

PAPER • OPEN ACCESS

A scalable maximum likelihood method for quantum state tomography

To cite this article: T Baumgratz *et al* 2013 *New J. Phys.* **15** 125004

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

You may also like

- [Cooperative emission spectra as an efficient key probe of qubits pair entanglement along with field state tomography: an effective response to nonlinearity and classical drive power](#)
M S Ateto
- [Quantum state tomography based on infidelity estimation](#)
Yong Wang, Lijun Liu, Tong Dou *et al.*
- [The controlled SWAP test for determining quantum entanglement](#)
Steph Foulds, Viv Kendon and Tim Spiller

A scalable maximum likelihood method for quantum state tomography

T Baumgratz¹, A Nüßeler, M Cramer and M B Plenio

Institut für Theoretische Physik, Albert-Einstein-Allee 11, Universität Ulm,
D-89069 Ulm, Germany

E-mail: tillmann.baumgratz@uni-ulm.de

New Journal of Physics **15** (2013) 125004 (22pp)

Received 15 August 2013

Published 4 December 2013

Online at <http://www.njp.org/>

doi:10.1088/1367-2630/15/12/125004

Abstract. The principle of maximum likelihood reconstruction has proven to yield satisfactory results in the context of quantum state tomography for many-body systems of moderate system sizes. Until recently, however, quantum state tomography has been considered to be infeasible for systems consisting of a large number of subsystems due to the exponential growth of the Hilbert space dimension with the number of constituents. Several reconstruction schemes have been proposed since then to overcome the two main obstacles in quantum many-body tomography: experiment time and post-processing resources. Here we discuss one strategy to address these limitations for the maximum likelihood principle by adopting a particular state representation to merge a well established reconstruction algorithm maximizing the likelihood with techniques known from quantum many-body theory.

¹ Author to whom any correspondence should be addressed.



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](#).

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

Contents

1. Introduction	2
2. Quantum state tomography via maximum likelihood estimation	3
3. Merging matrix product operators with the maximum likelihood algorithm	5
4. Numerical simulations	7
5. Conclusions	12
Acknowledgments	12
Appendix A. Implementation details for the maximum likelihood algorithm iterating over matrix product operators	13
Appendix B. R-local Hamiltonians are matrix product operators	17
Appendix C. Greenberger–Horne–Zeilinger-type states	19
References	21

1. Introduction

The ability to prepare and manipulate quantum mechanical states is crucial for implementing quantum information processing and hence for building quantum computers [1]. Thus, it is important to verify the algorithms realized on quantum mechanical systems by conducting measurements on the system. This is achieved by means of quantum state tomography [2, 3], quantum process tomography [4] and quantum detector tomography [5, 6], which together are capable of characterizing the three main stages of a quantum experiment. Here we concentrate on the task of quantum state tomography for an informationally incomplete set of observables. Deducing a quantum mechanical description of the system from these measurements is by no means trivial and what is the most appropriate method to invert the experimental data is still a matter of controversy [7, 8]. The strategies of quantum state tomography range from directly inverting the measurements [9] to statistical approaches known from classical probability theory such as maximum likelihood estimation [9, 10] and Bayesian inference [8, 11]. The latter methods were introduced into the field of quantum state tomography to circumvent the possibility of negative eigenvalues of the state estimation arising by inverting Born’s rule for relative frequencies, i.e. noisy data. In principle, maximum likelihood estimation and Bayesian inference divide the approach to statistical inference into two schools. The former is the method at hand for the frequentist statisticians where probabilities are interpreted as the infinite limit of relative frequencies obtained in a measurement process. Principally, the probabilities can be determined solely by repeating the measurements on the system infinitely often. For Bayesian statisticians, probabilities are a degree of belief in a certain event. The prior probability distribution imposed by the observer is then updated exploiting additional knowledge (i.e. measurements) of the system. Both principles have their advantages and drawbacks in the field of quantum state tomography. While maximum likelihood estimation is biased towards rank deficient states, no prior knowledge is required. In contrast, in Bayesian inference it is often not clear how to choose the prior distribution, whereas the estimates will have full rank [8, 11]. In this paper, we will focus on the frequentist approach to quantum state tomography.

Maximum likelihood estimation is currently the method of choice and has proven to yield satisfactory results for moderate system sizes [12]. However, as the system sizes realized in the

laboratory steadily increase [12–16], new techniques for performing quantum state tomography on large system sizes are required due to the exponential growth both in the number of required measurements and in the classical post-processing that is required to connect the measured data with a description of the quantum state. Recently, several scalable schemes for verifying the state in the laboratory by a fidelity estimation or learning the state via a concrete state reconstruction have been presented [17–26]. The efficiency of the scalable schemes stems from the exploitation of additional structure of the considered states such as low rank, local nature of correlations or a specific form of the state. If these assumptions are true, they allow for a reduction of the number of parameters that have to be determined experimentally and, simultaneously, for a reduction of the experiment time. To obtain an estimate of the state within these models, the scalable tomography schemes come along with efficiently implementable reconstruction algorithms to reduce the required post-processing resources. The algorithms are designed, for instance, to optimize a fit function over pure states [24], to adopt maximum likelihood estimation to permutationally invariant states, which can be represented efficiently [25], or to invert local measurements to obtain a (not necessarily positive) state estimate [26].

In this paper we strive to push the applicability of large scale state tomography further towards an efficient implementation of a maximum likelihood estimation scheme exploiting both the spirit of maximum likelihood reconstruction and adopting a specific representation of the states. We are considering one-dimensional quantum systems composed of N finite-dimensional d -level subsystems (e.g. qubits with $d = 2$). To realize the scalable tomography scheme we show how to implement a well established fixed point algorithm [7] maximizing the likelihood function by means of matrix product states and operators [27–29]. Note that in this paper we restrict to the iterative fixed point algorithm for maximizing the likelihood function. One could, however, also adopt a direct maximization of the likelihood function using convex optimization tools together with a suitable representation of the considered states (this has been demonstrated, e.g., for permutationally invariant systems in [25]). This paper is organized as follows. We start by reviewing the maximum likelihood approach to quantum state tomography in section 2. Then, we introduce the fixed point algorithm and discuss a modification of the latter when a pure state estimate is desired. In section 3 we show that all operations necessary for executing the maximum likelihood algorithm can be implemented by means of matrix product states and operators. Here, we solely discuss the individual computational steps and postpone details to the appendix. If a certain operator comprising the positive operator-valued measure (POVM) elements of the measurement setting has a matrix product operator representation of low bond-dimension, we show how this can be done efficiently. In the appendix we discuss how this operator can be recast to fit into the framework of matrix product states and operators for two specific measurement settings. In the last section of this paper, section 4, we present numerical results demonstrating the performance of the algorithm for numerically simulated states.

2. Quantum state tomography via maximum likelihood estimation

In this section we review the maximum likelihood approach to quantum state tomography and discuss a fixed point algorithm that maximizes the likelihood function [3, 7]. We further motivate a modification of the algorithm when a pure state estimate is desired. The latter reformulation reduces the computational cost of the algorithm since it only requires matrix–vector multiplications instead of matrix–matrix multiplications.

Quantum state tomography is a procedure for estimating the density matrix from repeated measurements on M identically prepared quantum systems. Let $\{\hat{\Pi}_i\}$, $i \in \Delta$, be the set of all POVM elements corresponding to the measurements performed on the system where Δ is an index set associated with these measurements². In the measurement process we record the number of times n_i the outcome $\hat{\Pi}_i$ is obtained for all $i \in \Delta$, i.e. $M = \sum_{i \in \Delta} n_i$. Now, let \mathcal{D} be the state space of the physical system under consideration. According to Born's rule, the conditional probability of measuring outcome i given state $\hat{\rho} \in \mathcal{D}$ is equal to $p_i = p(\hat{\Pi}_i | \hat{\rho}) = \text{tr}[\hat{\Pi}_i \hat{\rho}]$. The joint probability $p(\mathbf{n} | \hat{\rho})$ of registering the specific outcomes n_i , $i \in \Delta$, is given by

$$\mathcal{L}(\hat{\rho}) = p(\mathbf{n} | \hat{\rho}) = \prod_{i \in \Delta} p_i^{n_i} = \prod_{i \in \Delta} \text{tr}[\hat{\Pi}_i \hat{\rho}]^{n_i}. \quad (1)$$

The likelihood function $\mathcal{L}(\hat{\rho})$ is interpreted as the conditional probability of measuring on the state $\hat{\rho}$ when registering outcome \mathbf{n} , i.e. it is considered to be a function on \mathcal{D} for a specific outcome of a measurement series. The maximum likelihood estimate $\hat{\rho}_{\text{ML}}$ is the element of the state space \mathcal{D} which maximizes the likelihood function. Therefore, it is considered as the most likely state given a set of measurement outcomes.

Instead of maximizing the likelihood function $\mathcal{L}(\hat{\rho})$ directly, one maximizes the log-likelihood function

$$\log \mathcal{L}(\hat{\rho}) = \sum_{i \in \Delta} n_i \log(\text{tr}[\hat{\Pi}_i \hat{\rho}]), \quad (2)$$

which yields the same estimate $\hat{\rho}_{\text{ML}}$ due to the fact that the likelihood function $\mathcal{L}(\hat{\rho})$ is positive on the state space and the logarithm is strictly increasing. The resulting function reveals the property of being concave and since we are optimizing over a convex set we are left with a convex optimization problem [30]. It is well known that the state $\hat{\rho}_{\text{ML}}$ maximizing the log-likelihood satisfies $\hat{\rho}_{\text{ML}} = \hat{R}(\hat{\rho}_{\text{ML}})\hat{\rho}_{\text{ML}}$ where the positive operator $\hat{R}(\hat{\rho})$ is given by [3]

$$\hat{R}(\hat{\rho}) = \frac{1}{M} \sum_{i \in \Delta} \frac{n_i}{p_i} \hat{\Pi}_i = \frac{1}{M} \sum_{i \in \Delta} \frac{n_i}{\text{tr}[\hat{\Pi}_i \hat{\rho}]} \hat{\Pi}_i. \quad (3)$$

Since both the operator $\hat{R}(\hat{\rho})$ and $\hat{\rho}$ are Hermitian, the extremal equation $\hat{\rho}_{\text{ML}} = \hat{R}(\hat{\rho}_{\text{ML}})\hat{\rho}_{\text{ML}}$ can be transformed to yield a well-established fixed point algorithm [3]

$$\hat{\rho}_{k+1} = \mathcal{N}[\hat{R}(\hat{\rho}_k)\hat{\rho}_k \hat{R}(\hat{\rho}_k)], \quad (4)$$

where the operation $\hat{R}(\hat{\rho}_k)\hat{\rho}_k \hat{R}(\hat{\rho}_k)$ preserves the positivity of the current iteration, the function $\mathcal{N}: \mathcal{H}^{d^N \times d^N} \rightarrow \mathcal{D}$ mapping Hermitian matrices to states denotes normalization to trace one and the algorithm is usually initialized by an unbiased initial state such as the completely mixed state. Note that, although heuristically convergent, there is no analytical proof that the fixed point algorithm (4) converges to the maximum of the log-likelihood function. Diluting the operator $\hat{R}(\hat{\rho})$ to $(\mathbb{1} + \epsilon \hat{R}(\hat{\rho})) / (1 + \epsilon)$, one can show that for $\epsilon \ll 1$ the log-likelihood increases in each step of the fixed point algorithm at the expense of the rate of convergence [7]. Since the log-likelihood function is concave, this guarantees that the diluted fixed point algorithm converges to the maximum.

² We label operators using hats, i.e. $\hat{\rho}$ for a density matrix, rather than applying this notation to estimates of the underlying system.

Under the assumption that the state in the laboratory is pure, the fixed point algorithm can be rewritten to optimize only over pure states. In particular, iterating

$$|\psi_{k+1}\rangle = \mathcal{N}[\hat{R}(|\psi_k\rangle)|\psi_k\rangle] \quad (5)$$

will yield a pure estimate of the desired state. Note, however, that restricting the state space to rank one matrices alters the optimization problem to being non-convex. This implies that although the objective function, i.e. the log-likelihood, is still concave, the feasible set, i.e. the set of rank one matrices, is non-convex and hence it is not guaranteed that the global maximum is attained by increasing the log-likelihood in every step, both in the standard formulation and the diluted version of this pure state algorithm. Nevertheless, as we will see, iterating only over pure states yields satisfactory results in practice even for noisy data. Now, of course, in a realistic setting the state prepared in the laboratory will never be pure. A pure state estimate, however, can nevertheless be used to certify whether the (possibly mixed) state implemented in the laboratory is close to this pure estimate by means of the techniques described in [24]. With this, one can construct a lower bound on the fidelity of the pure state estimate with respect to the actual state only by means of the (experimentally determined) reduced density matrices.

3. Merging matrix product operators with the maximum likelihood algorithm

In this section, we introduce the family of matrix product states and operators [27–29] and further show that the iterative fixed point algorithm (4) can be implemented using this formalism. We only give a brief summary of the individual steps and postpone details to the appendix.

Let $\hat{\rho} \in \mathcal{D}$ be a state describing a physical system comprising N d -level subsystems. Then, $\hat{\rho}$ can be written as

$$\hat{\rho} = \sum_{\alpha_1, \dots, \alpha_N} P_1[\alpha_1] \cdots P_N[\alpha_N] \hat{P}_1^{(\alpha_1)} \cdots \hat{P}_N^{(\alpha_N)}, \quad (6)$$

where $\{\hat{P}_k^{(\alpha_k)}\}$ is an operator basis of $\mathbb{C}^{d \times d}$ for site k , $\alpha_k = 1, \dots, d^2$ enumerates the basis elements per site and $P_k[\alpha_k] \in \mathbb{C}^{D_k \times D_{k+1}}$ are complex matrices for $k = 1, \dots, N$, with $D_1 = 1$ and $D_{N+1} = 1$. Every state can be brought to this form and the exponentially growing dimension of the Hilbert space (with respect to the number of subsystems N) can be covered by exponentially growing matrix dimensions D_k . However, many interesting quantum states (in particular living on one-dimensional structures) can be represented with a moderate number of parameters. Examples, amongst others, are ground states and thermal states of gapped local Hamiltonians [31–36]. These states have in common that they can be approximated very well by states with low bond-dimension $D = \max_k D_k$. We call states of the form of equation (6) matrix product operators or matrix product states depending on $\{\hat{P}_k^{(\alpha_k)}\}$ being a basis of $\mathbb{C}^{d \times d}$ or \mathbb{C}^d [28, 29] (note that every matrix product state with bond-dimension D is a matrix product operator with bond-dimension at most D^2) and collect such states with maximum bond-dimension equal to D in the set \mathcal{M}_D .

Next, we list the operations required to run the algorithm and then proceed by pointing out the corresponding matrix product operator equivalents. In what follows, we restrict ourselves to a POVM which contains solely observables of tensor product form, i.e. $\hat{\Pi}_i = \hat{\pi}_1^i \otimes \cdots \otimes \hat{\pi}_N^i$ with $i \in \Delta$. The fixed point algorithm is commonly initialized with the completely mixed state. Then, in the k th iteration, the algorithm

1. determines the expectation values of the POVM elements $\{\hat{\Pi}_i\}$, $i \in \Delta$, i.e. $p_i = \text{tr}[\hat{\Pi}_i \hat{\rho}_k]$,
2. constructs the operator $\hat{R}(\hat{\rho}_k)$,
3. multiplies the current iteration $\hat{\rho}_k$ from both sides with $\hat{R}(\hat{\rho}_k)$ and
4. normalizes the output to obtain iteration $k + 1$.

Whenever $\hat{\rho}_k$, $\hat{R}(\hat{\rho}_k)$ and hence $\hat{R}(\hat{\rho}_k)\hat{\rho}_k\hat{R}(\hat{\rho}_k)$ may be written as matrix product operators of small bond-dimension, all these steps may be performed efficiently (for a detailed outline of the different operations that are required we refer to the literature [28, 29] and appendix A):

We now outline how the four steps above may be performed efficiently. Firstly, computing expectation values of product observables of the form $\hat{\Pi} = \hat{\pi}_1 \otimes \cdots \otimes \hat{\pi}_N$ is straightforward:

$$\text{tr}[\hat{\rho} \hat{\Pi}] = \prod_{k=1}^N E_k, \quad (7)$$

where $E_k = \sum_{\alpha_k=1}^{d^2} P_k[\alpha_k] \text{tr}[\hat{\pi}_k \hat{P}_k^{(\alpha_k)}]$.

Secondly, the operator $\hat{R}(\hat{\rho}_k)$ has to be determined. This operator acts on the state space and hence, generally, has the same dimension as $\hat{\rho}_k$. To efficiently implement the maximum likelihood algorithm we therefore need to represent this operator as a matrix product operator with small bond-dimension. Since $\hat{R}(\hat{\rho}_k)$ is basically a weighted sum of the POVM elements, this restricts the set of measurements for which this is possible. To give an example for which this can be achieved, consider POVM elements $\hat{\Pi}_k^i$ of the form

$$\hat{\Pi}_k^i = \mathbb{1} \otimes \cdots \otimes \mathbb{1} \otimes \hat{\pi}_k^i \otimes \cdots \otimes \hat{\pi}_{k+R-1}^i \otimes \mathbb{1} \otimes \cdots \otimes \mathbb{1} \quad (8)$$

where $k = 1, \dots, N - R + 1$ labels all blocks of R consecutive spins and i enumerates the set of local POVM elements. Key point for the efficient representation of the operator $\hat{R}(\hat{\rho}_k)$ is that with this specific kind of measurement set-up this operator is in fact a local Hamiltonian. But local Hamiltonians are matrix product operators of low bond-dimension and hence $\hat{R}(\hat{\rho}_k)$ can be stored efficiently in its matrix product operator representation (see appendix B for further details). Note again that every measurement setting where the operator $\hat{R}(\hat{\rho}_k)$ has an efficient matrix product operator representation with small bond-dimension is suitable for the efficient implementation of the maximum likelihood algorithm and the above form is merely an example. One may also add global observables to the POVM set, a scenario which we consider for the reconstruction of a Greenberger–Horne–Zeilinger (GHZ)-type state below, with a discussion of the construction of the operator $\hat{R}(\hat{\rho}_k)$ in appendix C.

Thirdly, to compute the next iteration of the algorithm (4) we need to multiply the current iteration $\hat{\rho}_k$ with the operator $\hat{R}(\hat{\rho}_k)$ from both sides. Both objects are stored as matrix product operators. Multiplying two matrix product operators with bond-dimensions D_1 and D_2 results in a matrix product operator of dimension $D_1 \cdot D_2$, i.e. bond-dimensions multiply. To keep the bond-dimension at a certain level (e.g. D), we compress the current iteration $\hat{R}(\hat{\rho}_k)\hat{\rho}_k\hat{R}(\hat{\rho}_k)$ that has been obtained after multiplication of the previous estimate $\hat{\rho}_k$ with the operator $\hat{R}(\hat{\rho}_k)$, i.e. we solve [28]

$$\hat{\xi}_k = \text{argmin}[\|\hat{\rho} - \hat{R}(\hat{\rho}_k)\hat{\rho}_k\hat{R}(\hat{\rho}_k)\|^2 \mid \hat{\rho} \in \mathcal{M}_D], \quad (9)$$

where $\|\cdot\|$ denotes the Hilbert–Schmidt norm and \mathcal{M}_D comprises all matrix product operators with bond-dimension D , see appendix A for a detailed discussion. Obviously, solving

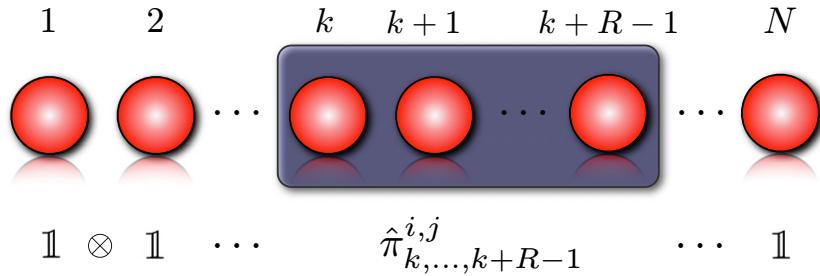


Figure 1. The considered measurements on blocks of R consecutive spins. For $\hat{\pi}_{k,\dots,k+R-1}^{i,j}$ see equation (11). Here, $k = 1, \dots, N - R + 1$ labels the block of size R , $i = 1, \dots, 3^R$ enumerates the basis rotation (all combinations of orientations along the X , Y and Z directions) and $j = 1, \dots, 2^R$ denotes the corresponding projectors per basis orientation on the eigenbasis of the respective Pauli spin operators.

equation (9) is a crucial point in the algorithm. If the operator is not compressible to a low bond-dimension, the algorithm will not be efficient. To recognize this issue in the execution of the algorithm, we choose a method that allows us to track the error made in the compression, i.e. the algorithm aborts when the minimum in equation (9) is not zero (or greater than a prescribed tolerance). Hence, if the norm difference does not converge to zero, we either choose a larger bond-dimension or abort the maximum likelihood algorithm (i.e. throw a flag and abort the tomography scheme).

Fourthly, in the final step of one iteration we have to normalize the current estimate to trace one. Again, this can be done efficiently since the trace of a matrix product operator can be computed directly using equation (7). Normalization is then equivalent to dividing each matrix by a fraction of the initial trace.

4. Numerical simulations

In this section we present results of our implementation of the maximum likelihood algorithm iterating over matrix product states and operators.³ The results suggest that applying maximum likelihood reconstruction to system sizes where the conventional implementations of full tomography fail due to (i) the number of required measurement settings and (ii) the limited post-processing resources is still manageable for appropriate POVM settings.

We are considering a one-dimensional chain of quantum systems implemented on N d -level systems with $d = 2$, i.e. qubits. We let the POVM elements be of the local form

$$\hat{\Pi}_k^{i,j} = \mathcal{W}(\mathbb{1}_{1,\dots,k-1} \otimes \hat{\pi}_{k,\dots,k+R-1}^{i,j} \otimes \mathbb{1}_{k+R,\dots,N}), \quad (10)$$

see figure 1. Here,

$$\hat{\pi}_{k,\dots,k+R-1}^{i,j} = |y_{k,\dots,k+R-1}^{i,j}\rangle\langle y_{k,\dots,k+R-1}^{i,j}|, \quad (11)$$

³ Due to the multiplication of the matrix product operators and a subsequent compression one has to be careful implementing the algorithm to maintain the Hermiticity in the current iteration.

with $k = 1, \dots, N - R + 1$ labelling the block of size R , $i = 1, \dots, 3^R$ enumerating the basis rotation (all combinations of orientations along the X , Y and Z directions) and $j = 1, \dots, 2^R$ denoting the corresponding projectors per basis orientation on the eigenbasis of the respective Pauli spin operators. This corresponds to measurements on blocks of R contiguous spins. For fixed R , the experimental effort, and hence the associated experiment time, grows linearly in the number of subsystems N reducing the exponential scaling of the total number of measurements from $M = 3^N m$ to a linearly scaling $M = 3^R m(N - R + 1)$. Here, m denotes the number of measurements per basis orientation.

We verify the performance of the maximum likelihood algorithm optimizing over matrix product operators (and matrix product states) for thermal states and ground states of next-neighbour Hamiltonians and for GHZ-type states. For the latter, we need to enlarge the above POVM, resulting in a set of measurements determining the GHZ-type state uniquely and a total number of measurements $M = 3^R m(N - R + 1) + Km$, where K is the cardinality of the set of additional (global) measurements.

We simulate the measurements by drawing m times from the exact multinomial distributions corresponding to the POVM elements specifying each basis orientation. Hence, the input to the reconstruction scheme are the relative frequencies $f_k^{i,j} = n_k^{i,j}/m$ for all i, j and k approximating the exact probability distributions. The reconstructed states $\hat{\rho}_{\text{rec}}$ are compared to the exact states by means of the renormalized Hilbert–Schmidt norm difference $D(\hat{\rho}, \hat{\rho}_{\text{rec}}) = \|\hat{\rho} - \hat{\rho}_{\text{rec}}\|^2/\|\hat{\rho}\|^2$, the fidelity $f(|\psi\rangle, \hat{\rho}_{\text{rec}}) = |\langle\psi|\hat{\rho}_{\text{rec}}|\psi\rangle|$ or $f(|\psi\rangle, |\psi_{\text{rec}}\rangle) = |\langle\psi|\psi_{\text{rec}}\rangle|^2$.

The first example outlines the performance of the algorithm (4) iterating over matrix product operators for thermal states of random next-neighbour Hamiltonians,

$$\hat{H} = \sum_{i=1}^{N-1} \hat{r}_{i,i+1}, \quad (12)$$

where the Hermitian matrices $\hat{r}_{i,i+1}$ act on sites i and $i + 1$. The real and imaginary parts of these matrices are drawn from a Gaussian distribution with zero mean and standard deviation one. Thermal states $\hat{\rho} = e^{-\beta \hat{H}}/Z$ are obtained by an imaginary time evolution using the time evolving block-decimation algorithm⁴ [37]. From these states, we compute the exact probability distributions corresponding to all contiguous blocks of size R for local tomographically complete measurements, simulate the measurements and reconstruct a state estimate $\hat{\rho}_{\text{rec}}$. Results are shown in figure 2. The plot indicates that, as expected, (i) the accuracy of the reconstruction scheme increases with increasing number of measurements and (ii) extending the range R over which measurements are performed decreases the error of the reconstructed estimate. Notably, the mean error for the reconstruction of thermal states corresponding to 45 different random Hamiltonians of the form of equation (12) with $R = 3$ and perfect measurements, i.e. $m = \infty$, is of order 10^{-3} .

Ground states of Hamiltonians of the form of equation (12) serve as our second example. We obtain the ground states by minimizing the expectation value of the Hamiltonian with respect to a matrix product state with given bond-dimension⁵. Sweeping through the chain and optimizing the matrices in the matrix product state site by site such that in each iteration the

⁴ We use a fourth order Trotter expansion with a step size in the imaginary time evolution of $\delta T = 10^{-3}$. We approximate the state as a matrix product operator with bond-dimension $D = 100$.

⁵ The random character of the Hamiltonian suggests that the eigenfunctions of the Hamiltonian are strongly localized and hence the ground state is very well approximated by a matrix product state of low bond-dimension.

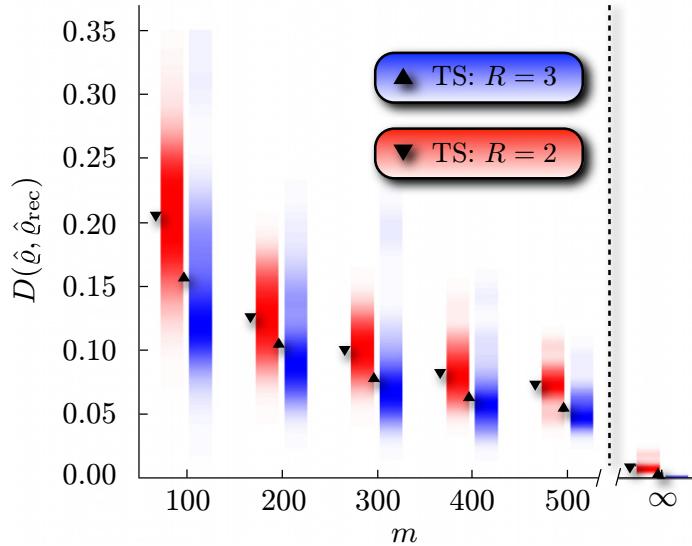


Figure 2. Performance of the iterative algorithm (4) for thermal states (TS) of random next-neighbour Hamiltonians of the form of equation (12) for $N = 16$ qubits with $\beta = 2$ after 1000 iterations. Input to the reconstruction scheme are relative frequencies corresponding to a tomographically complete set of measurements on all blocks of R contiguous sites. These relative frequencies are obtained by drawing from the exact probability distributions m times. Downward-pointing triangles: $R = 2$, upward-pointing triangles: $R = 3$. The densities illustrate the results for 45 random Hamiltonians and corresponding reconstructions while the triangles emphasize the mean value. All estimates have a bond-dimension of $D = 16$.

expectation value with respect to the Hamiltonian decreases, we will end up with a stationary point which serves as an approximation to the exact ground state⁶ [28]. Again, we compute the exact expectation values corresponding to local measurements on all blocks of R sites, simulate the measurements by drawing m times from the exact multinomial distributions and reconstruct the state using the maximum likelihood algorithm. Here, we are optimizing only over pure states using the iterative algorithm in equation (5). Figure 3 presents the results. Even for very small m we obtain a mean fidelity larger than 0.80 for $N = 20$ qubits. Moreover, for the exact probabilities ($m = \infty$) the mean fidelity is close to 1.00 after 5000 iterations indicating that iterating only over pure states does not get stuck in local minima. Note that the iterative algorithm is known to converge very slowly in comparison to other maximization techniques for the likelihood function [25].

Finally, let us consider GHZ-type states of the form

$$|\psi_N(\phi)\rangle = [|0\rangle^{\otimes N/2} |1\rangle^{\otimes N/2} + e^{i\phi} |1\rangle^{\otimes N/2} |0\rangle^{\otimes N/2}] / \sqrt{2} \quad (13)$$

where the number of sites N is even and $\phi \in [0; 2\pi)$ is a relative phase. These states are not uniquely determined by local measurements. The relative phase ϕ is a global feature and hence, besides local measurements, we need to perform some global measurements to fully determine

⁶ We approximate the ground state with matrix product states of bond-dimension $D = 20$.

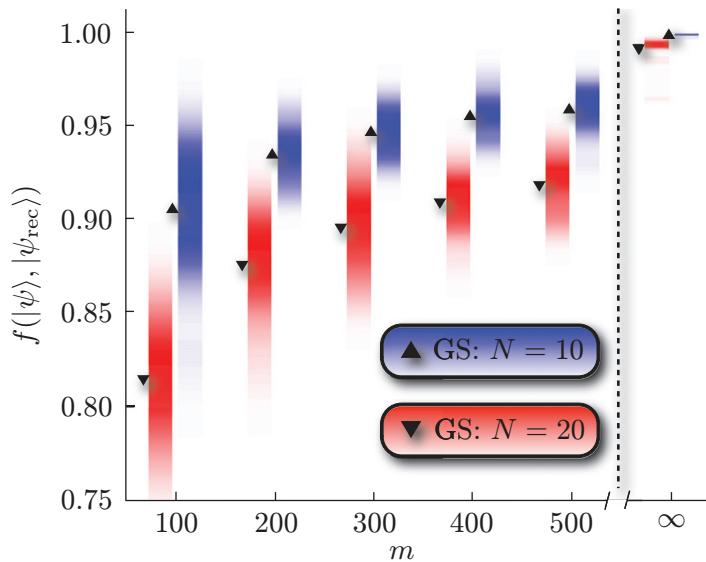


Figure 3. Results for the maximum likelihood algorithm iterating over matrix product states after 5000 iterations. The exact states are ground states (GS) of random next-neighbour Hamiltonians of the form of equation (12). Note that the reconstructed estimates have a bond-dimension of $D = 5$. Downward-pointing triangles: $N = 20$ sites, upward-pointing triangles: $N = 10$ sites. Measurements are simulated for all blocks of $R = 2$ contiguous sites by drawing from the exact multinomial distributions m times. For each m we generate 30 random Hamiltonians. The densities depict the distributions of the obtained fidelities while triangles indicate the mean value. Here, for $N = 20$, one iteration of the algorithm takes about one second on a standard laptop.

the state. This could be done by additionally measuring the operators $\hat{X}^{\otimes N}$ and $\hat{Y} \otimes \hat{X}^{\otimes N-1}$ since the expectation values

$$\langle \psi_N(\phi) | \hat{X}^{\otimes N} | \psi_N(\phi) \rangle = \cos(\phi) \quad (14)$$

and

$$\langle \psi_N(\phi) | \hat{Y} \otimes \hat{X}^{\otimes N-1} | \psi_N(\phi) \rangle = \sin(\phi) \quad (15)$$

fix the local phase. Note that these expectation values can be determined experimentally for large system sizes because only a simultaneous rotation of all spins is required plus a single site addressing to rotate one spin (e.g. the spin on the very left-hand side) along an orthogonal direction. We define the POVM $\{\hat{\Pi}_i\}$, $i = 1, \dots, 4$, with $\hat{\Pi}_1 = (\mathbb{1} + \hat{X}^{\otimes N})/4$, $\hat{\Pi}_2 = (\mathbb{1} - \hat{X}^{\otimes N})/4$, $\hat{\Pi}_3 = (\mathbb{1} + \hat{Y} \otimes \hat{X}^{\otimes N-1})/4$ and $\hat{\Pi}_4 = (\mathbb{1} - \hat{Y} \otimes \hat{X}^{\otimes N-1})/4$. These operators are positive and sum to one. Incorporating these operators into the POVM describing the local measurements can be done by a straightforward normalization of the operators. Let us denote the index set of the full POVM set by $\Delta = \Delta_1 \cup \Delta_2$ where Δ_1 comprises the elements corresponding to local measurements and Δ_2 comprises the four elements defined above. To apply the maximum likelihood algorithm for matrix product operators to this measurement

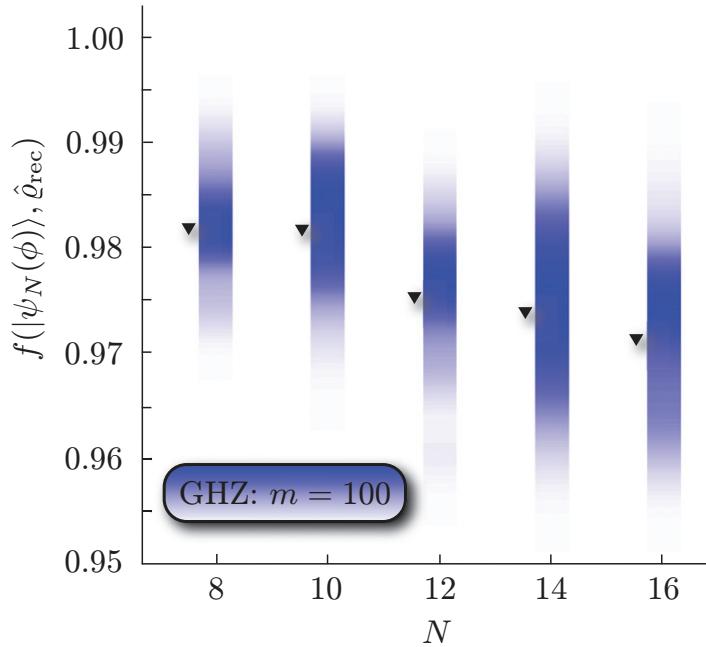


Figure 4. Results for GHZ-type states as shown in equation (13) where we choose $\phi = \pi/2$. Apart from the local measurements (on all blocks of size $R = 2$) global observables have to be measured given by $\{\hat{\Pi}_i\}$, $i \in \Delta_2$. Here, we draw $m = 100$ times from the exact probability distributions and reconstruct 30 states per N . Mean values are indicated by the downward-pointing triangles. The reconstructed estimates have bond-dimension $D = 10$. The plot shows the fidelity after 1000 iterations of the maximum likelihood algorithm.

scheme, it remains to show that the operator $\hat{R}(\hat{\rho})$, see equation (3), can be written as a matrix product operator with small bond-dimension D . In fact, we find

$$\hat{R}(\hat{\varrho}) = \hat{R}_1(\hat{\varrho}) + \hat{R}_2(\hat{\varrho}), \quad (16)$$

where $\hat{R}_k(\hat{\rho}) = \frac{1}{M} \sum_{i \in \Delta_k} \frac{n_i}{p_i} \hat{\Pi}_i$ for $k = 1, 2$. As before, $\hat{R}_1(\hat{\rho})$ corresponds to a matrix product operator with small bond-dimension D_1 due to its local character. Moreover, $\hat{R}_2(\hat{\rho})$ can be written as a matrix product operator with bond-dimension $D_2 = 2$ such that $\hat{R}(\hat{\rho})$ is a matrix product operator with bond-dimension $D = D_1 + 2$ (see appendix C for further details). We simulate measurements on the exact GHZ-type state (which is a matrix product state with bond-dimension $D_{\text{GHZ}} = 2$) for $R = 2$ and use the algorithm iterating over matrix product operators to obtain a state estimate. Results are shown in figure 4. This example illustrates that this method is also applicable to states that are not uniquely determined by local measurements only. In general, however, the experimentalist has to have some prior knowledge about the state he intends to implement on a physical system to decide whether additional measurements are necessary. The results suggest that even for a very small number of measurements where each considered basis rotation is measured $m = 100$ times, the algorithm is capable of reconstructing GHZ-type states very accurately.

5. Conclusions

We presented a scalable maximum likelihood algorithm for quantum state tomography. The only restrictions for the applicability is that the set of measurement operators is polynomial in the system size and that the individual POVM elements are of tensor product form, both of which are anyway generally desirable from an experimental point of view. The reconstruction technique relies on a well-established fixed point algorithm for maximizing the likelihood function. We have shown that this algorithm can be generalized to iterate over matrix product states and operators and hence yields a scalable reconstruction algorithm for quantum state tomography. Of course, for a general state where the measurements do not uniquely specify the state, the fixed point of the scalable maximum likelihood algorithm will not always come close to the true state. If the state is uniquely determined by the measured POVM elements, however, we provided numerical evidence that the algorithm chooses a state estimate which is close to the true state.

We observed that the convergence of the algorithm is very slow caused by the flat log-likelihood function when measurements are done only on a small subset of the full Hilbert space. Moreover, while our first and by no means optimal implementation is capable of dealing with measurements on blocks of size $R \geq 2$ the execution of the algorithm becomes rapidly slow with increasing R . An improvement of the implementation is left for future work. For instance, one could think of directly maximizing the concave objective function (e.g. by gradient descent methods) when the states are represented as matrix product states or in any other tailored efficient form. Concretely, restricting to permutationally invariant states and directly maximizing the likelihood, [25] has shown to yield superior performance when high accuracy is required.

The numerical results, however, suggest that the generalization of the iterative algorithm to matrix product operators is able to find state estimates which are close to the exact states already for a small number of measurements, which we verified for thermal and ground states of random next-neighbour Hamiltonians and for GHZ-type states. Furthermore, restricting the algorithm to optimize only over pure states has shown to yield satisfactory results under the assumption that the desired state is pure. This is remarkable since one iterates over a non-convex set. To certify the reconstructed pure estimate one could use the techniques described in [24] to find a lower bound on the fidelity of the state in the laboratory with respect to the pure state estimate. The results presented in this paper add to recent research for large scale quantum tomography and push the applicability of quantum state tomography to large system sizes.

Acknowledgments

We gratefully acknowledge R Rosenbach for the supply of results of the TEBD algorithm. Computations were performed on the bwGRiD⁷. This work was supported by the Alexander von Humboldt Foundation, the EU Integrating project SIQS, the EU STReP EQUAM and the BMBF Verbundprojekt QuOReP.

⁷ bwGRiD www.bw-grid.de (www.bw-grid.de), member of the German D-Grid initiative, funded by the Ministry for Education and Research and the Ministry for Science, Research and Arts Baden-Württemberg.

Appendix A. Implementation details for the maximum likelihood algorithm iterating over matrix product operators

In the first part of the appendix we review some basic properties and specific operations applied to matrix product operators (for a general overview we refer to [28, 29] where the ideas discussed in the remainder of this section of the appendix are presented in far more detail). We focus on the operations required for generalizing the maximum likelihood fixed point algorithm.

To set the notation, we recall the definition of a matrix product operator. For all sites $k = 1, \dots, N$ and $\alpha_k = 1, \dots, d^2$, let $\{P_k[\alpha_k]\}$ be a set of complex matrices of dimension $D_k \times D_{k+1}$ where $D_1 = 1 = D_{N+1}$. Further, let $\{\hat{P}_k^{(\alpha_k)}\}$ denote an operator basis for site k , e.g. for spin-1/2 particles the orthonormal Pauli spin basis, i.e. $\hat{P}_i^{(1)} = \mathbb{1}_i/\sqrt{2}$, $\hat{P}_i^{(2)} = \hat{X}_i/\sqrt{2}$, $\hat{P}_i^{(3)} = \hat{Y}_i/\sqrt{2}$ and $\hat{P}_i^{(4)} = \hat{Z}_i/\sqrt{2}$. The state

$$\hat{\varrho} = \sum_{\alpha_1, \dots, \alpha_N} P_1[\alpha_1] \cdots P_N[\alpha_N] \hat{P}_1^{(\alpha_1)} \cdots \hat{P}_N^{(\alpha_N)} \quad (\text{A.1})$$

is called matrix product operator if $D = \max_k D_k$ is low with respect to the state space dimension. Note, of course, that the notation of a low bond-dimension is a bit vague. The efficient simulation of a matrix product operator of bond-dimension D certainly depends on the computational resources available for the post-processing.

Basic matrix product operator manipulations. In the first section of appendix A we present a derivation of equation (7) in the main text, yielding an efficient strategy for computing expectation values of product observables with respect to matrix product operators. Further, we show how to multiply matrix product operators and give a specific representation of the completely mixed state. Let $\hat{\Pi} = \hat{\pi}_1 \otimes \cdots \otimes \hat{\pi}_N$ be a product observable acting on N d -level subsystems. The expectation value of $\hat{\Pi}$ with respect to the matrix product operator $\hat{\varrho}$ is determined via

$$\begin{aligned} \text{tr}[\hat{\varrho} \hat{\Pi}] &= \sum_{\alpha_1, \dots, \alpha_N} P_1[\alpha_1] \cdots P_N[\alpha_N] \text{tr}[\hat{P}_1^{(\alpha_1)} \cdots \hat{P}_N^{(\alpha_N)} \hat{\pi}_1 \cdots \hat{\pi}_N] \\ &= \sum_{\alpha_1, \dots, \alpha_N} P_1[\alpha_1] \cdots P_N[\alpha_N] \text{tr}[(\hat{P}_1^{(\alpha_1)} \hat{\pi}_1) \cdots (\hat{P}_N^{(\alpha_N)} \hat{\pi}_N)] \\ &= \sum_{\alpha_1, \dots, \alpha_N} P_1[\alpha_1] \cdots P_N[\alpha_N] \text{tr}[\hat{P}_1^{(\alpha_1)} \hat{\pi}_1] \cdots \text{tr}[\hat{P}_N^{(\alpha_N)} \hat{\pi}_N] \\ &= \prod_{k=1}^N E_k, \end{aligned} \quad (\text{A.2})$$

where $E_k = \sum_{\alpha_k=1}^{d^2} P_k[\alpha_k] \text{tr}[\hat{P}_k^{(\alpha_k)} \hat{\pi}_k] \in \mathbb{C}^{D_k \times D_{k+1}}$. In the derivation we used the fact that for systems i and $i+1$ we have $\hat{A}_i \hat{B}_{i+1} \hat{C}_i \hat{D}_{i+1} = \hat{A}_i \otimes \hat{B}_{i+1} \cdot \hat{C}_i \otimes \hat{D}_{i+1} = \hat{A}_i \cdot \hat{C}_i \otimes \hat{B}_{i+1} \cdot \hat{D}_{i+1} = \hat{A}_i \hat{C}_i \hat{B}_{i+1} \hat{D}_{i+1}$ and that $\text{tr}[A_i B_{i+1}] = \text{tr}[A_i \otimes B_{i+1}] = \text{tr}[A_i] \text{tr}[B_{i+1}]$.

Naively multiplying two matrix product operators with bond-dimensions D_1 and D_2 yields a matrix product operator with bond-dimension at most $D_1 \cdot D_2$. In particular, given the matrix product operators

$$\hat{\varrho} = \sum_{\alpha_1, \dots, \alpha_N} P_1[\alpha_1] \cdots P_N[\alpha_N] \hat{P}_1^{(\alpha_1)} \cdots \hat{P}_N^{(\alpha_N)} \quad (\text{A.3})$$

with bond-dimension D_1 and

$$\hat{\xi} = \sum_{\beta_1, \dots, \beta_N} Q_1[\beta_1] \cdots Q_N[\beta_N] \hat{P}_1^{(\beta_1)} \cdots \hat{P}_N^{(\beta_N)} \quad (\text{A.4})$$

with bond-dimension D_2 , we find

$$\begin{aligned} \hat{\varrho} \cdot \hat{\xi} &= \sum_{\alpha_1, \dots, \alpha_N} \sum_{\beta_1, \dots, \beta_N} P_1[\alpha_1] \cdots P_N[\alpha_N] Q_1[\beta_1] \cdots Q_N[\beta_N] \hat{P}_1^{(\alpha_1)} \cdots \hat{P}_N^{(\alpha_N)} \hat{P}_1^{(\beta_1)} \cdots \hat{P}_N^{(\beta_N)} \\ &= \sum_{\alpha_1, \dots, \alpha_N} \sum_{\beta_1, \dots, \beta_N} (P_1[\alpha_1] \otimes Q_1[\beta_1]) \cdots (P_N[\alpha_N] \otimes Q_N[\beta_N]) (\hat{P}_1^{(\alpha_1)} \hat{P}_1^{(\beta_1)}) \cdots (\hat{P}_N^{(\alpha_N)} \hat{P}_N^{(\beta_N)}). \end{aligned} \quad (\text{A.5})$$

Now, for every site $k = 1, \dots, N$, $\{\hat{P}_k^{(\alpha_k)}\}$ denotes an operator basis such that we have $\hat{P}_k^{(\alpha_k)} \hat{P}_k^{(\beta_k)} = \sum_{\gamma_k} c_{\gamma_k}^{\alpha_k, \beta_k} \hat{P}_k^{(\gamma_k)}$ with $c_{\gamma_k}^{\alpha_k, \beta_k} = \text{tr}[(\hat{P}_k^{(\gamma_k)})^\dagger \hat{P}_k^{(\alpha_k)} \hat{P}_k^{(\beta_k)}]$. Consequently, the product of two matrix product operators gives

$$\begin{aligned} \hat{\varrho} \cdot \hat{\xi} &= \sum_{\gamma_1, \dots, \gamma_N} \left(\sum_{\alpha_1, \beta_1} c_{\gamma_1}^{\alpha_1, \beta_1} P_1[\alpha_1] \otimes Q_1[\beta_1] \right) \cdots \left(\sum_{\alpha_N, \beta_N} c_{\gamma_N}^{\alpha_N, \beta_N} P_N[\alpha_N] \otimes Q_N[\beta_N] \right) \cdot \hat{P}_1^{(\gamma_1)} \cdots \hat{P}_N^{(\gamma_N)} \\ &= \sum_{\gamma_1, \dots, \gamma_N} R_1[\gamma_1] \cdots R_N[\gamma_N] \hat{P}_1^{(\gamma_1)} \cdots \hat{P}_N^{(\gamma_N)} \end{aligned} \quad (\text{A.6})$$

with $R_k[\gamma_k] = \sum_{\alpha_k, \beta_k} \text{tr}[(\hat{P}_k^{(\gamma_k)})^\dagger \hat{P}_k^{(\alpha_k)} \hat{P}_k^{(\beta_k)}] P_k[\alpha_k] \otimes Q_k[\beta_k]$. Hence, the product of two matrix product operators with bond-dimensions D_1 and D_2 is itself a matrix product operator with bond-dimension at most $D = D_1 \cdot D_2$.

The maximum likelihood algorithm is initialized by the completely mixed state. Choosing the orthonormal Pauli spin basis in the matrix product operator representation, this state corresponds to a matrix product operator with bond-dimension one. One possible choice of the 1×1 matrices defining the completely mixed state is

$$\begin{aligned} P_k[1] &= \frac{1}{N \cdot \sqrt{d^N}}, \\ P_k[\alpha_k] &= 0, \quad \text{for } \alpha_k = 2, \dots, d^2 \text{ and } k = 1, \dots, N. \end{aligned} \quad (\text{A.7})$$

Therefore, initializing the matrix product operator formulation of the maximum likelihood algorithm is straightforward.

Compressing matrix product operators. We have seen that the bond-dimension of the current iteration $\hat{\varrho}_k$ increases in each step of the algorithm due to the matrix product operator multiplication with the operator $\hat{R}(\hat{\varrho}_k)$. Hence, it is essential to introduce subroutines to keep the bond-dimension of the state estimate low. This is done by compressing the matrix product state to a state with smaller bond-dimension at each iteration of the tomography algorithm [28]. In the second section of appendix A we describe the compression of a matrix product operator $\hat{\varrho}$ with bond-dimension D_1 to a matrix product operator $\hat{\sigma}$ with bond-dimension $D_2 < D_1$.

For that, we show how to solve the optimization problem

$$\hat{\sigma} = \operatorname{argmin}[\|\hat{\rho} - \hat{\rho}'\|^2 \mid \hat{\rho}' \in \mathcal{M}_{D_2}], \quad (\text{A.8})$$

see equation (9) in the main text.

Before we describe the iterative compression scheme, let us first introduce a convenient representation of the matrix product operator. For this, note that the matrices $\{P_k[\alpha_k]\}$ are not unique since inserting $\mathbb{1} = UU^\dagger$ between sites k and $k+1$ will not alter the state, but the corresponding matrices $\{P_k[\alpha_k]\}$ and $\{P_{k+1}[\alpha_{k+1}]\}$. This freedom in the choice of the matrices allows (by a successive singular value decomposition starting from the left- and right-hand sides of the chain) to write the matrix product operator as

$$\hat{\rho} = \sum_{\alpha_1, \dots, \alpha_N} B_1[\alpha_1] \cdots B_N[\alpha_N] \hat{P}_1^{(\alpha_1)} \cdots \hat{P}_N^{(\alpha_N)}, \quad (\text{A.9})$$

where the matrices on sites $l = 1, \dots, k-1$ satisfy

$$\sum_{\alpha_l} B_l[\alpha_l]^\dagger B_l[\alpha_l] = \mathbb{1}_{D_{l+1} \times D_{l+1}} \quad (\text{A.10})$$

and where the matrices on sites $l = k+1, \dots, N$ satisfy

$$\sum_{\alpha_l} B_l[\alpha_l] B_l[\alpha_l]^\dagger = \mathbb{1}_{D_l \times D_l}. \quad (\text{A.11})$$

We call matrix product operators satisfying these relations k -normal matrix product operators [28]. Further, note that for k -normal matrix product operators the purity, i.e. the squared Hilbert–Schmidt norm $\|\hat{\rho}\|^2 = \operatorname{tr}[\hat{\rho}^\dagger \hat{\rho}]$, can be computed by

$$\|\hat{\rho}\|^2 = \sum_{\alpha_k} \operatorname{tr}[B_k[\alpha_k]^\dagger B_k[\alpha_k]]. \quad (\text{A.12})$$

Now, let

$$\hat{\rho} = \sum_{\beta_1, \dots, \beta_N} A_1[\beta_1] \cdots A_N[\beta_N] \hat{P}_1^{(\beta_1)} \cdots \hat{P}_N^{(\beta_N)} \quad (\text{A.13})$$

be a matrix product operator with bond-dimension D_1 (not necessarily in k -normal form) and $\hat{\sigma}$ be a matrix product operator with bond-dimension $D_2 < D_1$ satisfying

$$\hat{\sigma} = \operatorname{argmin}[\|\hat{\rho} - \hat{\rho}'\|^2 \mid \hat{\rho}' \in \mathcal{M}_{D_2}], \quad (\text{A.14})$$

i.e. $\hat{\sigma}$ is the matrix product operator with bond-dimension $D_2 < D_1$ closest to $\hat{\rho}$ with respect to the Hilbert–Schmidt norm (hence a compression of the latter). To determine $\hat{\sigma}$, we use a procedure where the minimization is performed iteratively by sweeping through the chain several times while optimizing the compressed matrices site by site [28], i.e. we minimize

$$\|\hat{\rho} - \hat{\rho}'\|^2 = \operatorname{tr}[\hat{\rho}^\dagger \hat{\rho}] - \operatorname{tr}[\hat{\rho}^\dagger \hat{\rho}'] - \operatorname{tr}[(\hat{\rho}')^\dagger \hat{\rho}] + \operatorname{tr}[(\hat{\rho}')^\dagger \hat{\rho}'] \quad (\text{A.15})$$

with respect to the complex conjugate matrices $\{B_k[\alpha_k]^*\}$ on site k of the state

$$\hat{\rho}' = \sum_{\alpha_1, \dots, \alpha_N} B_1[\alpha_1] \cdots B_N[\alpha_N] \hat{P}_1^{(\alpha_1)} \cdots \hat{P}_N^{(\alpha_N)} \quad (\text{A.16})$$

given in k -normal form while keeping all the other matrices fixed. Note that only the last two terms in (A.15) contribute to the minimization such that

$$\frac{\partial}{\partial (B_k[\alpha_k]^*)_{i_k, i_{k+1}}} \|\hat{\varrho} - \hat{\varrho}'\|^2 = \frac{\partial}{\partial (B_k[\alpha_k]^*)_{i_k, i_{k+1}}} (-\text{tr}[(\hat{\varrho}')^\dagger \hat{\varrho}] + \text{tr}[(\hat{\varrho}')^\dagger \hat{\varrho}']). \quad (\text{A.17})$$

Since the state $\hat{\varrho}'$ is given in k -normal form, we find

$$\frac{\partial}{\partial (B_k[\alpha_k]^*)_{i_k, i_{k+1}}} (\text{tr}[(\hat{\varrho}')^\dagger \hat{\varrho}']) = (B_k[\alpha_k])_{i_k, i_{k+1}} \quad (\text{A.18})$$

and

$$\frac{\partial}{\partial (B_k[\alpha_k]^*)_{i_k, i_{k+1}}} (-\text{tr}[(\hat{\varrho}')^\dagger \hat{\varrho}]) = -(L_{1, \dots, k-1} \cdot A_k[\alpha_k] \cdot R_{k+1, \dots, N})_{i_k, i_{k+1}}, \quad (\text{A.19})$$

where

$$L_{1, \dots, k-1} = \left(\sum_{\alpha_{k-1}} (B_{k-1}[\alpha_{k-1}])^\dagger \left(\dots \left(\sum_{\alpha_1} (B_1[\alpha_1])^\dagger A_1[\alpha_1] \right) \dots \right) A_{k-1}[\alpha_{k-1}] \right) \quad (\text{A.20})$$

and

$$R_{k+1, \dots, N} = \left(\sum_{\alpha_{k+1}} A_{k+1}[\alpha_{k+1}] \left(\dots \left(\sum_{\alpha_N} A_N[\alpha_N] (B_N[\alpha_N])^\dagger \right) \dots \right) (B_{k+1}[\alpha_{k+1}])^\dagger \right). \quad (\text{A.21})$$

Hence, the extremal point is given by

$$B_k[\alpha_k] = L_{1, \dots, k-1} \cdot A_k[\alpha_k] \cdot R_{k+1, \dots, N}. \quad (\text{A.22})$$

Finally, the iterative minimization of the norm difference can be performed by starting with a N -normal matrix product operator of dimension D_2 , updating the matrices $L_{1, \dots, N-1}$ and $\{B_N[\alpha_N]\}$ and a successive $(N-1)$ -normalization of the resulting matrix product operator. The matrices on site $N-1, \dots, 1$ are subsequently updated by the rule (A.22) while one needs to keep the matrix product operator in its k -normal form when optimizing on site k . Sweeping through the chain back and forth several times will lead to a convergence of this procedure. Since we are minimizing the norm distance in an iterative manner, this scheme might get stuck in local minima [28]. Therefore we monitor the norm difference by exploiting that after updating the matrices on site k we have

$$\|\hat{\varrho} - \hat{\varrho}'\|^2 = \|\hat{\varrho}\|^2 - \sum_{\alpha_k} \text{tr}[B_k[\alpha_k] (B_k[\alpha_k])^\dagger]. \quad (\text{A.23})$$

With this, we can either abort the algorithm if the norm difference does not converge to zero or increase the bond-dimension and redo the compression. To avoid the attraction of local minima, one can consider two sites instead of only one site in each step of the compression algorithm. Decomposing the optimized matrix products living on two sites by means of a singular value decomposition yields the matrices living on the single sites and hence the matrix product structure is preserved [28].

Appendix B. R-local Hamiltonians are matrix product operators

In this section of the appendix we are considering the specific measurement setting where all POVM elements act non-trivially only on subsets of R consecutive sites. In this scenario, the operator $\hat{R}(\hat{\varrho})$ has the structure of a R -local Hamiltonian and hence a matrix product operator representation where the bond-dimension can be related to the interaction range R .

It is crucial for the scalability of the matrix product operator formulation of the maximum likelihood algorithm, see equation (4) of the main text, that the operator [3]

$$\hat{R}(\hat{\varrho}) = \frac{1}{M} \sum_{i \in \Delta_1} \frac{n_i}{p_i} \hat{\Pi}_i = \frac{1}{M} \sum_{i \in \Delta_1} \frac{n_i}{\text{tr}[\hat{\Pi}_i \hat{\varrho}]} \hat{\Pi}_i \quad (\text{B.1})$$

can be written as a matrix product operator of low bond-dimension. Now, let the POVM elements $\{\hat{\Pi}_i\}$, $i \in \Delta_1$, only act non-trivially on a subset of R consecutive sites. Then, we find

$$\hat{R}(\hat{\varrho}) = \sum_{i=1}^{N-R+1} \sum_{\alpha_1, \dots, \alpha_R=1}^n c_i[\alpha_1, \dots, \alpha_R] \hat{S}_i^{(\alpha_1)} \dots \hat{S}_{i+R-1}^{(\alpha_R)}, \quad (\text{B.2})$$

where we consider n operators on each site and comprise all coefficients of the operator $\hat{R}(\hat{\varrho})$ in $c_i[\alpha_1, \dots, \alpha_R]$. Note, that the operator $\hat{S}_i^{(\alpha_1)} \dots \hat{S}_{i+R-1}^{(\alpha_R)}$ only acts on sites $i, i+1, \dots, i+R-1$ and that, for the moment, we do not require that $\{\hat{S}_k^{(\alpha_k)}\}$ is a basis on site $k = 1, \dots, N$. In this measurement setting, the operator $\hat{R}(\hat{\varrho})$ obeys the form of a R -local Hamiltonian. It is well known that R -local Hamiltonians are matrix product operators [28]. Before we show how to convert a R -local Hamiltonian into its matrix product operator representation, let us discuss an alternative representation of matrix product operators. Let $\hat{\varrho}$ be a matrix product operator of bond-dimension D . Then [28]

$$\begin{aligned} \hat{\varrho} &= \sum_{\alpha_1, \dots, \alpha_N} P_1[\alpha_1] \dots P_N[\alpha_N] \hat{P}_1^{(\alpha_1)} \dots \hat{P}_N^{(\alpha_N)} \\ &= \sum_{\alpha_1, \dots, \alpha_N} \sum_{i_2, \dots, i_N} (P_1[\alpha_1])_{1, i_2} (P_2[\alpha_2])_{i_2, i_3} \dots (P_N[\alpha_N])_{i_N, 1} \hat{P}_1^{(\alpha_1)} \dots \hat{P}_N^{(\alpha_N)} \\ &= \sum_{i_2, \dots, i_N} \left(\sum_{\alpha_1} (P_1[\alpha_1])_{1, i_2} \hat{P}_1^{(\alpha_1)} \right) \dots \left(\sum_{\alpha_N} (P_N[\alpha_N])_{i_N, 1} \hat{P}_1^{(\alpha_N)} \right) \\ &= \hat{\varrho}^{[1]} \dots \hat{\varrho}^{[N]} \end{aligned} \quad (\text{B.3})$$

where the operator-valued matrix $(\hat{\varrho}^{[k]})_{i_k, i_{k+1}} = \left(\sum_{\alpha_k} (P_k[\alpha_k])_{i_k, i_{k+1}} \hat{P}_k^{(\alpha_k)} \right)$ acts only on system k and where $i_1 = 1$ and $i_{N+1} = 1$. We will refer to this representation of the matrix product operators as the operator-valued matrix product operator representation. Note that, given the operator-valued matrices, one can find the matrices of the matrix product operator representation by inversion, i.e.

$$(P_k[\alpha_k])_{i_k, i_{k+1}} = \text{tr}[\hat{P}_k^{(\alpha_k)} (\hat{\varrho}^{[k]})_{i_k, i_{k+1}}]. \quad (\text{B.4})$$

We now set out to find an operator-valued matrix product operator representation of the R -local Hamiltonian $\hat{R}(\hat{\varrho})$, see equation (B.2). For this, we rewrite equation (B.2) as

$$\begin{aligned}\hat{R}(\hat{\varrho}) &= \sum_{i=1}^{N-R+1} \sum_{\alpha_1, \dots, \alpha_R=1}^n c_i[\alpha_1, \dots, \alpha_R] \hat{S}_i^{(\alpha_1)} \hat{S}_{i+1}^{(\alpha_2)} \dots \hat{S}_{i+R-1}^{(\alpha_R)} \\ &= \sum_{i=1}^{N-R+1} \sum_{\alpha_2, \dots, \alpha_R=1}^n \hat{Q}_i^{(\alpha_2, \dots, \alpha_R)} \hat{S}_{i+1}^{(\alpha_2)} \dots \hat{S}_{i+R-1}^{(\alpha_R)}\end{aligned}\quad (\text{B.5})$$

where $\hat{Q}_i^{(\alpha_2, \dots, \alpha_R)} = \sum_{\alpha_1=1}^n c_i[\alpha_1, \dots, \alpha_R] \hat{S}_i^{(\alpha_1)}$ for all $i = 1, \dots, N - R + 1$. Let us illustrate the operator-valued matrices for $\hat{R}(\hat{\varrho})$ in the case of next-neighbour interaction, i.e. we measure only on all blocks of two contiguous sites and hence $R = 2$. It is easily verified that the operator-valued matrices take on the form [28]

$$\hat{R}^{[1]} = \left[0 \hat{Q}_1^{(1)} \hat{Q}_1^{(2)} \dots \hat{Q}_1^{(n)} \mathbb{1} \right] \in \mathbb{C}^{d \times d(n+2)} \quad (\text{B.6})$$

for site 1,

$$\hat{R}^{[k]} = \begin{bmatrix} \mathbb{1} & 0 & \dots & 0 \\ \hat{S}^{(1)} & 0 & \dots & 0 \\ \hat{S}^{(2)} & \vdots & & \vdots \\ \vdots & & \ddots & \\ \hat{S}^{(n)} & 0 & \dots & 0 \\ 0 & \hat{Q}_k^{(1)} & \hat{Q}_k^{(2)} & \dots & \hat{Q}_k^{(n)} & \mathbb{1} \end{bmatrix} \in \mathbb{C}^{d(n+2) \times d(n+2)} \quad (\text{B.7})$$

for sites $2, \dots, N - 1$ and

$$\hat{R}^{[N]} = \begin{bmatrix} \mathbb{1} \\ \hat{S}^{(1)} \\ \hat{S}^{(2)} \\ \vdots \\ \hat{S}^{(n)} \\ 0 \end{bmatrix} \in \mathbb{C}^{d(n+2) \times d} \quad (\text{B.8})$$

for site N , such that

$$\hat{R}(\hat{\varrho}) = \hat{R}^{[1]} \dots \hat{R}^{[N]}. \quad (\text{B.9})$$

The rules to obtain the operator-valued matrices for R -local Hamiltonians can be generalized straightforwardly. The dimensions of the resulting operator-matrices grow according to $d(2 + \sum_{i=1}^{R-1} n^i)$ where d is the on-site dimension, n the number of considered operators per site (i.e. the cardinality of the set $\{\hat{S}^{(\alpha)}\}$) and R the number of consecutive sites in one block on which measurements are performed. Equation (B.4) allows to compute the matrices of the standard matrix product operator representation resulting in a bond-dimension of at most $d(2 + \sum_{i=1}^{R-1} n^i)$.

For the quantum state tomography of qubits ($d = 2$) we have $n = 6$ since every basis rotation (orientation along X , Y and Z) allows for two spin orientations (spin up or down in the respective basis). Consequently, for realistic quantum state tomography settings the bond-dimension of the operator $\hat{R}(\hat{\rho})$ still grows rapidly when measured on large block sizes R . Heuristically, it turns out that the so constructed matrix product operator representation of the operator $\hat{R}(\hat{\rho})$ is not optimal and that one can find a representation with smaller bond-dimension by compressing the corresponding matrices.

Note that with the strategy discussed above one constructs an exact matrix product operator representation of the operator $\hat{R}(\hat{\rho})$. Of course, to obtain a matrix product operator approximation one could simply add the individual terms of the operator $\hat{R}(\hat{\rho})$, which are often of matrix product operator structure or straightforwardly converted into the latter. Here, the bond-dimensions of the individual terms add to the overall bond-dimension. To keep the bond-dimension at a certain level, one could compress this operator as described in appendix A to keep the bond-dimension fixed if one exceeds a predetermined threshold. This provides a general strategy to obtain the matrix product operator representation of the operator $\hat{R}(\hat{\rho})$.

Appendix C. Greenberger–Horne–Zeilinger-type states

In this section of the appendix we discuss GHZ-type states of the form

$$|\psi_N(\phi)\rangle = [|0\rangle^{\otimes N/2} |1\rangle^{\otimes N/2} + e^{i\phi} |1\rangle^{\otimes N/2} |0\rangle^{\otimes N/2}] / \sqrt{2} \quad (\text{C.1})$$

in terms of matrix product states and show how to represent the operator $\hat{R}(\hat{\rho})$ comprising the POVM elements efficiently as a matrix product operator. Recall that for the GHZ-type state one needs to incorporate two global observables into the set of POVM elements, see equations (14) and (15) of the main text.

The GHZ-type state as a matrix product state. In this section we set out to find a matrix product state representation of the GHZ-type state. First, note that

$$|0\rangle^{\otimes N/2} |1\rangle^{\otimes N/2} / \sqrt{2} = \sum_{i_1, \dots, i_N} P_1[i_1] \cdots P_N[i_N] |i_1, \dots, i_N\rangle, \quad (\text{C.2})$$

where

$$\begin{aligned} P_1[1] &= \frac{1}{\sqrt{2}}, \\ P_k[1] &= 1, \quad \text{for } k = 2, \dots, N/2, \\ P_k[1] &= 0, \quad \text{for } k = N/2 + 1, \dots, N, \\ P_k[2] &= 0, \quad \text{for } k = 1, \dots, N/2, \\ P_k[2] &= 1, \quad \text{for } k = N/2 + 1, \dots, N. \end{aligned} \quad (\text{C.3})$$

Similarly, we can find a matrix product state representation with bond-dimension $D = 1$ for the second term in the GHZ-type state. Further, adding two matrix product states with

bond-dimension $D_1 = 1 = D_2$ results in a matrix product state with bond-dimension $D = 2$. Finally, we have for the GHZ-type state

$$\begin{aligned}
 P_1[1] &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \\
 P_k[1] &= \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \text{for } k = 2, \dots, N/2, \\
 P_k[1] &= \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad \text{for } k = N/2 + 1, \dots, N-1, \\
 P_N[1] &= \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \\
 P_1[2] &= \frac{e^{i\phi}}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \\
 P_k[2] &= \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad \text{for } k = 2, \dots, N/2, \\
 P_k[2] &= \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \text{for } k = N/2 + 1, \dots, N-1, \\
 P_N[2] &= \begin{pmatrix} 1 \\ 0 \end{pmatrix}.
 \end{aligned} \tag{C.4}$$

Matrix product operator structure of the \hat{R} -operator. For the GHZ-type states we suggest to measure on all blocks of R contiguous sites and additionally determine the expectation values of the observables $\hat{\Pi}_1 = (\mathbb{1} + \hat{X}^{\otimes N})/4$, $\hat{\Pi}_2 = (\mathbb{1} - \hat{X}^{\otimes N})/4$, $\hat{\Pi}_3 = (\mathbb{1} + \hat{Y} \otimes \hat{X}^{\otimes N-1})/4$ and $\hat{\Pi}_4 = (\mathbb{1} - \hat{Y} \otimes \hat{X}^{\otimes N-1})/4$. In this section of the appendix, we show that the operator

$$\hat{R}(\hat{\varrho}) = \frac{1}{M} \sum_{i \in \Delta} \frac{n_i}{p_i} \hat{\Pi}_i = \frac{1}{M} \sum_{i \in \Delta} \frac{n_i}{\text{tr}[\hat{\Pi}_i \hat{\varrho}]} \hat{\Pi}_i \tag{C.5}$$

can still be written efficiently in the matrix product operator formalism when incorporating these global observables. Here, M is the total number of measurements, Δ an index set and n_i denotes the number of times outcome $\hat{\Pi}_i$ is obtained for all $i \in \Delta$. Now, if Δ contains all indices corresponding to the local measurements comprised by Δ_1 plus the four indices contained in Δ_2 labelling the global POVM elements $\hat{\Pi}_1 = (\mathbb{1} + \hat{X}^{\otimes N})/4$, $\hat{\Pi}_2 = (\mathbb{1} - \hat{X}^{\otimes N})/4$, $\hat{\Pi}_3 = (\mathbb{1} + \hat{Y} \otimes \hat{X}^{\otimes N-1})/4$ and $\hat{\Pi}_4 = (\mathbb{1} - \hat{Y} \otimes \hat{X}^{\otimes N-1})/4$, the operator \hat{R} is given by

$$\hat{R}(\hat{\varrho}) = \hat{R}_1(\hat{\varrho}) + \hat{R}_2(\hat{\varrho}), \tag{C.6}$$

where $\hat{R}_1(\hat{\varrho})$ contains the local measurements and $\hat{R}_2(\hat{\varrho})$ the four global POVM elements. Now one agrees that $\hat{R}(\hat{\varrho})$ is indeed a matrix product operator with small bond-dimension. The first term is a matrix product operator with small bond-dimension D_1 due to its locality (see appendix B) and the global POVM elements are all matrix product operators with bond-dimensions equal to 1. Hence, $\hat{R}(\hat{\varrho})$ is a matrix product operator with bond-dimension equal

to $D_1 + 4$. In fact, we can even find a representation of $\hat{R}_2(\hat{\rho})$ with a smaller bond-dimension. We have

$$\begin{aligned}\hat{R}_2(\hat{\rho}) = \sum_{i=1}^4 \frac{f_i}{p_i} \hat{\Pi}_i &= \left(\frac{f_1}{p_1} + \frac{f_2}{p_2} + \frac{f_3}{p_3} + \frac{f_4}{p_4} \right) \cdot \mathbb{1}^{\otimes N}/4 + \left(\frac{f_1}{p_1} - \frac{f_2}{p_2} \right) \cdot \hat{X}^{\otimes N}/4 \\ &+ \left(\frac{f_3}{p_3} - \frac{f_4}{p_4} \right) \cdot \hat{Y} \otimes \hat{X}^{\otimes N-1}/4.\end{aligned}\quad (\text{C.7})$$

In this notation, the second and third terms can be combined into a matrix product operator with bond-dimension 1 such that $\hat{R}_2(\hat{\rho})$ has bond-dimension $D_2 = 2$. Consequently, $\hat{R}(\hat{\rho})$ is a matrix product operator with bond-dimension $D_1 + 2$. The matrices of the matrix product operator representation of $c_1 \cdot \hat{X}^{\otimes N} + c_2 \cdot \hat{Y} \otimes \hat{X}^{\otimes N-1}$ can be chosen as

$$\begin{aligned}P_k[1] &= 0, \quad \text{for } k = 1, \dots, N, \\ P_1[2] &= \sqrt{2}c_1 \\ P_k[2] &= \sqrt{2}, \quad \text{for } k = 2, \dots, N, \\ P_1[3] &= \sqrt{2}c_2 \\ P_k[3] &= 0, \quad \text{for } k = 2, \dots, N, \\ P_k[4] &= 0, \quad \text{for } k = 1, \dots, N,\end{aligned}\quad (\text{C.8})$$

where 1, 2, 3, 4 correspond to the normalized Pauli basis elements $\hat{P}^{(1)} = \mathbb{1}/\sqrt{2}$, $\hat{P}^{(2)} = \hat{X}/\sqrt{2}$, $\hat{P}^{(3)} = \hat{Y}/\sqrt{2}$ and $\hat{P}^{(4)} = \hat{Z}/\sqrt{2}$. Thus, we have

$$c_1 \hat{X}^{\otimes N} + c_2 \hat{Y} \otimes \hat{X}^{\otimes N-1} = \sum_{i=2}^3 P_1[i] P_2[2] \cdots P_L[2] \hat{P}_1^{(i)} \otimes \hat{P}_2^{(2)} \otimes \cdots \otimes \hat{P}_N^{(2)} \quad (\text{C.9})$$

such that $\hat{R}_2(\hat{\rho})$ is a sum of two matrix product operators with bond-dimensions equal to 1.

References

- [1] DiVincenzo D P 2000 *Fortschr. Phys.* **48** 771
- [2] Vogel K and Risken H 1989 *Phys. Rev. A* **40** 2847–9
- [3] Hradil Z, Řeháček J, Fiurášek J and Ježek M 2004 *Lecture Notes in Physics* **649** 59–112
- [4] Chuang I L and Nielsen M A 1997 *J. Mod. Opt.* **44** 2455
- [5] Lundeen J S, Feito A, Coldenstrodt-Ronge H, Pregnell K L, Silberhorn Ch, Ralph T C, Eisert J, Plenio M B and Walmsley I A 2009 *Nature Phys.* **5** 27
- [6] Zhang L, Coldenstrodt-Ronge H, Datta A, Puentes G, Lundeen J S, Jin X-M, Smith B J, Plenio M B and Walmsley I A 2012 *Nature Photon.* **6** 364
- [7] Řeháček J, Hradil Z, Knill E and Lvovsky A I 2007 *Phys. Rev. A* **75** 042108
- [8] Blume-Kohout R 2010 *New J. Phys.* **12** 043034
- [9] James D F V, Kwiat P G, Munro W J and White A G 2001 *Phys. Rev. A* **64** 052312
- [10] Hradil Z 1997 *Phys. Rev. A* **55** R1561
- [11] Rau J 2010 *Phys. Rev. A* **82** 012104
- [12] Häffner H *et al* 2005 *Nature* **438** 643–6
- [13] Leibfried D *et al* 2005 *Nature* **438** 639–42

- [14] Monz T, Schindler P, Barreiro J T, Chwalla M, Nigg D, Coish W A, Harlander M, Hänsel W, Hennrich M and Blatt R 2011 *Phys. Rev. Lett.* **106** 130506
- [15] Yao X-C, Wang T-X, Xu P, Lu H, Pan G-S, Bao X-H, Peng C-Z, Lu C-Y, Chen Y-A and Pan J-W 2012 *Nature Photon.* **6** 225
- [16] Islam R, Senko C, Campbell W C, Korenblit S, Smith J, Lee A, Edwards E E, Wang C-C J, Freericks J K and Monroe C 2013 *Science* **340** 583
- [17] Flammia S T and Liu Y-K 2011 *Phys. Rev. Lett.* **106** 230501
- [18] da Silva M P, Landon-Cardinal O and Poulin D 2011 *Phys. Rev. Lett.* **107** 210404
- [19] Gross D, Liu Y-K, Flammia S T, Becker S and Eisert J 2010 *Phys. Rev. Lett.* **105** 150401
- [20] Gross D 2011 *IEEE Trans. Inform. Theory* **57** 1548–66
- [21] Flammia S T, Gross D, Liu Y-K and Eisert J 2012 *New J. Phys.* **14** 095022
- [22] Landon-Cardinal O and Poulin D 2012 *New J. Phys.* **14** 085004
- [23] Tóth G, Wieczorek W, Gross D, Krischek R, Schwemmer C and Weinfurter H 2010 *Phys. Rev. Lett.* **105** 250403
- [24] Cramer M, Plenio M B, Flammia S T, Somma R, Gross D, Bartlett S D, Landon-Cardinal O, Poulin D and Liu Y-K 2010 *Nature Commun.* **1** 149
- [25] Moroder T, Hyllus P, Tóth G, Schwemmer C, Niggebaum A, Gaile S, Gühne O and Weinfurter H 2012 *New J. Phys.* **14** 105001
- [26] Baumgratz T, Gross D, Cramer M and Plenio M B 2013 *Phys. Rev. Lett.* **111** 020401
- [27] Orlitsky A, Orlitsky S and Rommer S 1995 *Phys. Rev. Lett.* **75** 3537
- [28] Schollwöck U 2011 *Ann. Phys.* **326** 96–192
- [29] Perez-Garcia D, Verstraete F, Wolf M M and Cirac J I 2007 *Quantum Inform. Comput.* **7** 401
- [30] Boyd S and Vandenberghe L 2009 *Convex Optimization* (Cambridge: Cambridge University Press)
- [31] Audenaert K, Eisert J, Plenio M B and Werner R F 2002 *Phys. Rev. A* **66** 042327
- [32] Plenio M B, Eisert J, Dreißig J and Cramer M 2005 *Phys. Rev. Lett.* **94** 060503
- [33] Hastings M B 2006 *Phys. Rev. B* **73** 085115
- [34] Hastings M B 2007 *J. Stat. Mech.* **2007** P08024
- [35] Eisert J, Cramer M and Plenio M B 2010 *Rev. Mod. Phys.* **82** 277
- [36] Brandão F G S L and Horodecki M 2012 arXiv:1206.2947
- [37] Zwolak M and Vidal G 2004 *Phys. Rev. Lett.* **93** 207205