Coarse-graining in retrodictive quantum state tomography

Dale Scerri, Erik M Gauger and George C Knee

1 SUPA, Institute of Photonics and Quantum Sciences, Heriot-Watt University, David Brewster Building, Edinburgh, EH14 4AS, United Kingdom
2 Department of Physics, University of Warwick, Coventry, CV4 7AL, United Kingdom
3 Author to whom correspondence should be addressed.
E-mail: ds32@hw.ac.uk

Keywords: quantum state tomography, quantum measurements, retrodiction, quantum retrodictive tomography, phase errors, quantum mechanics, quantum physics

Abstract
Quantum state tomography (QST) is a combination of experimental and data-processing methods for complete characterisation of a quantum system. However, it often operates in the highly idealised scenario of assuming perfect measurements. The errors implied by such an approach are entwined with other imperfections relating to the information processing protocol or application of interest. We consider the problem of retrodicting the quantum state of a system, existing prior to the application of random but known phase errors, allowing those errors to be separated and removed. The continuously random nature of the errors implies that effective measurement operators are never repeated. This is a feature of many physical scenarios such as photonic cluster state generation and has a drastically adverse effect on data-processing times. Utilising state-of-the-art approaches to quantum state tomography, we describe a novel and effective method to account for these errors, resulting in improved reconstruction fidelities. Furthermore, we show how the use of ‘coarse-graining’ introduced in this work can substantially reduce the computation time in several maximum likelihood algorithms, for only modest sacrifices in fidelity.

1. Introduction

Accurate quantum state reconstruction from finite data is a fundamental tool in quantum information science, as it allows for the calculation of any desired physical property of interest (such as coherence or entanglement), and is the most complete approach to the characterisation of quantum systems including atomic, molecular, and engineered platforms. Continued development of experimental tomography protocols and data-processing algorithms has improved both the accuracy and computational time required to produce state estimates in the face of the exponential complexity of quantum systems. Despite being a mature field of research, quantum state tomography suffers from outstanding problems, such as measurement errors caused, for example, by noisy detector readout or by mis-calibrated measurement apparatuses. Measurement errors may be systematic or random, and will tend to reduce the fidelity of the tomogram, with respect to the true state \( \rho \). If errors are known in a general quantum information processing protocol on a shot-by-shot basis, they may generally be compensated for by additional quantum control. The irreversible nature of the quantum detection process, however, means that post-measurement knowledge of errors is insufficient for such compensation.

A concrete example of such a situation comes from the field of photonic cluster state generation. A single emitter—e.g. a natural atom or quantum dot—will spontaneously undergo radiative decay at a random delay after excitation. The emitted photons are entangled with the emitter in such a way that repeated resonant control of the emitter’s spin state and further excitations causes the subsequent emission of a chain of photons to be generated in a linear cluster state [1–3]: a key resource [4] for measurement based quantum computation [5]. Such schemes rely on an external magnetic field orthogonal to the optical axis [1–3]. Due to the non-zero lifetime \( \tau_d \) of the emitter, the spin precesses at an angular frequency \( \omega_l \) for a random interval. We may thus think of nature applying a random phase to the spin, which is then transferred to the emitted photon but revealed to
the experimenter immediately upon detection. The task of estimating the density matrix $\rho$ of the photonic cluster state in the limit of $t \to 0$ is an instance of what we term 'retrodictive quantum state tomography'. A sw e show, retrodiction is made possible by re-defining the effective measurement operators using post-measurement knowledge of the random phases.

The situation may be modelled more generally by a semi-malevolent agent intervening in the experiment, applying random evolutions $r \to U_r \rho U_r^\dagger$ that are only revealed to the experimenter after they have made their measurements. For concreteness, we take $U_r = \cos \frac{\theta_0}{2} + i \sin \frac{\theta_0}{2} \sigma$ for $\sigma$, the usual Pauli operator, and $\rho$ as the system density matrix when no errors occur. Although the errors cannot be corrected in the sense of a fault tolerant quantum protocol, it is possible to retrodict the quantum state which existed before the errors were applied. Since the success probability of a fixed measurement operator $M$ is $\text{tr}(M [U_{\theta} \rho U_{\theta}^\dagger]) = \text{tr}((U_{\theta} M U_{\theta}^\dagger) \rho)$, moving from the Schrödinger to Heisenberg pictures, the situation becomes equivalent to performing tomography on an ideal preparation $\rho$ with random measurements—see figure 1. The retrodiction is useful because $\rho$ may still contain other sources of error, which may then be separately estimated [6, 7].

For the technique to work, it is necessary that the effective measurement operators are known: in the precessing spin example, this information is revealed by the arrival time of the photon, the angular precession frequency $\omega$, and the time-of-flight of the photon to the detector. Because of the continuous nature of the distribution over $\theta$, the measurement record has the following 'sparsity' feature: measurement operators will never be repeated, meaning that at most one click is attributed to each outcome. In this paper we show that retrodictive tomography is successful in spite of this feature; indeed, we show how standard tomographic techniques can be modified in order to account for these errors, resulting in accurate state reconstruction, which we characterise by the fidelity of the reconstructed density matrix. Furthermore, we go on to investigate the merits and demerits of coarse-graining—a technique which removes sparsity by introducing a finite number of discrete bins which the measurement results are aggregated into. Our numerical simulations reveal that fidelity degrades monotonically as the number of bins is reduced, but that this is accompanied by a drastic improvement in algorithm run-time. As well as being a choice available to the tomographer, coarse-graining can also be considered as one way of simulating imperfect knowledge about the errors $\theta$. Intuitively, a Bayesian shot-by-shot approach is a natural paradigm to tackle the sparse tomography problem, making use of prior knowledge to process additional data obtained as more measurements are performed. However, the binning approach (discussed in section 2) cannot be applied to this technique straightforwardly. Thus, the Bayesian approach will
2. Sparse and binned tomography

Define $p(\theta)$ as the distribution of random phases (supported on $[0, 2\pi]$), which dictates the distribution of effective measurement operators. This distribution depends on the physical scenario: in the example of frequency-encoded cluster-state generation in the hole-spin system in [2], when the precession time is much shorter than the emission time, $p(\theta) \approx 1/(2\pi)$. In such a case, the coherences of the reconstructed state would be completely washed out by conventional QST techniques (not making use of the knowledge of the errors $\theta$). For the more general case of photon emission from spin-bearing emitters, however, the exponential distribution $p(\theta) \propto e^{-\theta/\mu}$ (with the mean $\mu = \lambda^{-1}$, where $\lambda$ is the rate parameter) is more adequate to describe the spread of operators. Other distributions may be similarly treated—meaning that our analysis applies to a wider range of physical scenarios—although the measurement operators may then be clustered to a greater or lesser degree, having an affect on the accuracy of the retrodicted tomogram. In appendix D we give the phase distribution and corresponding parameters for some quantum information and computation schemes. The normal distribution $p(\theta) \propto e^{-\theta^2/2\lambda^2}$ ($\sigma$ being the standard deviation) is considered in section 5, while as $\mu \to \infty$, we recover the uniform distribution limit, i.e. $p(\theta) \to 1/(2\pi)$.

In the Schrödinger picture, we fix the four measurement operators $\{|\rangle \langle 1|, |1\rangle \langle 1|, |\rangle \langle 0|, |0\rangle \langle 1|\}$ and $\phi \in \{0, \pi\}$. Since emitted photons are measured independently, $m$-qubit states are tomographed by forming $m$-fold tensor products of all combinations of these projectors.

By using the Heisenberg picture (as in the previous section), the tomographic protocol is equivalent to reconstructing some unknown state $\rho$ with the following set of positive (projective) measurement operators

$$P = \{|\rangle \langle 1|, |1\rangle \langle 1|, |\rangle \langle 0|, |0\rangle \langle 1|\}, M_\theta = U_\theta^\dagger |\phi\rangle \langle \phi| U_\theta$$

where $U_\theta^\dagger |\phi\rangle = |\phi + \theta\rangle$ for $\theta (i = 1, \ldots, N/2)$ drawn from $p(\theta)$. Note that $U_\theta^\dagger |\rangle \langle 1| U_\theta = |\rangle \langle 1|$ and that the values of $\phi$ play less of a role as the spread of $\theta$ increases. Because $|\phi\rangle \langle \phi| + |\phi + \pi\rangle \langle \phi + \pi| = I$, this set may be considered a POVM (Positive Operator Valued Measure) upon appropriate normalisation (in the sense that the sum of all operators is proportional to the identity).

We generated pseudo-tomographic data for a fixed $\rho$ by drawing $N/2$ unique values of $\theta \in [0, 2\pi)$ from $p(\theta)$. We then simulate a single Bernoulli trial for each measurement operator, assigning the event to $M_\theta$ with probability $p_\mu = \text{tr}(\rho M_\theta)$, and to the orthogonal operator with the complementary probability. A further $N/2$ Bernoulli trials are assigned to $|\rangle \langle 1|, |\rangle \langle 0|, |\rangle \langle \pi|, |\rangle \langle \pi|, |\rangle \langle 0|, |\rangle \langle 1|$ in the same fashion. The measurement record then consists of a multiset of $N/2 + 2$ distinct measurement operators with (for the $N/2$ operators perpendicular to the ‘precession’ axis) multiplicities $n_\theta = 1$ (i.e. it is ‘sparse’). The pair of orthogonal operators parallel to the ‘precession’ axis have a joint multiplicity of $N/2$.

Optionally, we modify the measurement record by a process of coarse-graining or ‘binning’, resulting in a lower number $N_\theta < N/2$ of coarse-grained measurement operators perpendicular to the precession axis. Non-overlapping measurement bins are indexed by $j \in \{1, 2, \ldots, N_\theta\}$, and have multiplicities defined by

$$n_j = \sum _{i} n_i \text{rect}\left( \frac{N_i \theta_i}{2\pi} - j + \frac{1}{2} \right)$$

where $\text{rect}(x) = 1$ if $|x| < \frac{1}{2}$ and 0 otherwise, i.e. we simply accumulate the events according to the bin that they fall within, with the bins being intervals centred on $(2j - 1) \frac{2\pi}{N}$ and with width $2\pi/N_\theta$ (as shown graphically in figure 1). Each bin is assigned a measurement operator defined by

$$M_\theta = U_\theta^\dagger |\phi\rangle \langle \phi| U_\theta$$
where the angle $\tilde{\theta}_j$ is the mean of the random angles in the $j$th bin (that is $(j - 1) \frac{2\pi}{N_R} \leq \tilde{\theta}_j < \frac{2\pi}{N_R}$), that is

$$\tilde{\theta}_j = \frac{1}{N_j} \sum \theta_j \text{rec}\left(\frac{N_R \theta_j}{2\pi} - j + \frac{1}{2}\right),$$

(4)

Other binning schemes are possible [8]. We then run different reconstruction algorithms (to be introduced below) on the coarse-grained measurement record, to give a quantum state estimate or ‘tomogram’ $\rho_{est}$. The running time of the algorithm is noted, and the fidelity of the tomogram computed: $F(\rho_{est}, \rho) = \text{tr}\left[\sqrt{\sqrt{\rho} \rho_{est} \sqrt{\rho}}\right]$. The infidelity is $1 - F$, and is a measure of the distance between the true state and the retrodicted tomogram. The procedure was then repeated for distinct, randomly generated (but full rank) $\rho$, and we collected statistics to summarise the typical performance.

Counter to intuition, using sparse tomography without any binning works remarkably well. However, algorithm running time tends to scale badly with $N$ (since the calculation of the cost function and its gradient involves a contribution from each of the $N$ distinct operators). Hence our proposed coarse-grained approach. The remainder of the paper is dedicated to investigating the dependence of fidelity and run time on $N_R$, for different reconstruction algorithms. As $N_R \rightarrow \infty$, the sparse and coarse grained approaches are expected to give the same fidelities.

3. Non-adaptive Bayesian tomography

The Bayesian approach was introduced in the field of quantum tomography [9–14], and is an ongoing theoretical and experimental research topic [15–17]. This approach offers numerous advantages over other techniques, such as use of online information available to the experimentalist after each measurement. Furthermore, Bayesian inference was also shown to be optimal with respect to any strictly proper scoring rule derived from Bregman distances [15, 18, 19] (near-optimal if the infidelity is used as a loss function instead [20]), with the ability to track fidelity bounds online [20] (allowing for feedback to minimise number of required measurements), as well as giving robust region estimates [21] and allowing for model selection/averaging. Thus the Bayesian approach shall be used as a benchmark for the other techniques discussed in this work.

Our implementation follows closely the approaches used in [16] and [22]. For a Bayesian update scheme, we start with an initial prior probability density $p(\rho)$ over feasible state space (usually uninformed due to the absence of additional knowledge, resulting in a uniform prior). After obtaining a new measurement datum $D$, the posterior distribution $p(\rho|D)$ is then built using the likelihood function $L(\rho; D)$ as

$$p(\rho|D) \propto L(\rho; D)p(\rho).$$

(5)

Typically, Bayesian tomography schemes would then make use of the narrower posterior and additional criteria (for example, Shannon information [22]) to infer the next optimal measurement setting [16, 22]. However, when considering a finite lifetime for the emitter, any measurement basis prediction would be futile for the next measurement due to the random phase gained known only post-detection. Although we do not make use of any criteria to track the narrowing of the sample, one could still use the covariance of the thresholded posterior, in this case, to indicate when a sufficiently precise estimate has been found.

Despite the simple form of equation (5), the analytical evaluation of the posterior is seldom feasible, and hence the latter is typically replaced with an approximation. To this end, several Markov Chain Monte Carlo techniques (MCMC) have been adopted, including the Metropolis–Hastings algorithm [15]. However, these MCMC techniques tend to be computationally expensive, with decreasing acceptance probabilities at each sampling step, leading to more samples being discarded as additional data is obtained. Furthermore, these methods require the assumption of a normal posterior, which is not always the case in state tomography. The Sequential Monte Carlo technique (SMC) [23, 24], on the other hand, only requires the computation of a single term of the likelihood to update the weights of the approximate distribution with each measurement [22]. In this approach, adopting the notation in [22], the posterior after the $i$th measurement is approximated by a number $P$ of randomly sampled particles, $\{p_k\}$, and their corresponding weights $\{w_k^{(i)}\}$ as

$$p(\rho|\{D_i\}) \approx \sum_{p=1}^{P} w_k^{(i)} \delta(\rho - \rho_p).$$

(6)

Suppose our current (prior) knowledge is given by the dataset $\{D_i\} = \{\alpha_i; \ 1 \leq j \leq i, \ \alpha_i \in \mathcal{P}\}$, where the set $\mathcal{P}$ is defined in equation (1). If the next projection phase is, without loss of generality, $\theta_{i+1}$, (that is, $\alpha_{i+1} = M_{\theta_{i+1}}$), then, following [22] and using Bayes’ rule (equation (5)), we can write the approximation for the next posterior as

$$p(\rho|\{D_{i+1}\}) = p(\rho|\{D_i\} \cup \{M_{\theta_{i+1}}\})$$

(7)
where $P(M_{\theta_p}|\rho_p) = \text{Tr}(M_{\theta_p}\rho_p)$. In our numerical simulations, we do the first $N/2$ measurements along the $z$-axis (that is, using projection operators $\{|1\rangle \langle 1|, |\bar{1}\rangle \langle \bar{1}|\}$), followed by the remaining $N/2$ measurements along the Bloch equatorial plane. As more measurements are performed, narrowing the particle distribution, most of the original particle weights drop to zero, which can be remedied by resampling using the new posterior distribution \(\[\text{22}\]\). Finally, the Bayes estimator $\rho_{\text{est}}$ can be extracted from the mean of the final posterior approximation. In figure 2 we show the above steps graphically, emphasising the use of resampling to obtain an accurate posterior.

We numerically benchmarked the Bayesian technique, using a uniform prior\(^4\). An example is shown in figure 2, and further results are summarised in figure 3\(^5\). Despite the fact that we cannot decide which measurement to perform next, our random basis measurement can be seen to give a good convergence after 2000 measurements with 1000 particles.

\[\sum_{p=1}^{P} \frac{P(M_{\theta_p}|\rho_p)w^{(i)}_p}{\sum_{q=1}^{P} P(M_{\theta_q}|\rho_q)w^{(i)}_q} \delta(\rho - \rho_p) \]

\[= \sum_{p=1}^{P} w^{(i+1)}_p \delta(\rho - \rho_p) , \]

\[a\]

\[b\]

\[c\]

\[d\]

**Figure 2.** (a) Initial uninformed prior (orange), with the mean of the distribution shown in green, and the true state to be reconstructed in red. (b) Final posterior (orange) after 2000 measurements, where the marker size indicates the relative particle weights. (c) and (d) show the $\{ |\sigma_x\rangle, |\sigma_z\rangle\}$ projection of the prior and posterior, respectively, as a visual aid. As more measurements are performed, most of the original particle weights drop to zero, requiring resampling for a more accurate prediction without requiring an excessive number of particles to begin with.

4. **Maximum likelihood estimation**

A common, alternative, approach to state estimation is producing a tomogram $\rho_{\text{est}}$ which maximises the likelihood function. Naive approaches may result in an invalid tomogram (having, for example, negative eigenvalues). The search for the best fit to the data, therefore, should be constrained to the allowed state space of trace-one positive semidefinite matrices \(\[\text{27–29}\]\). Previously (in the Bayesian method) this was ensured by choosing a prior distribution supported only in the allowed state space. Here, the prior is not modelled, but we consider two alternative approaches: (A) the constraints are enforced by a non-linear parametrization of the density matrix and (B) the constraints are enforced periodically in the course of an iterative gradient descent procedure, allowing for temporary violations \(\[\text{29–31}\]\). Given a density matrix $\rho$, the likelihood function to be maximised has the form

---

\(^4\) Samples are drawn from a Ginibre ensemble (random matrices with normally distributed entries), and subsequently multiplied by their corresponding Hermitian conjugates to give positive matrices. Finally, these are normalised to unit trace \(\[\text{25, 26}\]\).

\(^5\) The data follow a skewed distribution due to the fidelity being capped by unity, and thus cannot be described by the use of symmetrical error bars. The use of box plots was thus preferred over the standard approach of mean and error bars.
with equality holding up to an irrelevant proportionality constant. For sparse tomography, the product would be over \( N \) exponentiated probabilities \( p_j \), with each \( n_j \) taking a binary value of either 0 or 1. Due to the monotonicity of the logarithm, maximising the likelihood function is identical to minimising the negative of its logarithm \[ \mathcal{C}(\rho) = -\log \mathcal{L}(\rho) = -\sum_{j=1}^{N} n_j \log(p_j) , \] where we took the normalising constant to identity. Recall that the sparse tomography limit is recovered when \( N_b = N \) and \( n_j = 1 \). In the limit of a large number of detections per measurement, the probability of obtaining the \( j \)th measurement can be approximated by a Gaussian distribution \[ 32, 33 \] with the estimated number of detections for the \( j \)th measurement given by \[ = n_j p_j . \] Since this approximation clearly fails for the sparse case due to the binary nature of the \( n_j \)'s, we do not make it.

### 4.1. Cholesky factorisation

In this section we implement a Cholesky-like decomposition of the density matrix in order to minimise equation (11) \[ 32, 33 \] allowing us to use Python’s SciPy least-squares solver on a 1D array\(^6\). One can easily show that any qubit density matrix \( \rho \) allows for a decomposition of the form \[ \rho = T^\dagger T / \text{Tr}[T^\dagger T] , \] where \( T \) is the lower triangular matrix given by \[ t_{0} \ t_{1} \ t_{2} \ t_{3} \ t_{4} \] with \( t = (t_0, t_1, t_2, t_3) \) being the array over which the minimisation search is performed. In particular, we can use this decomposition to calculate \[ \bar{n}_j \propto \mathcal{P}_j = \text{Tr}[(\phi + \theta_j)(\phi + \theta_j)^\dagger] / \text{Tr}[T^\dagger T]. \] Generalising this parametrisation to \( m \) qubits, we get

\[
T(t) = \begin{pmatrix}
  t_1 & 0 & \ldots & 0 \\
  t_2 & t_3 & \ldots & 0 \\
  \vdots & \vdots & \ddots & \vdots \\
  t_{2^{m-1}} & t_2 t_{2^{m-2}} & \ldots & t_{2^{m-1}} + it_{2^{m-2}} & t_{2^{m-1}} \end{pmatrix},
\]

and hence the search needs to be done over a real array of length \( 4^m \).

Having formulated a decomposition guaranteeing a valid density matrix, the problem can be recast to a least-squares minimisation problem \[ 32, 33 \] in order to find the minimum of the negative log likelihood, as the latter

\[^6\text{More formally, this least-squares solver uses the Trust Region Reflective technique (involving searching along directions reflected from the trust region bounds).}\]
may be written down as
\[
C(\rho) = \sum_{i=1}^{N} [f_i(t)]^2 ,
\]
where, for the general case of a multinomial probability distribution, we get using equation (11)
\[
f_i(t) = \sqrt{n_i \log(p_i)}
\]
\[
= \sqrt{n_i} \left( \log \left[ \text{Tr} \left[ \phi_i \phi_i^* T^\dagger(t) T(t) \right] \right] \right) - \log \left[ \text{Tr} \left[ T^\dagger(t) T(t) \right] \right] .
\]

Despite having multiple local minima, this optimization problem was shown to have a single global solution [34], meaning that all local minimizers lead to the same solution minimizing the negative log likelihood.

In Figure 4 we show the results for single qubit reconstruction. As expected, the fidelity of the reconstructed density matrix increases with number of Bloch sphere partitions. This is also the case for a two-qubit reconstruction, as we show in appendix A.

4.2. Projected gradient descent

Gradient descent algorithms rely on following the path of steepest descent of the cost function, in this case equation (11), starting from a well chosen initial estimate. If left unconstrained in the convex space of \(d \times d\) matrices (where \(d\) is the Hilbert space dimension), the resulting estimate \(\rho_{\text{est}}\) might lie outside the convex subspace of unit-trace, positive semidefinite matrices, leading to an unphysical estimate. Hence, projection back to the physical subspace, minimising distance as measured through of a matrix norm (such as projection of the spectrum onto the unit simplex [30, 31, 34]) is employed, giving rise to projected gradient descent (PGD) algorithms. Iterating this process leads to a convergence of the cost function to a minimum below a predefined threshold. A unique solution satisfying the appropriate constraints and minimising the cost function is then guaranteed as long as the latter is a continuously differentiable convex function of the density matrix.

Equation (11) is convex but not continuously differentiable, but this tends to not pose a problem in practice, as discussed in [29]. Choosing the projection of \(\rho\) to be of its spectrum onto the unit simplex (which we refer to as \(P_{\text{S}}\), the PGD algorithm update can be written as
\[
\rho_k = P_{\text{S}}[\rho_{k-1} - \nabla C(\rho_{k-1})] .
\]

As is commonplace, we supplement the PGD algorithm with a backtracking line search (PGDB) based on the Armijo–Goldstein condition to loosely optimise the maximum step size for each descent iteration \([30, 31, 34]\).

The estimate at the \(k\)th PGDB iteration can thus be written as
\[
\rho_k = (1 - \alpha)\rho_{k-1} + \alpha P_{\text{S}}[\rho_{k-1} - \nabla C(\rho_{k-1})] ,
\]
where \( \alpha \in [0, 1] \) is the line search parameter to be roughly optimised at each step. We assess the impact of binning on the PGDB algorithm, figure 5(a) showing the trade-off between computation time and fidelity for \( \mu = \pi/8 \). Figure 5(b), on the other hand shows the relation between computation time and fidelity for various number of bins \( N_b \) and events \( N \) for normally distributed phases, showing a similar trend to the exponentially spread phases. Appendix C shows a closer analysis of the exponential data, with first and third quartiles for the fidelity, and standard deviation errorbars for computation times.

As expected, within standard deviation error, the binned approach gives slightly lower fidelities than the sparse one. This difference, however, is well justified when considering the significant reduction in computation time shown in figure 5(a). The trends in figure 5(a), both for computation time and fidelity, are similar to those shown in figure 4 for the Cholesky method. However, our numerical simulations clearly show lower reconstruction times achieved using the PDGB technique. In figure 6, we show how the fidelity varies with increasing mean \( \mu \) for various values of the bin number \( N_b \).

5. Condition numbers

Using a single basis for reconstruction along the plane of precession, we see that the higher the spread of the distribution, the higher the fidelity one expects, as the effective rotated bases sample larger portions of the Bloch plane, whereas for lower spreads, the additional phase knowledge does not contribute considerably, and hence incomplete Pauli tomography (in which only x- and z- basis measurements are performed) is recovered. This can be seen in figure 7, showing the behaviour of the condition number \( \kappa(A) \) of the measurement matrix \( A \) for increasing \( N \), where \( A \) is given by

\[
A = \begin{pmatrix}
\operatorname{vec}(M_{R_0})^T \\
\vdots \\
\operatorname{vec}(M_{R_{N/2-2}})^T
\end{pmatrix},
\]

where the projectors \( \hat{I}_n \) make up the set \( \mathcal{P} \) in equation (1) [31, 35]. The condition number decreases significantly with increasing standard deviation of the distribution, meaning that sampling distributions with larger spreads results in a better conditioned measurement matrix.

6. Conclusion

QST is still an active area of experimental and theoretical research, allowing the reconstruction of quantum states from finite experimental data. In this work, we implemented several QST algorithms in the presence of phase errors which are only known after the system is measured. We showed, with a simple modification, how the unaffected state may be retrodicted using such knowledge. In appendix D, we make a connection between dimensionless parameters used in our work to characterise the random phase distributions, and the physical
parameters involved in some recent quantum information processing and computation protocols. Furthermore, we demonstrated that, at a small cost in fidelity, the reconstruction time can be significantly decreased. All data in this work was generated and visualised using Python and QuTiP package [25, 26].

Acknowledgments

We thank Cristian Bonato for insightful and stimulating discussions. DS thanks SUPA for financial support. GCK was supported by the Royal Commission for the Exhibition of 1851, and EMG acknowledges support from the Royal Society of Edinburgh and the Scottish Government.

Figure 6. (a) Full rank, sparse single qubit reconstruction infidelities for phases sampled from exponential distribution with various values of $\mu$ and experiment repetitions $N$. (b) Infidelities for various segment numbers $N_b$. In both (a) and (b), averages were performed over 1000 trials.

Figure 7. Sparse tomography condition number (shown above for $N = 2 \times 10^6$) decreases (improves) as the standard deviation of the normally distributed phases ($\sigma$) increases. The red bars indicate one sigma uncertainty. When $\sigma$ is high, we recover the limit of many measurements distributed evenly around the equator of the Bloch sphere. In this situation, we obtain the same condition number, $\kappa(A) = 2$, as in the case of complete Pauli measurements [35].
Appendix A. Two-qubit results

Figures A1 and A2 show the effect of particle filter sample sizes on a Bayesian two-qubit reconstruction, and the performance of the Cholesky method for a two-qubit reconstruction, respectively.

Figure A1. Bayesian reconstruction of random two-qubit state against particle filter sample sizes, averaged over 50 trials. In each case the number of measurements was taken to be $N = 100$ due to the computation time taken for higher sample sizes.

Figure A2. Full rank two-qubit reconstruction infidelity using the Cholesky method results for 4, 8 and 16 segments, with increasing number of measurements and averaged over 50 trials.

Appendix B. Pseudocodes

In this section we present the pseudocodes for the PGDB algorithm, and some subroutines used for the Bayesian approach taken from [19].

Algorithm 1. PGDB

1: $k = 0, \mu_{k=0} = 1$
2: Initial estimate $\mu_{\infty} \in \mathcal{S}$.
3: Given $\delta = 10^{-4}, \gamma = 10^{-7}, \mu_{\text{min}} = 10^{-4}, \mu_{\text{max}} = 10^{4}$
4: While $\sum_{\mu > 0} |C(\mu) - C(\mu_{\infty})| > \delta$ do
5: Calculate probability estimates
In table D1, we give the phase distributions that can be applied to various quantum computation and information schemes, along with the relevant parameters used to calculate their effective mean and Kok protocol performance. For all distributions shown, \( \mu \) is given by \( \mu = \frac{\omega}{2\pi \gamma} \), where \( \omega \) and \( \gamma \) are parameters to be defined shortly.

For the last four schemes in table D1, \( \omega \) denotes the emitter’s precession frequency, whereas for the Barrett and Kok protocol [36], \( \omega \) represents the local phase picked up by one of the two spins which spends additional time with the atom.
time in the excited state. For a typical GaAs quantum dot with excitonic energy \( \sim 1.35 \text{eV} \), this leads to an excitonic frequency of \( \omega \approx 3.265 \times 10^5 \text{GHz} \). Assuming a similar setup is used for the Lim et al \[37\] scheme, we arrive at a similar phase distribution. On the other hand, for the first four schemes, \( \gamma \) represents the decay rate of the emitter due to spontaneous emission, whereas for the Scerri et al scheme \[2\] (which makes use of photon-scattering), \( 1/\gamma \) is the loosely-optimised time between \( Y \)-rotations during which the emitter probabilistically scatters a photon. For the Lindner and Rudolph scheme \[1\], \( \mu \) was calculated using parameters which give an error rate of 0.2%, whereas for the Denning et al scheme \[38\], the mean was calculated based on values suggested for high Q-factor cavities.

### ORCID iDs

Dale Scerri @ [https://orcid.org/0000-0003-3640-2389](https://orcid.org/0000-0003-3640-2389)

### References

[1] Lindner N H and Rudolph T 2009 Proposal for pulsed on-demand sources of photonic cluster state strings Phys. Rev. Lett. 103 113602
[2] Scerri D, Malein R N E, Gerardot B D and Gauger E M 2018 Frequency-encoded linear cluster states with coherent Raman photons Phys. Rev. A 98 022318
[3] Schwartz I, Cogan D, Schmidgall E R, Don Y, Gantz L, Kenneth O, Lindner N H and Gershoni D 2016 Deterministic generation of a cluster state of entangled photons Science 354 (6311) 434–37
[4] Briegel H J and Rausendorf R 2001 Persistent entanglement in arrays of interacting particles Phys. Rev. Lett. 86 910–3
[5] Rausendorf R, Browne D E and Briegel H J 2003 Measurement-based quantum computation on cluster states Phys. Rev. A 68 022312
[6] Barnett S M, Pegg D T and Jeffers J 2000 Bayes’ theorem and quantum retrodiction J. Mod. Opt. 47 1779–89
[7] Barnett S M, Pegg D T, Jefferis J, Jedrkiewicz O and Loudon R 2000 Retrodiction for quantum optical communications Phys. Rev. A 62 022313

![Figure C1](image_url)

Figure C1. (a) Full rank, single qubit reconstruction using gradient descent, averaged over 1000 trials. The random phases were sampled from an exponential distribution with \( \mu = \pi/8 \). As expected the coarse grained approach returns slightly higher fidelities. (b) Algorithm running times for the unsorted, and coarse grained approaches. The unsorted approach scales linearly with number of measurement repetitions. The coarse grained approaches, within the standard deviation, do not scale with increased repetitions as the number of projective operators used for reconstruction is the same for all repetition numbers.

| Protocol | \( \omega/2\pi \) (GHz) | \( 1/\gamma \) (ns) | \( p(\theta) \) | \( \mu \) |
|----------|------------------|--------------|-------|-------|
| Barrett and Kok \[36\] | \( 3 \times 10^5 \) | 0.100 | QU | \( \rightarrow \infty \) |
| Lim et al \[37\] | \( 3 \times 10^5 \) | 0.100 | QU | \( \rightarrow \infty \) |
| Lindner and Rudolph \[1\] | 0.105 | 0.100 | E | 0.011 |
| Schwartz et al \[3\] | 0.200 | 0.330 | E | 0.066 |
| Denning et al \[38\] | 0.955 | 0.167 | E | 0.159 |
| Scerri et al \[2\] | 0.420 | 500 | QU | \( \rightarrow \infty \) |

Table D1. Table showing the phase distributions for various state-of-the-art protocols, where QU and E stand for ‘quasi-uniform’ and ‘exponential’, respectively. The mean varies from protocol to protocol, with some schemes having quasi-uniform phase distributions. For the latter, we expect retrodiction to significantly improve the reconstruction fidelity.
[8] Silva J L E, Glancy S and Vasconcelos H M 2018 Quadrature histograms in maximum-likelihood quantum state tomography Phys. Rev. A 98 022325

[9] Jones K R W 1991 Principles of quantum inference Ann. Phys., NY 207 170

[10] Jones K R W 1991 Quantum limits to information about states for finite dimensional hilbert space Journal of Physics A: Mathematical and General 24 121

[11] Slater P B 1995 Quantum coin-tossing in a Bayesian Jeffreys framework Phys. Lett. A 206 72

[12] Derka R, Bužek V, Adam G and Knight P L 1997 From quantum Bayesian inference to quantum tomography arXiv:quant-ph/9701029

[13] 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 266 496

[14] Schack R, Brun T A and Caves C M 2001 Quantum Bayes rule Phys. Rev. A 64 014305

[15] Blume-Kohout R 2010 Optimal, reliable estimation of quantum states New J. Phys. 12 043034

[16] Granade C, Combes J and Cory D G 2016 Practical Bayesian tomography New J. Phys. 18 033024

[17] Struchalin G I, Po gor elov I A, Straupe S S, Kravtsov K S, Radchenko I V and Kulik S P 2016 Experimental adaptive quantum tomography of two-qubit states Phys. Rev. A 93 012103

[18] Blume-Kohout R and Hayden P 2006 Accurate quantum state estimation via Keeping the experimentalist honest arXiv:quant-ph/0603116

[19] Granade C E, Ferrie C, Wiebe N and Cory D G 2012 Robust online Hamiltonian learning New J. Phys. 14 103013

[20] Kvetn R and Ferrie C 2015 Near-optimal quantum tomography: estimators and bounds New J. Phys. 17 123013

[21] Ferrie C 2014 High posterior density ellipsoids of quantum states New J. Phys. 16 023006

[22] Huszár F and Houlsby N M T 2012 Adaptive Bayesian quantum tomography Phys. Rev. A 85 052120

[23] Liu J S and Chen R 1998 Sequential monte carlo methods for dynamic systems J. Am. Stat. Assoc. 93 1032–44

[24] Liu J and West M 2001 Combined parameter and state estimation in simulation-based filtering Sequential Monte Carlo Methods in Practice ed A Doucet, N De Freitas and N J Gordon (New York: Springer)

[25] Johansson J R, Nation P D and Nori F 2012 QuTiP: an open-source Python framework for the dynamics of open quantum systems Comput. Phys. Commun. 183 1772

[26] 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 1240

[27] Kaznady M S and James D F V 2009 Numerical strategies for quantum tomography: alternatives to full optimization Phys. Rev. A 79 022109

[28] Teo Y S 2016 Introduction to Quantum-State Estimation (Singapore: World Scientific Publishing Co)

[29] Knee G C, Bolduc E, Leach J and Gauger E M 2018 Maximum-likelihood quantum process tomography via projected gradient descent Physical Review A 98 (6) 062336

[30] Gonçalves D S, Gomes–Ruggiero M A and Lavor C 2016 A projected gradient method for optimization over density matrices Optimization Methods and Software 31 328–41

[31] Bolduc E, Knee G C, Gauger E M and Leach J 2017 Projected gradient descent algorithms for quantum state tomography Npj Quantum Information 3 44

[32] James D F V, Kwiat P G, Munro W J and White A G 2001 Measurement of qubits Phys. Rev. A 64 052312

[33] Altepeter J B, Jeffrey E R and Kwiat P G 2005 Photonic state tomography Advances In Atomic, Molecular, and Optical Physics Vol 52 (New York: Academic) pp 105–9

[34] Gonçalves D S, Gomes–Ruggiero M A, Lavor C, Jiménez Farias O and Souto Ribeiro P H 2011 Local solutions of maximum likelihood estimation in quantum state tomography Quantum Information & Computation 12 775–90

[35] Miranowicz A, Bartkiewicz K, Peñina J, Koashi M, Imoto N and Nori F 2014 Optimal two-qubit tomography based on local and global measurements: maximal robustness against errors as described by condition numbers Phys. Rev. A 90 062123

[36] Barrett S D and Kok P 2005 Efficient high-fidelity quantum computation using matter qubits and linear optics Phys. Rev. A 71 060310

[37] Lim Y L, Beige A and Kwok L C 2005 Repeat–until-success linear optics distributed quantum computing Phys. Rev. Lett. 95 030505

[38] Denning E V, Iles-Smith J, McCutcheon D P S and Mork J 2017 Protocol for generating multiphoton entangled states from quantum dots in the presence of nuclear spin fluctuations Phys. Rev. A 96 062329