Quantum Process Tomography (QPT) refers to the use of measured data to estimate the dynamics of a quantum system. Unfortunately, in the general case, the dimension of the parameter space for QPT can be prohibitive, scaling exponentially with the number of qubits. This in turn places the parameter space for QPT can be prohibitive, scaling exponentially with the number of qubits. This in turn places the parameter space for QPT can be prohibitive, scaling exponentially with the number of qubits. This in turn places the parameter space for QPT can be prohibitive, scaling exponentially with the number of qubits.

Quantum process tomography (QPT) refers to the use of measured data to estimate the dynamics of a quantum system. Unfortunately, in the general case, the dimension of the parameter space for QPT can be prohibitive, scaling exponentially with the number of qubits. This in turn places the parameter space for QPT can be prohibitive, scaling exponentially with the number of qubits. This in turn places the parameter space for QPT can be prohibitive, scaling exponentially with the number of qubits. This in turn places the parameter space for QPT can be prohibitive, scaling exponentially with the number of qubits.

For an initially well designed but imperfect quantum information system, the process matrix is almost sparse in an appropriate basis. Existing theory and associated computational methods ($\ell_1$-norm minimization) for reconstructing sparse signals establish conditions under which the sparse signal can be perfectly reconstructed from a very limited number of measurements (resources). Although a direct extension to quantum process tomography of the $\ell_1$-norm minimization theory has not yet emerged, the numerical examples presented here, which apply $\ell_1$-norm minimization to quantum process tomography, show a significant reduction in resources to achieve a desired estimation accuracy over existing methods.

Quantum process tomography (QPT) refers to the use of measured data to estimate the dynamics of a quantum system. Unfortunately, in the general case, the dimension of the parameter space for QPT can be prohibitive, scaling exponentially with the number of qubits. This in turn places the parameter space for QPT can be prohibitive, scaling exponentially with the number of qubits. This in turn places the parameter space for QPT can be prohibitive, scaling exponentially with the number of qubits. This in turn places the parameter space for QPT can be prohibitive, scaling exponentially with the number of qubits.

The paper is organized as follows: QPT formalism is described next, followed by a discussion of the genesis of process matrix (almost) sparsity. A form of the $\ell_1$ minimization for QPT is then presented followed by numerical examples and some concluding remarks.

Quantum Process Tomography via $\ell_1$-norm Minimization

$\ell_1$-norm minimization – often referred to as Compressive Sensing – is applicable to a class of incomplete linear measurement equations ($y = Ax$, $A \in \mathbb{R}^{m \times n}$, $m \ll n$), constrained $\ell_1$-norm minimization (minimize $\|x\|_{\ell_1}$ subject to $y = Ax$), a convex optimization problem, can perfectly estimate the sparse variable $x$. These methods also work very well for systems which do not satisfy the theoretical conditions, i.e., for almost sparse variables and with measurement noise.

The underlying theory of $\ell_1$-minimization shows that under certain conditions on the matrix $A$, to realize perfect recovery, the number of measurements, $m$, scales with the product of the log of the number of variables $n$ and the sparsity. Since QPT parameters are linear in probability outcomes, and scale exponentially with the number of qubits, this approach heralds a possible linear scaling with qubits. The theory, however, has not as yet been extended to QPT.

The numerical examples here are not meant to lend support to this scaling as they are only presented for the two-qubit case. The examples do, however, show more than an order of magnitude savings in resources over a standard constrained least-squares estimation using a complete set of measurements, i.e., $\text{rank}(A) \geq n$.

The numerical examples here are not meant to lend support to this scaling as they are only presented for the two-qubit case. The examples do, however, show more than an order of magnitude savings in resources over a standard constrained least-squares estimation using a complete set of measurements, i.e., $\text{rank}(A) \geq n$.

For a two-qubit system, the process matrix is

$$\hat{\rho} = \sum_{\alpha,\beta} X_{\alpha\beta} \Gamma_{\alpha} \rho \Gamma_{\beta}^\dagger$$

where $\rho$, $\hat{\rho} \in \mathbb{C}^{n \times n}$ are the input and output state, respectively, of dimension $n$, $X_{\alpha\beta}$ are the elements of the $n^2 \times n^2$ process matrix $X$, and the matrices $\Gamma_{\alpha}$ form an orthonormal basis set for $n \times n$ complex matrices:

$$\{ \Gamma_{\alpha} \in \mathbb{C}^{n \times n} \mid \text{Tr} \Gamma_{\alpha}^2 = \delta_{\alpha\beta}, \alpha, \beta = 1, \ldots, n^2 \}$$

It is assumed that the quantum system to be estimated is completely positive and trace preserving (CPTP). The set of feasible process matrices is then restricted to the convex set $\mathcal{P} = \{ X \geq 0 \} (\text{positive semidefinite})$

$$\sum_{\alpha,\beta = 1}^{n^2} X_{\alpha\beta} \Gamma_{\alpha} \Gamma_{\beta}^\dagger = I_n$$

It follows from the condition that the number of real parameters in the process matrix is $n^4 - n^2$. For $q$ qubits $n = 2^q$, hence, scaling with parameters is exponential in the number of qubits.

Collecting data.— A common method for collecting data from a quantum system is via repeated identical experiments. Denote by $i = 1, \ldots, n_{\text{out}}$ the distinct outcomes, and by $k = 1, \ldots, n_{\text{cfg}}$ the experimental configurations, e.g., any “knobs” associated with state inputs and/or measurement devices. The measurement outcomes are recorded from identical experiments in each configuration $k$ repeated $N_k$ times. Let $N_{\text{tot}}$ denote the number of times out of $N_{\text{tot}}$ that outcome $i$ occurred in configuration $k$. The QPT data are the recorded outcome counts,

$$\{ N_{ik} \mid i = 1, \ldots, n_{\text{out}}, k = 1, \ldots, n_{\text{cfg}} \}$$

where $N = \sum_{k=1}^{n_{\text{cfg}}} N_k = \sum_{k=1}^{n_{\text{cfg}}} \sum_{i=1}^{n_{\text{out}}} N_{ik}$ is the total number of experiments.

Estimating the process matrix.— An empirical estimate of the probability of measuring outcome $i$ in configuration $k$ can be obtained from the empirical estimate as,

$$p_{ik}^{\text{emp}} = N_{ik} / N_k$$
From the Born Rule the model probability of outcome \( i \) given configuration \( k \) with observable \( M_{ik} \) is, \( p_{ik} = \text{Tr} \ M_{ik} \rho_k \), where from (1), \( \rho_k = \sum_{\alpha,\beta} \gamma_{\alpha\beta} \rho_{\alpha} \Gamma_{\beta} \). In terms of the process matrix \( X \), the Born rule then becomes,

\[
p_{ik}(X) = \text{Tr} \ G_{ik} X \quad (G_{ik})_{\alpha\beta} = \text{Tr} \ \Gamma_{\beta} M_{ik} \Gamma_{\alpha} \rho_k
\]  

(6)

The \( n_{\text{out}} n_{\text{cfg}} \) matrices \( G_{ik} \in \mathbb{C}^{n \times n} \) capture the effect of measurements in the matrix basis set \( \{ \} \). For each outcome \( i \), the complete set of configurations is the combination of all these matrices and the input states: \( \{ \rho_k, G_{ik} \} \).

A process matrix estimate can be obtained by minimizing the difference between the empirical probability estimates \( p_{ik}^{\text{emp}} \) and the model probabilities \( p_{ik}(X) \) subject to the feasibility constraint (3). Using a “least-squares” measure of probability error leads to estimating the process matrix by solving the optimization problem:

\[
\text{minimize} \quad V_{LS}(X) = \sum_{i,k} (p_{ik}^{\text{emp}} - p_{ik}(X))^2 \\
\text{subject to} \quad X \text{ satisfies (3)}
\]  

(7)

Because the outcomes of each experiment are independent, a maximum likelihood approach can also be considered, i.e.,

\[
\text{minimize} \quad V_{ML}(X) = -\sum_{i,k} N_{ik} \log p_{ik}(X) \\
\text{subject to} \quad X \text{ satisfies (3)}
\]  

(8)

Both (7) and (8) are convex optimization problems with the optimization variables being the elements of \( X \). The resulting solution (estimate) will always be CPTP (3). Unfortunately, as already mentioned, the dimension of the parameter space \( (n^4 - n^2, n = 2^q) \) can severely strain resources to the point of impracticality. To see this more clearly, let the linear relation in (6) between the \( n_{\text{out}} n_{\text{cfg}} \) model probability outcomes and the \( n^2 \) elements of the process matrix be represented by an \( n_{\text{out}} n_{\text{cfg}} \times n^2 \) matrix \( G \), i.e.,

\[
\tilde{p} = G \tilde{X}
\]  

(9)

where \( \tilde{p}, \tilde{X} \) are vectors formed from the \( p_{ik} \) and elements of \( X \), respectively. Accounting for the \( n^2 \) linear constraints in (9), \( X \) can be recovered from either (7) or (8) to within any desired accuracy by using enough data (\( N \) in (4) sufficiently large), provided that rank(\( G \)) \( \geq n_{\text{out}} n_{\text{cfg}} \geq n^2 - n^2 \). Therefore it would seem that the resources, \( n_{\text{out}} n_{\text{cfg}} \), must also scale exponentially with the number of qubits. This, however, is not the case when the process matrix is almost sparse and where the sparsity pattern is not known (17).

**Almost sparsity of the process matrix.**— With no noise the ideal channel \( \rho \rightarrow \hat{\rho} \) for a quantum information system is a unitary, i.e., \( \hat{\rho} = U \rho U^\dagger \). Let \( \{ \Gamma_\alpha \in \mathbb{C}^{n \times n} \}_{\alpha=1}^{n^2} \) denote the “Natural-Basis” for matrices in \( \mathbb{C}^{n \times n} \), i.e., each basis matrix has a single non-zero element of one. In this basis, the process matrix associated with the ideal unitary channel has the rank-1 form, \( X_{\text{ideal}} = x x^\dagger \) with \( x \in \mathbb{C}^n \), \( x^\dagger x = n \). A singular value decomposition (SVD) gives \( X_{\text{ideal}} = V \text{diag}(n,0,\ldots,0) V^\dagger \) with \( V \in \mathbb{C}^{n \times n} \) a unitary. An equivalent process matrix can be formed from the SVD in what is referred to here as the “Ideal/SVD-Basis,” \( \{ \Gamma_\alpha \}

\[
\Gamma_\alpha \in \mathbb{C}^{n \times n} \}_{\alpha=1}^{n^2} \}. The equivalent process matrix, in this basis, denoted by \( X_{\text{ideal}} \), is maximally sparse with a single non-zero element, specifically, \( (X_{\text{ideal}})_{11} = n \). As will always be the case, the actual channel will be a perturbation of the ideal unitary. If the noise source is small then the process matrix in the nominal basis will be almost sparse.

**Example: Noisy two-qubit memory.**— Consider a system which is ideally a two-qubit quantum memory, thus \( U = I_4, n = 4 \). Suppose the actual system is a perturbation of identity by independent bit-flip errors in each channel occurring with probability \( p_{bf} \). For \( p_{bf} = 0.05 \) and \( p_{bf} = 0.2 \), the respective channel fidelities are about 0.90 and 0.64, which for quantum information processing would need to be discovered by QPT and then corrected for the device to ever work. Referring to Fig.1 in the Natural-Basis, Fig.1(a), the ideal 16 x 16 process matrix has 16 non-zero elements out of 256, all of magnitude one. Using the Ideal/SVD-Basis the corresponding process matrix as shown in Fig.1(b) has a single non-zero element of magnitude \( n = 4 \) — it is clearly maximally sparse. Fig.1(c)-(d) and (e)-(f), respectively, show the effect of the two \( p_{bf} \) levels in the two basis sets. In the Ideal/SVD-basis Fig.1(d) and (f) show that the actual (noisy) process matrices are almost sparse.

**Sparsity minimization.**— A known heuristic for minimizing sparsity without knowing the sparsity pattern, and also a cri-
ing the benefit of using fewer resources, is to minimize the $\ell_1$-norm of the vector of variables $\mathbf{X}$ [5, 6, 9]. For QPT the equivalent $\ell_1$ norm is defined here as the sum of the absolute values of the real and imaginary parts of each element of the process matrix. There are many related approaches to incorporate this norm. For example, an estimate of $X$ can be obtained by solving the following convex optimization problem: [18]

$$\begin{align*}
\text{minimize} \quad & \|X\|_{\ell_1} = \sum_{\alpha,\beta} \left| \text{Re} X_{\alpha\beta} \right| + \left| \text{Im} X_{\alpha\beta} \right|
\text{subject to} & \quad V(X) \leq \sigma, X \text{ satisfies (3)}
\end{align*}$$

(10)

with, e.g., $V(X)$ from (7) or (8). The optimization parameter $\sigma$ is used to regulate the tradeoff between fitting $X$ to the data by minimizing $V(X)$ vs. minimizing the sparsity of $X$ via the $\ell_1$-norm. Selecting $\sigma$ is often done by averaging $V(X)$ over a series of surrogates for $X$ obtained from anticipated scenarios or iterating estimation and experiment design, e.g., [8].

In the examples to follow we use the modification of (10) suggested in [7], referred to there as “$\ell_1$-reweighted minimization.” In this approach a weighted $\ell_1$-norm is used with the weights determined iteratively. The algorithm described in [7] is:

**Initialize** $\sigma > 0, \varepsilon > 0, W = I_{n^4}$

**Repeat**

1. **Solve for $X$**
   
   $$\begin{align*}
   \text{minimize} \quad & \|WX\|_{\ell_1} \quad \text{subject to} \quad V(X) \leq \sigma, X \text{ satisfies (3)}
   
   \end{align*}$$

   (11)

2. **Update weights**

   $$W = \text{diag} \left( 1/(|x_1| + \varepsilon), \ldots, 1/(|x_{n^4}| + \varepsilon) \right)$$

   $$X = \tilde{X}$$

   (12)

**Until convergence** – the objective stops decreasing or a maximum number of iterations is reached.

In each of the examples to follow the procedure for QPT is:
(i) solve (7) to obtain $X_{\ell_2}$; (ii) set $\sigma = 1.3 V(X_{\ell_2})$; (iii) solve the reweighting algorithm (11)-(12) for $X_{\ell_1}$.

**Example:** QPT of noisy two-qubit memory.— For the systems from the example in Fig 1 the inputs and measurements are selected from the set of two-qubit states: $|0\rangle, |+\rangle = (|a\rangle + |b\rangle)/\sqrt{2}, |-\rangle = (|a\rangle - i|b\rangle)/\sqrt{2}$ with $a, b = 1, \ldots, 16$.

Specifically, the available set of states are the 16 columns of the matrices,

$$\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}, \quad \tilde{X} = \begin{bmatrix}
1 & 1 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 1 & 1
\end{bmatrix}, \quad \begin{bmatrix}
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0
\end{bmatrix}
$$

(13)

Considering only coincident input/measurement counts [10], the relevant probability outcomes [5] are,

$$\begin{align*}
p_{ab}(X) &= \phi_{ab}^\dagger X_{g_{ab}}, \quad X \in \mathbb{C}^{16 \times 16} \\
\langle g_{ab} \rangle_{a} &= \phi_{a}^\dagger \Gamma_{a} \phi_{b}, \quad a = 1, \ldots, 16
\end{align*}$$

(14)

with $\phi_{ab}$ $\phi_{b} (a, b) \in \{1, \ldots, 16\}$ the selected columns of (13).

Fig 2 shows the error in estimating the process matrix $\Delta X = X_{\text{true}} - X_{\text{est}}$ as measured by the RMS matrix norm $||\Delta X||_{\text{rms}} = (1/n) \text{Tr} \Delta X^\dagger \Delta X)^{1/2}$ vs. the number of experiments per input selected from the set (13) [19]. The results shown are from simulations described in the caption.

The benefit of $\ell_1$-minimization compared to the standard $\ell_2$-minimization is seen most clearly with small amounts of data from highly incomplete measurements. For example, for $p_{\text{rel}} = 0.05$ [Fig 2(a)], at $50 \times 10^3$ experiments per input for the 6-input/6-output configuration ($G \in \mathbb{C}^{256 \times 256}$) the $\ell_1$-minimization is 0.0019. Compare this to the $\ell_2$ error of 0.0012 at $500 \times 10^3$ experiments per input for the 16-input/16-output configuration ($G \in \mathbb{C}^{256 \times 256}$). The latter improvement can be attributed mostly to the 10-fold increase in the number of experiments per input. The additional resources to achieve this are significant, i.e., 16 inputs for $\ell_2$ vs. 6 for $\ell_1$, and additionally, an increase in the total number of experiments from $6 \times 50 \times 10^3$ to $16 \times 500 \times 10^3$. It is certainly not intuitive that to estimate the 240 parameters of the process matrix, the clearly incomplete set of measurements using only 36 outcomes (○ in Fig 2) could produce results not only similar
to, but for each number of experiments per input, even better than the full input case with all 256 combinations of inputs and measurements (□ in Fig.2). As seen the $\ell_1$ error is about 1/2 the $\ell_2$ error. Also, reweighting reduced the (unweighted) $\ell_1$ error by 1/2-1/3.

Comparing the estimation errors with the error between the actual and ideal (solid lines in Fig.2) suggests that at least $50 \times 10^3$ experiments per input are needed to achieve a sufficient post-QPT error correction towards the ideal unitary. Fig.2 also reveals that the estimation errors are very similar for both levels of bit-flip error, $p_{bf} \in \{0.05, 0.20\}$. This is explained by the Cramér-Rao bound which defines the asymptotic error of any unbiased estimator, i.e., the RMS decays as $\Delta/\sqrt{N}$. Here $\Delta$ is effectively the error between the empirical and actual probabilities which by definition is of order one; this provides a reasonable fit to the data in Fig.2.

Infinite data.— With infinite data the measurements are effectively noise-free, so the empirical probability estimates are equivalent to the true probabilities. Infinite data estimates are obtained by solving (7) and (11)-(12) with the constraint $V(X) \leq \sigma$ replaced by the linear equality constraint $p_{ik}(X) = p_{ik}(X_{\text{true}})$. For the numerical examples here, (14) gives the linear equality $g_{ab}(X - X_{\text{true}})g_{ab} = 0$.

In the examples, both $X_{\ell_1}$ from (11)-(12) and $X_{\ell_2}$ from (7) were numerically equal to $X_{\text{true}}$. This is to be expected for $X_{\ell_2}$ because of the complete set of 256 full rank measurements. Almost sparsity makes perfect estimation possible with the highly incomplete set of 36 measurements.

The infinite data case is useful for evaluating different configuration strategies in simulation, i.e., consider only those that result in a good estimate.

To stress the efficacy of $\ell_1$-minimization as a heuristic for sparsity, consider replacing the $\ell_1$ norm in (11)-(12) with the RMS norm $\|X\|_{\text{rms}}$, which is effectively the $\ell_2$ norm of $X$. Solving the 6-input/6-output case (□ in Fig.2) for $p_{bf} = 0.05$ with infinite data gives an RMS error of 0.11, which is considerably larger than the error between the actual and ideal of 0.03 (solid line in Fig.2(a)). The estimate gets even worse with finite data. This again emphasizes the advantage of $\ell_1$ minimization for sparse signal reconstruction [5, 6].

Conclusions.— The use of the $\ell_1$-norm minimization methods of Compressive Sensing [5, 6, 7] appear to apply equally well to sparse QPT. The examples of sparse process matrices presented here are meant to represent typical initial imperfect, designs. The numerical results illustrate how estimation resource tradeoffs can be obtained. Additionally, the findings suggest that QPT resources need not scale exponentially with qubits. In the ideal case, the theoretical question of showing linear scaling with sparsity using $\ell_1$ minimization for QPT remains open.

Because $\ell_1$ minimization uses considerably fewer resources than standard QPT, use in an on-line setting combined with optimal quantum error correction tuned to the specific QPT errors is compelling, e.g., (11, 12, 13). Another future direction is in conjunction with Hamiltonian parameter estimation. Here a bank of estimators can be applied to the data where each estimator is tuned via the Ideal/SVD-Basis to one of a number of finite samples of the unknown parameters. Such an approach may prove useful for a small number of parameters. In quantum metrology often a single uncertain parameter is to be estimated in an unknown noisy environment, e.g., (14, 15).

Acknowledgments.— Thanks to A. Gilchrist, I. Walmsley, D. Lidar, H. Rabitz, and M. Mohseni for suggestions and comments. The idea of applying $\ell_1$ minimization to QPT arose during discussions at [20].

[1] M. A. Nielsen and I. L. Chuang, Quantum Computation and Quantum Information (Cambridge University Press, Cambridge, UK, 2000).
[2] G. M. D’Ariano and P. L. Presti, Phys. Rev. Lett. 86, 4195 (2001).
[3] M. Mohseni, A. T. Rezakhani, and D. A. Lidar, Phys. Rev. A 77, 032322 (2008).
[4] J. Emerson et al., Science 317, 1893 (2007).
[5] D. Donoho, IEEE Trans. Inform. Theory 52, (2006).
[6] E. J. Candès, J. Romberg, and T. Tao, Comm. Pure Appl. Math. 59, 1207 (2006).
[7] E. J. Candès, M. B. Wakin, and S. Boyd, J. Fourier Anal. and Appl. 14, 877 (2008).
[8] R. L. Kosut, I. A. Walmsley, and H. Rabitz, quant-ph/0411093 (2004).
[9] S. Boyd and L. Vandenberghe, Convex Optimization (Cambridge University Press, Cambridge, UK, 2004).
[10] J. L. O’Brien et al., Phys. Rev. Lett. 93, 080502 (2004).
[11] M. Reimpell and R. F. Werner, Phys. Rev. Lett. 94, 080501 (2005).
[12] A. S. Fletcher, P. W. Shor, and M. Z. Win, Phys. Rev. A 75, 012338 (2007), quant-ph/0606035.
[13] R. L. Kosut, A. Shabani, and D. A. Lidar, Phys. Rev. Lett. 100, 020502 (2008), (arXiv:quant-ph/0703274).
[14] V. Giovannetti, S. Lloyd, and L. Maccone, Phys. Rev. Lett. 96, 010401 (2006).
[15] U. Dorner et al., quant-ph/0807.3659 (2008).
[16] M. Mohseni and A. T. Rezakhani, quant-ph/arXiv:0805.3188 (2008).
[17] A known sparsity pattern can arise from the underlying dynamics, thereby inherently increasing QPT efficiency [16].
[18] There are many alternatives to [10]. e.g., minimize $\|X\|_1 + \lambda V(X)$ subject to $X$ satisfies [2], or minimize $V(X)$ subject to $\|X\|_\ell_2 \leq s$, $X$ satisfies [3].
[19] The number of experiments per input/measurement configuration here is chosen uniformly. An optimal (non-uniform) choice which minimizes the Cramér-Rao lower bound can be cast as a convex optimization problem [8].
[20] Workshop on Quantum Estimation: Theory and Practice, Aug. 25-30, 2008, Perimeter Institute, Waterloo, Canada.