

Quantum Science and Technology

PAPER • OPEN ACCESS

Experimentally exploring compressed sensing quantum tomography

To cite this article: A Steffens *et al* 2017 *Quantum Sci. Technol.* **2** 025005

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

Related content

- [Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators](#)
Steven T Flammia, David Gross, Yi-Kai Liu *et al.*
- [Quantum tomographic reconstruction with error bars: a Kalman filter approach](#)
Koenraad M R Audenaert and Stefan Scheel
- [Errors in quantum tomography: diagnosing systematic versus statistical errors](#)
Nathan K Langford

Recent citations

- [Wigner tomography of multispin quantum states](#)
David Leiner *et al*
- [Extracting the physical sector of quantum states](#)
D Mogilevtsev *et al*



IOP | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

Quantum Science and Technology



PAPER

Experimentally exploring compressed sensing quantum tomography

OPEN ACCESS

RECEIVED

5 December 2016

REVISED

15 March 2017

ACCEPTED FOR PUBLICATION

3 April 2017

PUBLISHED

11 May 2017

Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/).

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



A Steffens^{1,5}, C A Riofrío¹, W McCutcheon², I Roth¹, B A Bell³, A McMillan², M S Tame⁴, J G Rarity² and J Eisert¹

¹ Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, D-14195 Berlin, Germany

² Quantum Engineering Technology Laboratory, Department of Electrical and Electronic Engineering, University of Bristol, Merchant Venturers Building, Woodland Road, Bristol, BS8 1UB, United Kingdom

³ Faculty of Science, University of Sydney NSW 2006, Australia

⁴ School of Chemistry and Physics, University of KwaZulu-Natal, Durban 4001, South Africa

⁵ Author to whom any correspondence should be addressed.

E-mail: adrian.steffens@fu-berlin.de

Keywords: quantum tomography, compressed sensing, quantum photonics

Abstract

In the light of the progress in quantum technologies, the task of verifying the correct functioning of processes and obtaining accurate tomographic information about quantum states becomes increasingly important. Compressed sensing, a machinery derived from the theory of signal processing, has emerged as a feasible tool to perform robust and significantly more resource-economical quantum state tomography for intermediate-sized quantum systems. In this work, we provide a comprehensive analysis of compressed sensing tomography in the regime in which tomographically complete data is available with reliable statistics from experimental observations of a multi-mode photonic architecture. Due to the fact that the data is known with high statistical significance, we are in a position to systematically explore the quality of reconstruction depending on the number of employed measurement settings, randomly selected from the complete set of data, and on different model assumptions. We present and test a complete prescription to perform efficient compressed sensing and are able to reliably use notions of model selection and cross validation to account for experimental imperfections and finite counting statistics. Thus, we establish compressed sensing as an effective tool for quantum state tomography, specifically suited for photonic systems.

Introduction

Quantum technologies have seen an enormous progress in recent years. Photonic architectures have matured from basic proof-of-principle schemes to intermediate scale quantum devices [1], while the robustness offered by integrated optical devices is poised to push these systems yet further [2, 3]. Similarly, systems of two-digit trapped ions [4] and other condensed-matter type systems such as superconducting devices are catching up at a remarkable pace [5]. Building upon this technological development, important primitives of quantum information science are being experimentally realised [6–10]. In light of these systems, it has become increasingly important to establish a toolbox for tomographic reconstruction that can keep up with this rapid development: The ironic situation that is emerging is that by now the state of large quantum systems can be manipulated with a high degree of control, but not easily reconstructed. Clearly, these technologies and the community require further advancement of their tools for state reconstruction. In this work, we discuss an explicit method to achieve such a reconstruction, thus contributing to this long-term goal. Specifically, we demonstrate a comprehensive exploration of the performance of state reconstruction in the photonic setting as one varies both the number of measurements and the noise model.

The framework of compressed sensing, a set of techniques originating from the context of classical signal processing [11, 12], has emerged as a key protagonist in closing the gap between technology and diagnostics [13–15]. The idea behind its functioning is rooted in the fact that a substantial amount of data encountered in realistic situations are structured and can be characterised by significantly fewer parameters than with ad hoc schemes. Approximately low-rank matrices are at the centre of the paradigm of matrix completion in

compressed sensing and correspond precisely to approximately pure quantum states. Since pure quantum states are widely regarded as the key resource for quantum information processing, such methods for reconstructing low-rank states are especially relevant. For even larger systems, tomographic tools based on basic variational sets are conceivable, with matrix product states [16, 17], their continuous analogues [18], and permutationally invariant states [19] providing prominent examples. The theory of such novel tools of reconstruction is progressing quickly. This applies, e.g., to new insights to the assignment of fair and rigorous confidence regions [20–23] as well as economical ways of performing instances of quantum process tomography [14, 24, 25].

Exciting steps toward using compressed sensing in experimental settings have been made [19, 24, 26, 27] in the regime in which one assumes knowledge about the basis in which sparsity is expected [24], assumes additional structure [19] or is in the highly tomographically incomplete regime [27]. In this work, we complement the picture for experimental tomography for medium-sized quantum systems. In its simplest formulation, compressed sensing tomography is based on a few random expectation values of suitable observables, from which approximately low-rank states can be accurately reconstructed [13]. This is suited for the situation in which expectation values can be obtained with good statistical significance, although acquiring many of them may be expensive. Still a missing piece in this picture, however, is the exploration of model selection techniques that have to be considered in the realm of experimental imperfections and finite counting statistics in order to make compressed sensing tomography a practical tool. Model selection allows to prevent over- and underfitting by controlling the dimensionality of the model of the system—in our case, the rank of the density matrix.

Here, we present a comprehensive analysis of experimental data from a multi-photon, multi-mode GHZ state source using tools of compressed sensing. Instead of working with expectation values of observables—as it is commonly done in this context, but may amount to information loss—our experimental setup allows us to obtain information on the individual projector level from the respective outcomes of each measurement setting. In contrast to complementing recent work [27], we are not tied to the regime of tomographically incomplete knowledge. This allows us to study the behaviour of the reconstruction for the entire range of measurement settings. We quantitatively explore model selection via cross validation and compare it to the model suggested by the anticipated noise statistics. With these tools, we provide a more systematic way to choose the appropriate parameters for compressed sensing quantum tomography. The results then provide the reader with the toolkit and understanding to effectively implement these methods for future quantum state tomography (QST) in general, and specifically for photonic systems.

This work is structured as follows. We start by reviewing concepts of quantum state tomography and discuss the specifics of compressed sensing in QST. We subsequently present our experimental setup consisting of a four-qubit photonic system, which is used as a test bed for our tomographical approach. We continue by discussing concepts of model selection in the context of QST and determine the appropriate model from the experimental data. With this, we perform compressed QST and study the performance of the reconstruction depending on the amount of collected data as well as the robustness of our method with respect to model mismatches.

Elements of quantum state tomography

Quantum state tomography is the most commonly used method to diagnose quantum information processing tasks. It is used to estimate the unknown quantum state of a system from data produced by measuring an ensemble of identically prepared systems. By fixing a basis, a general finite-dimensional quantum state can be identified with a positive semi-definite, unit-trace matrix, the density matrix

$$\rho \in \mathcal{S}_d = \{\chi \in \mathcal{H}_d : \chi \succcurlyeq 0, \text{tr}(\chi) = 1\}. \quad (1)$$

Here, $\mathcal{H}_d \subset \mathbb{C}^{d \times d}$ denotes the set of Hermitian matrices, and $\chi \succcurlyeq 0$ stands for a positive semi-definite matrix.

To determine the density matrix ρ of a quantum system, we need to prepare sufficiently many copies of the state from identical preparations, perform a measurement on each copy using one out of m different measurement settings—corresponding to different observables, i.e. Hermitian matrices $A^{(j)}$, $j = 1, \dots, m$ —and count the respective number of measurement outcomes. Ideal measurements are associated with unit rank projectors $\Pi_k^{(j)} = v_k^{(j)} v_k^{(j)\dagger}$, where $v_k^{(j)}$ is the k th normalised eigenvector of $A^{(j)}$. For each measurement setting j the specific outcome $k = 1, \dots, d$ occurs with probability

$$p_{j,k} := \text{tr}(\Pi_k^{(j)} \rho). \quad (2)$$

Completeness, i.e. the property that the projectors sum up to unity,

$$\sum_{k=1}^d \Pi_k^{(j)} = \mathbb{1}, \quad (3)$$

ensures normalisation for each measurement setting j , so that $\sum_{k=1}^d p_{j,k} = 1$. For each measurement setting j , the outcome k corresponds to a random variable $Y_{j,k}$. Repeated measurements are independent from each other,

and are performed on N_j copies of the state for each measurement setting j , yielding the respective integer-valued realisation $y_{j,k}$ as observed frequency with $\sum_{k=1}^d y_{j,k} = N_j$. Hence, for each measurement setting j , the probability of the random variables $(Y_{j,1}, \dots, Y_{j,d})$ to take the configuration of measurement outcomes $(y_{j,1}, \dots, y_{j,d})$ is given by

$$\frac{N_j!}{y_{j,1}! \cdots y_{j,d}!} p_{j,1}^{y_{j,1}} \cdots p_{j,d}^{y_{j,d}}, \quad (4)$$

following a multinomial distribution $\mathcal{M}(N_j, (p_{j,1}, \dots, p_{j,d}))$. Accordingly, we will obtain the k th outcome $N_j p_{j,k}$ times in expectation. We formalise the measurement process by introducing the linear operator

$$\mathcal{A} : \rho \mapsto (N_j \text{tr}(\Pi_k^{(j)} \rho))_{j,k}, \quad (5)$$

which maps density matrices in \mathcal{S}_d to matrices in $\mathbb{R}_+^{m \times d}$, corresponding to measurement outcomes $k = 1, \dots, d$ for different measurement settings $j = 1, \dots, m$. We emphasise that $\mathcal{A}(\rho)$ is not an experimental data matrix itself; according to the law of large numbers, the frequencies in each measurement realisation $\mathcal{Y} := (y_{j,k}) \in \mathbb{N}^{m \times d}$ from the experiment will converge to $\mathcal{A}(\rho)$ with growing number of measurements N_j , i.e. the expectation value $\mathbb{E}(Y_{j,k})$ of the random variable $Y_{j,k}$ is given by

$$\mathbb{E}(Y_{j,k}) = N_j \text{tr}(\Pi_k^{(j)} \rho) \quad (6)$$

for each j, k . Apart from additional systematic sources of error, e.g. due to experimental imperfections, the difference between \mathcal{Y} and $\mathcal{A}(\rho)$ is due to finite counting statistics, and in many settings, this is the largest contribution to the error.

The most straightforward approach to determine ρ from \mathcal{Y} would be to attempt to invert the linear system of equations

$$\mathcal{A}(\rho) = \mathcal{Y}. \quad (7)$$

In general, however, noise on the data \mathcal{Y} would render the *reconstructed* density matrix $\hat{\rho}$ unphysical ($\hat{\rho} \not\geq 0$). A generic (full rank) density matrix in \mathcal{S}_d is determined by $d^2 - 1$ independent real parameters. Hence, in general, one requires at least $d^2 - 1$ linearly independent equations in order to solve equation (7). This is also called *tomographic completeness* and corresponds to *informational completeness* with sufficient information to in principle capture full rank states. For further notions about informational completeness under prior information (e.g. the rank of the state is assumed to not be maximal) see [28, 29]. When dealing with significantly less information, specialised reconstruction techniques are important with compressed sensing being a natural choice, which we will discuss in the next section.

In our system, we will be concerned with local Pauli measurements on each subsystem of a multi-partite state. We measure an n -qubit system ($d = 2^n$) using m different measurement settings, each of which corresponds to an n -qubit Pauli operator

$$A^{(j)} = \bigotimes_{i=1}^n \sigma_i^{(j)}, \quad (8)$$

$j = 1, \dots, m$, with $\sigma_i^{(j)} \in \{\sigma_x, \sigma_y, \sigma_z\}$, where $\sigma_x, \sigma_y, \sigma_z$ are the Pauli matrices. This is often referred to as *Pauli basis measurement*. The projectors of the two-qubit operator $A^{(1)} := \sigma_z \otimes \sigma_z$, for example, are $\Pi_1^{(1)} = |0, 0\rangle\langle 0, 0|$, $\Pi_2^{(1)} = |0, 1\rangle\langle 0, 1|$, $\Pi_3^{(1)} = |1, 0\rangle\langle 1, 0|$, and $\Pi_4^{(1)} = |1, 1\rangle\langle 1, 1|$. The identity matrix can be excluded for each qubit since it has the same eigenvectors and hence corresponding projectors as σ_z and does not provide any additional information about the state. Note, that in a Pauli basis measurement, one obtains 2^n outcomes per measurement setting, as opposed to the *Pauli expectation value measurements*, in which we only use one expectation value per measurement setting. Pauli expectation value measurements (including those containing identity matrices) can easily be obtained from Pauli basis measurements by simply computing the expectation values from the projection data for each measurement setting. For n qubits, there exist $m_{\max} := 3^n$ different Pauli words in total, each with 2^n eigenvectors, which corresponds to a maximum of $3^n \cdot 2^n$ equations in equation (7). Each set of Pauli projectors $\{\Pi_k^{(j)}\}_{k=1}^d$ for fixed setting j' contains a subset of elements that is linearly independent from the projectors for all other settings. Hence, any number of smaller than m_{\max} measurement settings will lead to the loss of tomographic completeness. When performing QST on large systems, however, it is of practical necessity to employ as few measurement settings as possible (and often also only few repetitions per measurement setting). The key question arising in this context, therefore, is whether it is feasible to reconstruct an unknown state ρ with not only $m < m_{\max}$ measurement settings, but a significantly smaller subset. The need for minimising the number of measurement settings is particularly pressing in architectures such as linear optical ones, since high repetition rates and good statistics are available, while it can be tedious or costly to alter the measurement setting. This is indeed the case in many practically relevant situations using compressed sensing schemes, which will be discussed in the next section.

Compressed sensing for quantum state tomography

By parameter counting, a state with rank $r < d$ can be completely characterised by fewer than d^2 parameters, that is $\sim rd$ and informational completeness could in principle be achieved using correspondingly fewer measurement settings. However, it is far from obvious how to acquire these parameters using fewer measurement settings and how to do so in a robust fashion—this is the starting point for compressed sensing [11, 30]. Originally conceived for reconstructing sparse vectors, the concept was extended to the recovery of low-rank matrices [31, 32] and adapted to the problem of QST [13, 33]. Here, one again considers structured problems in which one can exploit the fact that in many useful settings approximately low-rank states are of interest. This is a reasonable assumption, since most quantum information experiments aim at preparing pure states.

In order to obtain a general complex-valued low-rank matrix from measurements \mathcal{A} , naively, one would search within the set of low-rank matrices for the one that matches the measurement constraint, solving

$$\min_{\chi \in \mathbb{C}^{d \times d}} \text{rank}(\chi) \quad \text{s.t.} \quad \mathcal{A}(\chi) = \mathcal{Y}. \quad (9)$$

The key idea for compressed sensing in matrix recovery is to relax this NP-hard problem [34] into the closest convex optimisation problem [35]

$$\min_{\chi \in \mathbb{C}^{d \times d}} \|\chi\|_* \quad \text{s.t.} \quad \mathcal{A}(\chi) = \mathcal{Y}. \quad (10)$$

We denote the nuclear norm (better known as the trace norm in the context of reconstructions in quantum mechanics) of a matrix χ by $\|\chi\|_* := \text{tr}(\sqrt{\chi^\dagger \chi})$. Such problems are well known to be efficiently solvable [36].

The crucial question in compressed sensing is how many measurements are required to satisfactorily reconstruct the sought-after matrix. Many proofs rely on randomised measurements schemes: In [37], it has been shown that for a general map $\mathcal{A} : \mathbb{R}^{d \times d} \rightarrow \mathbb{R}^M$ with Gaussian entries, $M \gtrsim 3r(2d - r)$ copies of ϱ are provably sufficient for the recovery of ϱ . Building on this and closer to our situation is the recovery guarantee presented in [38], in which $M \geq crd$ copies are needed with some constant $c > 0$, for $\mathcal{A} : \mathcal{S}_d \rightarrow \mathbb{R}^M$, $\varrho \mapsto (\text{tr}(\Pi^{(j)} \varrho))_{j=1, \dots, M}$, mapping density matrices from \mathcal{S}_d to vectors in \mathbb{R}^M , with $\Pi^{(j)} = \nu^{(j)} \nu^{(j)\dagger}$, and $\nu^{(j)}$ a Gaussian vector for each j . In practice, numerical computations outperform these theoretical bounds. However, there is a fundamental lower bound for the number of copies, $M = 4r(d - r) - 1$, using a theoretically optimal POVM with M elements [39]. Note that—in the mindset of measurement settings and outcomes—the number of outcomes k per measurement setting j scales with the dimension of the Hilbert space d . Since M corresponds to $m d$, the number of measurement settings scales just with the rank, i.e. $m = c r$.

It is in general harder to prove comparable results for deterministic measurements—in our setting with $\nu^{(j)}$ being eigenvectors of Pauli operators. To bridge this gap, notions of partial derandomisation have been introduced, where $\nu^{(j)}$ are not Gaussian, but drawn from spherical designs—certain finite subsets of the d -dimensional complex sphere—leading to similar statements [38]. Spherical designs, in turn, can be related to eigenvectors of n -qubit Pauli operators [40]. Apart from results on the level of expectation values [41], less has been proven for products of single-qubit eigenvectors, the setting at hand—strikingly in contrast to the great success of the procedure in practice. These results remain stable when taking noise into account.

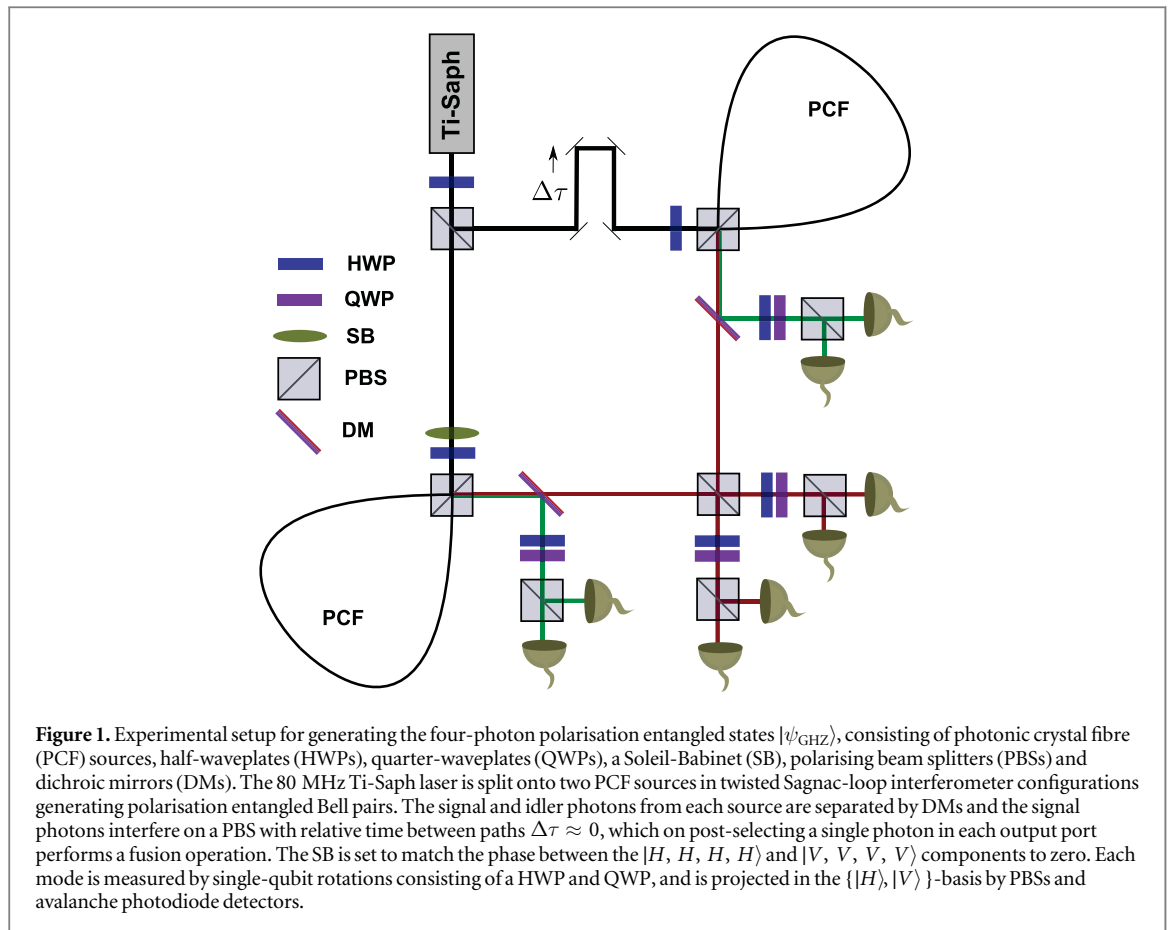
The measured data can be written as

$$\mathcal{Y} = \mathcal{A}(\varrho) + \mathcal{N}(\varrho) = (N_j \text{tr}(\Pi_k^{(j)} \varrho))_{j,k} + (\eta_{j,k})_{j,k}, \quad (11)$$

with \mathcal{N} and $\eta_{j,k}$ representing the noise due to finite counting statistics. For positive semi-definite matrices such as quantum states, the nuclear norm of a matrix reduces to the trace of the matrix. Consequently, relaxing the equality constraint in equation (10) and including the positivity constraint, we arrive at the semi-definite programme (SDP) [34]

$$\min_{\chi \geq 0} \text{tr} \chi \quad \text{s.t.} \quad \|\mathcal{A}(\chi) - \mathcal{Y}\|_2^2 < \varepsilon, \quad (12)$$

for some yet-to-be-determined $\varepsilon > 0$ and $\|\cdot\|_2$ representing the entrywise two-norm. This is exactly the problem we aim to solve in order to achieve efficient QST. SDPs, being convex programmes, feature a rich theory, and numerical implementation is easily achievable [42, 36]. Note that the procedure minimises the trace, which at first sight might seem contradictory to the requirement for density matrices to have unit trace. However, the unit trace requirement is implicitly included in the data constraint since the probabilities in the map \mathcal{A} are normalised. Perfect data would lead to an optimiser with trace exactly equal to one. In turn, a relaxation of this constraint leads to a relaxation of the unit trace requirement. As a result, generically for not too small ε , the optimal χ , denoted by $\hat{\chi}$, will be subnormalised, due to its location on the part of the boundary of the ε -ball with the lowest trace. In order to obtain a physically meaningful reconstruction $\hat{\varrho} \in \mathcal{S}_d$, we find in our simulations that renormalising via



$$\hat{\chi} \mapsto \hat{\varrho} := \frac{\hat{\chi}}{\text{tr}(\hat{\chi})} \quad (13)$$

produces the highest fidelity results. To carry out the optimisation procedure, we employ the convex optimisation solver SDPT3 4.0 [43] together with CVX [44]. For higher Hilbert space dimensions, methods like singular value thresholding [45] come into play, which typically are faster, but less accurate.

Experimental setup

The experiment is designed to prepare the four-qubit GHZ state associated with the state vector

$$|\psi_{\text{GHZ}}\rangle = \frac{1}{\sqrt{2}}(|H, H, H, H\rangle + |V, V, V, V\rangle) \quad (14)$$

with the qubits encoded in the polarisation degree of freedom of four photons. Here, $|H\rangle$ and $|V\rangle$ represent horizontally and vertically polarised photons, respectively, hence effectively spanning a two-dimensional Hilbert space. The experimental setup, building upon the one outlined in [46], is shown in figure 1 and consists of two Bell pair sources which undergo a parity check or post-selected fusion [8, 47–52] to probabilistically generate the GHZ state. Both the photon pairs, generated by spontaneous four-wave-mixing in microstructured fibres, and the fusion operation are successful only probabilistically, but in a heralded fashion, i.e. a classical signal is available signifying success of the preparation. Successful generation of the state is determined by post-selecting only four-photon coincident events which occur at a rate of approximately 1–2 Hz. The post-selected data is effectively free from dark counts—noise generated by single-photon detectors firing erroneously in the absence of a photon. This is due to the fact that the rate at which dark counts in n modes occur in the coincidence window decreases exponentially with n , i.e. four simultaneous dark counts are negligibly rare. Due to additional experimental imperfections, however, the prepared state is non-ideal. The main cause of deviation between the actually prepared state and the target state arises from the distinguishability of photons partaking in the fusion operation and inherent mixedness from the parasitic effects in the pair generation [53]. These tend to cause the generated state to resemble a partially dephased GHZ state [8]. Measurements on the state then proceed using single-qubit rotations (waveplates) and projections (polarising beam splitters and single-photon detection with avalanche photo-diodes) using well-characterised bulk-optical elements allowing high-fidelity measurements to be performed.

As stated above, in order to achieve a tomographically complete basis for n qubits, one requires $m_{\max} = 3^n$ measurement settings. In our system of four qubits, $n = 4$, we have measured a tomographically complete set of 81 local Pauli operators. For each measurement setting, around 650 four-coincident events are accumulated within an integration time of six minutes. Evidently, given the exponential scaling of the tomographically complete set of measurement settings, achieving such reliable statistics for larger states ($n > 4$) is increasingly demanding on resources and quickly becomes infeasible.

Model selection

The starting point for carrying out compressed sensing quantum tomography is the question of determining an appropriate value for ε in the optimisation procedure equation (12). Essentially, larger values of ε result in greater relaxation of the data fitting constraint, leading to lower-rank estimates $\hat{\varrho}$; while smaller ε values will yield $\hat{\varrho}$ matrices with larger rank, which better fit the particular data set. Depending on the underlying state and the particular instance of noise in the data, the choice of ε might result in underfitting with too coarse a model, or in overfitting—i.e. including parts of the noise into the model of the state. Both extremes in general lead to states that fail to correctly predict future data. In the most severe cases, it could happen that using the same measurement prescription \mathcal{A} and the same data \mathcal{Y} , the optimisation procedure in equation (12) yields a full rank or a rank-one matrix, depending on the choice of ε . Worse still, too small a value of ε can make the optimisation procedure unfeasible, whereby there is no feasible state that would result in data sufficiently close to that measured. The task of determining the appropriate model—in our case, the value of ε —that is statistically faithful to the data via an appropriate choice of the respective external parameters is called *model selection* (see e.g. [54]). Several ideas of model selection have a rigorous mathematical underpinning: Particularly well known is the Akaike information criterion (AIC) [55], providing a measure of the relative quality of statistical models for a given set of data. For a collection of models compatible with a given data set, this criterion gives an estimate for the relative quality of each model. Similarly frequently employed is the Bayesian information criterion (BIC) [56]. Direct application of AIC and BIC to quantum tomography—an approach followed in [15]—is problematic for larger systems since it requires rank-restricted maximum likelihood estimation, leading to non-convex optimisation, which scales unfavourably with the system size. This is due to the fact that these techniques are discrete in the sense that they explicitly restrict the rank of the density matrix. In the compressed sensing mindset, the parameter that controls the rank in a continuous fashion is ε . As we mentioned above, this is at the centre of our discussion.

For sufficiently small noise, a promising ansatz for identifying a suitable ε is to use the data to compute the estimate $\hat{\varepsilon}(\mathcal{Y})$ according to the expectation value of

$$\|\mathcal{A}(\chi) - \mathcal{Y}\|_2^2 = \|\mathcal{N}(\varrho)\|_2^2. \quad (15)$$

Assuming the noise is solely due to finite counting statistics, i.e. the deviations from measurement outcomes from the expected variance of the multinomial distribution, we obtain

$$\mathbb{E}(\|\mathcal{N}(\varrho)\|_2^2) = \sum_{j,k} \mathbb{E}(\eta_{j,k}^2) = \sum_{j,k} \mathbb{V}(\eta_{j,k}) = \sum_{j,k} N_j p_{j,k} (1 - p_{j,k}), \quad (16)$$

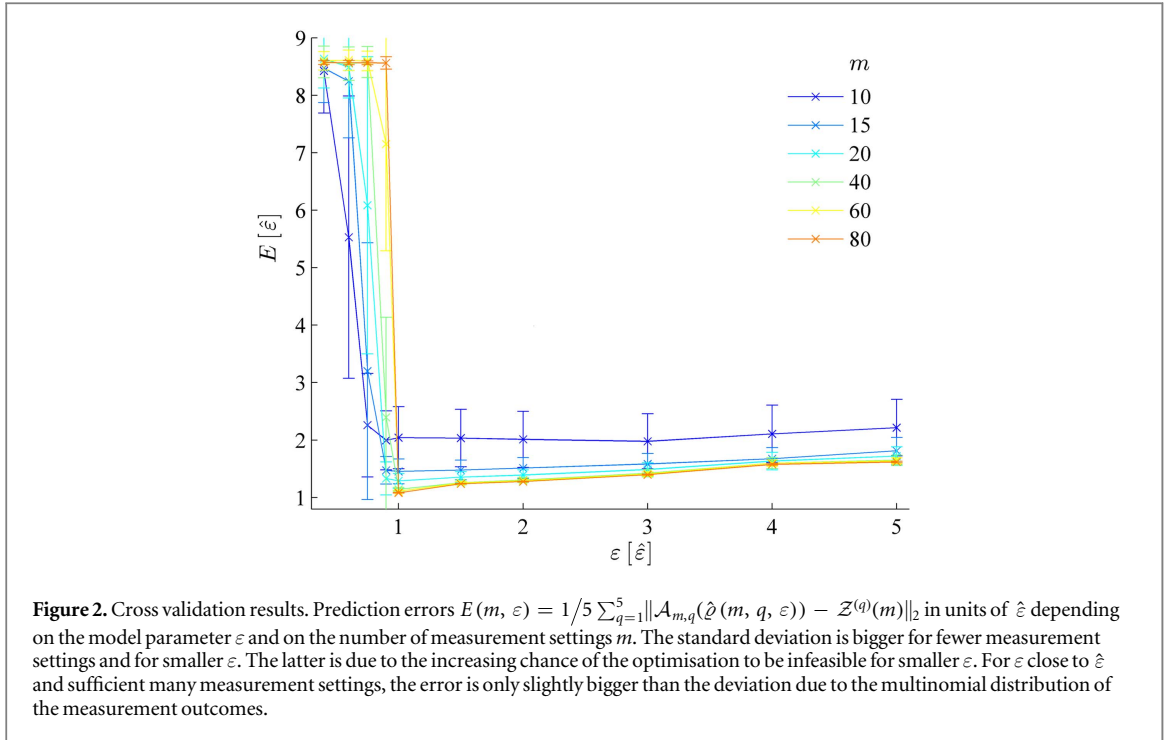
with variance \mathbb{V} . The second step follows from $\mathbb{E}(\eta_{j,k}) = 0$ for each j and k . In order to compute $\hat{\varepsilon}$ from the data, we need to approximate $p_{j,k}$ as $y_{j,k}/N_j$, which is reasonable for sufficiently large N_j according to the law of large numbers. By equation (16), we obtain the estimate

$$\hat{\varepsilon}(\mathcal{Y}) := \sum_{j=1}^m \sum_{k=1}^d y_{j,k} (1 - y_{j,k}/N_j). \quad (17)$$

This choice of $\hat{\varepsilon} = \hat{\varepsilon}(\mathcal{Y})$ scales linearly with m , the number of measurements in the data set \mathcal{Y} . Note that $\hat{\varepsilon}$ depends on the noise model, which in several cases may not be sufficiently established. In our case, however, the noise model is known to a high degree, which allows us to study and compare different methods for estimating the parameter ε .

Complementarily, we employ a straightforward, well-established model selection technique based on cross validation (see e.g. [57]), which is more scalable than the use of AIC or BIC in our case. Because of its generality and independence from the noise model, it can be used in a variety of scenarios. The method works as follows: The data is partitioned into independent training and testing sets. Different models, i.e. different values for ε , are built from the training data and used to predict the testing data. The sought-after parameters—in our case ε —then result from the model corresponding to the smallest error with respect to the testing data.

Specifically, we randomly draw $m = 10, 15, 20, 40, 60, 80$ out of the $m_{\max} = 81$ measurement settings without replacement, corresponding to different levels of limited experimental knowledge. The respective data sets $\mathcal{Y}(m) \in \mathbb{R}_+^{m \times d}$ are then partitioned into five subsets $\mathcal{Z}^{(1)}(m), \dots, \mathcal{Z}^{(5)}(m) \in \mathbb{R}_+^{m/5 \times d}$. The optimisation in equation (12) is performed with respect to every possible union of four subsets $\bigcup_{i=1, i \neq q}^5 \mathcal{Z}^{(i)}(m)$, $q = 1, \dots, 5$,



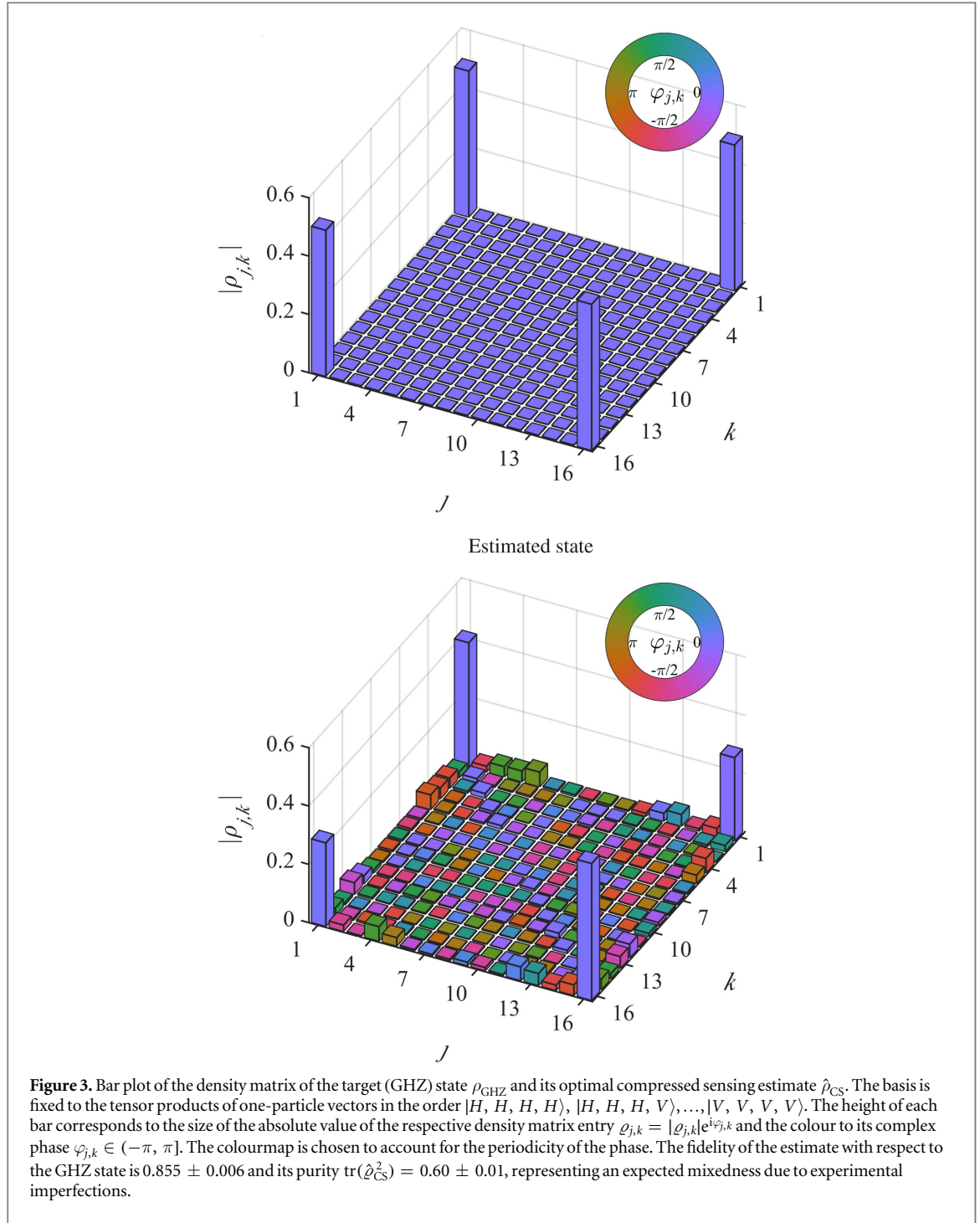
and different ε parameters. Each reconstruction yields an estimate $\hat{\rho}(m, q, \varepsilon)$ and the remaining subset $\mathcal{Z}^{(q)}(m)$ is used as a testing set. The state estimate $\hat{\rho}(m, q, \varepsilon)$ is used to compute the *predicted* measurement data $\mathcal{A}_{m,q}(\hat{\rho}(m, q, \varepsilon))$ and compare these with the corresponding subset of the *experimental* measurement data $\mathcal{Z}^{(q)}(m)$ ($\mathcal{A}_{m,q} : \mathcal{S}_d \rightarrow \mathbb{R}_+^{m/5 \times d}$ being the reduction of the operator \mathcal{A} to the subsets of measurement settings corresponding to m and q). The resulting distance $\|\mathcal{A}_{m,q}(\hat{\rho}(m, q, \varepsilon)) - \mathcal{Z}^{(q)}(m)\|_2$, between the *predicted* and *measured* data, also known as the prediction error or predicted risk, is averaged over q (fivefold cross validation), yielding an estimate for the averaged prediction error (testing set error)

$$E(m, \varepsilon) = \frac{1}{5} \sum_{q=1}^5 \|\mathcal{A}_{m,q}(\hat{\rho}(m, q, \varepsilon)) - \mathcal{Z}^{(q)}(m)\|_2. \quad (18)$$

If the corresponding optimisation problem is infeasible for a certain combination of ε , m , and q (i.e. the set of density matrices that satisfy the constraint in equation (12) is empty), the prediction error is set to $\|\mathcal{Z}^{(q)}(m)\|_2$. For averaging, each point (m, ε) is sampled 50 times.

The mean values and standard deviations of the prediction error depending on the model parameter are depicted in figure 2. We see that for values of ε around $\hat{\varepsilon}$ the error is smallest, which is consistent with our ansatz and allows us to gain confidence in the assumption that the measurement data can be effectively modelled by a multinomial distribution. The more measurement settings are considered, the clearer the choice of the optimal ε becomes, with both the prediction error and its variance attaining their minima close to $\varepsilon = \hat{\varepsilon}$. For those values of ε close to $\hat{\varepsilon}$ and sufficiently many measurement settings, the prediction error $E(m, \varepsilon)$ is only slightly bigger than the error estimate for the data ε . Here, the error arises primarily from raw multinomial noise, ε , present in the testing set itself and cannot be overcome with improved reconstruction methods. Where fewer measurement settings are considered, less information about the state is available, resulting in large testing set errors as well as greater variance of the state estimates, although the smallest prediction errors are still seen for ε close to $\hat{\varepsilon}$. As ε decreases below $\hat{\varepsilon}$, the chance of the optimisation being infeasible increases, causing the prediction errors to effectively increase with a greater spread attributed to different optimisation runs. As ε increases above $\hat{\varepsilon}$, the data fitting constraint is weakened, resulting in too coarse model fits and a gradually increasing prediction error.

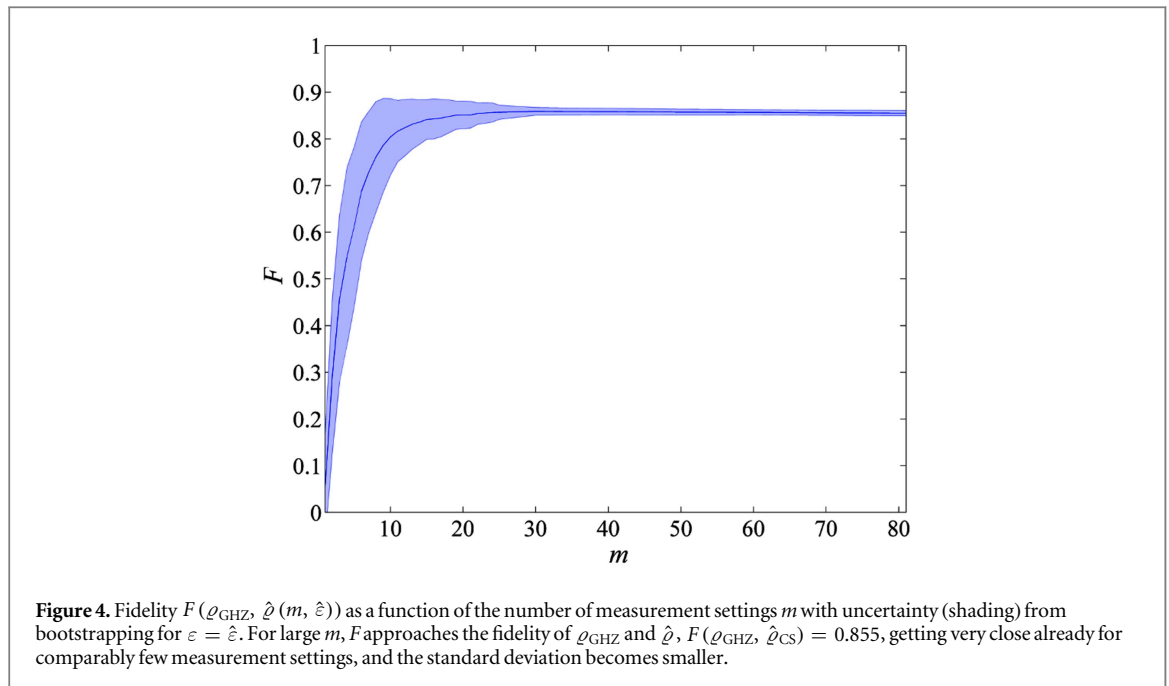
Using equation (17) instead of cross validation has the advantage of much less computational effort and is useful in a scenario with good statistics for each measurement setting. Moreover, cross validation relies on partially discarding data, which could aggravate the issues of having too little data, yielding poorer estimates for ε . However, equation (16) relies on the assumption of a well identified error model—in our case, multinomial noise, as verified by cross validation. In cases in which the error model is not known, cross validation can provide a more robust estimate of ε .



Compressed sensing tomography of the GHZ state

Having verified that the optimal value for ε is close to that computed from equation (17), we use it as input for the compressed sensing tomography of the experimental state and compute the optimal estimate $\hat{\varrho}_{\text{CS}} := \hat{\varrho}(m_{\text{max}}, \hat{\varepsilon})$ of the *a priori* unknown experimentally prepared state ϱ . The good statistics available in our experiment allow us to estimate ϱ with comparably high accuracy. In general, due to experimental imperfections, ϱ (and hence $\hat{\varrho}$) will deviate from the target state $\varrho_{\text{GHZ}} := |\psi_{\text{GHZ}}\rangle\langle\psi_{\text{GHZ}}|$, see figure 3 for a pictorial representation. There, we show a comparison between the density matrices of the target state and the optimal compressed sensing estimate using bar plots.

The standard figure of merit to determine the performance of tomography is the quantum fidelity F of two states χ and σ , which is defined as $F(\chi, \sigma) = \text{tr}((\sqrt{\chi}\sigma\sqrt{\chi})^{1/2})$ [6]. We find that the fidelity between the GHZ state ϱ_{GHZ} and the estimated state $\hat{\varrho}_{\text{CS}}$ is



$$F(\varrho_{\text{GHZ}}, \hat{\varrho}_{\text{CS}}) = 0.855 \pm 0.006. \quad (19)$$

The uncertainty of the fidelity is determined by using the optimal compressed sensing estimate, $\hat{\varrho}$, as input for the generation of simulated data—parametric bootstrapping [57]—and taking the empirical standard deviation of the fidelity values. This uncertainty determines the robustness of the method. Obtaining a closed expression for proper error bounds from the data with respect to positivity constraints is hard [23, 58], while bootstrapping and taking the empirical standard deviation gives a good estimate of uncertainty [57].

To build confidence, we also computed the maximum likelihood estimate [59], $\hat{\varrho}_{\text{MLE}}$, using the same data to obtain a fidelity with respect to the target state of $F(\varrho_{\text{GHZ}}, \hat{\varrho}_{\text{MLE}}) = 0.843 \pm 0.004$, which shows that the estimators yield similar results; as will other estimators such as least squares with positivity constraint. Additionally, since we have measured a tomographically complete set of observables and the statistical properties of the measured data are sufficiently understood, we are able to provide an estimate of the fidelity with respect to the target state directly from the measured data without the need of performing tomography and an estimate of the corresponding error bound, see the [appendix](#) for details. With this, we obtain a fidelity of 0.845 ± 0.005 , which again is in good agreement with the results computed from the compressed sensing estimate. We note that the standard technique for estimating the fidelity of a state with respect to a specific target state requires estimating only the expectation values of a set of operators that form a decomposition of the target state. For a four-qubit GHZ state, this requires a minimum of nine specific Pauli basis measurements, as explained in the [appendix](#). In contrast, using compressed sensing tomography, even a random set of measurement settings produces fidelity estimates with respect to the GHZ state, which quickly approach the maximum at around 25 measurement settings. Furthermore, these measurement settings suffice to compute the fidelities with respect to arbitrary states, since they allow for the estimation of the entire state.

Compressed sensing is about employing provably fewer measurement settings than with standard methods, while still producing satisfactory reconstructions, i.e. to effectively sense in a compressive way. Along these lines, we explore how varying the number of measurement settings m affects the fidelity. This is shown in figure 4. In order to make the results independent from specific measurement settings, we randomly draw without replacement m out of m_{max} different settings 200 times and average over the resulting fidelities, thus providing a value for a typically expected fidelity for each m . As one would expect intuitively, we can see that the value of the fidelity increases monotonically with the number of measurement settings and converges to the fidelity of the estimate from tomographically complete data. The shaded region represents the uncertainty (\pm standard deviation) in the fidelity computed via bootstrapping and displays the decreasing uncertainty with increasing numbers of measurement settings. The fidelity already falls within the error bars of its final value for comparably small m .

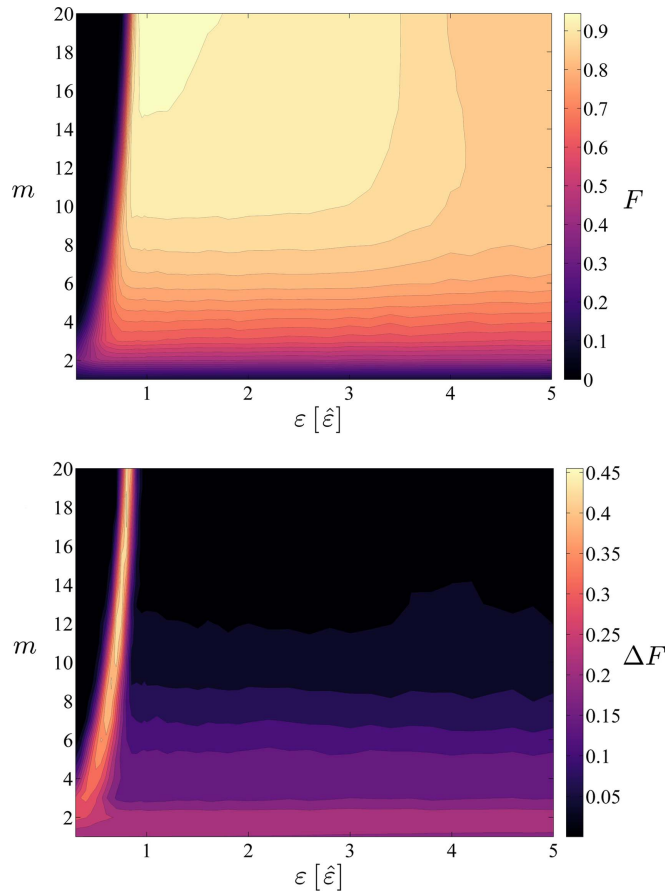


Figure 5. Fidelity $F(\hat{\rho}_{\text{CS}}, \hat{\rho}(m, \varepsilon))$ depending on the number of measurement settings m and the model parameter ε (top) and corresponding standard deviation ΔF (bottom) obtained via bootstrapping. Since in compressed sensing we are more interested in the regime of few measurement settings and the fidelities do not change significantly for larger m , we restrict ourselves to the region with $m \leq 20$. The data are generated randomly from $\hat{\rho}_{\text{CS}}$ and the measurement settings per m are drawn randomly as well. The fidelities are averaged over different data realisations and measurement settings. The highest fidelities are achieved for $\varepsilon \approx \hat{\varepsilon}$ with rapid decrease for $\varepsilon < \hat{\varepsilon}$ where the fraction of infeasible optimisations increases. Note that the higher the fidelity, the lower the standard deviation.

Deviations from the optimal parameter

In this section, we study the effect that misestimating ε has in the performance of the reconstruction of the state. We carry out this task by numerical simulation: Using the compressed sensing state estimate $\hat{\rho}_{\text{CS}}$, we simulate measurement data, which we subsequently input to our compressed sensing reconstruction procedure, varying both ε , m and randomly drawing measurement settings without replacement. If the corresponding optimisation problem is infeasible and yields no estimate, the fidelity F is set to zero. The fidelities $F(\hat{\rho}_{\text{CS}}, \hat{\rho}(m, \varepsilon))$ are averaged over data and measurement settings (500 different data sets and different measurement settings per m and ε).

The results for varying m and ε in units of $\hat{\varepsilon}$ are shown in figure 5. We compare the reconstructed states to $\hat{\rho}_{\text{CS}}$, which we used to generate the simulated data. We see that as m increases, the fidelity converges to unity at $\varepsilon = \hat{\varepsilon}$ (where $\hat{\rho}_{\text{CS}}$ is defined). We are interested in how quickly our reconstructed state approaches the optimal $\hat{\rho}_{\text{CS}}$ with fewer measurement settings, particularly if ε is misestimated. For instance, we see that we can obtain average fidelities of more than 0.8 for only 6 measurement settings. Figure 5 (top) again illustrates that $\varepsilon = \hat{\varepsilon}$ is the best choice as the fidelities around this region (and away from pathologically small numbers of measurement settings, $m > 3$) are the highest. Moreover, we also see that with increasing m , the standard deviation ΔF of the fidelity becomes smaller for $\varepsilon \geq \hat{\varepsilon}$. For $\varepsilon < \hat{\varepsilon}$, infeasibilities of the optimisation equation (12) that appear for certain choices of measurement settings lead to large standard deviations, which can be seen by the ridge in the area left of $\varepsilon = \hat{\varepsilon}$ in figure 5 (bottom). The ridge as well as the region of infeasibility gets close to $\varepsilon = \hat{\varepsilon}$ for large m , which is reasonable since more information (i.e. more constraints) puts greater restrictions on the optimisation problems. If fewer measurement settings are considered, as in the highly tomographically incomplete regime, overestimation of ε is less detrimental and state estimates still perform well, i.e. the fidelity is relatively constant for $\hat{\varepsilon} \lesssim \varepsilon \lesssim 3 \hat{\varepsilon}$. However, as m increases, the reconstruction becomes more strongly dependant on the choice of ε . Generally, we see that the higher the fidelity, the lower the standard deviation.

Discussion

In this work, we have experimentally explored the compressed sensing paradigm for quantum state tomography as applied to the photonic setting. We have explicitly laid out a method for applying these techniques and reconstructed the state of a four-photon system with tomographically complete data available, observing a high fidelity of the reconstructed state with respect to the target state. The presence of noise in the data requires that one carefully chooses appropriate constraints on the optimisation. In current applications, these parameters are usually obtained in an ad hoc way. We have provided a prescription to establish the parameters in a more systematic way by modelling the noise and performing cross validation, which is a general method for model selection. The quality of the data, being afflicted with noise predominantly attributed to finite counting statistics, allows us to model the noise via a multinomial distribution. This is a situation commonly expected for photonic experiments with post-selected data. In fact, we observe a great agreement between estimating the model parameter from theoretical noise modelling and cross validation.

Having established the appropriate model, we have been able to perform state reconstruction with tomographically incomplete data, which rapidly converges to the highest fidelity estimate as the number of measurement settings increases. As a validity check, we have also run different estimators on the full data and obtained similar results, showing that our compressed sensing procedure yields reasonable estimates. As is predicted by the mathematical theory of compressed sensing, we have found that the number of measurement settings needed for a satisfactory estimate of the underlying state is much smaller than the number of measurements necessary for tomographic completeness. We have also carried out a comprehensive bootstrapping analysis to build confidence in the robustness of our method. In fact, we have observed that the uncertainty in the fidelity quickly decreases with increasing number of measurement settings.

Furthermore, we have studied the robustness of our method with respect to improper model selection and the effects on the reconstruction. We have found that for several choices of models and different numbers of measurement settings, the performance of the reconstruction can vary dramatically. For small numbers of measurement settings, our method depends less strongly on the model. In contrast, for large numbers of measurement settings, it is imperative to determine the appropriate model for optimal performance.

Our results confirm that compressed sensing in conjunction with suitable model selection gives rise to reliable procedures for state reconstruction leading to effective tomography with tomographically incomplete data. These techniques can be applied to a wide range of experimental settings and provide a means to identify and verify appropriate models thereby paving the way for the future of practical quantum state tomography. With this, we contribute to establishing compressed sensing as a practical tool for quantum state tomography in the low-information regime.

Acknowledgments

We thank the DFG (EI 519/7-1, EI 519/9-1 within SPP 1798 CoSIP), the Templeton Foundation, the EU (AQuS, SIQS, RAQUEL), EU FP7 grant 600838 QWAD, U.S. Army Research Office (ARO) grant W911NF-14-1-0133, the BMBF (Q.com), the Freie Universität Berlin within the Excellence Initiative of the German Research Foundation, the South African Research Foundation, the Fritz Haber Institute of the Max Planck Society, and the German National Academic Foundation (Studienstiftung des Deutschen Volkes) for support.

Appendix: fidelity estimation with error bound

In this section, we provide more detail to the fidelity estimation with an error bound without the need of resorting to quantum state tomography. In the Pauli operator basis

$$\{O_l : O_l \in \bigotimes_{j=1}^n \{\mathbb{1}, \sigma_x, \sigma_y, \sigma_z\}\}, \quad (\text{A1})$$

we can estimate from the measured probabilities $\hat{p}_{j,k} = y_{j,k}/N_j$ the expansion coefficients

$$\xi_{\varrho}^l = \text{tr}(\varrho O_l / \sqrt{d}) \quad (\text{A2})$$

of the prepared state ϱ by a linear transformation Ω ,

$$\xi_{\varrho} = \Omega \hat{p}. \quad (\text{A3})$$

For convenience, we denote by \hat{p} the row-vectorisation of the matrix with entries $\hat{p}_{j,k}$. The fidelity with respect to a pure target state ϱ_T can be written in terms of the expansion coefficients as

$$F^2(\varrho_T, \varrho) = \sum_l \xi_{\varrho_T}^l \xi_{\varrho}^l = \xi_{\varrho_T}^T \Omega \hat{p}. \quad (\text{A4})$$

The frequency of the d different outcomes for the j th measurement setting is described by a multinomial distribution. The covariance matrix is given for each multinomial distribution by

$$\text{Cov}(Y_{j,k}, Y_{j,l}) = N_j(p_{j,k} \delta_{i,j} - p_{j,k} p_{j,l}). \quad (\text{A5})$$

Since different measurement settings correspond to mutually orthonormal operators, the frequencies of different measurement settings are uncorrelated, i.e. $\text{Cov}(Y_{i,k}, Y_{j,l}) = 0$ for $i \neq j$. Therefore the covariance matrix for the probabilities \hat{p} can be estimated from the data as

$$\text{Cov}(\hat{p}_{j,k}, \hat{p}_{j,l}) = N_j^{-1}(\hat{p}_{j,k} \delta_{i,j} - \hat{p}_{j,k} \hat{p}_{j,l}). \quad (\text{A6})$$

By means of linear error propagation, the variance of the fidelity is given by

$$\text{Var}(F^2) = \xi_{\varrho_T}^T \Omega \text{Cov}(\hat{p}, \hat{p}) \Omega^T \xi_{\varrho_T}, \quad (\text{A7})$$

which yields an estimate of the statistical error of the fidelity estimate from the data

$$\Delta F^2(\varrho, \varrho_T) = \sqrt{\text{Var}(F^2)}. \quad (\text{A8})$$

In particular, in order to estimate the fidelity with respect to the GHZ state, only nine Pauli basis measurements contribute. This can be seen from the expansion of the GHZ density matrix in the Pauli operator basis

$$\varrho_{\text{GHZ}} = \frac{1}{16} \left[\sum_{\sigma \in \{\mathbb{1}, \sigma_x, \sigma_y, \sigma_z\}} \sigma^{\otimes 4} + \sum_{\text{Perm.}} \mathbb{1} \otimes \mathbb{1} \otimes \sigma_z \otimes \sigma_z + \sum_{\text{Perm.}} \sigma_x \otimes \sigma_x \otimes \sigma_y \otimes \sigma_y \right], \quad (\text{A9})$$

where the last two sums run over all six distinct orders of the four factors of the tensor product.

To estimate the fidelity (A4), only the 16 Pauli coefficients of the prepared state are required that correspond to the operators of the expansion (A9). From the measurement outcomes of the measurement setting $\sigma_z^{\otimes 4}$, all coefficients of operators containing only the identity $\mathbb{1}$ and σ_z can be estimated. Thus, only nine Pauli basis measurements are necessary to estimate the fidelity.

Note that it is also possible to employ the measurement outcomes of all other measurement settings in the estimation of coefficients of terms that include the identity in equation (A9). In principle, it is thereby possible to further reduce the statistical error of the estimate of those coefficients. However, for the data set considered in this work, using more than nine measurement settings does not significantly alter the fidelity estimate.

References

- [1] O'Brien J L, Furusawa A and Vucovic J 2009 Photonic quantum technologies *Nature Phot.* **3** 687–95
- [2] Metcalf B J et al 2013 Multi-photon quantum interference in a multi-port integrated photonic device *Nature Comm.* **4** 1356
- [3] Peruzzo A, Laing A, Politi A, Rudolph T and O'Brien J L 2011 Multimode quantum interference of photons in multiport integrated devices *Nature Comm.* **2** 224
- [4] Monz T, Schindler P, Barreiro J T, Chwalla M, Nigg D, Coish W A, Harlander M, Haensel W, Hennrich M and Blatt R 2011 14-qubit entanglement: creation and coherence *Phys. Rev. Lett.* **106** 130506
- [5] Barends R et al 2014 Logic gates at the surface code threshold: superconducting qubits poised for fault-tolerant quantum computing *Nature* **508** 500–3
- [6] Nielsen M A and Chuang I L 2000 *Quantum Computation and Quantum Information* (Cambridge, UK: Cambridge University Press)
- [7] Pirandola S, Eisert J, Weedbrook C, Furusawa A and Braunstein S L 2015 Advances in quantum teleportation *Nature Photon.* **9** 641–52
- [8] Bell B, Clark A S, Tame M S, Halder M, Fulconis J, Wadsworth W J and Rarity J G 2012 Experimental characterization of photonic fusion using fiber sources *New J. Phys.* **14** 023021
- [9] Bell B A, Herrera-Martí D A, Tame M S, Markham D, Wadsworth W J and Rarity J G 2014 Experimental demonstration of a graph state quantum error-correction code *Nature Comm.* **5** 3658
- [10] Bell B A, Markham D, Herrera-Martí D A, Marin A, Wadsworth W J, Rarity J G and Tame M S 2014 Experimental demonstration of graph-state quantum secret sharing *Nat. Commun.* **5** 5480
- [11] Candès E J, Romberg J and Tao T 2006 Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information *IEEE Trans. Inf. Theor.* **52** 489–509
- [12] Foucart S and Rauhut H 2013 *A Mathematical Introduction to Compressive Sensing* (Heidelberg: Springer)
- [13] 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
- [14] 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
- [15] Guta M, Kypraios T and Dryden I 2012 Rank-based model selection for multiple ions quantum tomography *New J. Phys.* **14** 105002
- [16] 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 *Nature Comm.* **1** 149
- [17] Huebener R, Mari A and Eisert J 2013 A Wick's theorem for matrix product states *Phys. Rev. Lett.* **110** 040401
- [18] Steffens A, Friesdorf M, Langen T, Rauer B, Schweigler T, Hübener R, Schmiedmayer J, Riofrío C A and Eisert J 2015 Towards experimental quantum field tomography with ultracold atoms *Nature Comm.* **6** 7663
- [19] Schwemmer C, Toth G, Niggelbaum A, Moroder T, Gross D, Gühne O and Weinfurter H 2014 Experimental comparison of efficient tomography schemes for a six-qubit state *Phys. Rev. Lett.* **113** 040503

- [20] Blume-Kohout R 2012 Robust error bars for quantum tomography arXiv:1202.5270
- [21] Christandl M and Renner R 2012 Reliable quantum state tomography *Phys. Rev. Lett.* **109** 120403
- [22] Faist P and Renner R 2016 Practical and reliable error bars in quantum tomography *Phys. Rev. Lett.* **117** 010404
- [23] Carpentier A, Eisert J, Gross D and Nickl R 2015 Uncertainty quantification for matrix compressed sensing and quantum tomography problems ArXiv e-prints arXiv:1504.03234
- [24] Rodionov A V, Veitia A, Barends R, Kelly J, Sank D, Wenner J, Martinis J M, Kosut R L and Korotkov A N 2014 Compressed sensing quantum process tomography for superconducting quantum gates *Phys. Rev. B* **90** 144504
- [25] Kliesch M, Kueng R, Eisert J and Gross D 2016 Improving compressed sensing with the diamond norm *IEEE Trans. Inf. Th.* **62** 7445–63
- [26] Shabani A, Kosut R L, Mohseni M, Rabitz H, Broome M A, Almeida M P, Fedrizzi A and White A G 2011 Efficient measurement of quantum dynamics via compressive sensing *Phys. Rev. Lett.* **106** 100401
- [27] Riofrío C A, Gross D, Flammia S T, Monz T, Nigg D, Blatt R and Eisert J 2016 Experimental quantum compressed sensing for a seven-qubit system arXiv:1608.02263
- [28] Carmeli C, Heinosaari T, Schultz J and Toigo A 2014 Tasks and premises in quantum state determination *J. Phys. A: Mathematical and Theoretical* **47** 075302
- [29] Kalev A, Kosut R L and Deutsch I H 2015 Quantum tomography protocols with positivity are compressed sensing protocols *NPJ Quantum Information* **1** 15018
- [30] Donoho D L 2006 Compressed sensing *IEEE Trans. Inf. Theor.* **52** 1289–306
- [31] Candès E J and Recht B 2012 Exact matrix completion via convex optimization *Commun. ACM* **55** 111–9
- [32] Candès E J and Tao T 2010 The power of convex relaxation: near-optimal matrix completion *IEEE Trans. Inf. Theor.* **56** 2053–80
- [33] Gross D 2011 Recovering low-rank matrices from few coefficients in any basis *IEEE Trans. Inf. Theor.* **57** 1548–66
- [34] Vandenberghe L and Boyd S 1996 Semidefinite programming *SIAM Rev.* **38** 49–95
- [35] Fazel M 2002 Matrix rank minimization with applications *PhD Thesis* Stanford University
- [36] Vandenberghe L and Boyd S 2004 *Convex Optimization* (New York: Cambridge University Press)
- [37] Tropp J A 2015 Convex recovery of a structured signal from independent random linear measurements *Sampling Theory, a Renaissance* (Verlag: Birkhaeuser)
- [38] Kueng R, Rauhut H and Terstiege U 2014 Low rank matrix recovery from rank one measurements arXiv:1410.6913
- [39] Heinosaari T, Mazzarella L and Wolf M M 2013 Quantum tomography under prior information *Commun. Math. Phys.* **318** 355–74
- [40] Kueng R and Gross D 2015 Qubit stabilizer states are complex projective 3-designs arXiv:1510.02767
- [41] Liu Y-K 2011 Universal low-rank matrix recovery from Pauli measurements *Adv. Neural Inf. Process. Syst.* **24** 1638–46
- [42] Rockafellar R T 1997 *Convex analysis Princeton Landmarks in Mathematics and Physics* (Princeton, NJ: Princeton University Press)
- [43] Toh K C, Todd M J and Tütüncü R H 1999 SDPT3—a matlab software package for semidefinite programming *Optimization Methods and Software* **11** 545–81
- [44] Grant M and S Boyd 2014 CVX: matlab software for disciplined convex programming, version 2.1, cvxr.com/cvx
- [45] Cai J-F, Candès E J and Shen Z 2010 A singular value thresholding algorithm for matrix completion *SIAM J. Opt.* **20** 1956–82
- [46] Bell B A, Tame M S, Clark A S, Nock R W, Wadsworth W J and Rarity J G 2013 Experimental characterization of universal one-way quantum computing *New J. Phys.* **15** 053030
- [47] Browne D E and Rudolph T 2005 Resource-efficient linear optical quantum computation *Phys. Rev. Lett.* **95** 010501
- [48] Gross D, Kielsing K and Eisert J 2006 Potential and limits to cluster state quantum computing using probabilistic gates *Phys. Rev. A* **74** 042343
- [49] Pittman T B, Jacobs B C and Franson J D 2001 Probabilistic quantum logic operations using polarizing beam splitters *Phys. Rev. A* **64** 062311
- [50] Pan J-W, Simon C, Brukner C and Zeilinger A 2001 *Nature* **410** 1067–70
- [51] Pan J-W, Gasparoni S, Ursin R, Weihs G and Zeilinger A 2003 Experimental entanglement purification of arbitrary unknown states *Nature* **423** 417–22
- [52] Bodiya T P and Duan L-M 2006 Scalable generation of graph-state entanglement through realistic linear optics *Phys. Rev. Lett.* **97** 143601
- [53] Bell B, McMillan A, McCutcheon W and Rarity J 2015 Effects of self- and cross-phase modulation on photon purity for four-wave-mixing photon pair sources *Phys. Rev. A* **92** 053849
- [54] Burnham K P and Anderson D R 2003 *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (New York: Springer)
- [55] Akaike H 1974 A new look at the statistical model identification *IEEE Trans. Aut. Cont.* **19** 716–23
- [56] Schwarz G E 1978 Estimating the dimension of a model *Ann. Stat.* **6** 461–64
- [57] Efron B and Tibshirani R J 1994 *An Introduction to the Bootstrap* (New York: Chapman and Hall/CRC Press)
- [58] Suess D, Rudnicki Ł and Gross D 2016 Error regions in quantum state tomography: computational complexity caused by geometry of quantum states arXiv:1608.00374
- [59] Hradil Z 1997 Quantum-state estimation *Phys. Rev. A* **55** R1561–4