Permutationally Invariant Quantum Tomography

Géza Tóth,[1,2,3] Witlef Wieczorek,[4,5,*] David Gross,[6] Roland Krischek,[4,5] Christian Schwemmer,[4,5] and Harald Weinfurter[4,5]

[1]*Department of Theoretical Physics, The University of the Basque Country, P.O. Box 644, E-48080 Bilbao, Spain*
[2]*IKERBASQUE, Basque Foundation for Science, E-48011 Bilbao, Spain*
[3]*Research Institute for Solid State Physics and Optics,*
Hungarian Academy of Sciences, P.O. Box 49, H-1525 Budapest, Hungary
[4]*Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Strasse 1, D-85748 Garching, Germany*
[5]*Fakultät für Physik, Ludwig-Maximilians-Universität, D-80797 München, Germany*
[6]*Institute for Theoretical Physics, Leibniz University Hannover, D-30167 Hannover, Germany*
(Dated: February 14, 2022)

We present a scalable method for the tomography of large multiqubit quantum registers. It acquires information about the permutationally invariant part of the density operator, which is a good approximation to the true state in many, relevant cases. Our method gives the best measurement strategy to minimize the experimental effort as well as the uncertainties of the reconstructed density matrix. We apply our method to the experimental tomography of a photonic four-qubit symmetric Dicke state.

Because of the the rapid development of quantum experiments, it is now possible to create highly entangled multiqubit states using photons [1–5], trapped ions [6], and cold atoms [7]. So far, the largest implementations that allow for an individual readout of the particles involve on the order of 10 qubits. This number will soon be overcome, for example, by using several degrees of freedom within each particle to store quantum information [8]. Thus, a new regime will be reached in which a complete state tomography is impossible even from the point of view of the storage place needed on a classical computer. At this point the question arises: Can we still extract useful information about the quantum state created?

In this Letter we propose permutationally invariant (PI) tomography in multiqubit quantum experiments [9]. Concretely, instead of the density matrix ϱ, we propose to determine the PI part of the density matrix defined as

$$\varrho_{\text{PI}} = \frac{1}{N!} \sum_k \Pi_k \varrho \Pi_k, \qquad (1)$$

where Π_k are all the permutations of the qubits. Reconstructing ϱ_{PI} has been considered theoretically for spin systems (see, e.g., Ref. [10]). Recently it has been pointed out that photons in a single mode optical fiber will always be in a PI state and that there is only a small set of measurements needed for their characterization [11, 12].

Here, we develop a provably optimal scheme, which is feasible for large multiqubit systems: For our method, the measurement effort increases only *quadratically* with the size of the system. Our approach is further motivated by the fact that almost all multipartite experiments are done with PI quantum states [1–4, 6]. Thus, the density matrix obtained from PI tomography is expected to be close to the one of the experimentally achieved state. The expectation values of symmetric operators, such as some entanglement witnesses, and fidelities with respect to symmetric states are the same for both density matrices and are thus obtained exactly from PI tomography [2–4]. Finally, if ϱ_{PI} is entangled, so is the state ϱ of the system, which makes PI tomography a useful and efficient tool for entanglement detection.

Below, we summarize the four main contributions of this Letter. We restrict our attention to the case of N qubits – higher-dimensional systems can be treated similarly.

1. In most experiments, the qubits can be individually addressed whereas nonlocal quantities cannot be measured directly. The experimental effort is then characterized by the number of local measurement settings needed, where "setting" refers to the choice of one observable per qubit, and repeated von Neumann measurements in the observables' eigenbases [13]. Here, we compute the minimal number of measurement settings required to recover ϱ_{PI}.

2. The requirement that the number of settings be minimal does not uniquely specify the tomographic protocol. On the one hand, there are infinitely many possible choices for the local settings that are both minimal and give sufficient information to find ϱ_{PI}. On the other hand, for each given setting, there are many ways of estimating the unknown density operator from the collected data. We present a systematic method to find the optimal scheme through statistical error analysis.

3. Next, we turn to the important problem of gauging the information loss incurred due to restricting attention to the PI part of the density matrix. We describe an easy test measurement that can be used to judge the applicability of PI tomography *before* it is implemented.

4. Finally, we demonstrate that these techniques are viable in practice by applying them to a photonic experiment observing a four-qubit symmetric Dicke state.

Minimizing the number of settings. We will now present our first main result.
Observation 1. For a system of N qubits, permutationally invariant tomography can be performed with

$$\mathcal{D}_N = \binom{N+2}{N} = \frac{1}{2}(N^2 + 3N + 2) \qquad (2)$$

local settings. It is not possible to perform such a tomography

with fewer settings.

Proof. First, we need to understand the information obtainable from a single measurement setting. We assume that for every given setting, the same basis is measured at every site [14]. Measuring a local basis $\{|\phi_1\rangle, |\phi_2\rangle\}$ is equivalent to estimating the expectation value of the traceless operator $A = |\phi_1\rangle\langle\phi_1| - |\phi_2\rangle\langle\phi_2|$. Merely by measuring $A^{\otimes N}$, it is possible to obtain all the N expectation values

$$\langle (A^{\otimes(N-n)} \otimes \mathbb{1}^{\otimes n})_{\mathrm{PI}}\rangle, \qquad (n = 0, \ldots, N-1) \quad (3)$$

and, conversely, that is all the information obtainable about ϱ_{PI} from a single setting.

Next, we will use the fact that any PI density operator can be written as a linear combination of the pairwise orthogonal operators $(X^{\otimes k} \otimes Y^{\otimes l} \otimes Z^{\otimes m} \otimes \mathbb{1}^{\otimes n})_{\mathrm{PI}}$, where X, Y and Z are the Pauli matrices. We consider the space spanned by these operators for one specific value of n. Simple counting shows that its dimension is $\mathcal{D}_{(N-n)}$. The same space is spanned by $\mathcal{D}_{(N-n)}$ generic operators of the type $(A^{\otimes(N-n)} \otimes \mathbb{1}^{\otimes n})_{\mathrm{PI}}$. We draw two conclusions: First, any setting gives at most one expectation value for every such space. Hence the number of settings cannot be smaller than the largest dimension, which is $\mathcal{D}_N$. Second, a generic choice of $\mathcal{D}_N$ settings is sufficient to recover the correlations in each of these spaces, and hence completely characterizes ϱ_{PI}. This concludes the proof [15].

The proof implies that there are real coefficients $c_j^{(k,l,m)}$ such that

$$\langle (X^{\otimes k} \otimes Y^{\otimes l} \otimes Z^{\otimes m} \otimes \mathbb{1}^{\otimes n})_{\mathrm{PI}}\rangle =$$
$$\sum_{j=1}^{\mathcal{D}_N} c_j^{(k,l,m)} \langle (A_j^{\otimes(N-n)} \otimes \mathbb{1}^{\otimes n})_{\mathrm{PI}}\rangle. \quad (4)$$

We will refer to the numbers on the left-hand side of Eq. (4) as the elements of the *generalized Bloch vector*. The expectation values on the right-hand side can be obtained by measuring the settings with A_j for $j = 1, 2, ..., \mathcal{D}_N$.

Minimizing uncertainties. We now have to determine the optimal scheme for PI tomography. To this end, we define our measure of statistical uncertainty as the sum of the variances of all the Bloch vector elements

$$(\mathcal{E}_{\mathrm{total}})^2 =$$
$$\sum_{k+l+m+n=N} \mathcal{E}^2\left[(X^{\otimes k} \otimes Y^{\otimes l} \otimes Z^{\otimes m} \otimes \mathbb{1}^{\otimes n})_{\mathrm{PI}}\right] \times$$
$$\left(\frac{N!}{k!l!m!n!}\right), \quad (5)$$

where the term with the factorials is the number of different permutations of $X^{\otimes k} \otimes Y^{\otimes l} \otimes Z^{\otimes m} \otimes \mathbb{1}^{\otimes n}$. Based on Eq. (4), the variance of a single Bloch vector element is

$$\mathcal{E}^2[(X^{\otimes k} \otimes Y^{\otimes l} \otimes Z^{\otimes m} \otimes \mathbb{1}^{\otimes n})_{\mathrm{PI}}]$$
$$= \sum_{j=1}^{\mathcal{D}_N} |c_j^{(k,l,m)}|^2 \mathcal{E}^2[(A_j^{\otimes(N-n)} \otimes \mathbb{1}^{\otimes n})_{\mathrm{PI}}]. \quad (6)$$

Equation (5) can be minimized by changing the A_j matrices and the $c_j^{(k,l,m)}$ coefficients. We consider the coefficients first. For any Bloch vector element, finding $c_j^{(k,l,m)}$'s that minimize the variance Eq. (6) subject to the constraint that equality holds in Eq. (4) is a least squares problem. It has an analytic solution obtained as follows: Write the operator on the left-hand side of Eq. (6) as a vector $\vec{v}$ (with respect to some basis). Likewise, write the operators on the right-hand side as $\vec{v}_j$ and define a matrix $V = [\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_{\mathcal{D}_N}]$. Then Eq. (4) can be cast into the form $\vec{v} = V\vec{c}$, where $\vec{c}$ is a vector of the $c_j^{(k,l,m)}$ values for given (k, l, m). If E is the diagonal matrix with entries $E_{j,j}^2 = \mathcal{E}^2[(A_j^{\otimes(N-n)} \otimes \mathbb{1}^{\otimes n})_{\mathrm{PI}}]$, then the optimal solution is $\vec{c} = E^{-2}V^T(VE^{-2}V^T)^{-1}\vec{v}$, where the inverse is taken over the range [16].

Equipped with a method for obtaining the optimal $c_j^{(k,l,m)}$'s for every fixed set of observables A_j, it remains to find the best settings to measure. Every qubit observable can be defined by the measurement directions $\vec{a}_j$ using $A_j = a_{j,x}X + a_{j,y}Y + a_{j,z}Z$. Thus, the task is to identify $\mathcal{D}_N$ measurement directions on the Bloch sphere minimizing the variance. In general, finding the globally optimal solution of high-dimensional problems is difficult. In our case, however, $\mathcal{E}_{\mathrm{total}}$ seems to penalize an inhomogeneous distribution of the $\vec{a}_j$ vectors; thus, using evenly distributed vectors as an initial guess, usual minimization procedures can be used to decrease $\mathcal{E}_{\mathrm{total}}$ and obtain satisfactory results [16].

The variance $\mathcal{E}^2[(A_j^{\otimes(N-n)} \otimes \mathbb{1}^{\otimes n})_{\mathrm{PI}}]$ of the observed quantities depends on the physical implementation. In the photonic setup below, we assume Poissonian distributed counts. It follows that (see also Refs. [17, 18])

$$\mathcal{E}^2[(A_j^{\otimes(N-n)} \otimes \mathbb{1}^{\otimes n})_{\mathrm{PI}}] = \frac{[\Delta(A_j^{\otimes(N-n)} \otimes \mathbb{1}^{\otimes n})_{\mathrm{PI}}]_{\varrho_0}^2}{\lambda_j - 1}, \quad (7)$$

where $(\Delta A)_\varrho^2 = \langle A^2\rangle_\varrho - \langle A\rangle_\varrho^2$, ϱ_0 is the state of the system and λ_j is the parameter of the Poissonian distribution, which equals the expected value of the total number of counts for the setting j. The variance depends on the unknown state. If we have preliminary knowledge of the likely form of ϱ_0, we should use that information in the optimization. Otherwise, ϱ_0 can be set to the completely mixed state. For the latter, straightforward calculation shows that $\mathcal{E}^2[(A_j^{\otimes(N-n)} \otimes \mathbb{1}^{\otimes n})_{\mathrm{PI}}] = \binom{N}{n}^{-1}/(\lambda_j - 1)$. For another implementation, such as trapped ions, our scheme for PI tomography can be used after replacing Eq. (7) by a formula giving the variance for that implementation.

Estimating the information loss due to symmetrization. It is important to know how close the PI quantum state is to the state of the system as PI tomography should serve as an alternative of full state tomography for experiments aiming at the preparation of PI states.

Observation 2. The fidelity between the original state and the permutationally invariant state, $F(\varrho, \varrho_{\mathrm{PI}})$, can be estimated from below as $F(\varrho, \varrho_{\mathrm{PI}}) \geq \langle P_{\mathrm{s}}\rangle_\varrho^2$, where $P_{\mathrm{s}} =$
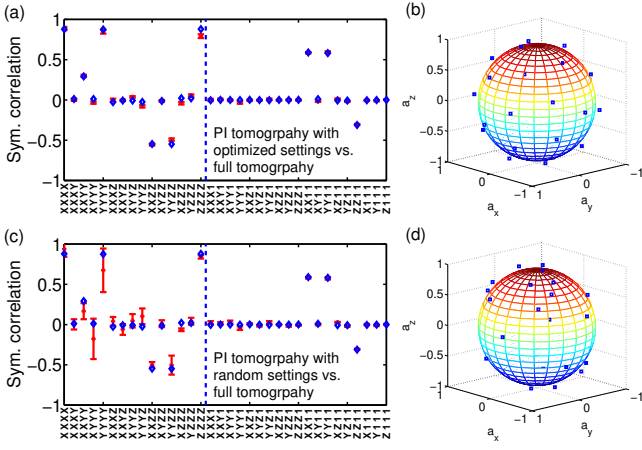
Figure 1: (Color online) (a) Comparison of the 34 symmetrized correlations coming from (crosses with error bars) 15 permutationally invariant measurement settings with optimized A_j matrices for $N = 4$ qubits and (diamonds) from full tomography requiring 81 local settings. The average uncertainty of all symmetrized correlations obtained from full tomography is ± 0.022, and is not shown in the figure. The labels refer to symmetrized correlations of the form given in the left-hand side of Eq. (4). The results corresponding to the 15 full four-qubit correlations are left from the vertical dashed line. (b) Measurement directions. A point at (a_x, a_y, a_z) corresponds to measuring operator $a_x X + a_y Y + a_z Z$. (c) Results for randomly chosen A_j matrices and (d) corresponding measurement directions.
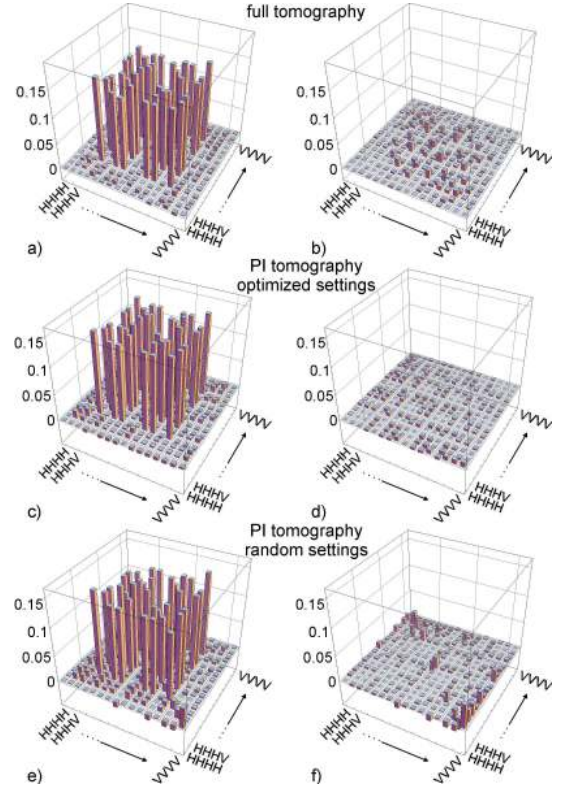

Figure 2: (Color online) (a) The real and (b) imaginary parts of the density matrix coming from full tomography. (c),(d) The same for permutationally invariant tomography with optimized and (e),(f) random measurement directions, respectively.

$\sum_{n=0}^{N} |D_N^{(n)}\rangle\langle D_N^{(n)}|$ is the projector to the N-qubit symmetric subspace, and the symmetric Dicke state is defined as $|D_N^{(n)}\rangle = \binom{N}{n}^{-\frac{1}{2}} \sum_k \mathcal{P}_k(|0\rangle^{\otimes(N-n)} \otimes |1\rangle^{\otimes n})$, where the summation is over all the different permutations of the qubits. Observation 2 can be proved based on Ref. [19] and elementary matrix manipulations. Note that Observation 2 makes it possible to estimate $F(\varrho, \varrho_{\mathrm{PI}})$ based on knowing only ϱ_{PI}.

Lower bounds on the fidelity to symmetric Dicke states, i.e., $\mathrm{Tr}(|D_N^{(n)}\rangle\langle D_N^{(n)}|\varrho)$ can efficiently be obtained by measuring X, Y and Z on all qubits, i.e., measuring only three local settings independent of N [20]. With the same measurements, one can also obtain a lower bound on the overlap between the state and the symmetric subspace. For four qubits, this can be done based on $P_\mathrm{s} \geq [(J_x^4 + J_y^4 + J_z^4) - (J_x^2 + J_y^2 + J_z^2)]/18$, where $J_x = (1/2)\sum_k X_k$, $J_y = (1/2)\sum_k Y_k$, etc. Operators for estimating $\langle P_\mathrm{s}\rangle$ for $N = 6, 8$ are given in Ref. [16]. This allows one to judge how suitable the quantum state is for PI tomography *before* such a tomography is carried out.

Experimental results. We demonstrate the method and the benefits of our algorithm for PI tomography for a 4-qubit symmetric Dicke state with two excitations $|D_4^{(2)}\rangle$. First, we optimize the $\vec{a}_j$'s and the $c_j^{(k,l,m)}$'s for $\varrho_0 = \mathbb{1}/16$ and only for the uncertainty of full four-qubit correlation terms, which means that when computing $\mathcal{E}_{\mathrm{total}}$, we carry out the summation in Eq. (5) only for the terms with $n = 0$. With simple numerical optimization, we were looking for the set of A_j basis matrices that minimize the uncertainty of the full correlation terms. Then, we also looked for the basis matrices that minimize the

sum of the squared error of all the Bloch vector elements and considered also density matrices different from white noise, such as a pure Dicke state mixed with noise. We find that the gain in terms of decreasing the uncertainties is negligible in our case and that it is sufficient to optimize for $\varrho_0 = \mathbb{1}/16$ and for the full correlation terms. To demonstrate the benefits of the optimization of the measurement directions, we also compare the results with those obtained with randomly distributed basis matrices.

The Dicke state was observed in a photonic system. Essentially, four photons emitted by the second-order collinear type-II spontaneous parametric down-conversion process were symmetrically distributed into four spatial modes. Upon detection of one photon from each of the outputs, the state $|D_4^{(2)}\rangle$ is observed. Polarization analysis in each mode is used to characterize the experimentally observed state. We collected data for each setting for 5 minutes, with an average count rate of 410 per minute. The experimental setup has been described in detail in Refs. [2, 3].

First, to check the applicability of the PI tomography, we apply our tools described above requiring only the measurement of the three settings, $X^{\otimes 4}, Y^{\otimes 4}$ and $Z^{\otimes 4}$. We determine the expectation value of the projector to the symmetric subspace, yielding $\langle P_\mathrm{s}\rangle \geq 0.905 \pm 0.015$. Based on Observation 2, we obtain $F(\varrho, \varrho_{PI}) \geq 0.819 \pm 0.028$. These results show

that the state is close to be PI and has a large overlap with the symmetric subspace. Thus, it makes sense to apply PI tomography.

For PI tomography of a four-qubit system, the measurement of 15 settings is needed. We used Eq. (4) to obtain the Bloch vector elements from the experimentally measured quantities. This way, we could obtain all the 34 symmetric correlations of the form $(X^{\otimes k} \otimes Y^{\otimes l} \otimes Z^{\otimes m} \otimes \mathbb{1}^{\otimes n})_{\mathrm{PI}}$. In Fig. 1, we give the values of the correlations for optimized and for randomly chosen measurement directions, compared to the results obtained from full tomography, which needed 81 measurement settings. As can be seen in Fig. 1, the uncertainty for the optimized settings is considerably smaller than the one for the randomly chosen settings. Moreover, the results from the optimized settings fit very well the results of the full tomography. In Fig. 2, we compare the density matrices obtained from full tomography [Fig. 2(a)], from PI tomography for optimized [Fig. 2(b)] and for random measurement directions [Fig. 2(c)]. Because of noise, the fidelity of the result of the full tomography with respect to $|D_4^{(2)}\rangle$ is 0.873 ± 0.005, which is similar to the fidelity of the results of the PI tomography with optimized settings, 0.852 ± 0.009 [21]. In contrast, for the method using random measurement directions, the fidelity is 0.814 ± 0.059, for which the uncertainty is the largest compared to all previous fidelity values. Finally, we also computed the fidelity of the results with respect to the PI density matrix obtained from full tomography [22]. The results of the PI tomography with optimized settings shows a good agreement with full tomography, the fidelity is 0.947, which is quite close to the fidelity between the results of full tomography and its PI part, 0.964. On the other hand, for the PI tomography with random settings the corresponding fidelity is much lower, 0.880. Overall, the PI tomography shows a good agreement with the full tomography for this particular experiment. However, a reasonable choice of measurement directions is needed to obtain uncertainties in the reconstructed Bloch vector elements similar to the ones from full tomography.

Finally, let us comment on how our method can be extended to lager systems. Permutationally invariant operators can be represented efficiently on a digital computer in the basis of $(X^{\otimes k} \otimes Y^{\otimes l} \otimes Z^{\otimes m} \otimes \mathbb{1}^{\otimes n})_{\mathrm{PI}}$ operators. We determined the optimal A_j operators for PI tomography for systems with $N = 6, 8, ..., 14$ qubits. To have the same maximum uncertainty of the Bloch vector elements as for the $N = 4$ case, one has to increase the counts per setting by less than 50% [16].

In summary, we presented a scalable method for permutationally invariant tomography, which can be used in place of full state tomography in experiments that aim at preparing permutationally invariant many-qubit states. For our approach, the same operator has to be measured on all qubits, which is a clear advantage in some experiments. We showed how to choose the measurements such that the uncertainty in the reconstructed density matrix is the smallest possible. This paves the way of characterizing permutationally invariant states of many qubits in various physical systems. Moreover, this work also shows that, given some knowledge or justifiable assumptions, there is a way to obtain scalable state tomography for multiqubit entangled states.

* Present address: Faculty of Physics, University of Vienna, Boltzmanngasse 5, A-1090 Wien, Austria

[1] J.-W. Pan *et al.*, Nature (London) **403**, 515 (2000); M. Bourennane *et al.*, Phys. Rev. Lett. **92**, 087902 (2004); N. Kiesel *et al.*, *ibid.* **95**, 210502 (2005).

[2] N. Kiesel *et al.*, Phys. Rev. Lett. **98**, 063604 (2007).

[3] W. Wieczorek *et al.*, Phys. Rev. Lett. **103**, 020504 (2009); R. Krischek *et al.*, Nature Photon. **4**, 170 (2010).

[4] R. Prevedel *et al.*, Phys. Rev. Lett. **103**, 020503 (2009).

[5] W. Wieczorek *et al.*, Phys. Rev. Lett. **101**, 010503 (2008).

[6] C. A. Sackett *et al.*, Nature (London) **404**, 256 (2000); H. Häffner *et al.*, Nature (London) **438**, 643 (2005).

[7] O. Mandel *et al.*, Nature (London) **425**, 937 (2003).

[8] C. Cinelli *et al.*, Phys. Rev. Lett. **95**, 240405 (2005); G. Vallone *et al.*, *ibid.* **98**, 180502 (2007); W.-B. Gao *et al.*, Nature Phys. **6**, 331 (2010).

[9] For other approaches see D. Gross *et al.*, Phys. Rev. Lett. **105**, 150401 (2010); M. Cramer and M.B. Plenio, arXiv:1002.3780; S.T. Flammia *et al.*, arXiv:1002.3839; O. Landon-Cardinal *et al.*, arXiv:1002.4632.

[10] G. M. D'Ariano *et al.*, J. Opt. B **5**, 77 (2003).

[11] R.B.A. Adamson *et al.*, Phys. Rev. Lett. **98**, 043601 (2007).

[12] L. K. Shalm *et al.*, Nature **457**, 67 (2009).

[13] O. Gühne and G. Tóth, Phys. Rep. **474**, 1 (2009).

[14] Otherwise more than $\mathcal{D}_N$ settings are necessary [16].

[15] This is connected to a general idea: It is expected that the determination of an operator within a subspace whose dimension depends polynomially on N needs a number of settings increasing also polynomially with N.

[16] See supplementary material at http://link.aps.org/ supplemental/10.1103/PhysRevLett.000.000000 for additional derivations and experimental results.

[17] C. Schmid, Ph.D. Thesis, Ludwig-Maximillian-Universität, Munich, Germany, 2008; D.F.V. James *et. al*, Phys. Rev. A **64**, 052312 (2001).

[18] B. Jungnitsch *et. al*, Phys. Rev. Lett. **104**, 210401 (2010).

[19] J.A. Miszczak *et al.*, J. Quant. Inf. Comp. **9**, 0103 (2009).

[20] G. Tóth *et al.*, New J. Phys. **11**, 083002 (2009).

[21] Expectation values are obtained directly from the measured data, rather than from ρ_{PI}.

[22] Values without error are deduced from fitted matrices obtained via maximum likelihood estimation [17].

Supplementary Material

The supplement contains some derivations to help to understand the details of the proofs of the main text. It also contains some additional experimental results.

Proof of that we have to measure the same operator on all qubits. From the proof of Observation 1, we know that at least $\mathcal{D}_N$ measurements are needed to get the expectation values of all the $\mathcal{D}_N$ independent symmetric full N-particle correlations. What if we measure $\mathcal{D}_N$ settings, but several of them are not $\{A_j, A_j, ..., A_j\}$-type, but $\{A_j^{(1)}, A_j^{(2)}, ..., A_j^{(N)}\}$-type, i.e., we do not measure the same operator on all qubits? Each setting makes it possible to get a single operator containing full N-qubit correlations. Let us denote this operator by M_k for $k = 1, 2, ..., \mathcal{D}_N$. Then, we know the expectation value of any operator of the space defined by the M_k operators. However, not all M_k's are permutationally invariant. Thus, the size of the PI subspace of the space of the M_k operators is less than $\mathcal{D}_N$. We do not have $\mathcal{D}_N$ linearly independent symmetric operators in this space. Thus, $\mathcal{D}_N$ measurement settings are sufficient to measure ϱ_{PI} only if we have settings of the type $\{A_j, A_j, ..., A_j\}$.

Derivation of Eq. (7). The eigen-decomposition of the correlation term is

$$(A_j^{\otimes(N-n)} \otimes \mathbb{1}^{\otimes n})_{\mathrm{PI}} = \sum_k \Lambda_{j,n,k} |\Phi_{j,k}\rangle\langle\Phi_{j,k}|. \quad (S1)$$

The individual counts $N_C(A_j)_k$ follow a Poissonian distribution $f(n_c, \lambda_{j,k})$, where $\lambda_{j,k}$ are the parameters of the Poissonian distributions and $\sum_k \lambda_{j,k} = \lambda_j$. The conditional variance, knowing that the total count is $N_C(A_j)$, is

$$\mathcal{E}^2[(A_j^{\otimes(N-n)} \otimes \mathbb{1}^{\otimes n})_{\mathrm{PI}} | N_C(A_j)] = \frac{[\Delta(A_j^{\otimes(N-n)} \otimes \mathbb{1}^{\otimes n})_{\mathrm{PI}}]^2}{N_C(A_j)}. \quad (S2)$$

After straightforward algebra, the variance is obtained as

$$\mathcal{E}^2[(A_j^{\otimes(N-n)} \otimes \mathbb{1}^{\otimes n})_{\mathrm{PI}}]$$
$$= \sum_m f(m, \lambda_j) \mathcal{E}^2[(A_j^{\otimes(N-n)} \otimes \mathbb{1}^{\otimes n})_{\mathrm{PI}} | N_C(A_j) = m]$$
$$= \frac{[\Delta(A_j^{\otimes(N-n)} \otimes \mathbb{1}^{\otimes n})_{\mathrm{PI}}]^2}{\lambda_j - 1}. \quad (S3)$$

Similar results can be obtained through assuming Poissonian measurement statistics and Gaussian error propagation [S1,S2]. If $\varrho_0 = \mathbb{1}/2^N$, then $\Delta(A_j^{\otimes(N-n)} \otimes \mathbb{1}^{\otimes n})_{\mathrm{PI}}$ is independent from the choice of A_j. By substituting $A_j = Z$, straightforward calculations gives

$$\mathcal{E}^2[(A_j^{\otimes(N-n)} \otimes \mathbb{1}^{\otimes n})_{\mathrm{PI}}] = \frac{\binom{N}{n}^{-1}}{\lambda_j - 1}. \quad (S4)$$

Obtaining the formula for $c_j^{(k,l,m)}$ for the smallest error. We look for $c_j^{(k,l,m)}$ for which the squared uncertainty given in Eq. (6) is the smallest. In the following, we use the definition given in the main text for $\vec{c}$, $\vec{v}$, V and E. Thus, V is matrix mapping a large space $\mathbb{R}^l$ to a small space $\mathbb{R}^s$. Let E be a non-singular diagonal matrix in the small space. We have to solve

$$\min_{\vec{c}} \|E\vec{c}\|^2 \qquad \text{s.t.} \qquad V\vec{c} = \vec{v}, \quad (S5)$$

where $\|\vec{a}\|$ is the Euclidean norm of $\vec{a}$. Using Lagrangian multipliers, we write down the condition for a minimum fulfilling the constraints $V\vec{c} = \vec{v}$

$$\nabla_{\vec{c}}\Big\{\vec{c}^T E^2 \vec{c} + \sum_{i=1}^s \lambda_i [(V\vec{c})_i - w_i]\Big\} = 0. \quad (S6)$$

Hence, the condition for a local (and, due to convexity, global) minimum is

$$\vec{c} = \frac{1}{2} E^{-2} V^T \vec{\lambda}, \quad (S7)$$

where $\lambda \in \mathbb{R}^s$ is the vector of multipliers. In other words, we have a minimum if and only if $\vec{c} \in \mathrm{range}\, E^{-2} V^T$. Because the range of V^T is an s-dimensional subspace in $\mathbb{R}^l$, there is a *unique* $\vec{c}$ in that range such that $V\vec{c} = \vec{v}$. A solution in a closed form can be obtained as

$$c = E^{-2} V^T (V E^{-2} V^T)^{-1} \vec{v}. \quad (S8)$$

Simple calculation shows that the $V\vec{c} = \vec{v}$ condition holds

$$Vc = V E^{-2} V^T (V E^{-2} V^T)^{-1} \vec{v} = \vec{v}. \quad (S9)$$

Proof of Observation 2. The eigenstates of $\vec{J}^2 = J_x^2 + J_y^2 + J_z^2$ are usually labelled by $|j, m, \alpha\rangle$, where $\vec{J}^2 |j, m, \alpha\rangle = j(j+1)|j, m, \alpha\rangle$, $J_z|j, m, \alpha\rangle = m|j, m, \alpha\rangle$, and α is used to label the different eigenstates having the same j and m [S3]. Let $P_{j,\alpha}$ denote the projector to the subspace of a given j and α. The number of subspaces is denoted by N_{SS}, and, for a given N, it can be calculated from group theory. Moreover, $P_{\mathrm{s}} \equiv P_{N/2,1}$. Using this notation, $\varrho_{\mathrm{PI}} = \sum_{j,\alpha} P_{j,\alpha} \varrho P_{j,\alpha} = (P_{\mathrm{s}} \varrho P_{\mathrm{s}}) + \sum_{j<N/2,\alpha} (P_{j,\alpha} \varrho P_{j,\alpha})$. In the basis of $\vec{J}^2$ eigenstates, ϱ_{PI} can be written as a block diagonal matrix

$$\varrho_{\mathrm{PI}} = \bigoplus_{j,\alpha} (\langle P_{j,\alpha}\rangle_\varrho \hat{\varrho}_{j,\alpha}), \quad (S10)$$

where $\hat{\varrho}_{j,\alpha}$ are density matrices of size $(2j+1) \times (2j+1)$. In another context,

$$\varrho_{\mathrm{PI}} = \sum_{j,\alpha} \langle P_{j,\alpha}\rangle_\varrho \varrho_{j,\alpha}, \quad (S11)$$

where $\varrho_{j,\alpha} = P_{j,\alpha} \varrho P_{j,\alpha} / \mathrm{Tr}(P_{j,\alpha} \varrho P_{j,\alpha})$. Based on that, we obtain

$$F(\varrho, \varrho_{j,\alpha}) = \langle P_{j,\alpha}\rangle_\varrho. \quad (S12)$$

Then, due to the separate concavity of the fidelity, i.e., $F(\varrho, p_1\varrho_1 + p_2\varrho_2) \geq p_1 F(\varrho, \varrho_1) + p_2 F(\varrho, \varrho_2)$, we obtain

Table S1: Fidelities to the 4-qubit Dicke states.

| measurement	$|D_4^{(0)}\rangle$	$|D_4^{(1)}\rangle$	$|D_4^{(2)}\rangle$	$|D_4^{(3)}\rangle$	$|D_4^{(4)}\rangle$	Σ
full tomography	-0.001 ± 0.002	0.023 ± 0.004	0.873 ± 0.005	0.026 ± 0.004	0.002 ± 0.002	0.922
full tomography (max-like)	0.001	0.021	0.869	0.023	0	0.914
PI tomography	-0.001 ± 0.002	0.040 ± 0.007	0.852 ± 0.009	0.036 ± 0.007	-0.002 ± 0.002	0.925
PI tomography (max-like)	0.003	0.038	0.850	0.037	0	0.928
PI tomography (ran)	0.000 ± 0.002	0.055 ± 0.027	0.814 ± 0.059	0.023 ± 0.027	0.001 ± 0.002	0.893
PI tomography (ran,max-like)	0.004	0.050	0.816	0.020	0.007	0.897
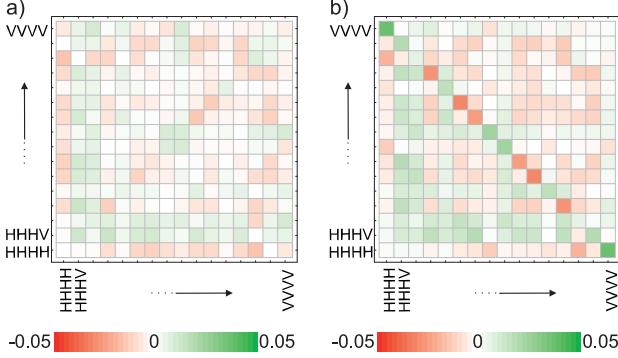
Figure S1: (a) The difference of the real part of the density matrices from optimized settings and the one of full tomography. (b) The difference of the density matrices from random settings and the one of full tomography. For the former, no clear structure is observed, whereas for the latter the largest difference is observed for the antidiagonal elements.

$F(\varrho, \varrho_{\mathrm{PI}}) \geq \langle P_s \rangle_\varrho F(\varrho, \varrho_s) + \sum_{j<N/2,\alpha} \langle P_{j,\alpha} \rangle_\varrho F(\varrho, \varrho_{j,\alpha})$. Substituting Eq. (S12) into this inequality, we obtain $F(\varrho, \varrho_{\mathrm{PI}}) \geq \langle P_s \rangle_\varrho^2 + \sum_{j<N/2,\alpha} \langle P_{j,\alpha} \rangle_\varrho^2$. Using the fact that $\langle P_s \rangle_\varrho + \sum_{j<N/2,\alpha} \langle P_{j,\alpha} \rangle_\varrho = 1$, we obtain

$$F(\varrho, \varrho_{\mathrm{PI}}) \geq \langle P_s \rangle_\varrho^2 + \frac{(1 - \langle P_s \rangle_\varrho)^2}{N_{\mathrm{SS}} - 1}. \quad (S13)$$

In many practical situations, the state ϱ is almost symmetric and N is large. In such cases the second term in Eq. (S13) is negligible. Thus, a somewhat weaker bound presented in Observation 2 can be used.

Numerical optimization used to minimize $\mathcal{E}_{\mathrm{total}}$. The measurement directions minimizing $\mathcal{E}_{\mathrm{total}}$ can be obtained as follows. Let us represent the measurement directions by three-dimensional vectors $\{\vec{a}_j\}_{j=1}^{\mathcal{D}_N}$. The operators can be obtained as $A_j = a_{j,x}X + a_{j,y}Y + a_{j,z}Z$.

First, we need an initial guess. This can come from a set of randomly chosen vectors representing the measurement directions. One can also use the result of a minimization for some measure that characterizes how equally the vectors are distributed. Such a measure is defined by

$$\mathcal{F}(\{v_j\}) = \sum_{k,l} (\vec{v}_k \cdot \vec{v}_l)^{2m}, \quad (S14)$$

where $\vec{v}_k$ represent the measurement directions and $\cdot$ is the scalar product and m is an integer. Such cost functions, called frame potentials, appear in the theory of t-designs essentially for the same purpose.

After we obtain the initial guess from such a procedure, we start an optimization for decreasing $\mathcal{E}_{\mathrm{total}}$. At each iteration of the method, we change the measurement directions by rotating them with a small random angle around a randomly chosen axis. If the change decreases $\mathcal{E}_{\mathrm{total}}$, then we keep the new measurement directions, while if it does not then we discard it. We repeat this procedure until $\mathcal{E}_{\mathrm{total}}$ does not change significantly.

Three-setting witness for estimating the fidelity The three-setting witness for detecting genuine multipartite entanglement in the vicinity of the Dicke state is [S4]

$$\mathcal{W}_{\mathrm{D}(4,2)}^{(\mathrm{P3})} = 2 \cdot \mathbb{1} + \tfrac{1}{6}(J_x^2 + J_y^2 - J_x^4 - J_y^4) + \tfrac{31}{12}J_z^2 - \tfrac{7}{12}J_z^4. \quad (S15)$$

For this witness we have [S4]

$$\mathcal{W}_{\mathrm{D}(4,2)}^{(\mathrm{P3})} - 3\mathcal{W}_{\mathrm{D}(4,2)}^{(\mathrm{P})} \geq 0, \quad (S16)$$

where the projector witness is defined as

$$\mathcal{W}_{\mathrm{D}(4,2)}^{(\mathrm{P})} = \tfrac{2}{3} \cdot \mathbb{1} - |D_4^{(2)}\rangle\langle D_4^{(2)}|. \quad (S17)$$

Hence, the fidelity with respect to the state $|D_4^{(2)}\rangle$ is bounded from below as [S4]

$$F_{\mathrm{D}(4,2)} \geq \tfrac{2}{3} - \tfrac{1}{3}\langle \mathcal{W}_{\mathrm{D}(4,2)}^{(\mathrm{P3})} \rangle. \quad (S18)$$

Fidelities with respect to the four-qubit Dicke states. In Table S1 we summarize the results for full tomography (full) and for permutationally invariant tomography (PI) for random (ran) and optimized (opt) directions. To obtain a physical density matrix with non-negative eigenvalues we perform a maximum-likelihood fit (max-like) of the measured data. In Fig. S1, the differences between the density matrix obtained from full tomography and the ones obtained from permutationally invariant tomography can be seen.

Efficient representation of permutationally invariant operators on a digital computer. Every PI operator O can be decomposed as

$$O = \sum_{k+l+m+n=N} c_{k,l,m,n}^{(O)} (X^{\otimes k} \otimes Y^{\otimes l} \otimes Z^{\otimes m} \otimes \mathbb{1}^{\otimes n})_{\mathrm{PI}}. \quad (S19)$$
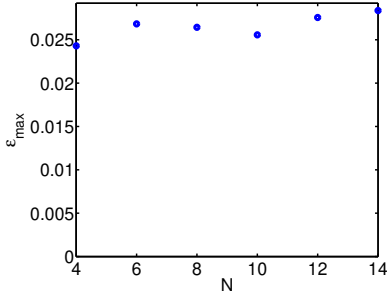
Figure S2: The maximum uncertainty of the Bloch vector elements defined in Eq. (S21) for the optimal measurement settings as a function of the number of qubits, N, for $N = 4, 6, 8, 10, 12$ and 14.

Such a decomposition for operators of the form $(A^{\otimes(N-n)} \otimes \mathbb{1}^{\otimes n})_{\mathrm{PI}}$ with $A = a_x X + a_y Y + a_z Z$ is given by

$$\sum_{k,l,m} a_x^k a_y^l a_z^m \frac{(k+l+m)!}{k!l!m!} (X^{\otimes k} \otimes Y^{\otimes l} \otimes Z^{\otimes m} \otimes \mathbb{1}^{\otimes n})_{\mathrm{PI}},$$
(S20)

where the summation is carried out such that $k+l+m+n = N$.

Results for larger systems. We determined the optimal A_j for PI tomography for $N = 4, 6, ..., 14$. In Fig. S2, we plot the maximal uncertainty of the Bloch vector elements

$$\epsilon_{\max} = \max_{k,l,m,n} \mathcal{E}[(X^{\otimes k} \otimes Y^{\otimes l} \otimes Z^{\otimes m} \otimes \mathbb{1}^{\otimes n})_{\mathrm{PI}}] \quad (\text{S21})$$

for the total count realized in the experiment $\lambda_j = \lambda = 2050$ as a function of N, when the state of the system is $\varrho_0 = \mathbb{1}/2^N$. It increases slowly with N. Thus, for large N the number of counts per measurement setting does not have to increase very much in order to keep the maximal uncertainty of the Bloch vector elements the same as for the $N = 4$ case. In particular, for $N = 14$, a total count of 2797 per setting yields the same maximal uncertainty as we had for the $N = 4$ case.

An upper bound on the uncertainty of PI tomography for ϱ_0 different from the white noise can be obtained by using $[\Delta(A_j^{\otimes(N-n)} \otimes \mathbb{1}^{\otimes n})_{\mathrm{PI}}]_{\varrho_0}^2 = 1$ for error calculations. According to numerics, for optimal A_j for $N = 4, 6, ..., 14$, $\epsilon_{\max}$ remains the same as in the case of white noise, since for the full correlation terms with $n = 0$ the upper bound equals the value for white noise, and the full correlations terms contribute to the noise of the Bloch vector elements with the largest uncertainty. Thus, the total count per setting will not increase more with the number of qubits even for states different from the completely mixed state.

The operators that give a bound on $\langle P_s \rangle$ with three settings for $N = 6$ and 8 are the following

$$P_s^{(6)} \geq \frac{2}{225}(Q_2 + J_z^2) - \frac{1}{90}(Q_4 + J_z^4) + \frac{1}{450}(Q_6 + J_z^6),$$
$$P_s^{(8)} \geq -0.001616 Q_2 + 0.002200 Q_4 - 0.0006286 Q_6$$
$$+ \ 0.00004490 Q_8 + 0.003265 J_z^2 - 0.004444 J_z^4$$
$$+ \ 0.001270 J_z^6 - 0.00009070 J_z^8,$$
(S22)

where $Q_n = J_x^n + J_y^n$. They were determined using semidefinite programming, with a method similar to one used for obtaining three-setting witnesses in Ref. [S4]. They have an expectation value $+1$ for the Dicke states $|D_6^{(3)}\rangle$ and $|D_8^{(4)}\rangle$, respectively. Moreover, their expectation value give the highest possible lower bound on $\langle P_s \rangle$ for states of the form

$$\varrho_{\mathrm{noisy}}(p) = p\frac{\mathbb{1}}{2^N} + (1-p)|D_N^{(N/2)}\rangle\langle D_N^{(N/2)}| \quad (\text{S23})$$

among the operators that are constructed as a linear combination of the operators J_l^n. The validity of the relations in Eq. (S22) can easily be checked by direct calculation.

Bounding the differences between elements of ϱ and ϱ_{PI} based on the fidelity. For any pure state $|\Psi\rangle$, it is possible to bound the difference between $|\langle\Psi|\varrho_{\mathrm{PI}}|\Psi\rangle|$ and $|\langle\Psi|\varrho|\Psi\rangle|$ as

$$|\langle\Psi|\varrho|\Psi\rangle - \langle\Psi|\varrho_{\mathrm{PI}}|\Psi\rangle| \leq \sqrt{1 - F(\varrho, \varrho_{\mathrm{PI}})}. \quad (\text{S24})$$

Thus, if the fidelity is close to 1, then $\langle\Psi|\varrho|\Psi\rangle \approx \langle\Psi|\varrho_{\mathrm{PI}}|\Psi\rangle$, even if $|\Psi\rangle$ is non-symmetric. If $|\Psi\rangle$ is an element of the product basis, e.g., $|0011\rangle$, then Eq. (S24) is a bound on the difference between the corresponding diagonal elements of ϱ and ϱ_{PI}.

Eq. (S24) can be proved as follows: There is a well-known relation between the trace norm and the fidelity [S5]

$$\frac{1}{2}||\varrho - \varrho_{\mathrm{PI}}||_{\mathrm{tr}} \leq \sqrt{1 - F(\varrho, \varrho_{\mathrm{PI}})}. \quad (\text{S25})$$

Moreover, for a projector P and density matrices ϱ_k we have [S6]

$$|\mathrm{Tr}(P\varrho_1) - \mathrm{Tr}(P\varrho_2)| \leq \frac{1}{2}||\varrho_1 - \varrho_2||_{\mathrm{tr}}. \quad (\text{S26})$$

Combining Eq. (S25) and Eq. (S26), leads to Eq. (S24).

References

[S1] C. Schmid, Ph.D. Thesis, Ludwig-Maximillian-Universität, Munich, Germany, 2008; D.F.V. James *et. al,* Phys. Rev. A **64**, 052312 (2001).

[S2] B. Jungnitsch *et. al*, Phys. Rev. Lett. **104**, 210401 (2010).

[S3] See, for example, J.I. Cirac *et. al*, Phys. Rev. Lett. **82**, 4344 (1999).

[S4] G. Tóth *et al.*, New J. Phys. **11**, 083002 (2009).

[S5] J.A. Miszczak *et al.*, J. Quant. Inf. Comp. **9**, 0103 (2009).

[S6] See Eqs. (9.18) and (9.22) of M.A. Nielsen and I.L. Chuang, *Quantum computation and quantum information* (Cambridge University Press, Cambridge, 2000).