

Practical Bayesian tomography

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2016 *New J. Phys.* 18 033024

(<http://iopscience.iop.org/1367-2630/18/3/033024>)

[View the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 61.149.19.4

This content was downloaded on 15/05/2017 at 12:49

Please note that [terms and conditions apply](#).

You may also be interested in:

[High posterior density ellipsoids of quantum states](#)

Christopher Ferrie

[Optimal, reliable estimation of quantum states](#)

Robin Blume-Kohout

[Robust online Hamiltonian learning](#)

Christopher E Granade, Christopher Ferrie, Nathan Wiebe et al.

[Quantum tomographic reconstruction with error bars: a Kalman filter approach](#)

Koenraad M R Audenaert and Stefan Scheel

[Quantum bootstrapping via compressed quantum Hamiltonian learning](#)

Nathan Wiebe, Christopher Granade and D G Cory

[Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators](#)

Steven T Flammia, David Gross, Yi-Kai Liu et al.

[Open-system dynamics of entanglement:a key issues review](#)

Leandro Aolita, Fernando de Melo and Luiz Davidovich

[Near-optimal quantum tomography: estimators and bounds](#)

Richard Kueng and Christopher Ferrie

[Accelerated randomized benchmarking](#)

Christopher Granade, Christopher Ferrie and D G Cory

New Journal of Physics

The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft 

IOP Institute of Physics

Published in partnership
with: Deutsche Physikalische
Gesellschaft and the Institute
of Physics



OPEN ACCESS

PAPER

Practical Bayesian tomography

Christopher Granade^{1,2}, Joshua Combes^{3,4,5} and D G Cory^{3,5,6,7}

¹ Centre for Engineered Quantum Systems, University of Sydney, Sydney, NSW, Australia

² School of Physics, University of Sydney, Sydney, NSW, Australia

³ Institute for Quantum Computing, University of Waterloo, Waterloo, ON, Canada

⁴ Department of Applied Mathematics, University of Waterloo, Waterloo, ON, Canada

⁵ Perimeter Institute for Theoretical Physics, 31 Caroline St. N, Waterloo, Ontario, N2L 2Y5, Canada

⁶ Department of Chemistry, University of Waterloo, Waterloo, ON, Canada

⁷ Canadian Institute for Advanced Research, Toronto, ON, Canada

E-mail: cgranade@cgranade.com

Original content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence.

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



Abstract

In recent years, Bayesian methods have been proposed as a solution to a wide range of issues in quantum state and process tomography. State-of-the-art Bayesian tomography solutions suffer from three problems: numerical intractability, a lack of informative prior distributions, and an inability to track time-dependent processes. Here, we address all three problems. First, we use modern statistical methods, as pioneered by Huszár and Houslsby (2012 *Phys. Rev. A* **85** 052120) and by Ferrie (2014 *New J. Phys.* **16** 093035), to make Bayesian tomography numerically tractable. Our approach allows for practical computation of Bayesian point and region estimators for quantum states and channels. Second, we propose the first priors on quantum states and channels that allow for including useful experimental insight. Finally, we develop a method that allows tracking of time-dependent states and estimates the drift and diffusion processes affecting a state. We provide source code and animated visual examples for our methods.

1. Introduction

Quantum state and process tomography are important methods for diagnosing and characterizing imperfections in small quantum systems. By fixing problems in models and implementations, and by having a well-characterized system, we may hope to compose multiple systems to build reliable larger quantum systems. These larger systems require a scalable approach to characterization, such as using the matrix-product state ansatz [3] or information locality [4, 5]. Quantum tomography has seen many improvements since its inception [6]. In particular, tomography has enjoyed advances in providing maximum-likelihood estimators [7], region estimators [8–10], model selection [11, 12], hedging [13], and compressed sensing [14, 15].

These techniques, though powerful, do not take advantage of prior information available to experimentalists. Such prior information can include knowledge gained in building the experiment, or in performing similar experiments, as well as knowledge gained from the calibration leading up to an experiment of interest. A class of techniques that allow one to include prior information is called Bayesian estimation.

Bayesian techniques in the context of quantum tomography were first suggested by Jones [16], Slater [17], Derka *et al* [18], Bužek *et al* [19], and Schack *et al* [20]. In addition to the inclusion of prior information, Bayesian estimation also naturally includes several other experimental advantages, such as optimality [21–23], adaptive experimental design [1, 24], robust region estimates [25] and model selection criteria [2, 26]. These advantages arise from the fact that Bayesian methods provide a complete characterization of the current state of an experimentalist's knowledge after each datum.

Given the many proposals for and advantages of Bayesian methods, the lack of adoption of Bayesian methods in experimental tomography, with the exception of some recent work [24, 27], seems to be primarily a practical problem. Bayesian methods are rarely analytically tractable. Further numerical implementations of

Bayesian methods (for inference of many variables) are a small scale software engineering project and computationally expensive. Bayesian reasoning, in the classical statistics literature, is typically realized with efficient well-known classical algorithms which go by the names of particle filtering [28] and sequential Monte Carlo (SMC) [29]. Even though these algorithms can be numerically efficient, for multiple variables they are non-trivial to implement and optimize. In the context of quantum state tomography SMC has been applied to adaptive tomography [1] and model selection [2], generalizing and dramatically simplifying earlier efforts based on Kalman filtering [30]. However, much of the code developed for this application is either specialized to particular cases, or has not been released to or adopted by the community. Releasing reusable code is critical not only for practicality in experiments, but also for producing reproducible research results [31].

Another difference between the application of Bayesian methods in classical statistics and its application to quantum state and processes tomography is the lack of choice in priors. Experimentalists spend many hours designing, testing and calibrating their platforms. It would be nice to include the prior information from this lengthy process into a quantum estimation procedure. Classically one can choose from many different priors, which could be ‘informative’ or ‘uninformative,’ depending on the domain of the probability distribution. In the quantum setting, many canonical priors are unitarily invariant and have some uninformative distribution over purity [32]. The lack of alternatives can be explained both by the lack of theoretical knowledge on how to construct such priors and more importantly there is no general purpose software available that could make use of such priors once constructed.

Finally, there is a critical issue facing the tomography community, independent of the respective merits of frequentist or Bayesian approaches. It is the fact that the output of realistic quantum sources vary in time, both deterministically (‘drift’) and stochastically (‘diffusion’). Incorporating time-dependence in quantum characterization protocols has been a topic of significant recent interest [12, 33–37]. For the most part, model selection has been used in tomographic experiments to detect drift or diffusion. While useful, this does not yield a protocol for incorporating that drift (or diffusion) into the estimation protocol once it has been detected. Moreover, many of the current solutions are not general purpose, nor are they practical in a range of important applications. For instance, see the proposal of Blume-Kohout *et al* [38], which requires quantum memory on the order of the number of samples collected so that the DeFinetti conditions are met. This is clearly impractical, and further precludes adaptivity.

In this work, we address all three issues. The unifying theme is the use of the SMC algorithm to perform tomography.

To address the issue of adoption we implement SMC-based tomography using an open-source library for Python [39] that integrates with the widely used QuTiP library for quantum information [40], and can be used with modern instrument control software [41]. The techniques that we introduce in this paper can readily be applied in experimental practice—we provide a tutorial on software implementations in appendix F.

Second, we give a constructive method for defining priors that represent initial estimates informed by experimental insight. We use SMC to avoid the need to write an analytic form for our prior. This is especially useful in the context of quantum state and process tomography, where analytic expressions are available for only a few distributions. Instead, SMC only requires that we provide a means to sample from priors. To do so, we first draw sample states or channels from a reference or ‘flat’ prior, then transform each sample by a quantum operation drawn from an ensemble. The only input required to define this ensemble is a prior estimate of the state. Our method then guarantees that this becomes the mean of our new prior.

Third, we incorporate tracking of drifting and diffusing evolution into state tomography by using techniques from computer vision which perturb a sample. Each such perturbation is drawn from a Gaussian whose mean and variance represent deterministic and stochastic evolution respectively.

This article is structured as follows. We begin by reviewing Bayesian tomography in section 2. Then we describe prior work on priors for quantum states and channels in section 3. In section 4 we transform these priors into new priors for states and channels that include experimental insight. Using Bayesian tomographic methods, we then describe how to track quantum states and channels as a function of time in section 5. In section 6 we illustrate these ideas and more, then conclude in section 7.

2. Bayesian tomography

To infer the state of a quantum system of dimension D , we perform a set of measurements on N identically and independently prepared quantum systems. As is usually the case, we will restrict to measuring each system separately. In general, one could consider performing a different generalized measurement on each system, and the kind of generalized measurement could adaptively depend on all prior measurements. This can be included in the expressions below at the cost of additional notational baggage.

Consider a single positive operator valued measure (POVM) $\{E_k\}$ whose K elements represent the outcomes of measurements. If we perform this generalized measurement on all N systems it results in a string of measurement results $\mathcal{M} = \{M_1, M_2, \dots, M_N\}$, where M_i is the POVM element obtained in the i th trial. Let n_k be the number times that E_k is observed in \mathcal{M} , i.e. frequency of E_k . The statistical information measurement outcomes have about the preparation is described by a *likelihood function*; that is, a probability distribution over measurement records, conditioned on a hypothesis ρ about the state. In particular, by using Born's rule $\Pr(E_k|\rho) = \text{Tr}(E_k\rho)$ to write out the likelihood, we obtain that

$$\mathcal{L}(\rho) = \Pr(\mathcal{M}|\rho) = \prod_{i=1}^N \text{Tr}[M_i\rho] \quad (1a)$$

$$= \text{Tr}[E_1\rho]^{n_1} \text{Tr}[E_2\rho]^{n_2} \dots \text{Tr}[E_K\rho]^{n_K} \quad (1b)$$

Moreover, by using the Choi–Jamiołkowski isomorphism, we can associate a state $J(\Lambda)/D$ with each quantum channel Λ . As we detail in appendix A, preparation and measurement in a process tomography experiment can be written as a measurement of the state $J(\Lambda)/D$, such that (1a) also includes this case.

In general, a likelihood function completely models an experiment by specifying the probability of observing any measurement, conditioned on the hypothesis we would like to learn. Thus, the likelihood function serves as the basis for subsequent estimation.

For instance, in maximum likelihood estimation (MLE), an estimate $\hat{\rho}$ of the state ρ is formed by

$$\hat{\rho}_{\text{MLE}} := \arg \max_{\rho} \mathcal{L}(\rho). \quad (2)$$

Here, however, we use the likelihood function to instead perform Bayesian inference as described by Jones [16], Slater [17], Derka *et al* [18], Bužek *et al* [19], and Schack *et al* [20]. Below we follow Blume-Kohout's [22] presentation closely. We begin by using Bayes' rule

$$\Pr(\rho|\mathcal{M})d\rho = \frac{\Pr(\mathcal{M}|\rho)\Pr(\rho)d\rho}{\Pr(\mathcal{M})}, \quad (3)$$

where $\pi(\rho) := \Pr(\rho)d\rho$ is called the *prior* distribution, and

$$\mathcal{N} = \Pr(\mathcal{M}) = \int d\rho \rho \Pr(\mathcal{M}|\rho)\Pr(\rho) \quad (4)$$

is a normalization constant. We shall write $\rho \sim \pi$ to indicate that the random variable ρ is drawn from the prior π .

In the next two Sections, we will return to the question of how to choose the prior distribution $\Pr(\rho)d\rho$. Henceforth, we drop the measure $d\rho$ unless we are integrating a distribution, as we will later use a numerical algorithm which approximates this continuous distribution by a discrete distribution.

The Bayesian mean estimate (BME) is then given by the expectation

$$\hat{\rho}_{\text{BME}}(\mathcal{M}) := \mathbb{E}_{\rho}[\rho|\mathcal{M}] = \int d\rho \rho \Pr(\rho|\mathcal{M}), \quad (5)$$

where \mathbb{E} indicates an expectation value, and where conditional bars and the subscript denote the distribution the expectation is taken over e.g. $\mathbb{E}_x[f(x)|y] = \sum_x f(x)\Pr(x|y)$ denotes the conditional expectation of $f(x)$ given y .

The Bayesian mean estimator is an optimal estimator for any strictly proper scoring rule on states [21]. These scoring rules arise from Bregman divergences [42] such as the Kullback–Leibler divergence or the quadratic loss $L_Q(\rho, \hat{\rho}) := \text{Tr}[(\rho - \hat{\rho})Q[\rho - \hat{\rho}]]$, where Q is a positive semidefinite superoperator⁸. As we will discuss in more detail below, the error incurred by the BME is well-characterized by spread of samples from the posterior distribution. Importantly, if one uses infidelity as a loss function, the BME remains approximately optimal, even though the infidelity is not a Bregman divergence [23].

To make the problem of estimating states and channels more concrete, it is helpful to specify a real-valued parameterization of the tomographic model. We start by considering the space of linear operators acting on a D -dimensional Hilbert space. We represent an operator with an ‘operator ket’ $|A\rangle\langle A|$ corresponding to A , while the dual vector is the corresponding ‘bra’ $\langle A|$ and represents A^\dagger . This vector space has the Hilbert–Schmidt inner product $\langle\langle A|B\rangle\rangle = \text{Tr}[A^\dagger B]$. In the D -dimensional Hilbert space, a state matrix can be represented as

$$\rho = \frac{1}{D}\mathbb{1} + \sum_{\alpha=1}^{D^2-1} x_\alpha B_\alpha, \quad (6)$$

where $x_\alpha = \langle\langle B_\alpha|\rho\rangle\rangle$ for a basis of Hermitian operators $\{B_\alpha\}$ that is orthonormal under the Hilbert–Schmidt inner product $\langle\langle B_\alpha|B_\beta\rangle\rangle = \text{Tr}[B_\alpha^\dagger B_\beta] = \delta_{\alpha,\beta}$. For simplicity, we choose $B_0 = \mathbb{1}/\sqrt{D}$ to be the only traceful

⁸This definition of the quadratic loss follows from the typical definition by interpreting the scale matrix as a superoperator.

element. The corresponding operator ket representation of ρ is

$$|\rho\rangle\langle\rho| = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_{D^2-1} \end{pmatrix} = \begin{pmatrix} \langle\langle B_0|\rho\rangle\rangle \\ \langle\langle B_1|\rho\rangle\rangle \\ \langle\langle B_2|\rho\rangle\rangle \\ \vdots \\ \langle\langle B_{D^2-1}|\rho\rangle\rangle \end{pmatrix} = \begin{pmatrix} 1/\sqrt{D} \\ \text{Tr}[B_1\rho] \\ \text{Tr}[B_2\rho] \\ \vdots \\ \text{Tr}[B_{D^2-1}\rho] \end{pmatrix}. \quad (7)$$

As a consequence of ρ being considered as a random variable, each parameter x_i is also a random variable. That is, we have represented ρ in terms of a vector-valued random variable $|\rho\rangle\langle\rho|$; because we have chosen a Hermitian basis, $|\rho\rangle\langle\rho|$ is also a real vector. Moreover, each parameter x_i is then by definition equal to the mean value $\langle B_i \rangle$ of the observable B_i , taken over possible measurement outcomes. In general, tensor products of Pauli matrices and generalized Gell-Mann matrices can be used as such a Hermitian basis for multiple finite-dimensional quantum systems.

Having chosen a parameterization in terms of observables, we can now reason about the *error* in parameters of $\hat{\rho}$. Suppose that for a given posterior, ρ is normally distributed about the estimated state $\hat{\rho}$, then the distribution over ρ is fully described by its mean, i.e. (5), and covariance $\sigma_{i,j}$ between \hat{x}_i and \hat{x}_j (the i and j components of $\hat{\rho}$)

$$\sigma_{i,j} = \text{Cov}_{i,j}(|\rho\rangle\langle\rho|) = \mathbb{E}_\rho[(x_i - \mathbb{E}_\rho[x_i])(x_j - \mathbb{E}_\rho[x_j])]. \quad (8)$$

The covariance matrix for the posterior is then

$$\Sigma\rho = \mathbb{E}_\rho[|\rho - \mathbb{E}[\rho]\rangle\langle\rho - \mathbb{E}[\rho]|] = \text{Cov}(|\rho\rangle\langle\rho|) = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & \sigma_{1,1} & \sigma_{1,2} & \dots & \sigma_{1,D^2-1} \\ 0 & \sigma_{1,2} & \sigma_{2,2} & & \sigma_{2,D^2-1} \\ \vdots & & & \ddots & \vdots \\ 0 & \sigma_{1,D^2-1} & \sigma_{2,D^2-1} & \dots & \sigma_{D^2-1,D^2-1} \end{pmatrix}. \quad (9)$$

Because we have written the covariance matrix in the basis of $|\rho\rangle\langle\rho|$, we can apply $\Sigma\rho$ as a superoperator on linear operators. The action of $\Sigma\rho$ on an observable X then gives the variance of the value taken on by measurements of X in terms of the law of total variance as

$$\begin{aligned} \mathbb{V}[X] &= \mathbb{V}_\rho[\langle X \rangle_\rho] + \langle X^2 \rangle_{\mathbb{E}[\rho]} - \langle X \rangle_{\mathbb{E}[\rho]}^2 \\ &= \langle\langle X | \Sigma\rho | X \rangle\rangle + \langle\langle X | X | \hat{\rho} \rangle\rangle - \langle X \rangle_{\mathbb{E}[\rho]}^2, \end{aligned} \quad (10)$$

where \mathbb{V}_ρ is the variance over the state ρ and where $\langle \cdot \rangle_\rho$ is the expectation value over measurement outcomes, conditioned on the state ρ . Note that although we have used a coordinate form to arrive at the above expression, it is basis independent.

As discussed in detail by Blume-Kohout [22], the covariance matrix $\Delta\rho$ describes a credible region (ellipsoid) up to a scaling parameter Z corresponding to the level of the region [43]. Specifically, the eigenvectors and eigenvalues of $\Delta\rho$ are the principal axes and lengths of the axes respectively. Minimizing an appropriate norm of $\Sigma\rho$ thus provides a natural objective function for adaptively designing tomographic experiments, as we will discuss further in section 6.

Returning to the problem of finding posteriors, we note that in practice, the integral in (5) is rarely analytically tractable. Thus, Blume-Kohout suggested an approximation such as the Metropolis–Hastings algorithm could be used instead [22]. Rejection sampling methods such as Metropolis–Hastings tend to be prohibitively expensive, however, and suffer from vanishingly small acceptance probabilities as data is collected. Though there have been recent advances in rejection sampling for Bayesian inference [44], the assumption of a normal posterior is difficult to use in the context of quantum tomography. Consequently, we instead follow the approach of Huszár and Houlsby [1], and later Ferrie [25], and use the SMC algorithm [29], which does not rely on the assumption of a normal posterior. A brief review of SMC can be found in appendix B and the references therein.

SMC offers the advantage that we need not explicitly write down a prior, but instead treat Bayes' rule as a transition kernel that transforms prior hypotheses called *particles* [45] into samples from the posterior. These particles are then used to approximate the integral (5). For tomography, each particle represents a particular hypothesis about the state ρ , so that the prior $\text{Pr}(\rho)d\rho$ can be written under the SMC approximation as [1]

$$\text{Pr}(\rho) \approx \sum_{p \in \text{particles}} w_p \delta(\rho - \rho_p), \quad (11)$$

where w_p reflects the relative plausibility of the corresponding state conditioned on all available evidence. Initially, w_p is taken to be uniform, as the density of samples carries information about prior plausibility. After

updating the particles based on experimental observations, we can readily calculate the Bayesian mean estimator and posterior covariance matrix by summing over the particle approximation.

Before moving on, we wish to point out that SMC also allows for more sophisticated credible region estimators—we focus here on the covariance region estimator for simplicity. In particular, any set of particles P_α such that $\sum_{p \in P_\alpha} w_p \geq 1 - \alpha$ forms an α -credible region. Taking a convex hull over such a region then provides a region which naturally includes the convexity of state space, and a minimum-volume enclosing ellipsoid over a credible region yields a compact description [25]. Both of these credible region estimators are included in the open-source package that we rely on for numerical implementations, QInfer [39]. Thus, we inherit a variety of practical data-driven credible region estimators. For the remainder of this work, we choose to use to use naive ‘ 3σ ’ covariance ellipsoids for the purpose of illustration, by which we mean that we take the ellipsoid defined by the covariance matrix and scale it by a factor of $Z = 3$.

3. Default priors: the sampling of states and channels

In SMC, we need to be able to draw samples from a prior, see (11). In this Section we briefly review how to draw samples from several well-established priors [16]. Loosely speaking, these priors are useful as they define a notion of uniformity over states and channels, and do not posit any prior estimate other than the maximally mixed state. Following the advice of Wasserman [46], we will term these well-established priors as *default priors*, as each of them is a reasonable choice to adopt as a prior in lieu of more detailed information.

Later, in section 4, we will take priors as algorithmic *inputs* that define what states are feasible for a given tomographic experiment. We will refer to priors that can be used in this way as *fiducial*. From the perspective of the assumptions our algorithm makes, we require inputs to have the property that

$$\mathbb{E}_{\rho \sim \pi}[\rho] = \int \rho \pi(\rho) = \int \rho \Pr(\rho) d\rho = \frac{1}{D}; \quad (12)$$

(fiducial prior)

that is, that the mean of the prior is the maximally mixed state. All of the default priors described in this section are fiducial in the sense given in (12). Similarly, we say that a prior is *insightful* if its mean is anything other than a maximally mixed state. In section 4, we consider priors that are insightful by our definition, that is

$$\mathbb{E}_{\rho \sim \pi}[\rho] = \int \rho \pi(\rho) = \int \rho \Pr(\rho) d\rho = \rho_\mu, \quad (13)$$

(insightful prior)

where $\rho_\mu \neq 1/D$.

In constructing default priors, we will make repeated use of random complex-valued matrices with entries sampled from normal distributions. Such matrices form the Ginibre matrix ensemble [47]. We provide pseudocode for all default sampling algorithms in appendix C. For brevity, we refer to algorithms by the initials of their authors; for instance, we refer to algorithm 1 as the ZS algorithm after Życzkowski and Sommers [48].

3.1. Priors on states

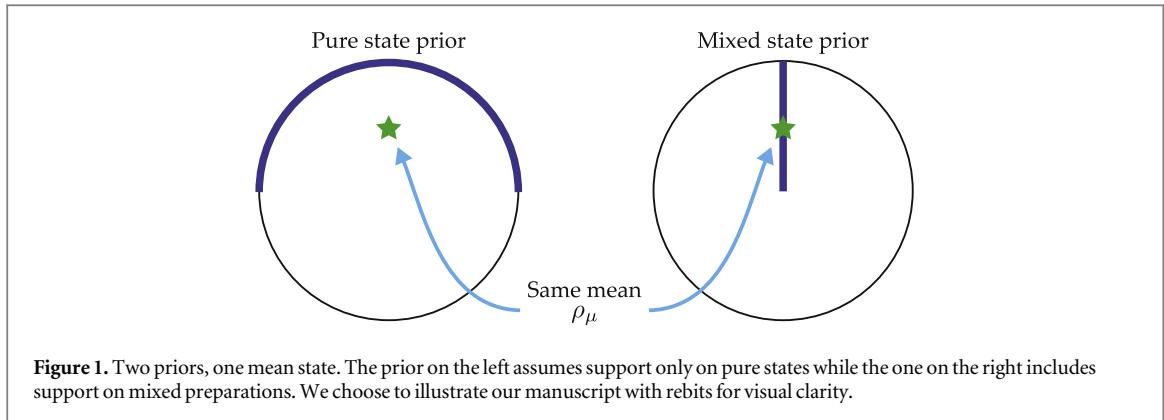
For pure quantum states, the canonical default prior is the Haar measure. One can easily sample states from this measure by sampling a vector in \mathbb{C}^D with Gaussian-distributed entries, then renormalizing. Alternatively, one can sample unitary matrices uniformly according to the Haar measure as detailed in the Mezzardi algorithm [49], see algorithm 2. A random pure state is then a Haar-random unitary applied to a fiducial state. Note that the Haar measure is *fiducial*; that is, it makes a prediction of the maximally mixed state, in the sense of (12).

Generalizing to mixed states, we consider two well-known ensembles of random states. First, we consider states drawn from the Ginibre ensemble, a generalization of the Hilbert–Schmidt ensemble that allows for restrictions on rank. Second, we consider the ensemble of states drawn from the Bures measure.

Samples from both ensembles are neatly captured by a single equation

$$\rho_{\text{sample}} = \frac{(\mathbb{1} + U)AA^\dagger(\mathbb{1} + U^\dagger)}{\text{Tr}[(\mathbb{1} + U)AA^\dagger(\mathbb{1} + U^\dagger)]}, \quad (14)$$

where U is either the identity or a Haar-random unitary, and A is a $D \times K$ Ginibre matrix. If U is taken to be the identity, then the state is drawn from the Ginibre ensemble with rank K and is unitarily invariant. This means the prior will only have support on states with rank less than or equal to K . If A is taken to be a $D \times D$ Ginibre matrix and U is Haar-random, then the state is drawn from the Bures measure. Thus, $\rho_{\text{sample}} \sim \text{Ginibre}(D, K)$ or $\rho_{\text{sample}} \sim \text{Bures}(D)$, respectively. These samples can then serve as an fiducial prior for SMC, in the sense described by equations (11) and (12), or can be transformed into samples from a prior that is insightful, as described in section 4.



These procedures are given by algorithm 3 and algorithm 4 respectively⁹.

3.2. Priors on channels

In developing applications to quantum process tomography (QPT), we use the fact that learning the Choi states of unknown channels is a special case of state tomography, as derived in appendix A. Thus, it is also useful to consider prior distributions over the Choi states of completely positive trace preserving (CPTP) quantum maps. In particular, for process tomography, we will use the measure derived by BCSZ [50] to draw samples from a prior over quantum channels that is fiducial. The resulting algorithm is unitarily invariant and supported over all channels of a given Kraus rank; that is the minimal number of Kraus operators required to specify the channel.

As detailed by Bruzda *et al* [50], to generate a channel $\Lambda : \mathbb{C}^D \rightarrow \mathbb{C}^D$, the BCSZ algorithm (see algorithm 5) begins by selecting a Ginibre-random density operator ρ of dimension D^2 and a fixed rank K . For notational simplicity, we let ρ be an operator acting on the bipartite Hilbert space $\mathbb{C}^D \otimes \mathbb{C}^D$. The trace-preserving condition is then imposed by letting $Y = \text{Tr}_2 \rho$ be the partial trace over the second copy of \mathbb{C}^D , then transforming ρ into the Choi state of the sampled channel Λ_{sample} by

$$J(\Lambda_{\text{sample}})/D = (Y^{-1/2} \otimes \mathbb{I}) \rho (Y^{-1/2} \otimes \mathbb{I}). \quad (15)$$

It is easy to verify that channels Λ_{sample} sampled in this way are indeed trace-preserving and completely positive. Moreover, the transformation above preserves the property that $\mathbb{E}[J(\Lambda_{\text{sample}})/D] = \mathbb{I}/D$, such that the mean of the BCSZ distribution is the completely depolarizing channel. This condition on channel priors is precisely that given as (12). We will show below that the BCSZ distribution suffices to construct a prior over channels that is insightful.

4. Insightful priors for states and channels

Our basic technique is to transform the samples drawn from fiducial priors, in particular the default priors described in section 3, to insightful priors by applying a channel Φ to the fiducial prior

$$\underbrace{\rho_{\text{sample}}}_{\text{insightful}} = \Phi(\underbrace{\rho_{\text{sample}}}_{\text{fiducial}}). \quad (16)$$

The algebraic gymnastics below simply determine how to construct insightful priors with a given mean ρ_μ .

Here, we seek to use the default priors from section 3 to construct a prior $\pi(\rho)$ over states with that has a desired mean $\rho_\mu = \mathbb{E}_{\rho \sim \pi}[\rho]$, and introduces little other information. The choice of the mean could be informed by, for example, experimental design or previous experimental estimates. Critically, a prior π is *not* uniquely specified by the first moment of ρ over π , $\rho_\mu = \mathbb{E}_{\rho \sim \pi}[\rho]$. Rather, the mean state ρ_μ only is a complete specification of observables measured against a state drawn from the prior. Indeed, different sets of assumptions can result in the same mean state, as illustrated in figure 1. Thus, additional constraints are required to select an appropriate prior.

We also require that π has support over all feasible states; for instance, those states of the appropriate dimension [10], possibly subject to rank restrictions. By making this demand, our tomography procedure can recover from bad priors, given sufficient data; we will show the robustness of our algorithm in later in this

⁹ More generally, Osipov *et al* [47] show that by linearly interpolating between \mathbb{I} and U in (14), one obtains a continuous family of distributions with the Hilbert–Schmidt and Bures ensembles as its extrema.

section as well as in section 6. Finally, we demand that our insightful priors can be sampled efficiently with the dimension of the state under consideration.

4.1. Construction of insightful priors

To achieve the desiderata that all feasible states are supported, that we can sample efficiently, and that the prior mean is ρ_μ , we proceed in two steps. First, we sample ρ_f from an fiducial prior $\phi(\rho)$ i.e. $\rho_f \sim \phi$. The sample from the fiducial prior is then transformed to a sample from the insightful prior under a generalized amplitude damping channel (GAD)

$$\rho_{\text{sample}} = \Phi(\rho_f | \epsilon, \rho_*) = (1 - \epsilon)\rho_f + \epsilon\rho_* \text{Tr}[\rho_f], \quad (17)$$

where $\epsilon \in [0, 1]$ is the damping parameter and ρ_* is the fixed point of the map. For $\epsilon = 0$, the map is the identity channel, while for $\epsilon = 1$, the map damps to the mixed state ρ_* . In our method ϵ is not a fixed number but is drawn from an ensemble described by the beta distribution, i.e. $\epsilon \sim \text{Beta}(\alpha, \beta)$.

Thus to determine the mean of $\pi(\rho)$ we must determine ρ_μ given $\rho \sim \phi$ and $\epsilon \sim \text{Beta}(\alpha, \beta)$, that is

$$\begin{aligned} \rho_\mu &:= \mathbb{E}_{\rho, \epsilon} [\Phi(\rho | \epsilon, \rho_*)] \\ &= \mathbb{E}_\epsilon \left[(1 - \epsilon) \frac{\mathbf{1}}{D} + \epsilon \rho_* \right] \\ &= \frac{\beta}{\alpha + \beta} \frac{\mathbf{1}}{D} + \frac{\alpha}{\alpha + \beta} \rho_*, \end{aligned} \quad (18)$$

where on the second line we have used the fact that $\mathbb{E}_{\rho \sim \pi}[\rho] = \mathbf{1}/D$ for priors that are fiducial, and on the third line we have used $\mathbb{E}_{\epsilon \sim \text{Beta}}[\epsilon] = \alpha/(\alpha + \beta)$. Inverting this relationship tells one how to choose the fixed point of the channel (17) to obtain a given mean

$$\rho_* = \frac{\alpha + \beta}{\alpha} \left(\rho_\mu - \frac{\beta}{\alpha + \beta} \frac{\mathbf{1}}{D} \right). \quad (19)$$

Clearly we need more than the first moment ρ_μ to specify the prior; we must determine α and β to complete the specification. The first constraint on α and β comes from the positivity of ρ_* , which is a valid state only if

$$\frac{\alpha + \beta}{\alpha} \left(\lambda_i - \frac{\beta}{\alpha + \beta} \frac{\mathbf{1}}{D} \right) > 0 \quad \forall i, \quad (20)$$

where λ_i are the eigenvalues of ρ_μ . Thus, the minimum eigenvalue λ_{\min} of ρ_μ partially constrains α and β by $\lambda_{\min} > \beta/[D(\alpha + \beta)]$. In order to completely determine the parameters of the beta distribution, we adopt the principle that the action of the channel (17) should be minimized. We use this principle as an efficient heuristic motivated by analogy with maximum entropy methods. In other words, *the insightful prior is as uninformative as possible given the constraint of the chosen mean, and with respect to a particular default prior*.

We therefore choose to minimize the expected value $\mathbb{E}_{\epsilon \sim \text{Beta}}[\epsilon] = \alpha/(\alpha + \beta)$, such that π is the closest GAD-transformed distribution to ϕ with the given mean ρ_μ . This minimization gives

$$\alpha = 1 \quad \text{and} \quad \beta = \frac{D\lambda_{\min}}{1 - D\lambda_{\min}}. \quad (21)$$

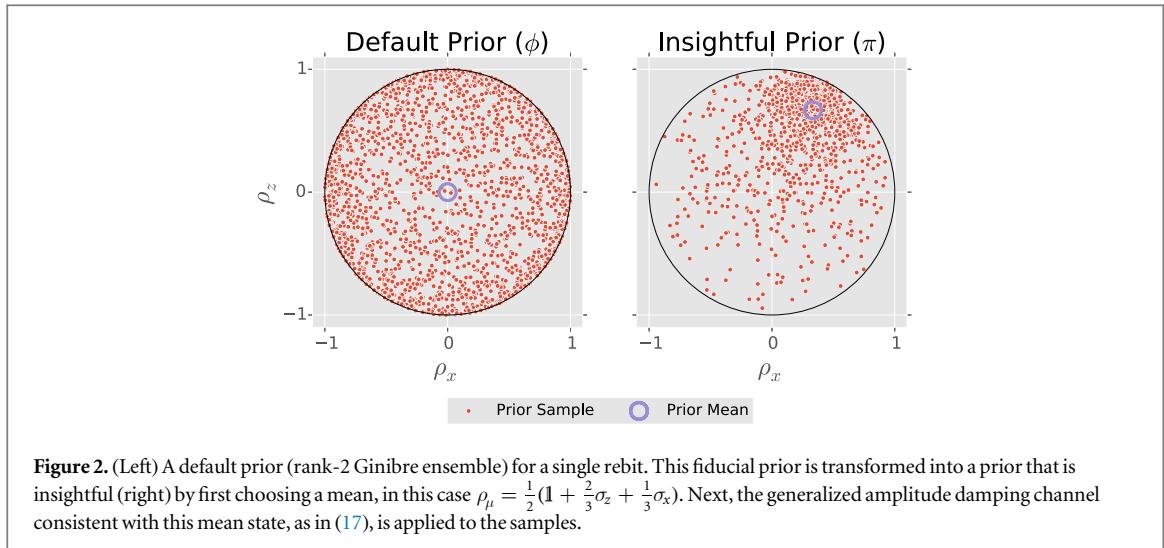
This construction naturally specializes to provide a procedure for estimating the bias of a coin, as discussed in appendix D.

4.2. Convexity and robustness of insightful priors

Note that, because $\epsilon = 0$ is in the support of all beta distributions, our prior ensures that its support is at least that of the given fiducial prior, $\text{supp } \pi \supseteq \text{supp } \phi$. For the default priors listed in section 3, the prior constructed by our procedure has support over all states of the appropriate dimension. In general, the fiducial prior *defines* the states that we consider to be valid, as can be seen from the convexity of our construction.

That is, if ρ_μ is a convex combination over states in the support of the fiducial prior, then $\text{supp } \pi$ is the convex closure of $\text{supp } \phi$. On the other hand, if ρ_μ lies outside of the support of the fiducial prior, then our algorithm chooses ρ_* to lie outside as well, such that the support of the insightful prior is bigger than that of the convex closure of the fiducial support.

As our procedure preserves the support of the fiducial prior in both cases, our procedure also defines insightful priors for rebits and channels by using the real Ginibre and BCSZ priors as fiducial priors, respectively. In particular, using the BCSZ prior as the fiducial prior for a GAD-transformed distribution, we then obtain a prior that is insightful and is supported over all completely positive and trace-preserving maps of a given dimension. Together with the Choi–Jamiołkowski isomorphism described in appendix A, we can apply SMC to process tomography with little further effort.



More exotic fiducial priors, such as distributions over stabilizer states or mundane states in the sense described by Veitch *et al* (zero-mana) [51], can also be used, provided that they can be efficiently sampled and have the maximally mixed state as their mean. For example, a prior that is insightful for mixed rebits is given in figure 2.

Before proceeding, however, we note that our notion of a fiducial prior does not imply that such priors are *uninformative*—indeed, we have seen that they serve to define what states are considered valid at all. Indeed, as Wasserman states [46], ‘*by definition, a prior represents information. So it should come as no surprise that a prior cannot represent lack of information.*’ As a further example, consider a prior over states of a given purity r ; for a qubit, such states form a shell inside of the Bloch sphere. This prior conveys information about what states are considered feasible at all, but still reports as its initial estimate that the state of interest is maximally mixed, i.e. it is fiducial, and can be used as input to our algorithm. Indeed, were one to do so, our algorithm would use this specification of what a feasible state is to define an insightful prior that is limited to a convex hull of the ball of states of purity no greater than r and the desired mean state (provided $\text{Tr}(\rho_\mu^2) \leq r$, such that ρ_μ lies within the given purity ball). Taking the case as $r \rightarrow 1/D$ (that is, a δ -function prior supported only at the maximally mixed state), the situation becomes more extreme, in that the insightful prior is then supported only on the line connecting the maximally mixed state to ρ_* . For this reason, the default priors given in section 3 are chosen to have support over all states of a given dimension and rank, making them especially useful inputs to our algorithm.

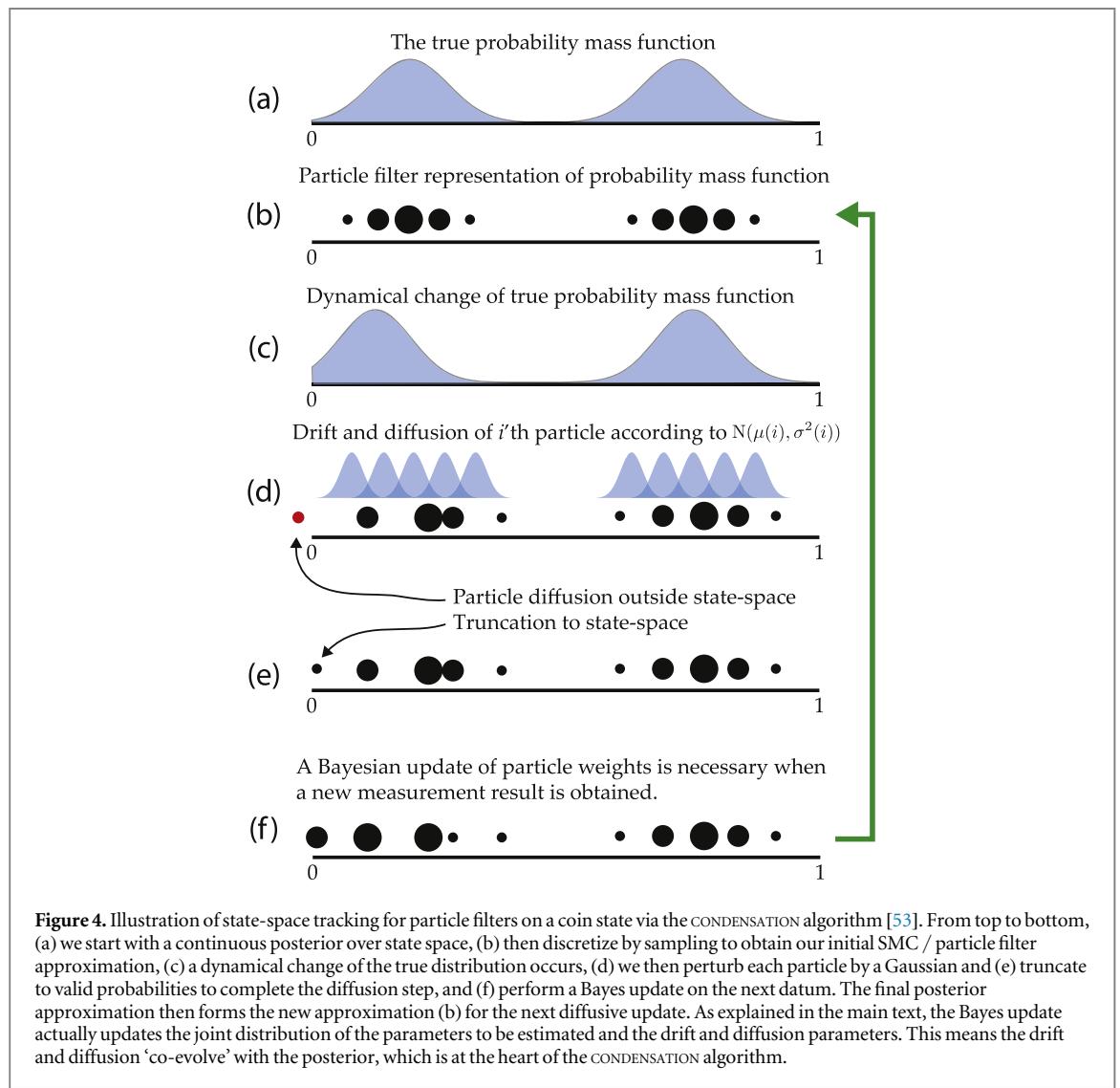
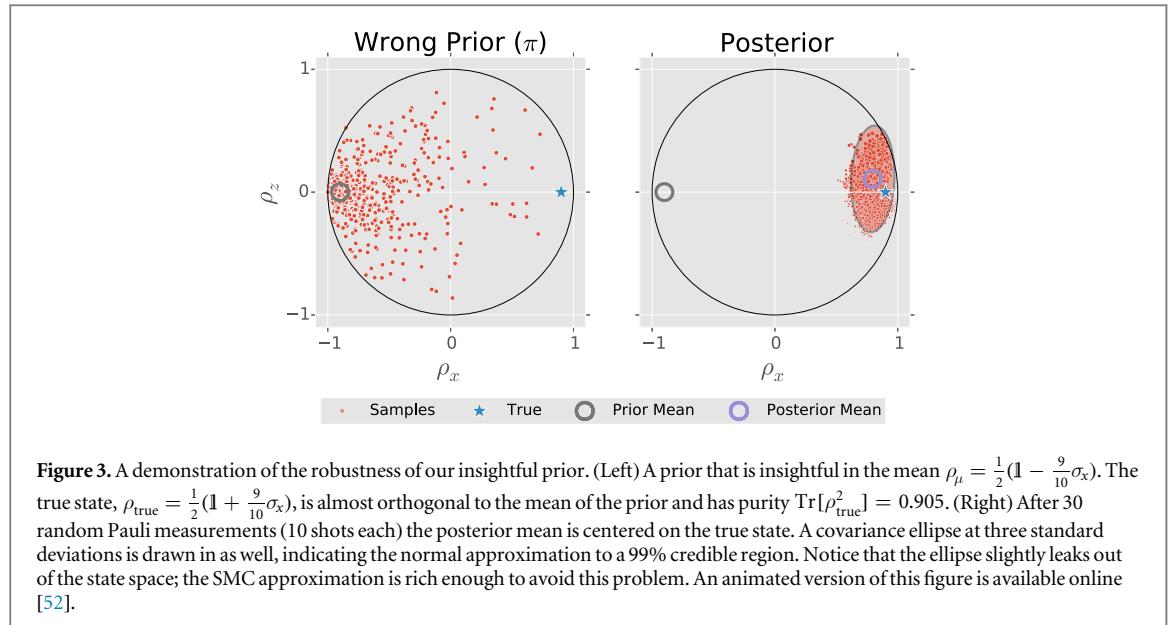
Finally, it is vital that the prior we have suggested is robust. In figure 3 we choose the mean of our insightful prior to be almost orthogonal to the true state. After 300 random Pauli measurements, the posterior has support on the true state and the mean of the posterior is approximately the true state. Thus, even if the mean of the prior is woefully wrong our procedure is robust in that it provides a reliable estimate. This robustness to a bad initial prior requires additional data to be collected, such that useful prior information can accelerate experiments but will not, in general, lead to wrong conclusions. We explore this robustness further in section 6.

5. Tomographic state tracking

In this section we use a generalization of particle filtering methods to track a stochastic processes. In the context of quantum state tomography, *state-space methods* allow us to characterize a stochastically evolving source without having to ignore all previous data. We call the resulting method *tomographic state tracking*.

In particular we use the CONDENSATION algorithm, which interlaces Bayes updates with updates to move SMC particles (using drift and diffusing of the particles), to follow a stochastic process [53]. This technique has since been applied in a variety of other classical contexts [29, 54] as well as in quantum information [37]. Such methods, collectively known as *state-space* particle filtering, are useful for following the evolution of a stochastic process observed through a noisy measurement.

We now briefly explain the CONDENSATION algorithm, readers interested in further details are directed to the original paper [53]. Consider figure 4 which illustrates the CONDENSATION algorithm for tracking a coin with a dynamical bias $\text{Pr}(\text{Heads}) = p(t)$. The current posterior of the coin bias, i.e. figure 4(a), is given an SMC representation in figure 4(b). In figure 4(c), the true probability mass function changes—at this point, we step



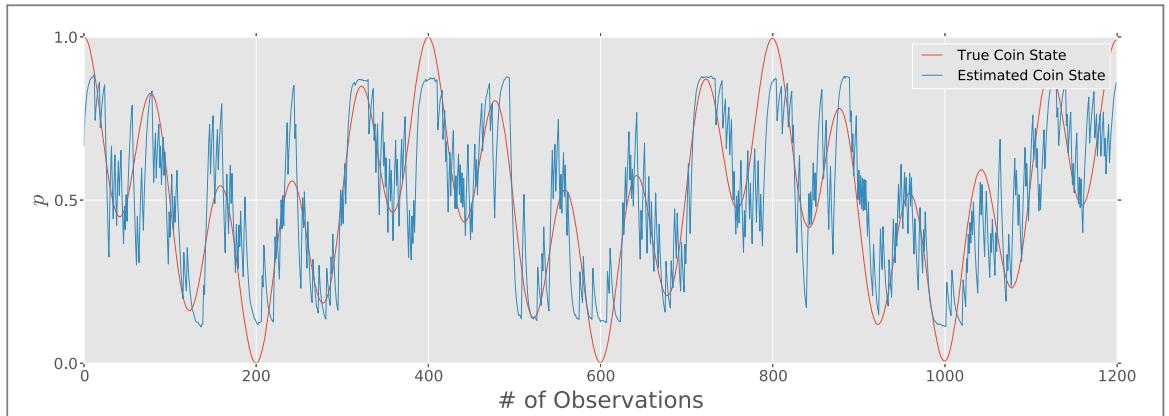


Figure 5. Diffusive tracking of a coin state (bias). This figure illustrates that the CONDENSATION algorithm can track a deterministic process even when ‘drift’ terms in the CONDENSATION algorithm are not updated. Here, we consider a coin with the true coin state evolving as a two-tone sinusoidal function $p(t_k) = \frac{1}{4}[2 + \cos(2\pi f_1 t_k) + \cos(2\pi f_2 t_k)]$, sampled at discrete times $t_k = k\Delta t$ where Δt is the time between samples. In this figure, the frequencies are $f_1 = 1/80$ and $f_2 = 1/294$.

through the CONDENSATION algorithm to track this change. Each SMC particle is perturbed by a Gaussian. In particular, the i th particle is perturbed by a Gaussian with mean $\mu(i)$ and variance $\sigma^2(i)$.

In keeping with the terminology used by Isard and Blake [53], the mean of the perturbation is termed the *drift*, and allows one to track deterministic evolution of a probability mass function; this becomes evident if $\sigma^2(i) = 0$ for all i . Similarly, the variance of the perturbation is termed *diffusion* and will allow the algorithm to track a stochastic process. As we will detail further below, both the drift and diffusion parameters can be learned online, such that we do not require them to be known *a priori*.

Sometimes the perturbation by the Gaussian will cause the particles to fall outside of the state space, in this case the unit interval, see e.g. figure 4(d). In this situation, we modify the CONDENSATION algorithm to truncate particles to be valid probabilities, completing the Gaussian perturbation step, see figure 4(e). Finally, we obtain the next datum and perform a Bayes update on the next datum. The final posterior approximation, figure 4(f), then forms the new approximation (b) for the next diffusive update.

Interestingly, the CONDENSATION algorithm starts with a joint distribution over the parameters to be estimated. For the coin case, these are the bias $p(t)$, and the distribution parameters for drift and diffusion $N(\mu(i), \sigma^2(i))$. Thus, when the Bayes update occurs the drift and diffusion parameters are updated as well, even though the likelihood does not explicitly depend on these parameters, which is referred to as co-evolution. It is this co-evolution that powers the tracking capabilities of the CONDENSATION algorithm.

As the learning of deterministic evolution of states is well-understood [43, 55, 56], we will suppose that the state under study is with respect to a frame that has already been well-characterized. Thus, the dominant remaining dynamics of the state under study are stochastic, such that we need not consider drift updates in our state-space model. Even with this assumption our model can still track deterministic evolution, however, provided that the diffusion is strong enough to include the true evolution with high probability (see figure 5 for an example of tracking deterministic evolution with diffusion alone).

Concretely, we update particle i by $\rho_i(t_k) \mapsto \rho_i(t_{k+1})$ by first adding drift and diffusion terms to find a step $\Delta\rho_i(t_k)$ to take in state-space, then truncating the negative eigenvalues of $\rho_i(t_k) + \Delta\rho_i(t_k)$. For each particle ρ_i , we let $\Delta\rho_i(t_k) = \Delta\mu + \Delta\eta$, where $\Delta\mu = \Delta\mu(t_k)$ is a deterministic drift, and where each traceless component of $\Delta\eta$ is drawn from a Gaussian $N(0, \sigma^2)$ with standard deviation σ . As stated above, we work in a frame where the deterministic part has been taken out so that $\Delta\mu = 0$ for all particles and for all time. The diffusion standard deviation σ is then taken to be a function of evolution time and the new model parameter η , $\sigma = \sqrt{t_{k+1} - t_k}\eta$. This allows the evolution rate to ‘co-evolve’ with the state model ρ , as described above.

Diffusion is completed by finding the spectral decomposition of $\rho_i(t_k) + \Delta\rho_i(t_k)$, then truncating and renormalizing. In particular, let $\rho_i(t_k) + \Delta\rho_i(t_k) = \sum_j \lambda_{i,j} |\psi_{i,j}\rangle \langle \psi_{i,j}|$. Then

$$\rho_i(t_{k+1}) = \frac{1}{N_i} \sum_j \begin{cases} \lambda_{i,j} |\psi_{i,j}\rangle \langle \psi_{i,j}| & \text{if } \lambda_{i,j} \geq 0, \\ 0 & \text{if } \lambda_{i,j} < 0, \end{cases} \quad (22)$$

where N_i is chosen such that $\text{Tr}(\rho_i(t_{k+1})) = 1$. This truncation rule avoids expensive optimization to find the closest state consistent with a given drift and diffusion update $\Delta\rho_i(t_k)$, while still generalizing methods known to be effective and efficient for coin estimation.

In order to determine the limitations of state space tracking, we considered tracking a single tone cosine $p(t_k) = \cos(2\pi fk\Delta t)$, where $t_k = k\Delta t$ is a discrete time and Δt is the time between samples. Recall that we are

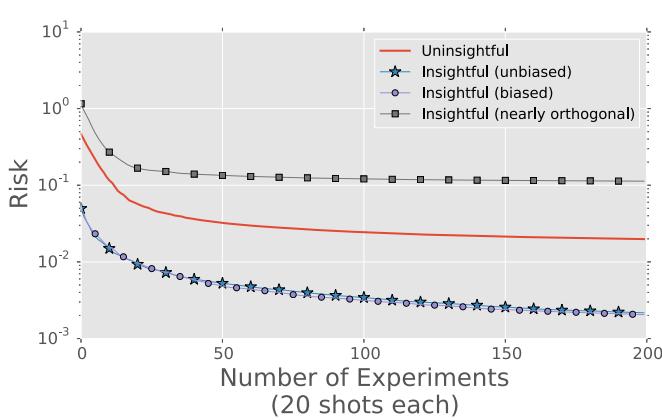


Figure 6. Risk of SMC-based tomography for a qutrit, using three different priors, with true states drawn from the generalized amplitude damping distribution for $\rho_\mu = \text{diag}(0.9, 0.05, 0.05)$. The default prior is taken to be the full-rank qutrit Ginibre distribution, while the insightful prior is taken to be identical to the actual distribution over true states and the biased prior uses the mean $\rho_\mu = \text{diag}(0.87, 0.065, 0.065)$. The nearly orthogonal prior is taken to be insightful with the mean $\rho_\mu = \text{diag}(0.065, 0.065, 0.87)$. Risk is measured with respect to the quadratic loss function on vectorized density operators, $L(\hat{\rho}, \rho) = |||\hat{\rho}\rangle\rangle - |\rho\rangle\rangle||_2$. The risk is averaged over 1200 trials. Measurements are chosen to be rank-1 projectors onto randomly drawn single-qutrit stabilizer states. Each such stabilizer state is then measured 20 times.

choosing not to include deterministic evolution (i.e. ‘drift’) in our model, thus the following observation only applies to purely diffusive tracking. We found our algorithm could track a frequency up to $f_{\text{track}} = (1/10) \times (1/\Delta t)$. At higher frequencies, our approach failed to track the oscillatory behavior of $p(t)$, in that it would report $p(t_k) = 1/2$ for all time. This modality can be tested using model selection [12, 37], such that more appropriate algorithm can be used in that case. A more sophisticated, though less quantitative, analysis of this failure can be found in appendix E.

6. Numerical examples

We now show examples of Bayesian state and process tomography using SMC, with priors that are respectively default and insightful. These examples were generated using QInfer [39].

For state tomography, we demonstrate our methods by learning qutrit states, as shown in figure 6. We demonstrate the performance of our algorithm in this case by reporting the *risk*, defined as the expected quadratic loss over repetitions of the algorithm

$$r(\hat{\rho}, \pi) = \mathbb{E}_{\rho \sim \pi} [|||\hat{\rho}\rangle\rangle - |\rho\rangle\rangle||_2]. \quad (23)$$

We use each of a default, insightful and unbiased, and an insightful but biased prior. In all three cases, we draw the ‘true’ states ρ from the prior that matches the insightful and unbiased prior.

In figure 6, we also verify that our method is robust for the qutrit example by considering an insightful prior whose mean is nearly orthogonal to the true state. Notably, in well over half of the 1200 trials considered in that case, QInfer reported that the algorithm was likely to fail, heralding the impact of the ‘bad’ prior.

We then proceed to consider state-space quantum state tomography, as detailed in section 5. We demonstrate the performance of our state-space method in an animation, available online [57], also see figure 7 for a snapshot.

Finally, we demonstrate the application of our method to learning quantum channels acting on a qubit. In figure 8, we show an example of a single simulated QPT experiment, where the true channel and the insightful prior (generated with a BCSZ fiducial prior) agree. The preparation and measurement settings are chosen to be elements of the Pauli basis. Specifically, 20% of the experiments use random Pauli preparation and measurements, while 80% of the experiment use settings that maximize $\langle\langle \rho_i^\text{T}, E_i | \Sigma \rho | \rho_i^\text{T}, E_i \rangle\rangle$ out of 50 randomly proposed Pauli preparations and measurements; that is, adaptively chosen to overlap with the principal components of the current posterior. The resulting posterior distribution characterizes the uncertainty remaining about the ‘true’ channel, as shown in figure 8. In particular, we note the principal components of the posterior covariance matrix are themselves quantum maps which describe the directions of maximal uncertainty in the final posterior. For this example, our error is dominated by uncertainty about the contribution of the identity and Hadamard channels, as is made clear by visual inspection in figure 9 (bottom right).

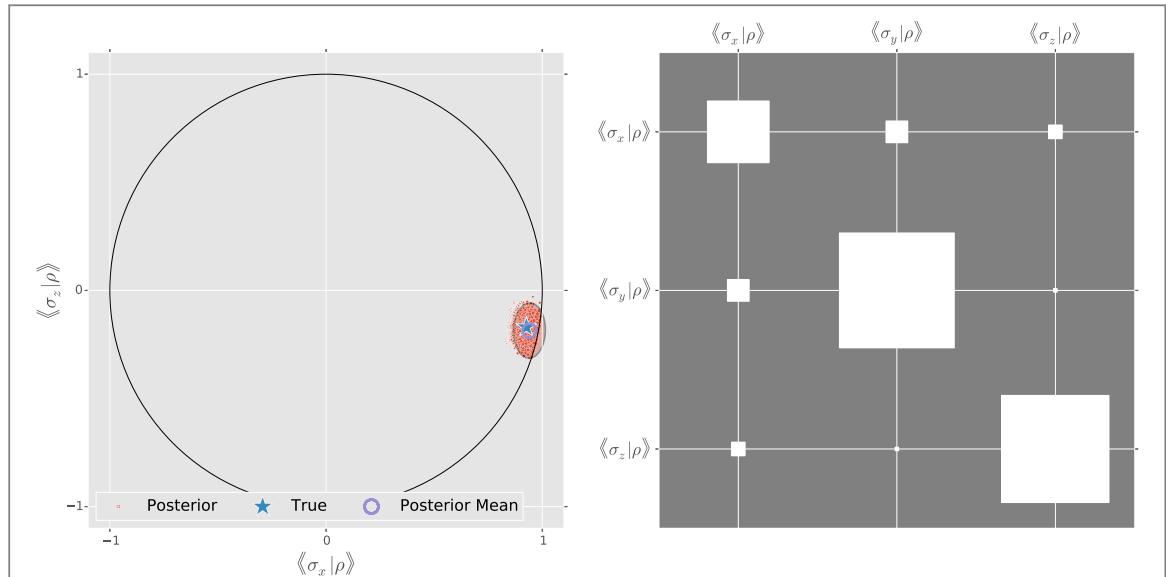


Figure 7. Snapshot of ‘Video 1’ in the supplementary material which demonstrates the use of tomographic state tracking to estimate a diffusing quantum state. Shown at left is the projection of the true state (star), the sequential Monte Carlo particle cloud (red dots), and covariance approximation to the 99% credible region (red circle) to the x-z plane. Shown at right is a Hinton diagram of the current posterior covariance matrix. At each time step, each element of the vectorized true state is perturbed by a Gaussian with mean 0 and standard deviation 0.0045, and then is truncated to lie within the space of valid quantum states. The prior over this diffusion rate is taken to be a log-normal distribution with mean 0.006. Each measurement consists of 25 shots along a random Pauli axis.

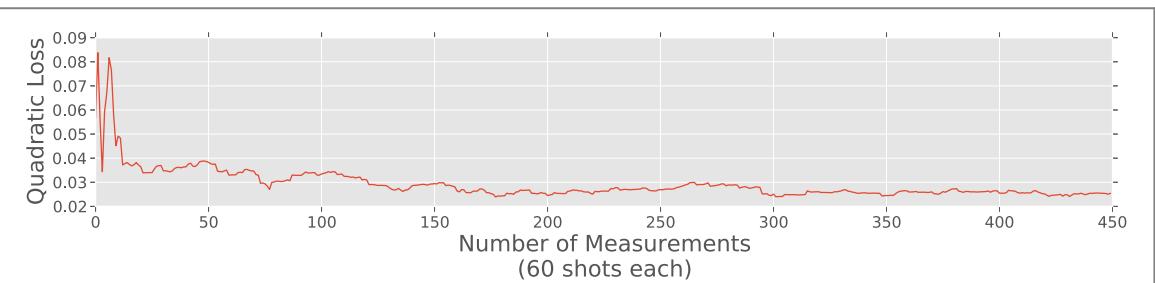


Figure 8. Example quadratic loss between estimated and true channels for SMC-based tomography for a qubit channel $\Lambda[\rho] = 0.7\rho + 0.3H\rho H$, where H is a Hadamard. The prior for both the SMC estimator and the true channel is taken to be the insightful prior with a mean taken to be a 90% mixture of the true channel and 10% of a completely depolarizing channel. Out of the 450 experiments, 20% are chosen to be random Pauli preparation and measurements, while 80% are chosen adaptively. In particular, adaptation proceeds by proposing 50 different random pairs of Pauli preparations and measurements, then selecting the one which has the maximum expectation value $\langle\Sigma\rangle = \langle P, M | \Sigma | P, M \rangle$, where Σ is the current covariance superoperator.

7. Conclusions

In this work, we have provided a new prior distribution over quantum states and channels that allows for including experimental insight, a software implementation for numerically approximating Bayesian tomography, and a method for tracking time-dependent states. Together, our advances make Bayesian quantum tomography practical for current and future experiments. In particular, our methods allow for exploiting well-known benefits of Bayesian methods, including credible region estimation, hyperparameterization and model selection.

We note, however, that our insightful prior on states and channels is completely specified by its first moment. An interesting and open problem would thus be to develop a prior on states and channels that is completely specified by its first and second moments.

Finally, with respect to the tomographic state tracking methods presented in section 5, it is worth noting that van Enk and Blume-Kohout [12] suggested that model selection could be used to determine if a source was drifting or diffusing. In this manuscript we have provided a way that allows one to track a source that is drifting or diffusing. It is also possible to combine the approaches and use model selection to determine when tracking is necessary or when a static model is sufficient as demonstrated by Granade [37].

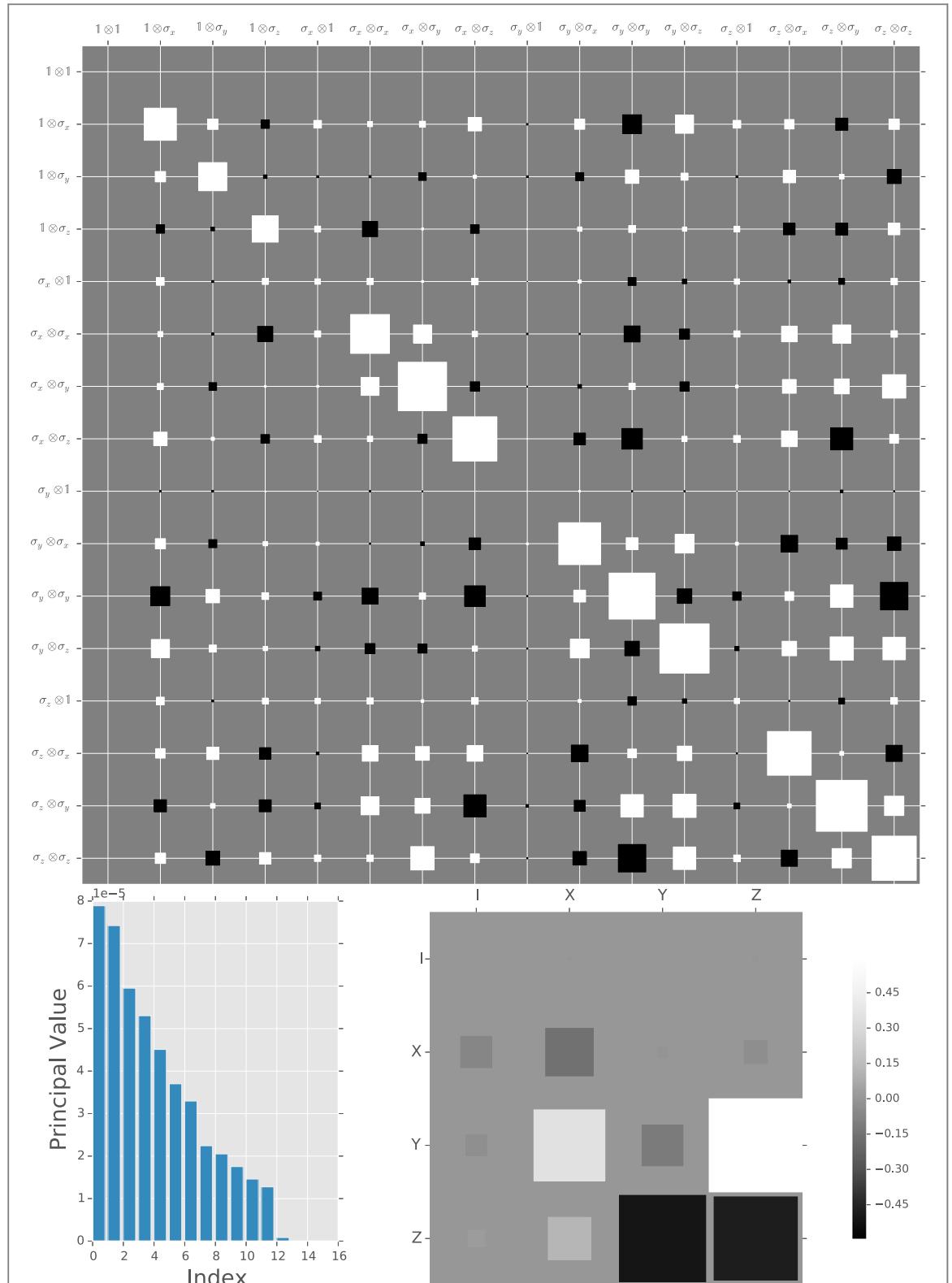


Figure 9. (Top) Covariance matrix for process tomography experiment shown in figure 8. (Bottom left) Spectrum of covariance matrix (principal values). (Bottom right) Largest eigenchannel of the covariance matrix (principal channel), indicating the direction in which the channel is least certain.

In short, with constructive methods for sampling from insightful priors, and with modern statistical methods, Bayesian state and process tomography are made practical for current experimental needs. This in turn allows us to explore new questions in tomography, and thus better characterize and diagnose quantum information processing systems.

Acknowledgments

The authors acknowledge Sarah Kaiser, Chris Ferrie, Brendon Higgins, and Richard Küng for discussions. The authors acknowledge funding from Industry Canada, CERC, NSERC, the Province of Ontario, and the CFI is gratefully acknowledged. CG acknowledges Connor Jarvis for noticing an important typographical error. CG also acknowledges Ian Hincks, Nathan Wiebe, and Yuval Sanders for contributions to and testing of QInfer, as well as support from the US Army Research Office via grant numbers W911NF-14-1-0098 and W911NF-14-1-010. The authors would like to note independent related work by Faist and Renner [58] and by Struchalin *et al* [27]. <http://cgranade.com/> Literate source code for the figures, animations and tutorials appearing in this work is available at <http://goo.gl/fRqnIn>.

Appendix A. Choi–Jamiołkowski isomorphism and QPT

In order to interpret (1a) as specifying probabilities for QPT [59] as well as state tomography, we use the Choi–Jamiołkowski isomorphism and represent a hypothesis about the channel Λ as a state

$$J(\Lambda) := (\mathbb{1} \otimes \Lambda)(\lvert \mathbb{1} \rangle \langle \mathbb{1} \rvert). \quad (\text{A1})$$

The formalism of using the Choi–Jamiołkowski isomorphism [60, 61] to describe open quantum processes has also been recently detailed in detail by Wood *et al* [62] using tensor network diagrams with particular regard to quantum tomography and its generalizations [63, 64].

In the case that QPT is carried out by ancilla-assisted process tomography [65], we are done, as $J(\Lambda)$ is, up to normalization, the state used in AAPT. On the other hand, we can also use that for any linear operator X , $\Lambda(X) = \text{Tr}_1[J(\Lambda)(\mathbb{1} \otimes X^T)]$ to represent preparation followed by evolution under Λ and measurement as a measurement of $J(\Lambda)/D$ [66]. In particular, suppose that a QPT experiment is performed in which each observation consists of a measurement effect E and a preparation ρ ,

$$\Pr(E'|J(\Lambda)) = \text{Tr}[E\Lambda(\rho)], \quad (\text{A2a})$$

$$= \text{Tr}[(\mathbb{1} \otimes E)J(\Lambda)(\rho^T \otimes \mathbb{1})], \quad (\text{A2b})$$

$$= \langle\langle \rho^T, E|J(\Lambda)\rangle\rangle, \quad (\text{A2c})$$

$$= \langle\langle P, E|J(\Lambda)/D\rangle\rangle, \quad (\text{A2d})$$

where $P := \rho^T \cdot D$ and $E' = P \otimes E$. This now has the form of a measurement $P \otimes E$ being made on a state $J(\Lambda)/D$, such that (1a) completely describes the outcomes of both in-place and ancilla-assisted QPT experiments. Though we used the column-stacking basis to derive this equivalence, implicitly defining a convention for the Choi state, we can insert a unitary operator on the space of vectorized operators to observe that (A2d) is not dependent on the choice of basis.

The likelihood after N measurements is

$$\mathcal{L}(J(\Lambda)) = \Pr(\mathcal{M}|J(\Lambda)) = \prod_{i=1}^N \Pr(E'_i|J(\Lambda)), \quad (\text{A3})$$

where \mathcal{M} is the string of measurement results and E'_i is the observation of the i 'th generalized measurement. Note that (A3) is of the same form as (1a), such that quantum process tomography can be treated as a special case of quantum state tomography, provided that we use an appropriate fiducial prior.

Appendix B. Review of SMC

The purpose of this appendix is to provide a minimal summary and point the interested reader to the relevant literature on the particle filtering and SMC algorithms as applied to Bayesian inference. There is a wealth of information in the references that need not be reproduced here.

The basic idea of SMC is to approximate a distribution over some parameters \mathbf{x} conditioned on the j th data d_j , i.e. $\Pr(\mathbf{x}|d_j)$. The approximation uses a weighted sum of delta-functions, specifically

$$\Pr(\mathbf{x}|d_j) \approx \sum_{k=1}^N w_k(d_j) \delta(\mathbf{x} - \mathbf{x}_k), \quad (\text{B1})$$

where the N objects \mathbf{x}_k are called *particles* and $w_k(d_j)$ is the *weight* of the k th particle. Here (B1) should be compared to (11) of the main text; where we suppressed many technical and notational details. As the number of particles N increases the quality of the approximation improves. The initial particles are chosen by sampling the prior distribution over \mathbf{x} , and are taken to have uniform weight.

The next step is to update the posterior distribution given new data, say the $j + 1$ th datum is d_{j+1} . Then using the likelihood function the particle weights are updated from the previous weights ($w_k(d_j)$) using

$$w_k(d_{j+1}) \propto \Pr(d_{j+1}|\mathbf{x}_k) w_k(d_j). \quad (\text{B2})$$

Clearly, the particle weights after this update need to renormalized such that $\sum_k w_k(d_{j+1}) = 1$.

As updates proceed, the concentration of weights on the most plausible particles causes the distribution to be samples by a much smaller number of *effective* particles, as measured by the effective sample size $n_{\text{ess}} = 1/\sum_k w_k^2(d_{j+1})$. Thus, periodically particle locations have to be perturbed and the weights reset to uniform, restoring numerical stability. This is called resampling. A video demonstrating the Bayesian update and resampling steps for a simple Hamiltonian learning problem is available online [67].

In the remainder of this appendix we provide references which can be used to obtain more details on SMC. Doucet *et al* have provided a more through review from the perspective of classical statistics [28, 45]. In the quantum domain summaries of the SMC algorithm can be found in references [4, 25, 26, 43, 68, 69]. Finally, Svensson has provided a useful video tutorial illustrating the application of particle filters in a simplified radar application [70].

For our application (quantum tomography), Huszár and Houlsby [1] and by Ferrie [2] suggested SMC as numerical tool for implementing (5). Of particular interest is Ferrie's recent work on tomographic region estimators with SMC [25]. Reference [37] has provided a summary of recent applications in quantum information.

Appendix C. Sampling default priors

In this appendix we review the algorithms necessary to sample from default priors, provided by References [47–50].

Algorithm 1. ZS algorithm [48] for sampling from the Ginibre matrix ensemble.

Input: Dimension D and K .
Output: An $D \times K$ matrix, drawn from the Ginibre ensemble.
function GINIBREMATRIX(D, K)
 Generate an $D \times D$ matrix G with elements drawn independently from $(N(0, 1) + i \cdot N(0, 1))$.
 return G
end function

Algorithm 2. Mezzardi algorithm [49] for sampling from the Haar measure.

Input: Dimension (D).
Output: A unitary operator $U \in \mathbb{C}^{D \times D}$, drawn from the Haar measure.
function HAARUNITARY D
 $Z \sim \text{GINIBREMATRIX}(D, D)$
 $Q, R \leftarrow$ QR-decomposition of Z
 $\Lambda \leftarrow \text{diag}(R)$
 $\lambda_{ii} \leftarrow \lambda_{ii}/|\lambda_{ii}|$
 return $U \leftarrow QA$
end function

Algorithm 3. ZS algorithm [48] for sampling states from the Ginibre ensemble.

Input: Dimension D and rank $K \leq D$.
Output: A $D \times D$ positive semidefinite matrix of rank K drawn from the Ginibre ensemble.
function GINIBRESTATE(D, K)
 $A \sim \text{GINIBREMATRIX}(D, K)$
 $\rho \leftarrow AA^\dagger/\text{Tr}(AA^\dagger)$
 return ρ
end function

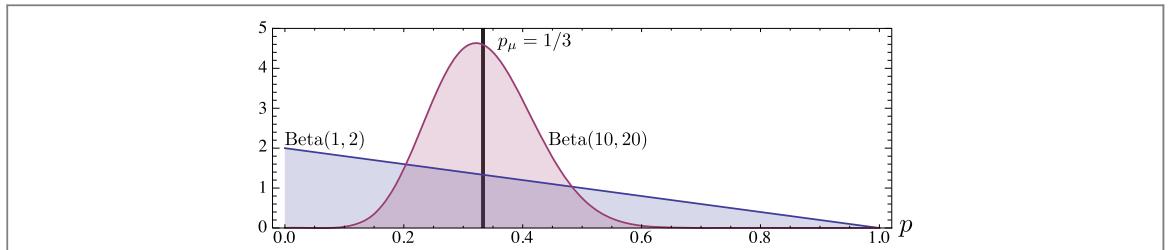


Figure D1. Two different priors on coins, each with the same prior mean: $p_\mu = 1/3$. This figure should be compared to figure 1.

Algorithm 4. OSZ algorithm [47] for sampling from Bures measure.

Input: Dimension D .
Output: An $D \times D$ positive semidefinite matrix of rank K drawn from the Bures measure.

```

function BURESSTATE( $D$ )
     $A \sim \text{GINIBREMATRIX}(D, D)$ 
     $U \sim \text{HAARUNITARY}(D)$ 
     $\rho \leftarrow (\mathbb{1} + U)AA^\dagger(\mathbb{1} + U^\dagger)/\text{Tr}[(\mathbb{1} + U)AA^\dagger(\mathbb{1} + U^\dagger)]$ 
    return  $\rho$ 
end function
```

Algorithm 5. BCSZ algorithm [50] for sampling from a uniform ensemble of CPTP maps.

Input: Dimension D , Kraus rank K
Output: A $D^2 \times D^2$ Choi matrix $J(\Lambda)$ of a channel drawn from the BCSZ distribution.

```

function BCSZCHANNEL( $D$ )
     $X \sim \text{GINIBREMATRIX}(D^2, K)$ 
     $\rho \leftarrow XX^\dagger$ 
     $Y \leftarrow \text{Tr}_2\rho$  ▷  $\text{Tr}_2$  indicates the partial trace over the second copy of  $\mathbb{C}^D$ .
     $Z \leftarrow (\mathbb{1} \otimes Y^{-1/2})\rho(\mathbb{1} \otimes Y^{-1/2})$ 
    return  $ZD$ 
end function
```

Appendix D. GADFLI prior for coins

A prior distribution over the bias of a coin is a special case of a qubit and rebit prior. It corresponds to a prior over one axis of the Bloch sphere. Without loss of generality, we will take that axis to be the \hat{z} -axis, such that the density operator for a coin is given by

$$\rho = \begin{pmatrix} p & 0 \\ 0 & 1-p \end{pmatrix} = \frac{\mathbb{1}}{2} + z \frac{Z}{2} = \frac{1}{2} \begin{pmatrix} 1+z & 0 \\ 0 & 1-z \end{pmatrix}, \quad (\text{D1})$$

where $z = 2p - 1$ is the \hat{z} -axis coordinate on the Bloch sphere at which the coin state is positioned.

Given the result of a coin toss $r \in \{0, 1\}$, we assign a likelihood for that outcome conditioned on the coin p as

$$\Pr(r|p) = p^{1-r}(1-p)^r, \quad (\text{D2})$$

where the probability of heads is $\Pr(0|p) = p$. Using the density operator definition of a coin, we define that the minimum eigenvalue of a coin is $\lambda_{\min} = \min(p, 1-p)$.

We draw samples of the coin state p from our GAD prior by first choosing $\epsilon \sim \text{Beta}(\alpha, \beta)$ and p to be sampled from a fiducial prior such that $\mathbb{E}[p] = 1/2$. Through the rest of this example, we use $p \sim \text{Uniform}(0, 1)$. The GAD-prior sample p' is obtained by transitioning p with a linear function Φ given by

$$p' = \Phi[p|\epsilon, p_*] = (1 - \epsilon)p + \epsilon p_*. \quad (\text{D3})$$

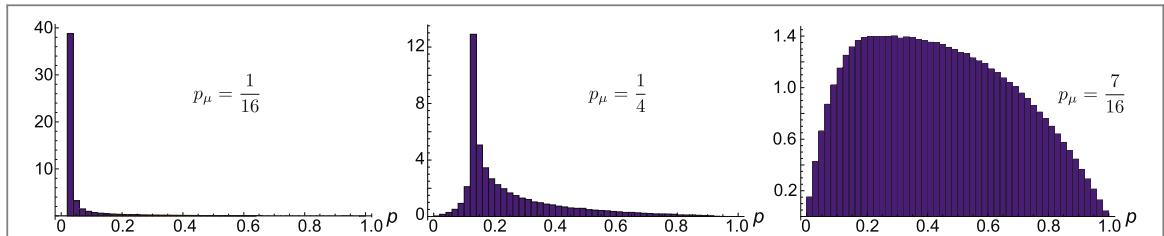


Figure D2. A histogram of the generalized amplitude damping (GAD) prior for the coin, with $p_\mu \in \{1/16, 4/16, 15/16\}$. The histogram was generated with 4000 000 samples from the GAD prior. The Gad prior is symmetric about $p_\mu = 1/2$ even though $p_\mu = 1/2$ is not an allowed value.

The expectation over this GAD prior then gives the Bayesian mean estimator before any data is collected. Using that p is fiducial (has mean 1/2) and that $\mathbb{E}[\text{Beta}(\alpha, \beta)] = \alpha/(\alpha + \beta)$, we find that

$$p_\mu := \mathbb{E}_{p,\epsilon} [\Phi(p|\epsilon, p_*)] = \frac{\beta}{\alpha + \beta} \cdot \frac{1}{2} + \frac{\alpha}{\alpha + \beta} p_*, \quad (\text{D4})$$

and $p_\mu \neq 1/2$. Inverting, we find that the fixed point p_* needed to guarantee that the prior mean is p_μ is given by

$$p_* = \frac{\alpha + \beta}{\alpha} \left(p_\mu - \frac{\beta}{\alpha + \beta} \cdot \frac{1}{2} \right), \quad (\text{D5})$$

which is analogous to equation (19). In order for p_* to be a valid coin, this constrains

$$p_\mu > \frac{\beta}{\alpha + \beta} \cdot \frac{1}{2}. \quad (\text{D6})$$

The mean of a Beta distribution does not fully specify the distribution, see figure D1. Because of this we find α and β consistent with this condition and such that the mean

$$\mathbb{E}_\epsilon [\epsilon] = \frac{\alpha}{\alpha + \beta} \quad (\text{D7})$$

is minimized. This represents that the GAD channel used to define samples does as little as possible to transform fiducial (uniform) samples to our prior. Minimizing subject to the constraints that $\alpha > 0$, $\beta > 0$ and $p_* > 0$, we obtain that

$$\alpha = 1 \quad \text{and} \quad \beta = \begin{cases} \frac{1}{2p_\mu - 1} - 1 & \text{if } p_\mu > 1/2, \\ \frac{2p_\mu}{1 - 2p_\mu} & \text{if } p_\mu < 1/2. \end{cases} \quad (\text{D8})$$

We can write this in terms of λ_{\min} to obtain the final GAD-prior condition

$$\alpha = 1 \quad \text{and} \quad \beta = \frac{2\lambda_{\min}}{1 - 2\lambda_{\min}}, \quad (\text{D9})$$

which should be compared to (21). Using this condition, we obtain the coin priors shown in figure D2 .

Appendix E. An estimate of the tracking frequency

We now give an order of magnitude estimate for maximum frequency that is ‘smoothly’ trackable by our algorithm. It is possible our algorithm can track higher frequencies but the evolution is likely to be fairly discontinuous.

Our estimate for the tracking frequency comes from sampling arguments in tracking a bias of a coin, with probability $\Pr(0|p) = p$ for heads. We will assume the coin’s bias changes from $\Pr(0|p) = p$ to $\Pr(0|p) = p \pm \epsilon$ where $\epsilon \ll 1$. The question is how quickly can we detect such a change, i.e. how many equally spaced in time samples does it take to notice the difference of $\pm \epsilon$.

We must specify an error tolerance— ϵ —for our sensing protocol and a confidence interval— Z —for our estimate of p . It turns out if we use the trace distance between two coins $\rho = \text{diag}(p, 1 - p)$ and $\sigma = \text{diag}(p \pm \epsilon, 1 - p \pm \epsilon)$ then $D(\rho, \sigma) = (1/2)\text{Tr}|\rho - \sigma| = \epsilon$. These two specifications will in turn (approximately) determine the number of samples required and therefore the bandwidth of our tracking protocol. We choose $Z = 1.9599$ which corresponds to a 95% level of confidence for our estimator and the maximum acceptable error $|p_{\text{true}} - p_{\text{est}}| < \epsilon$. Using the standard deviation of the Bernoulli distribution $\sqrt{p(1 - p)/N}$, we have $\epsilon = Z\sqrt{p(1 - p)/N}$. The largest standard deviation (worst case) is when $p = 1/2$ this

gives $\epsilon = Z/(2\sqrt{N})$. Rearranging for N gives $N = Z^2/(4\epsilon^2)$. To obtain this many samples we must measure for a time $T_{\text{meas}} = \Delta t N$, where Δt is the time between measurements. Thus T_{meas} is effectively the time between our samples of $p(t)$. Naïve arguments from the Nyquist–Shannon sampling theorem imply that we can not determine frequency components of $p(t)$ greater than $f_{\text{max}} = 1/(2T_{\text{meas}})$, which is called the detection frequency bandwidth. The implication is we can track, for example, $p(t) = 1/2 + \epsilon \sin(f_{\text{max}}t)$.

Appendix F. QInfer tomography tutorial

In this appendix, we demonstrate the use of QInfer for Bayesian state and process tomography. In particular, we show how to estimate states and channels given data synthesized from a description of a true state, and discuss how to obtain region estimates, covariance superoperators and other useful functions of tomography posteriors. We then discuss how to apply these techniques in experimental systems.

The tomography implementation in QInfer is based on QuTiP and NumPy, so we start by importing everything here.

```
In [1]: from IPython.display import display
        import numpy as np
        import qutip as qt
        import qinfer as qi
```

As a first step, we define a *basis* for performing tomography; the choice of basis is largely arbitrary, but depending on the experiment, some bases may be more or less convenient. Here, we focus on the example of the single-qubit Pauli basis $B = \{\mathbb{1}, \sigma_x, \sigma_y, \sigma_z\}$.

```
In [2]: basis = qi.tomography.pauli_basis(1)
        display(basis)
```

<TomographyBasis dims=[2] at 362504488>

We will get a lot of use out of the Pauli basis, so we also define some useful shorthand.

```
In [3]: I, X, Y, Z = qt.qeye(2), qt.sigmax(), qt.sigmay(), qt.sigmaz()
```

Basis objects are responsible for converting between QuTiP's rich `Qobj` format and the unstructured model parameter representation used by QInfer.

```
In [4]: display(basis.state_to_modelparams(I / 2 + X / 2))

array([ 0.70710678,  0.70710678,  0.          ,  0.          ])
```

```
In [5]: display(basis.modelparams_to_state(np.array([1, 0, 0, 1]) / np.sqrt(2)))
```

Quantum object: dims = [[2], [2]], shape = [2, 2], type = oper, isherm = True

$$\begin{pmatrix} 1.000 & 0.0 \\ 0.0 & 0.0 \end{pmatrix}$$

Having defined a basis, we then define the core object describing a tomography experiment, the *model*. In QInfer, models encapsulate the likelihood function, experimental parameters and other useful metadata about the experimental properties being estimated. In our case, we use `TomographyModel` to describe the single-shot experiment, and `BinomialModel` to describe batches of the single-shot experiment.

```
In [6]: model = qi.BinomialModel(qi.tomography.TomographyModel(basis))
        display(model)
```

<qinfer.derived_models.BinomialModel at 0x159b6048>

A `Model` defines a vector of model parameters; for a single qubit `TomographyModel`, this is a vector of length 4, each describing a different element of the Hermitian operator basis. Each `Model` also defines experiment parameters as a NumPy record array. A record then describes a single measurement of the model.

```
In [7]: display(model.expparams_dtype)
```

[('meas', float, 4), ('n_meas', 'uint')]

In this case, the experiment parameters record has two *fields*: `meas` and `n_meas`. The first is a vector of four floats corresponding to $|M\rangle\rangle = (\langle\langle B_0|M\rangle\rangle, \langle\langle B_1|M\rangle\rangle, \langle\langle B_2|M\rangle\rangle, \langle\langle B_3|M\rangle\rangle)$. The second is an unsigned integer (`uint`) describing how many times that measurement is performed. For instance, measuring $(\mathbb{1} + \sigma_z)/2$ 40 times is given by the array:

```
In [8]: expparams = np.array([
    # Each tuple, marked with (), defines a single record.
    (
        # Within each tuple, fields are separated by commas.
        # The fields follow in the order given by the model,
        # so the first field is meas, a length-4 vector.
        [1 / np.sqrt(2), 0, 0, 1 / np.sqrt(2)],
        # The second field is then the number of measurements.
        40
    )
],
# We finish building the array by passing along the right data
# type to NumPy. This is somewhat of a QInfer idiom.
dtype=model.expparams_dtype)
display(expparams)

array([[0.7071067811865475, 0.0, 0.0, 0.7071067811865475], 40L]),
dtype=[('meas', '<f8', (4,)), ('n_meas', '<u4')])
```

The fields of a record array can be obtained by indexing. For instance, the `['meas']` field is then a 1×4 array, with the first index allowing for a sequence of measurements to be described at once.

```
In [9]: display(expparams['meas'])
```

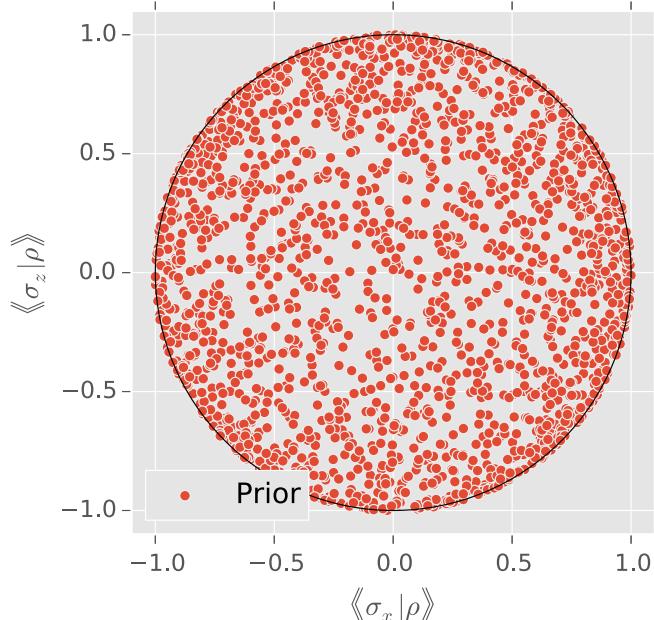
```
array([[ 0.70710678, 0.0, 0.0, 0.70710678]])
```

Note that by convention, `meas` is normalized to $1/\sqrt{d}$.

Often, we will not construct experiments directly, but will instead rely on QInfer's heuristics (described below). In any case, once we have a model, the next step is to create a prior. QInfer comes with several useful fiducial priors, as well as insightful priors constructed from amplitude damping channels. For instance, to create a Hilbert–Schmidt uniform prior constrained to rebits, we use the `GinibreReditDistribution`:

```
In [10]: fiducial_prior = qi.tomography.GinibreReditDistribution(basis)
```

```
In [11]: qi.tomography.plotting_tools.plot_rebit_prior(fiducial_prior, rebit_axes=[1, 3])
```



Here, we have told QInfer that we wish to treat σ_x and σ_z as our rebit axes using the `rebit_axes=[1, 3]` argument.

Insightful priors can be constructed by specifying a fiducial prior and a QuTiP `Qobj` representing the desired mean.

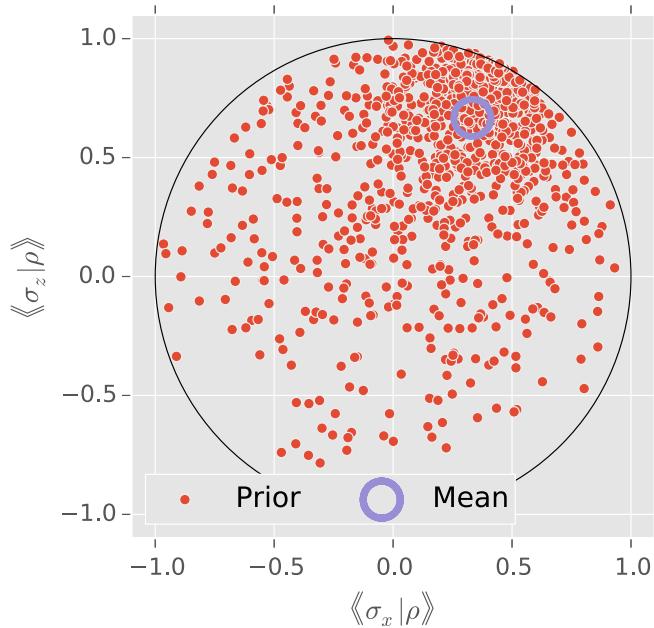
```
In [12]: prior_mean = (I + (2/3) * Z + (1/3) * X) / 2
display(prior_mean)
```

Quantum object: dims = [[2], [2]], shape = [2, 2], type = oper, isherm = True

$$\begin{pmatrix} 0.833 & 0.167 \\ 0.167 & 0.167 \end{pmatrix}$$

In [13]: `prior = qi.tomography.GADFLIDistribution(fiducial_prior, prior_mean)`

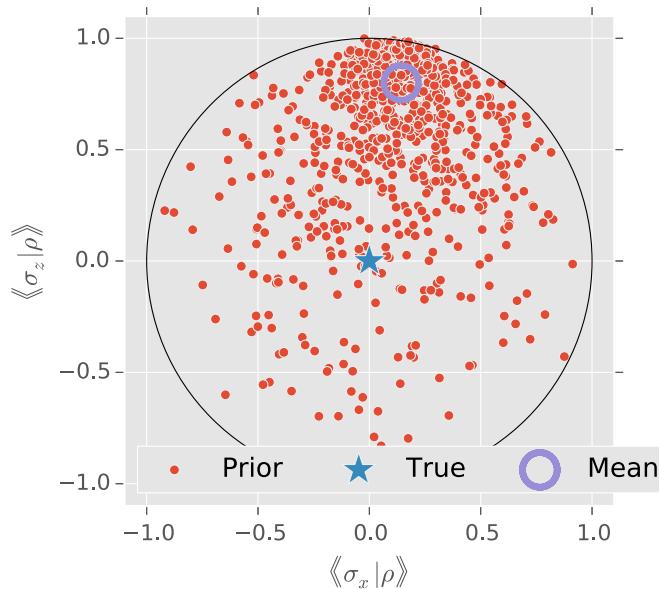
In [14]: `qi.tomography.plotting_tools.plot_rebit_prior(prior, rebit_axes=[1, 3])`



Having constructed a prior and a model, we can now continue to perform Bayesian inference using SMC. We demonstrate using the true state $\rho = 1/2 + (2/3)\sigma_z/2$ with the prior mean $\rho_\mu = 1/2 + (4/5)\sigma_z + (1/7)\sigma_x$.

```
In [15]: basis = qi.tomography.pauli_basis(1)
model = qi.BinomialModel(qi.tomography.TomographyModel(basis))
true_state = basis.state_to_modelparams(
    I / 2 + (2 / 3) * Z / 2
)[np.newaxis, :]
fiducial_prior = qi.tomography.GinibreReditDistribution(basis)
prior = qi.tomography.GADFLIDistribution(fiducial_prior,
    I / 2 + (4 / 5) * Z / 2 + (1 / 7) * X / 2
)
```

In [16]: `qi.tomography.plotting_tools.plot_rebit_prior(prior, true_state=true_state, rebit_axes=[1, 3])`



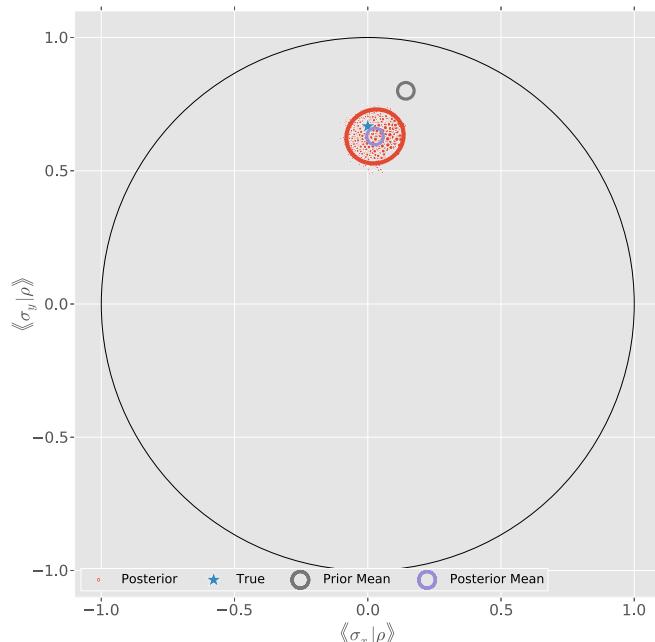
The updater and heuristic classes track the posterior and the random-measurement experiment design, respectively.

```
In [17]: updater = qi.smc.SMCUpdater(model, 2000, prior)
heuristic = qi.tomography.RandomPauliHeuristic(updater, other_fields={'n_meas': 40})
```

We synthesize data for the true state, then feed it into the updater in order to obtain our final posterior.

```
In [18]: for idx_exp in xrange(50):
    experiment = heuristic()
    datum = model.simulate_experiment(true_state, experiment)
    updater.update(datum, experiment)
```

```
In [19]: plt.figure(figsize=(10, 10))
qi.tomography.plotting_tools.plot_rebit_posterior(
    updater, prior, true_state,
    rebit_axes=[1, 3]
)
```



We can use our tomography basis object to read out the estimated final state as a QuTiP Qobj.

```
In [20]: est_mean = basis.modelparams_to_state(updater.est_mean())
display(est_mean)
```

Quantum object: dims = [[2], [2]], shape = [2, 2], type = oper, isherm = True

$$\begin{pmatrix} 0.814 & 0.014 \\ 0.014 & 0.186 \end{pmatrix}$$

As discussed in the main text, the posterior can also be described by the covariance superoperator $\Sigma\rho = \text{Cov}(|\rho\rangle\rangle)$. We demonstrate by showing the Choi matrix $J(\text{Cov}(|\rho\rangle\rangle))$.

```
In [21]: cov_superop = basis.covariance_mtx_to_superop(updater.est_covariance_mtx())
display(qt.to_choi(cov_superop))
display(Latex(r"\Sigma\rho = \text{Cov}(|\rho\rangle\rangle) = \text{Tr}(\Sigma\rho) = \text{Tr}(\Sigma\rho) / \text{Tr}(\rho)"))
```

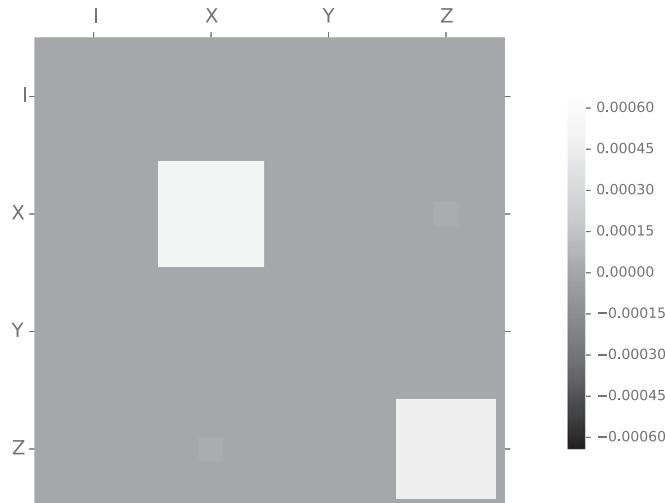
Quantum object: dims = [[[2], [2]], [[2], [2]]], shape = [4, 4], type = super, isherm = True, superrep = choi

$$\begin{pmatrix} 2.854 \times 10^{-4} & 1.622 \times 10^{-5} & 1.622 \times 10^{-5} & 3.210 \times 10^{-4} \\ 1.622 \times 10^{-5} & -2.854 \times 10^{-4} & 3.210 \times 10^{-4} & -1.622 \times 10^{-5} \\ 1.622 \times 10^{-5} & 3.210 \times 10^{-4} & -2.854 \times 10^{-4} & -1.622 \times 10^{-5} \\ 3.210 \times 10^{-4} & -1.622 \times 10^{-5} & -1.622 \times 10^{-5} & 2.854 \times 10^{-4} \end{pmatrix}$$

$$\|\Sigma\rho\|_{\text{Tr}} = 0.0012$$

Here, we use the Hinton diagram plotting functionality provided by QuTiP to depict the covariance in each observable that we obtain from the posterior.

```
In [22]: display(qt.visualization.hinton(cov_superop))
(<matplotlib.figure.Figure at 0x18218a20>,
 <matplotlib.axes._subplots.AxesSubplot at 0x18209f98>)
```



References

- [1] Huszár F and Housley N M T 2012 Adaptive Bayesian quantum tomography *Phys. Rev. A* **85** 052120
- [2] Ferrie C 2014 Quantum model averaging *New J. Phys.* **16** 093035
- [3] 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
- [4] Wiebe N, Granade C and Cory D G 2015 Quantum bootstrapping via compressed quantum Hamiltonian learning *New J. Phys.* **17** 022005
- [5] Holzapfel M, Baumgratz T, Cramer M and Plenio M B 2015 Scalable reconstruction of unitary processes and Hamiltonians *Phys. Rev. A* **91** 042129
- [6] Newton R G and Young B-I 1968 Measurability of the spin density matrix *Ann. Phys., NY* **49** 393
Band W and Park J L 1970 The empirical determination of quantum states *Found. Phys.* **1** 133
Band W and Park J L 1979 Quantum state determination: quorum for a particle in one dimension *Am. J. Phys.* **47** 188
- [7] Hradil Z 1997 Quantum-state estimation *Phys. Rev.* **55** R1561
- [8] Christandl M and Renner R 2012 Reliable quantum state tomography *Phys. Rev. Lett.* **109** 120403
- [9] Blume-Kohout R 2012 Robust error bars for quantum tomography (arXiv:1202.5270)
- [10] Shang J, Ng H K, Sehravat A, Li X and Englert B-G 2013 Optimal error regions for quantum state estimation *New J. Phys.* **15** 123026
- [11] Guţă M, Kypraios T and Dryden I 2012 Rank-based model selection for multiple ions quantum tomography *New J. Phys.* **14** 105002
- [12] 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
- [13] Blume-Kohout R 2010 Hedged maximum likelihood estimation *Phys. Rev. Lett.* **105** 200504
- [14] 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
- [15] Flammia S T, Gross D, Liu Y-K and Eisert J 2012 Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators *New J. Phys.* **14** 095022
- [16] Jones K 1991 Principles of quantum inference *Ann. Phys., N.Y.* **207** 140
Jones K 1991 Quantum limits to information about states for finite dimensional Hilbert space *J. Phys. A: Math. Gen.* **24** 121
Jones K 1994 Fundamental limits upon the measurement of state-vectors *Phys. Rev. A* **50** 3682
- [17] Slater P B 1995 Quantum coin-tossing in a Bayesian Jeffreys framework *Phys. Lett. A* **266** 66
- [18] Derka R, Bužek V, Adam G and Knight P L 1996 From quantum Bayesian inference to quantum tomography *J. Fine Mech. Opt.* **11–12** 341 (arXiv:quant-ph/9701029)
- [19] Bužek V, Derka R, Adam G and Knight P L 1998 Reconstruction of quantum states of spin systems: from quantum Bayesian inference to quantum tomography *Ann. Phys., NY* **266** 454
- [20] Schack R, Brun T A and Caves C M 2001 Quantum Bayes rule *Phys. Rev.* **64** 014305
- [21] Blume-Kohout R and Hayden P 2006 Accurate quantum state estimation via ‘Keeping the experimenter honest’ (arXiv:quant-ph/0603116)
- [22] Blume-Kohout R 2010b Optimal, reliable estimation of quantum states *New J. Phys.* **12** 043034
- [23] Ferrie C and Kueng R 2015 Have you been using the wrong estimator? These guys bound average fidelity using this one weird trick von Neumann didn’t want you to know *New J. Phys.* **17** 123013
- [24] Kravtsov K S, Straupe S S, Radchenko I V, Housley N M T, Huszar F and Kulik S P 2013 Experimental adaptive Bayesian tomography *Phys. Rev. A* **87** 062122
- [25] Ferrie C 2014b High posterior density ellipsoids of quantum states *New J. Phys.* **16** 023006
- [26] Wiebe N, Granade C, Ferrie C and Cory D 2014 Quantum Hamiltonian learning using imperfect quantum resources *Phys. Rev. A* **90** 042314
- [27] Struchalin G, Pogorelov I, Straupe S, Kravtsov K, Radchenko I and Kulik S 2015 Experimental adaptive quantum tomography of two-qubit states *Phys. Rev. A* **93** 012103

- [28] Doucet A and Johansen A M 2009 *A Tutorial on Particle Filtering and Smoothing: Fifteen Years Later* (Oxford: Oxford University Press)
- [29] Doucet A, Godsill S and Andrieu C 2000 On sequential Monte Carlo sampling methods for Bayesian filtering *Stat. Comput.* **10** 197
- [30] Audenaert K M R and Scheel S 2009 Quantum tomographic reconstruction with error bars: a Kalman filter approach *New J. Phys.* **11** 023028
- [31] Ince D C, Hatton L and Graham-Cumming J 2012 The case for open computer programs *Nature* **482** 485
- [32] Borot G and Nadal J 2012 Purity distribution for generalized random bures mixed states *J. Phys. A: Math. Theor.* **45** 075209
- [33] Schwarz L and van Enk S J 2011 Detecting the drift of quantum sources: not the de Finetti theorem *Phys. Rev. Lett.* **106** 180501
- [34] Langford N K 2013 Errors in quantum tomography: diagnosing systematic versus statistical errors *New J. Phys.* **15** 035003
- [35] Shulman M D, Harvey S P, Nichol J M, Bartlett S D, Doherty A C, Umansky V and Yacoby A 2014 Suppressing qubit dephasing using real-time Hamiltonian estimation *Nat. Commun.* **5** 5156
- [36] Fogarty M A, Veldhorst M, Harper R, Yang C H, Bartlett S D, Flammia S T and Dzurak A S 2015 Nonexponential fidelity decay in randomized benchmarking with low-frequency noise *Phys. Rev. A* **92** 022326
- [37] Granade C E 2015 Characterization, verification and control for large quantum systems *PhD Thesis* University of Waterloo (<http://hdl.handle.net/10012/9217>)
- [38] The third paragraph on page 22 of Blume-Kohout R, Gamble J K, Nielsen E, Maunz P, Scholten T and Rudinger K 2015 Turbocharging quantum tomography *Technical Report* Sandia National Laboratories (SNL-NM), Albuquerque, NM (United States) (SAND2015-0224)
- [39] Granade C, Ferrie C et al 2012 QInfer: library for statistical inference in quantum information (<http://github.com/cferrie/python-qinfer>)
- [40] Johansson J R, Nation P D and Nori F 2013 QuTiP 2: a Python framework for the dynamics of open quantum systems *Comput. Phys. Commun.* **184** 1234
- [41] Casagrande S and Granade C 2013 InstrumentKit: Python package for interacting with laboratory equipment (<https://github.com/Galvant/InstrumentKit>)
- [42] Banerjee A, Guo X and Wang H 2005 On the optimality of conditional expectation as a Bregman predictor *IEEE Trans. Inf. Theory* **51** 2664
- [43] Granade C E, Ferrie C, Wiebe N and Cory D G 2012b Robust online Hamiltonian learning *New J. Phys.* **14** 103013
- [44] Wiebe N, Granade C, Kapoor A and Svore K M 2015 Bayesian interference via rejection filtering (arXiv:1511.06458)
- [45] Doucet A and Johansen A M 2011 A tutorial on particle filtering and smoothing: fifteen years later *Handbook of Nonlinear Filtering* (Oxford: Oxford University Press)
- [46] Wasserman L 2013 LOST CAUSES IN STATISTICS: II. Noninformative Priors (<http://normaldeviate.wordpress.com/2013/07/13/lost-causes-in-statistics-ii-noninformative-priors>)
- [47] Osipov V A, Sommers H-J and Zyczkowski K 2010 Random Bures mixed states and the distribution of their purity *J. Phys. A: Math. Theor.* **43** 055302
- [48] Zyczkowski K and Sommers H-J 2001 Induced measures in the space of mixed quantum states *J. Phys. A: Math. Gen.* **34** 7111
- [49] Mezzadri F 2007 How to generate random matrices from the classical compact groups *Not. Am. Math. Soc.* **54** 592
- [50] Bruzda W, Cappellini V, Sommers H-J and Życzkowski K 2009 Random quantum operations *Phys. Lett. A* **373** 320
- [51] Veitch V, Mousavian S A H, Gottesman D and Emerson J 2014 The resource theory of stabilizer quantum computation *New J. Phys.* **16** 013009
- [52] Granade C, Combes J and Cory D G 2015a Practical Bayesian tomography supplementary video: recovery from ‘bad’ priors (<https://goo.gl/gpb94w>)
- [53] Isard M and Blake A 1998 CONDENSATION-conditional density propagation for visual tracking *Int. J. Comput. Vis.* **29** 5
- [54] Jasra A and Doucet A 2009 Sequential Monte Carlo methods for diffusion processes *Proc. R. Soc. A* **465** 3709
- [55] Schirmer S G and Oi D K L 2010 Quantum system identification by Bayesian analysis of noisy data: Beyond Hamiltonian tomography *Laser Phys.* **20** 1203
- [56] Sergeevich A, Chandran A, Combes J, Bartlett S D and Wiseman H M 2011 Characterization of a qubit Hamiltonian using adaptive measurements in a fixed basis *Phys. Rev. A* **84** 052315
- [57] Granade C, Combes J and Cory D G 2015 Practical Bayesian tomography supplementary video: state-space state tomography (<https://goo.gl/mkibti>)
- [58] Faist P and Renner R 2015 Practical, reliable error bars in quantum tomography (arXiv:1509.06763)
- [59] Chuang I L and Nielsen M A 1997 Prescription for experimental determination of the dynamics of a quantum black box *J. Mod. Opt.* **44** 2455
- [60] Choi M-D 1975 Completely positive linear maps on complex matrices *Linear Algebr. Appl.* **10** 285
- [61] Jamiołkowski A 1972 Linear transformations which preserve trace and positive semidefiniteness of operators *Rep. Math. Phys.* **3** 275
- [62] Wood C J, Biamonte J D and Cory D G 2015 Tensor networks and graphical calculus for open quantum systems *Quantum Inf. Comput.* **15** 0579 (www.rintonpress.com/journals/qiconline.html#v15n910)
- [63] Wood C 2015 Initialization and characterization of open quantum systems *PhD Thesis* University of Waterloo (<http://hdl.handle.net/10012/9557>)
- [64] Ringbauer M, Wood C J, Modi K, Gilchrist A, White A G and Fedrizzi A 2015 Characterizing quantum dynamics with initial system-environment correlations *Phys. Rev. Lett.* **114** 090402
- [65] Altepeter J B, Branning D, Jeffrey E, Wei T C, Kwiat P G, Thew R T, O’Brien J L, Nielsen M A and White A G 2003 Ancilla-assisted quantum process tomography *Phys. Rev. Lett.* **90** 193601
- [66] Watrous J 2013 CS 766: *Theory of Quantum Information (Lecture Notes)* (<http://as.uwaterloo.ca/~watrous/TQI/>)
- [67] Granade C 2015 Robust online Hamiltonian learning: multi- \cos^2 model resampling (<https://goo.gl/CR9Dtr>)
- [68] Stenberg M P, Sanders Y R and Wilhelm F K 2014 Efficient estimation of resonant coupling between quantum systems *Phys. Rev. Lett.* **113** 210404
- [69] Wiebe N, Granade C, Ferrie C and Cory D 2014b Hamiltonian learning and certification using quantum resources *Phys. Rev. Lett.* **112** 190501
- [70] Svensson A 2013 The particle filter explained without equations (<https://goo.gl/VjZJVm>)