final term $\Pr(D|C)$ can simply be thought of as a normalization factor. Since each measurement is statistically independent given \vec{x} , the processing of the data can be done on- or off-line. That is, we can sequentially update the probability distribution as the data arrive or post-process it afterward.

In many scenarios the mean of the posterior distribution²

$$\hat{\vec{x}}(D; C) = \mathbb{E}_{\vec{x}|D; C}[\vec{x}] \quad (3)$$

is the optimal *point* estimator [20]. But we desire more than point estimates. In this scenario, the most powerful region estimators are HPD regions. Let $\mathbb{1}_X$ be the indicator function of the set X . That is

$$\mathbb{1}_X(\vec{x}) = \begin{cases} 1 & \text{if } \vec{x} \in X, \\ 0 & \text{if } \vec{x} \notin X. \end{cases} \quad (4)$$

Then, the probability that \vec{x} lies in some set X is

$$\Pr(\vec{x} \in X|D; C) = \mathbb{E}_{\vec{x}|D; C}[\mathbb{1}_X(\vec{x})]. \quad (5)$$

The set X is an α -credible region if

$$\Pr(\vec{x} \in X|D; C) \geq 1 - \alpha. \quad (6)$$

² We use the following notation for expectation values with respect to the distribution of the random variable, say, X : $\mathbb{E}_X[f(X)] = \int f(X)\Pr(X) dx$.

That is, a set is α -credible region if no more than α probability mass exists outside the region or, equivalently, at least $1 - \alpha$ probability mass is contained within the region. The set X is an HPD α -credible region if

$$X = \{\vec{x} : \Pr(\vec{x}|D; C) \geq y\}, \quad (7)$$

where y is the largest number such that X is a α -credible interval. Under natural notions of ‘size’, HPD α -credible are the *smallest* (i.e. most powerful) α -credible regions.

Being optimal, HPD regions are the solutions to computationally intractable optimization problems—at least if we require a deterministic algorithm. In this work, we will explore a Monte Carlo algorithm to numerically approximate HPD regions. In particular, we use *posterior covariance ellipsoids* (PCEs). These are defined via the mean and covariance matrix of the posterior probability distribution as follows³:

$$X_{\text{PCE}} = \{\vec{x} : (\vec{x} - \mathbb{E}_{\vec{x}|D;C}[\vec{x}])^T \text{Cov}_{\vec{x}|D;C}[\vec{x}]^{-1} (\vec{x} - \mathbb{E}_{\vec{x}|D;C}[\vec{x}]) \leq Z_\alpha^2\}, \quad (8)$$

where Z_α^2 is the α -quantile of the χ_d^2 distribution [21], the values of which are readily available in countless tables. Note that this is simply the covariance ellipse under a Gaussian approximation to the posterior $\vec{x}|D; C \sim \mathcal{N}(\mathbb{E}_{\vec{x}|D;C}[\vec{x}], \text{Cov}_{\vec{x}|D;C}[\vec{x}])$. In addition to being HPD credible regions in the asymptotic limit, PCEs are computationally tractable and, even for modest numbers of experiments, they are remarkably close in size and coverage probability to true HPD regions. The remainder of this work is devoted to detailing the algorithm and demonstrating the above claims via simulated experiments on qubits.

3. Construction of the regions

Our numerical algorithm fits within the subclass of Monte Carlo methods called *sequential Monte Carlo* or SMC⁴. We approximate the posterior distribution by a weighted sum of delta-functions

$$\Pr(\vec{x}|D; C) \approx \sum_{j=1}^n w_j(D; C) \delta(\vec{x} - \vec{x}_j), \quad (9)$$

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

$$w_j(d_{l+1}; c_{l+1}) \propto \Pr(d_{l+1}|\vec{x}_j; c_{l+1}) w_j(d_l; c_l) \quad (10)$$

followed by a normalization step. The elements of the set $\{\vec{x}_j\}_{j=0}^n$ are called *particles* and are initially chosen by sampling the prior distribution $\Pr(\vec{x})$ and setting each weight $w_j = 1/n$. In the equations above, n is the number of particles and controls the accuracy of the approximation.

Note that the approximation in equation (9) is not a particularly good one *per se* (we are approximating a continuous function by a discrete one after all). However, it does allow us to calculate some quantities of interest with arbitrary accuracy. Like all Monte Carlo algorithms, it was designed to approximate expectation values, such that

$$\mathbb{E}_{\vec{x}|D;C}[f(\vec{x})] \approx \sum_{j=1}^n w_j(D; C) f(\vec{x}_j). \quad (11)$$

³ The covariance matrix of a vector random variable X is denoted $\text{Cov}_X[X] = \mathbb{E}_X[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T]$.

⁴ See [13] and references therein for a more detailed description of the algorithm and [22] for an implementation in software.

In other words, it allows us to efficiently evaluate difficult multidimensional integrals with respect to the measure defined by the posterior distribution.

The SMC approximation provides a nearly trivial method to approximate HPD credible regions, which surprisingly has been overlooked. Since the SMC approximate distribution (9) is a discrete distribution, the credible regions will be (at least initially) discrete sets. In particular, the HPD α -credible set, \hat{X}_{SMC} , is defined by the following construction:

- (1) Sort the particles according to their weights: $\{\vec{x}_j : w_i \geq w_l \text{ for } i < l\}$.
- (2) Collect particles (starting with the highest weighted) until the sum of the collected particle weights is at least $1 - \alpha$. The resulting set is \hat{X}_{SMC} .

The proof that this an HPD α -credible set is as follows. Assuming the particles are sorted as above. Begin with the highest weighted particle \vec{x}_1 with weight w_1 . Then, the set $X_1 = \vec{x}_1$ clearly has weight w_1 and the largest k satisfying equation (7), in the definition of HPD credible regions, is $k = w_1$. Now take the set $X_{1,2} = \{\vec{x}_1, \vec{x}_2\}$ with weight $w_1 + w_2$. The largest k is now $k = w_2$. Iterate this process until we reach the first weight w_{n_α} such that set $X_{1,2,\dots,n_\alpha} = \{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_{n_\alpha}\}$ satisfies $\sum_{j=1}^{n_\alpha} w_j \geq 1 - \alpha$. This set will have largest $k = w_{n_\alpha}$. The set $X_{1,2,\dots,n_\alpha}$ is clearly an α -credible set but it is also an HPD α -credible set since any $k > w_{n_\alpha}$ will result in a set excluding all particles \vec{x}_j with $j \geq n_\alpha$ and necessarily have weight less than $1 - \alpha$.

The immediate problem with \hat{X}_{SMC} is that it is a discrete set of points and while it is HPD α -credible set for the SMC approximated distribution, *any* discrete set has zero measure according the true posterior $\Pr(\vec{x}|D; C)$. The resolution is quite simple. Suppose we have some region \hat{X} which contains \hat{X}_{SMC} . Then, according to the SMC approximation (11)

$$\Pr(\vec{x} \in \hat{X}) = \mathbb{E}_x[\mathbb{1}_{\hat{X}}(x)] \approx \sum_j w_j \mathbb{1}_{\hat{X}}(\vec{x}_j) \geq 1 - \alpha \quad (12)$$

since $\mathbb{1}_{\hat{X}}(\vec{x}_j) = 1$ for all \vec{x}_j in \hat{X}_{SMC} . Thus, any region enclosing the points in \hat{X}_{SMC} will be a α -credible region.

But we do not want just any α -credible region. The HPD requirement is conceptually similar to asking for the region to be as small as possible while maintaining $1 - \alpha$ weight. If we assume (relaxed later on) the credible regions are convex, then we seek the smallest convex set containing \hat{X}_{SMC} . This defines the *convex hull* of \hat{X}_{SMC}

$$\hat{X}_{\text{hull}} = \text{hull}(\hat{X}_{\text{SMC}}). \quad (13)$$

Since \hat{X}_{hull} is a convex polytope in \mathbb{R}^{d^2-1} , it can be most compactly represented by a list of its vertices, which in the absolute worst cases is the entire set of SMC particles. That is, we require $n(d^2 - 1)$ numbers to specify \hat{X}_{hull} . Although certain classes of convex polytopes contain many symmetries and are easy to conceptualize geometrically, specifying the vertices of the convex hull of a random set of points is not most convenient representation for credible regions.

The most efficient way to describe this hull is the smallest ball containing it since this would be described by a location and single radial parameter. However, a ball would not account for large covariances in the posterior distribution. To account for these covariances, we will use ellipsoidal regions where \vec{c} and \mathbf{A} define an ellipsoid via the set of states satisfying $(\vec{x} - \vec{c})^T \mathbf{A}^{-1} (\vec{x} - \vec{c}) \leq 1$. In other words, \vec{c} is the center of the ellipsoid and \mathbf{A} its covariance matrix. Crucially, we want the smallest ellipse containing the hull, the so-called *minimum*

volume enclosing ellipse (MVEE)

$$(\vec{c}, \mathbf{A}) := \hat{X}_{\text{HPD}} = \text{HPD}(\hat{X}_{\text{hull}}). \quad (14)$$

To numerically construct the MVEE, we use the algorithm of Khachiyan [23].

To summarize, \hat{X}_{HPD} is the numerical approximation to the HPD credible region. The posterior covariance regions, \hat{X}_{PCE} , defined earlier in equation (8), are far less computationally intensive to construct than \hat{X}_{HPD} and are expected to be HPD credible regions in the asymptotic limit. In order to show that they are also approximately HPD credible regions for finite data, we compare \hat{X}_{HPD} and \hat{X}_{PCE} in a number of examples and also look at the performance in cases where limited computational resources prohibit constructing \hat{X}_{HPD} over many simulations.

Finally, we note that to compare the sizes of various ellipsoids, we will use the volume

$$\text{Vol}(\mathbf{A}) = \frac{\pi^{(d-1)/2}}{\Gamma(\frac{d}{2} + 1)} \det(\mathbf{A})^{-1/2}, \quad (15)$$

where d is the dimension of the parameter space and $\Gamma(n) = (n-1)!$ is the well-known Gamma function.

4. Examples

Consider repeated preparations of a qubit subjected to random Pauli measurements. We label the Pauli operators $\{\sigma_0, \sigma_1, \sigma_2, \sigma_3\}$ such that an arbitrary operator can be written

$$\rho = \frac{1}{2}(\sigma_0 + x_1\sigma_1 + x_2\sigma_2 + x_3\sigma_3) \quad (16)$$

and

$$x_k = \text{Tr}(\rho\sigma_k). \quad (17)$$

For many qubits, the situation is similar. The reconstruction is given by

$$\rho = \frac{1}{2^n} \sum_k x_k \sigma_k, \quad (18)$$

where

$$\sigma_k = \sigma_{k_1} \otimes \sigma_{k_2} \otimes \cdots \otimes \sigma_{k_n} \quad (19)$$

and we index by $k = k_1 + 4k_2 + 4^2k_3 + \cdots + 4^{n-1}k_n$. Then the parameterization is equivalent to that in equation (17). Since each Pauli is idempotent, $\sigma_k^2 = \mathbb{1}$, each individual measurement has two possible outcomes which we label $d_k \in \{+1, -1\}$ for the +1 and -1 eigenvalues. The likelihood function of a single measurement is then

$$\Pr(\pm 1 | \vec{x}, \sigma_k) = \frac{1 \pm x_k}{2}. \quad (20)$$

We also consider the effect of errors. We will not assume a particular model for the errors since any error model, for our two outcome measurements, manifests as a bit flip, or equivalently, randomization channel. For simplicity we assume the process is symmetric so we have a single parameter η , called the *visibility*, which has the following effect on the likelihood function:

$$\Pr(\pm 1 | \vec{x}, \eta, \sigma_k) = \eta \frac{1 \pm x_k}{2} + \frac{1 - \eta}{2}. \quad (21)$$

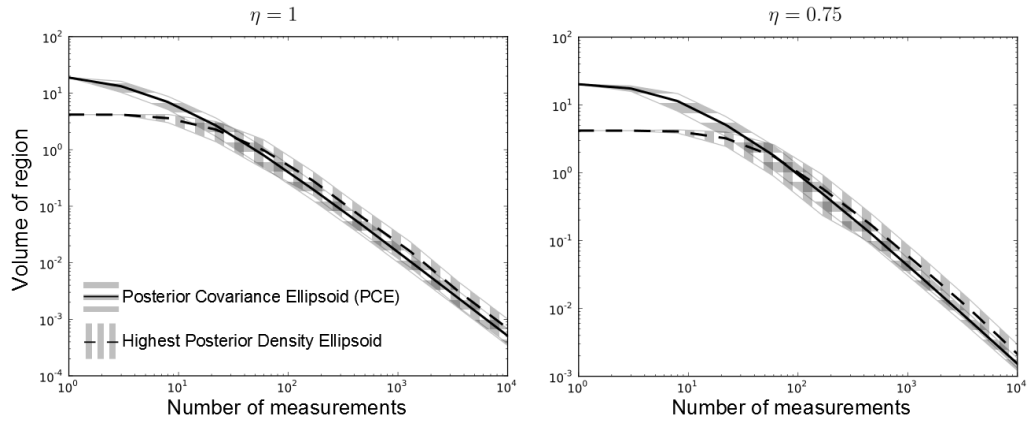


Figure 1. The volume of the constructed ellipsoid for (left) perfect measurements and (right) limited visibility measurements. For all constructed regions, $n = 100\,000$ particles were used. The results are from 100 simulations, where the line represents the mean and the shaded areas are those volumes one standard deviation from the mean.

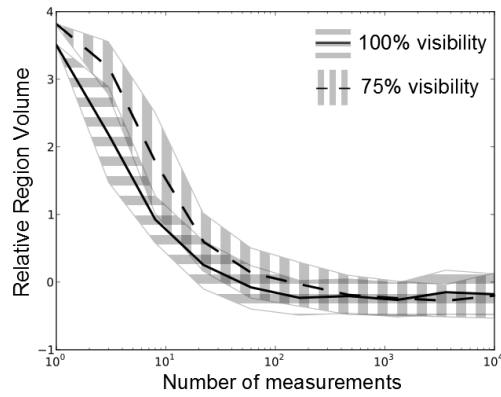


Figure 2. The relative size $(\text{Vol}(\hat{X}_{\text{PCE}}) - \text{Vol}(\hat{X}_{\text{HPD}})) / \text{Vol}(\hat{X}_{\text{HPD}})$ of the 95% credible regions, where \hat{X}_{HPD} is a numerical approximation to the HPD 95% credible region. For both strategies, $n = 100\,000$ particles were used. The results are from 100 simulations, where the line represents the mean and the shaded area are ratios one standard deviation from the mean. Note that the PCE is on average 10% smaller than the HPD region after about 100 measurements.

We consider two cases: the visibility is known and fixed or the visibility is unknown but still fixed run-to-run. In the former case, the task is to compute the PCE \hat{X}_{PCE} for the state only, while in the latter case the task is to compute the PCE over the *joint distribution* of (\vec{x}, η) . If only a region of states is desired, we can orthogonally project the PCE onto the subspace of the parameter space defining the state (and similarly for η). Hence, we will have separate marginal PCEs for the state and visibility separately.

Examples of how the regions are constructed are presented in figures 8 and 9.

5. Results

We first look at a comparison of the sizes of \hat{X}_{HPD} and \hat{X}_{PCE} . These results are presented in figures 1 and 2. In figures 1 and 2, the size of the two classes of regions is compared. Initially, the volume of the PCE is not smaller than that of the entire parameter space, which is to be expected

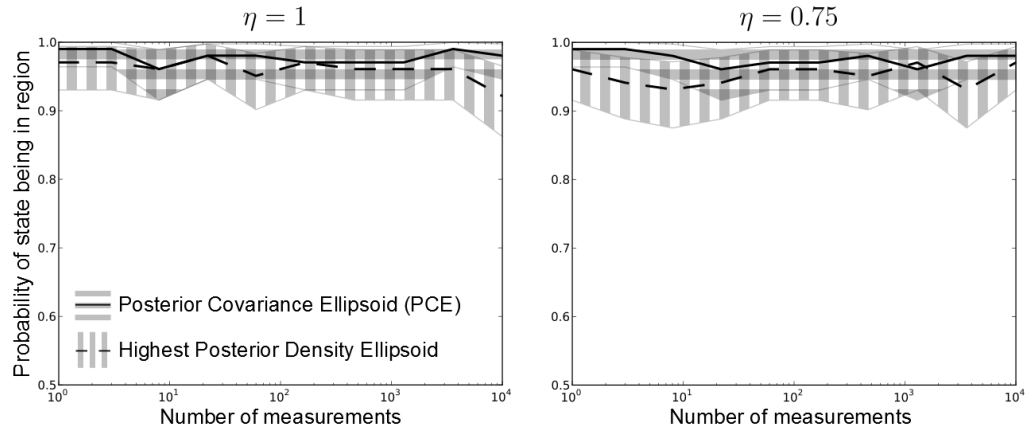


Figure 3. The probability of the state lying in the constructed ellipsoid for (left) perfect measurements and (right) limited visibility measurements. In all cases, the target was a 95% credible region. For all constructed regions, $n = 100\,000$ particles were used. The results are from 100 simulations, where the line represents the mean and the shaded area is (just to be meta) the HPD 95% credible region of the probability (derived from the beta distribution and a uniform prior).

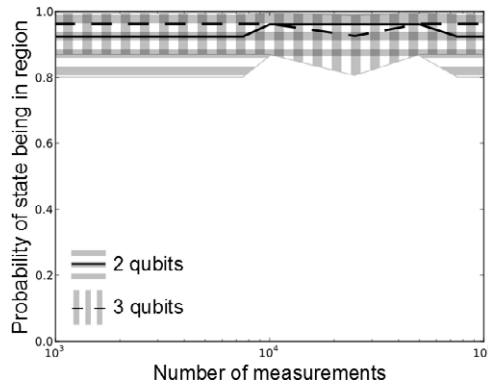


Figure 4. The probability of the state lying in the constructed posterior covariance ellipsoid (approximating the 95% HPD region) for the two and three qubit model described in the text and using a visibility parameter $\eta = 0.9$. For all constructed regions, $n = 100\,000$ particles were used. The results are from 25 simulations, where the line represents the mean and the shaded area is the HPD 95% credible region of the probability (derived from the beta distribution and a uniform prior).

since it is motivated from asymptotic normality. However, it rapidly converges in volume to, and becomes slightly smaller than, \hat{X}_{HPD} . Both sets of regions decrease in size at the same rate as a function of the number of measurements. These features are necessary in order to claim that the PCEs are approximate HPD credible regions. However, it is also necessary to confirm that the regions contain the true state with the correct probability. In figure 3, the probability (estimated from many simulations) of the state lying in the region is shown. We see that both regions correctly contain the state at the target level (95% in this case).

The astute reader will notice that, on average, the PCE is both smaller and contains the state with higher probability than the HPD. But the HPD was defined as the ‘smallest’ region with a fixed level of reliability so how can this be? The answer is both trivial and complicated.

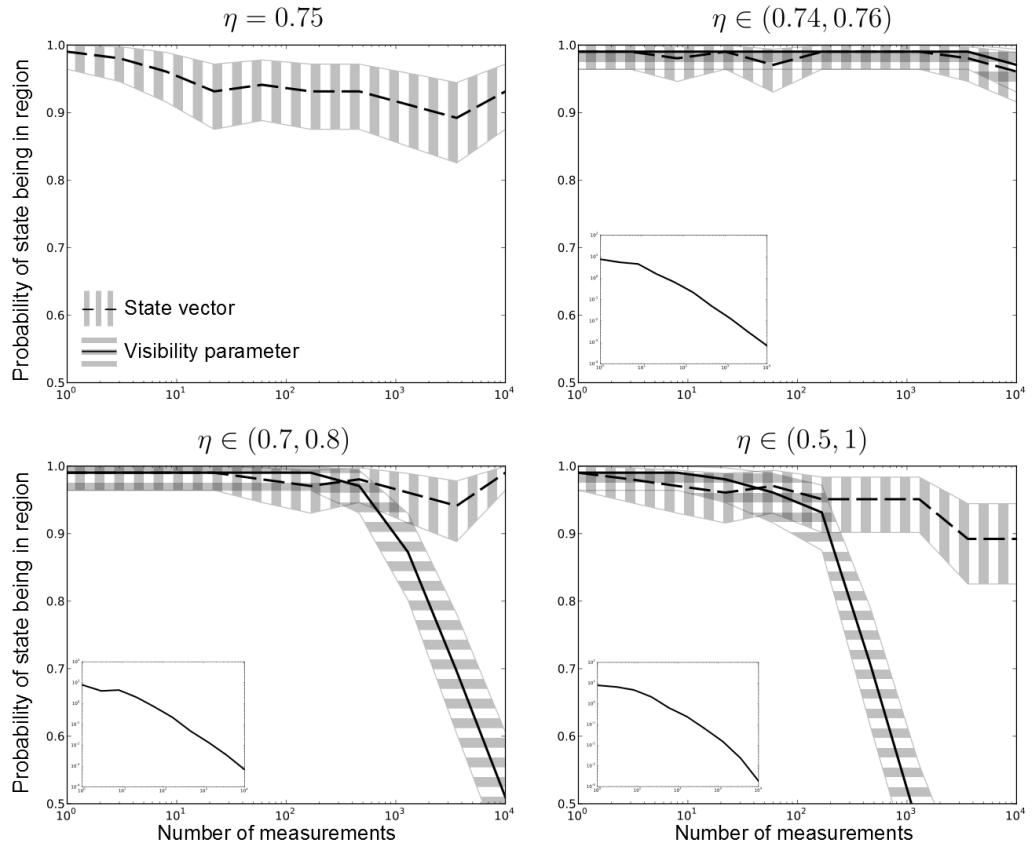


Figure 5. The probability of the state lying in the constructed (95% credible) posterior covariance ellipse for varying levels of knowledge of the visibility parameter. In the upper left, the visibility is known (this is identical to figure 3). In all other figures, the visibility is unknown (but known to lie in the specified interval—with a uniform prior). For both the state and the visibility parameter, a marginal posterior covariance ellipse is constructed and the tested against the true state and parameter. The inset in each of the later three figures shows the volume of the ellipsoid of states.

It is complicated because the accuracy of the numerical construction of the regions contains many layers of approximation. On the other hand, a trivial response is that the results are consistent with our expectation up to the numerical and statistical errors incurred by the approximation. In any case, the results demonstrate that the PCE is close to the HPD (and indeed better than a naive numerical approximation to the latter). This is important because, as opposed to all other region estimators, PCE regions are computationally tractable.

That the PCEs remain valid in higher dimensions is shown in figure 4. In figure 4 the probability for the state to lie in the constructed PCE region is shown to be consistent with the target of 95% containment probability for two and three-qubits subjected to random Pauli measurements.

In the above mentioned cases, the visibility η was assumed to be known. In figure 5, the case of unknown visibility is considered. When the visibility is known relatively accurately, the PCE captures the state and visibility accurately. However, as the initial variance in the prior on the visibility increases, the ability of the PCE to capture both the state and visibility

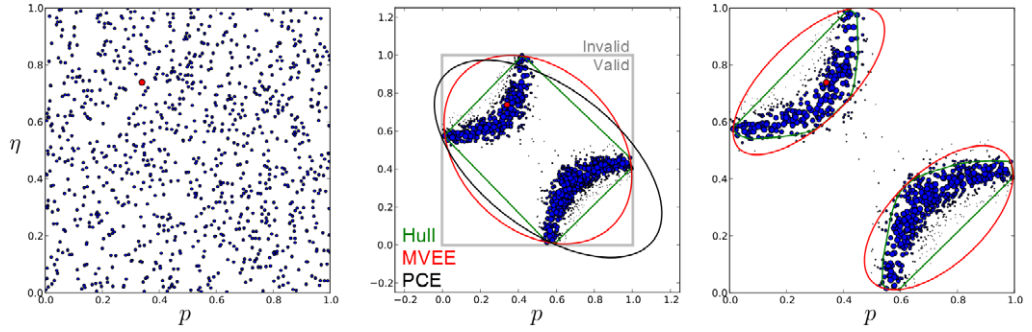


Figure 6. Construction of regions for qubit (known to be of the form $\rho = p|0\rangle\langle 0| + (1 - p)|0\rangle\langle 0|$) with unknown visibility subjected to 1000 randomly selected Z measurements. On the left, we have the initial 1000 particles (blue) randomly select according to a uniform prior and the randomly generated ‘true’ state (red). In the middle figure, we have the posterior SMC particle cloud after 1000 randomly selected Z measurements along with the following regions: the green line is the convex hull of those highest weighted particles comprising at least 95% of the particle weight (this is \hat{X}_{hull}); the red ellipse is \hat{X}_{HPD} , the smallest ellipse containing \hat{X}_{hull} and, in black is the ellipse defined by the estimated covariance matrix of the particle cloud, \hat{X}_{PCE} . When the posterior is disjoint, all regions poorly approximate the HPD credible region. On the right, the same distribution of particles is shown along with the convex hull and MVEE regions after the modes of the distribution have been identified via the DBSCAN clustering algorithm.

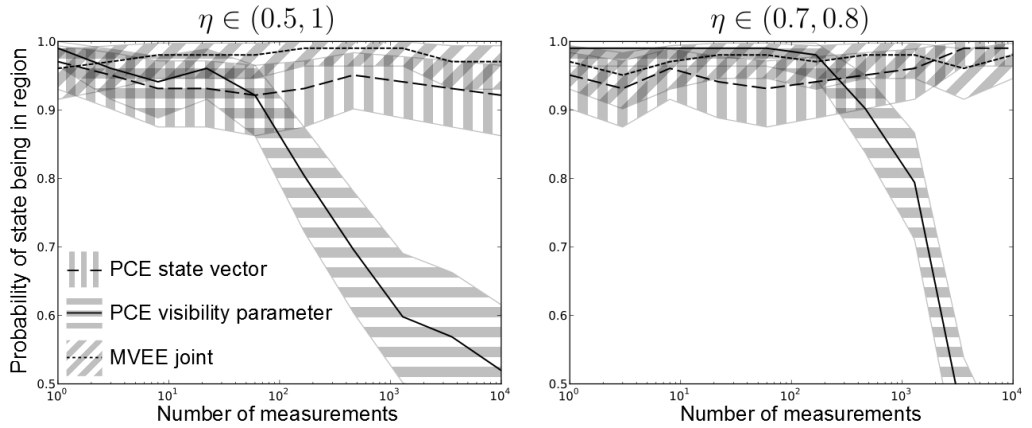


Figure 7. The probability of the state lying in the (95% credible) posterior covariance ellipse \hat{X}_{PCE} or the MVEE \hat{X}_{HPD} for varying levels of knowledge of the visibility parameter. The visibility is unknown (but known to lie in the specified interval—with a uniform prior). For both the state and the visibility parameter, a marginal posterior covariance ellipse is constructed and the tested against the true state and parameter.

diminishes. Surprisingly, the PCE still finds the state even when it cannot resolve the visibility accurately.

It may seem at first counterintuitive that the accuracy of the regions of states is higher in the case of unknown visibility (comparing the upper left plot of figure 5 to the rest). But, we must remember that the actual region constructed by the algorithm is an ellipsoid in the larger

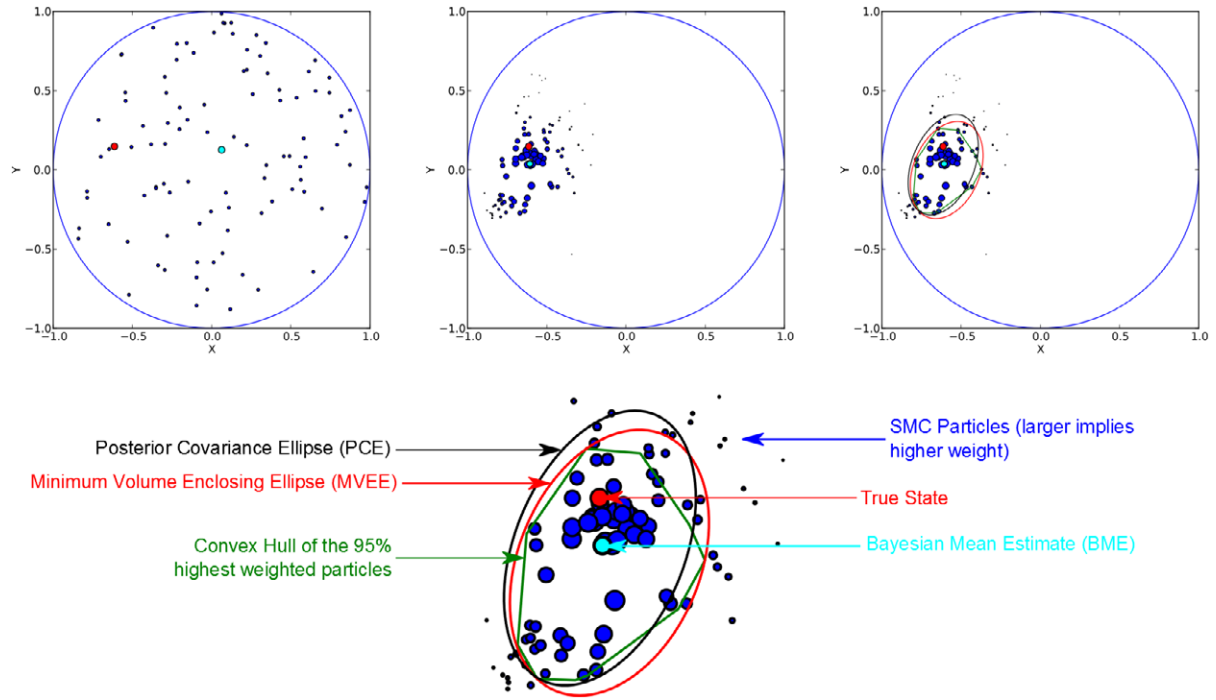


Figure 8. Construction of regions for a rebit subjected to 100 randomly selected X and Y . On the left, we have the initial 100 particles (blue) randomly select according to the Hilbert–Schmidt prior and the randomly generated ‘true’ state (red). In the middle figure, we have the posterior SMC particle cloud after 100 randomly selected X and Y measurements and the estimated state, the mean of distribution, is shown in teal. The larger weighted particles are represented as larger dots. On the right, the same distribution of particles is presented along with the regions discussed in the text. The green line is the convex hull of those highest weighted particles comprising at least 95% of the particle weight (this is \hat{X}_{hull}). The red ellipse is \hat{X}_{HPD} , the smallest ellipse containing \hat{X}_{hull} . Finally, in black, is the ellipse defined by the estimated covariance matrix of the particle cloud, \hat{X}_{PCE} . These objects are blown up below the figure to show details.

space which includes the visibility. The region of states reported is a projection of this object, a *marginal* PCE. Thus, it can happen that a covariance matrix with large covariance fails to capture the state while the ranges constructed by orthogonal projections do⁵.

The problem of failing to capture the state is easily identified to be the assumption that this posterior has a single mode with a convex HPD credible region. To illustrate the problem graphically, we need to reduce the dimensionality. To this end, we assume the state is of the form $\rho = p|0\rangle\langle 0| + (1 - p)|1\rangle\langle 1|$, with p unknown and to be estimated along with the visibility η . A typical example of a posterior distribution and the possible regions is shown in figure 6. There are two things to note: (i) the posterior distribution has two modes and (ii) even within each mode, the distribution is not well approximated by a Gaussian. Both of these are due to the degeneracy in the posterior distribution of (p, η) arising from the symmetry in of p and η in the

⁵ To get an intuitive two-dimensional picture, draw an ellipse and project it on to the axes. All the states in the smallest rectangle containing the ellipse are contained in the *marginal* ranges but not necessarily in the original ellipse.

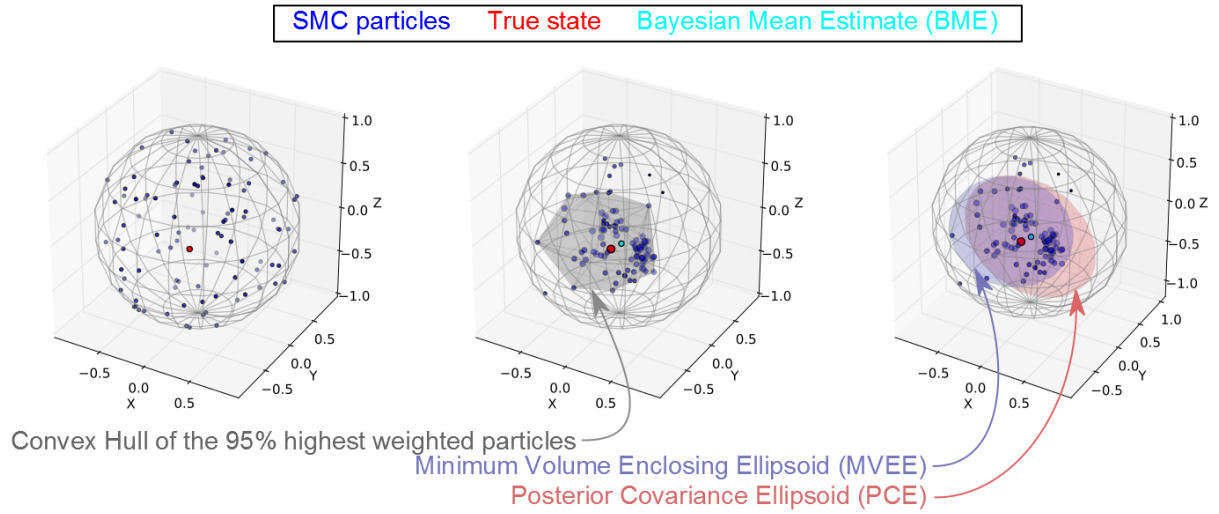


Figure 9. Construction of regions for qubit subjected to 20 randomly selected Pauli measurements using the SMC approximation with 100 particles. On the left, we have the initial 100 particles (blue) randomly select according to the Hilbert–Schmidt prior and the randomly generated ‘true’ state (red). In the middle and right figure, we have the posterior SMC particle cloud after 20 randomly selected Pauli measurements and the estimated state, the mean of distribution, is shown in teal. The larger weighted particles are represented as larger dots. In the middle figure, the gray object is the convex hull of those highest weighted particles comprising at least 95% of the particle weight (this is \hat{X}_{hull}). In the right figure, the blue ellipsoid is \hat{X}_{HPD} , the smallest ellipse containing \hat{X}_{hull} while the red ellipsoid is the PCE \hat{X}_{PCE} .

likelihood function. For example, an outcome could equally well be explained by a large p and small η as a small η and large p .

The problem of many modes in the posterior can be resolved by reporting disjoint ellipsoidal regions, one for each mode. The discrete set of highest weighted SMC points X_{SMC} naturally find themselves within the modes. Given this set, the task is to then identify which particles belong to which modes. In machine learning parlance, this is the problem of *clustering*. Many solutions to this problem exist, each with its own set of advantages and drawbacks. Here we have used DBSCAN [24, 25], as it seems to require the fewest number of assumptions⁶.

When the visibility is known fairly well only the first problem, disjoint regions, is automatically resolved. In other words, it is not likely that the PCE will be the optimal region estimator unless the visibility is relatively well-known. Practically, when the noise is known with some—but not perfect—accuracy, the MVEE region \hat{X}_{HPD} still behaves properly even when the PCE region \hat{X}_{PCE} does not. This is demonstrated in figure 7 where we see that \hat{X}_{HPD} contains the true state with the correct probability but \hat{X}_{PCE} does not. In practice then, the recommendation is to identify whether the problem specifies convex credible regions. If so, then the PCE is the appropriate choice; if not, then a clustering algorithm should be used to identify the modes of the distribution first.

⁶ This clustering algorithm was suggested to the author by Chris Granade (as noted in [13]).

6. Conclusion

For the three qubit data shown in figure 4, the constructed PCEs were ellipsoids in a $4^3 - 1 = 63$ dimensional parameters space. That these simulations were performed on an average laptop computer lends credence to the claim that the regions discussed here, constructed via the SMC algorithm, are a computationally attractive solution to the problem of providing region estimators of quantum states. For the implementation given in [22], further optimization of the code itself would be required to move beyond 100-or-so parameters for conventional computing resources. Alternative to this is the ‘throw money at it’ solution and run the code on high performance computing machines.

The key obstacle to a fully scalable solution is the *curse of dimensionality*, which is ever-present in quantum theory. The problem that the parameter space grows exponentially presents a challenge in representing the quantum state and simulating the dynamics of the quantum system. Both of these obstacles can be overcome within the computational framework presented here. To address these problems, we follow the already many ingenious methods reducing the complexity of identifying and characterizing quantum states and processes. These include identifying stabilizer states [26], tomography for matrix product states [27], tomography for permutationally invariant states [28], learning local Hamiltonians [29], tomography for low-rank states via compressed sensing [5] and tomography for multi-scale entangled states [30]. These techniques employ efficient simulation algorithms which propagate efficient representations of the state vector to calculate the probabilities defining the likelihood function. Physical constraints and careful engineering lead to such drastic dimensional reductions that we can use along with additional prior information in the Bayesian method.

The above mentioned methods aimed at reducing the complexity of estimating parameters has relied on the notion of *strong* simulation, where the likelihood function is computed exactly. On the other hand is *weak* simulation, where the likelihood is not computed but is instead sampled from. The distinction between strong and weak simulation has been a topic of recent interest in quantum computational complexity where it has been shown, for example, that there exists subtheories of quantum mechanics which admit efficient weak simulation but do not allow for efficient strong simulation [31, 32]. It has recently been shown that the Bayesian sequential Monte Carlo algorithm can also be used in the case where one only has access to a weak simulator [33].

Finally, we might find ourselves with a quantum system which does not admit efficient strong nor weak simulation. In that case, it still may be possible to efficiently characterize the system using a trusted *quantum* simulator. For the case of estimating dynamical parameters, it has been shown that the Bayesian SMC approach can also perform estimation tasks efficiently using a quantum simulator [14].

In this work we have considered supplementing point estimates of quantum states with *regions* of state space. These regions contain the true state with a pre-determined probability and within the tolerance of the numerical algorithm. The numerical algorithm has tunable parameters which trades accuracy for computational efficiency and thus can be determined based on desired optimality and available computational resources. When the noise is known with relatively high accuracy, the optimal regions are the *PCEs*. When the noise is unknown, more complex techniques are available to construct ellipsoids which capture the state. In any case, the constructed regions are ellipsoids which are easily described and conceptualized.

In the context of classical statistics, quantum state estimation can simply be thought of as overly ambitious parameter estimation. That is, quantum state estimation is just classical

parameter estimation with a specific model and, perhaps, oddly appearing constraints. The point is that the framework presented here for region estimation is suitable to any parameter estimation problem. In particular, we have already shown that additional noise on the measurements can be estimated *simultaneously* with the unknown state. More generally, the framework possesses a beautiful modularity which allows arbitrary statistical models to be learned.

Acknowledgments

The algorithm presented here has been implemented in a software package called *Qinfer* [22] using the Python programming language and the scientific computation library *SciPy* [34]. The author thanks Chris Granade for many discussions, timely advice and the majority of contributions necessary to make this open-source software a reality, which has indeed proven useful in many of our collaborations. This work was supported by the National Science Foundation grant numbers PHY-1212445 and PHY-1314763, by Office of Naval Research grant number N00014-11-1-0082 and by the Canadian Government through the NSERC PDF program.

References

- [1] Blume-Kohout R 2010 *New J. Phys.* **12** 043034
- [2] Home J P, Hanneke D, Jost J D, Amini J M, Leibfried D and Wineland D J 2009 *Science* **325** 1227
- [3] Blume-Kohout R 2012 Robust error bars for quantum tomography arXiv:1202.5270
- [4] Řeháček J, Mogilevtsev D and Hradil Z 2008 *New J. Phys.* **10** 043022
- [5] Flammia S T, Gross D, Liu Y-K and Eisert J 2012 *New J. Phys.* **14** 095022
- [6] Sugiyama T, Turner P S and Murao M 2011 *Phys. Rev. A* **83** 012105
- [7] Sugiyama T, Turner P S and Murao M 2013 *Phys. Rev. Lett.* **111** 160406
- [8] Christandl M and Renner R 2012 *Phys. Rev. Lett.* **109** 120403
- [9] Shang J, Ng H K, Sehrawat A, Li X and Englert B-G 2013 *New J. Phys.* **15** 123026
- [10] Arrazola J M, Gittsovich O, Donohue J M, Lavoie J, Resch K J and Lütkenhaus N 2013 *Phys. Rev. A* **87** 062331
- [11] Huszár F and Houlshby N M T 2012 *Phys. Rev. A* **85** 052120
- [12] Chase B A and Geremia J M 2009 *Phys. Rev. A* **79** 022314
- [13] Granade C E, Ferrie C, Wiebe N and Cory D G 2012 *New J. Phys.* **14** 103013
- [14] Wiebe N, Granade C E, Ferrie C and Cory D G 2013 Hamiltonian learning and certification using quantum resources arXiv:1309.0876
- [15] Rosset D, Schöbitz R F, Bancal J D, Gisin N and Liang Y C 2012 *Phys. Rev. A* **86** 062325
- [16] Moroder T, Kleinmann M, Schindler P, Monz T, Gühne O and Blatt R 2013 *Phys. Rev. Lett.* **110** 180401
- [17] Langford N K 2013 *New J. Phys.* **15** 035003
- [18] van Enk S J and Blume-Kohout R 2013 *New J. Phys.* **15** 025024
- [19] Schwarz L and van Enk S 2013 Error models in quantum computation: an application of model selection arXiv:1307.0858
- [20] Blume-Kohout R and Hayden P 2006 arXiv:quant-ph/0603116
- [21] Robert C P 2007 *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation* 2nd edn (*Springer Texts in Statistics*) (New York: Springer)
- [22] Ferrie C and Granade C 2012 QInfer: library for statistical inference in quantum information (<http://github.com/csferrie/python-qinfer>)
- [23] Todd M J and Yildirim E A 2007 *Discrete Appl. Math.* **155** 1731

- [24] Ester M, Kriegel H-P, Jorg S and Xu X 1996 *Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining (KDD-96)* (Palo Alto, CA: AAAI) pp 226–31 (<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.71.1980>)
- [25] Pedregosa F *et al* 2011 *J. Mach. Learn. Res.* **12** 2825
- [26] Gottesman D 2008 <http://pirsa.org/08080052/>
- [27] 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 *Nature Commun.* **1** 149
- [28] Tóth G, Wieczorek W, Gross D, Krischek R, Schwemmer C and Weinfurter H 2010 *Phys. Rev. Lett.* **105** 250403
- [29] da Silva M P, Cardinal O L and Poulin D 2011 *Phys. Rev. Lett.* **107** 210404
- [30] Landon-Cardinal O and Poulin D 2012 Practical learning method for multi-scale entangled states arXiv:1204.0792
- [31] Aaronson S and Arkhipov A 2010 arXiv:1011.3245
- [32] Van den Nest M 2011 *Quantum Inform. Comput.* **11** 784–812
- [33] Ferrie C and Granade C E 2013 Likelihood-free quantum inference: tomography without the Born rule arXiv:1304.5828
- [34] Jones E, Oliphant T and Peterson P 2001 SciPy: open source scientific tools for Python (www.scipy.org/)