

## State learning from pairs of states

Pranjal Agarwal , <sup>1,2,\*</sup> Nada Ali , <sup>1,2,†</sup> Camilla Polvara , <sup>2,3,‡</sup> Martin-Ibjörn Trappe , <sup>4,§</sup>Berthold-Georg Englert , <sup>5,4,6,||</sup> and Mark Hillery <sup>1</sup>*Department of Physics and Astronomy, Hunter College of the City University of New York, 695 Park Avenue, New York, New York 10065, USA*<sup>2</sup>*Physics Program, Graduate Center of the City University of New York, 365 Fifth Avenue, New York, New York 10016, USA*<sup>3</sup>*College of Staten Island of the City University of New York, 2800 Victory Blvd., Staten Island, New York 10314, USA*<sup>4</sup>*Centre for Quantum Technologies, National University of Singapore, Singapore 117543, Singapore*<sup>5</sup>*Key Laboratory of Advanced Optoelectronic Quantum Architecture and Measurement of the Ministry of Education, School of Physics, Beijing Institute of Technology, Beijing 100081, China*<sup>6</sup>*Department of Physics, National University of Singapore, Singapore 117542, Singapore*

(Received 2 September 2024; accepted 29 May 2025; published 24 June 2025)

Suppose you receive a sequence of qubits where each qubit is guaranteed to be in one of two pure states, but you do not know what those states are. Your task is to determine the states. This can be viewed as a kind of quantum state learning, or quantum state estimation. The problem is that, without more information, all that can be determined is the density matrix of the sequence and, in general, density matrices can be decomposed into pure states in many different ways. To solve the problem, additional information, either classical or quantum, is required. We show that if an additional copy of each qubit is supplied (that is, one receives pairs of qubits, both in the same state, rather than single qubits) the task can be accomplished. This is possible because the mixed two-qubit state has a unique decomposition into pure product states. For illustration purposes, we numerically simulate the symmetric, informationally complete measurement of a sequence of qubit pairs and show that the unknown states and their respective probabilities of occurrence can be inferred from the data with high accuracy. Finally, we propose an experiment that employs a product measurement and can be realized with existing technology, and we demonstrate how the data tell us the states and their probabilities. We find that it is enough to detect a few thousand qubit pairs.

DOI: [10.1103/tyf3-1zdd](https://doi.org/10.1103/tyf3-1zdd)

## I. INTRODUCTION

Suppose we have a collection of data, about which we have very little information, and we are interested in learning something about it. If the data are classical, an example of such a problem is unsupervised machine learning. In this scenario, the objective is to classify data into clusters with the idea that the data within a cluster are related. There is no training phase in which sample data with their classifications are provided; in the unsupervised case, there are only the data to work with. Quantum algorithms have been applied to obtain speedups of the unsupervised learning of classical data [1–4] (for reviews of quantum machine learning, see [5]). In these works, the classical data are converted into quantum states, which can then be processed by a quantum computer.

Quantum learning (or quantum estimation) is related to classical machine learning, but because the objects to be learned are quantum, new elements come into play. One can learn a number of different quantum objects, unitary operators [6], and measurements [7–10], for example. In many cases,

there is a training set. In the case of a unitary operator, one is allowed a certain number of uses of the operator, and in the case of a measurement, one is given examples of the states one wants the measurement to distinguish. What we want to do here is to see what can be done in the case of learning a set of unknown quantum states, in particular, determining what states are in the set. In most approaches to quantum unsupervised machine learning, one has access to unitary operators that produce the data by acting on a reference quantum state [4]. What, however, can be done if this is not the case and one has access only to the raw data, that is, just the quantum systems themselves? The first treatment of this kind of quantum learning was given in [11]. There one is given a sequence of  $N$  particles, each in one of two unknown states,  $|\psi_0\rangle$  and  $|\psi_1\rangle$ , and one wants to determine the sequence. For any individual qubit in the sequence, you do not know which state it is in. The output of this procedure is classical, a sequence of 0s and 1s, corresponding to the labels of the states, of length  $N$  that is the best guess for the sequence of states.

A second approach was taken in [12]. There the setup was the same as in [11], but the objective was to use the data to construct a positive operator-valued measure (POVM) that would distinguish the two states. The fundamental problem is that all one can measure is the density matrix of the ensemble that describes the sequence

<sup>\*</sup>Contact author: [pagarwal@gradcenter.cuny.edu](mailto:pagarwal@gradcenter.cuny.edu)<sup>†</sup>Contact author: [nali@gradcenter.cuny.edu](mailto:nali@gradcenter.cuny.edu)<sup>‡</sup>Contact author: [camipolv@gmail.com](mailto:camipolv@gmail.com)<sup>§</sup>Contact author: [martin.trappe@quantumlah.org](mailto:martin.trappe@quantumlah.org)<sup>||</sup>Contact author: [berge@bit.edu.cn](mailto:berge@bit.edu.cn), [englert@u.nus.edu](mailto:englert@u.nus.edu)<sup>¶</sup>Contact author: [mhillery@hunter.cuny.edu](mailto:mhillery@hunter.cuny.edu)

$$\rho_{1qb} = p_0|\psi_0\rangle\langle\psi_0| + p_1|\psi_1\rangle\langle\psi_1|, \quad (1)$$

if  $|\psi_0\rangle$  appears with probability  $p_0$  and  $|\psi_1\rangle$  appears with probability  $p_1$ . A rank-two mixed state density matrix can be decomposed into a sum of pure states in many different ways. That means that additional information, either classical or quantum, is required to determine  $|\psi_0\rangle$  and  $|\psi_1\rangle$ . Several examples of additional classical information were explored in [12]. In one example, it was specified that the two states lie on a known circle of the Bloch sphere, e.g., the intersection of the  $x, z$  plane and the sphere, and that  $p_0 = p_1 = 1/2$ . Under those highly restrictive conditions, it is possible to construct a POVM that will discriminate the states.

A simple example of how extra classical information allows one to determine which of two ensembles with the same density matrix one has is the following [13,14]. We have  $2N$  spin-1/2 particles, which are promised to be in one of two ensembles. In the first,  $N$  spins point in the  $+x$  direction and  $N$  point in the  $-x$  direction, and in the second,  $N$  spins point in the  $+z$  direction and  $N$  point in the  $-z$  direction. These ensembles are described by the same single-particle density matrix, but they can be distinguished with high probability by measuring all the particles in the  $z$  direction. If you find that  $N$  of the particles point in the  $+z$  direction, the ensemble is with high probability the second, and if you find that some number other than  $N$  particles point in the  $+z$  direction, the ensemble is definitely the first.

In this paper, we are going to use a model similar to that employed in [11,12], but our objective will be to determine the states. Our main tool will be a form of state tomography. This is a form of state learning [15]. In state learning, one receives many copies of the state to be learned and this state is guaranteed to be a member of a certain set of states. One then performs measurements on the copies, and the result is a sufficiently accurate description of the state. The difference in our case is that there are two states, not one, that one is trying to learn, and the copies are scrambled; any given copy could represent either of the quantum states, and you do not know which. In addition, the two states could be any pure states, and you do not know with what probability they occur. This task sounds formidable, but, as we shall see, if the states are received in pairs, where the members of each pair are identical, tomography can be applied to determine the states.

## II. PAIRS

In this section, we will show that if the states are received in pairs, the situation is much improved, and all you have to know is that the two-qubit density matrix is an incoherent superposition of two pure product states. Receiving an extra copy of each state means that extra quantum information is being provided. To be more specific, consider the following scenario. You are sent a stream of pairs of qubits, and each pair is in the state  $|\psi_0\rangle \otimes |\psi_0\rangle$  or  $|\psi_1\rangle \otimes |\psi_1\rangle$  with

$$\begin{aligned} |\psi_0\rangle &= a_0|0\rangle + a_1|1\rangle, \\ |\psi_1\rangle &= b_0|0\rangle + b_1|1\rangle, \end{aligned} \quad (2)$$

where  $|0\rangle, |1\rangle$  are single-qubit orthonormal kets that serve as the reference basis (“computational basis”). The  $|\psi_0\rangle$  pair occurs with probability  $p_0$  and the  $|\psi_1\rangle$  pair occurs with probability  $p_1 = 1 - p_0$ . You do not know what  $|\psi_0\rangle$  and  $|\psi_1\rangle$  are,

and for any given pair, you do not know which kind of pair it is. You also do not know  $p_0$  and  $p_1$ . The task is to find  $|\psi_0\rangle, |\psi_1\rangle, p_0$ , and  $p_1$ .

One can then, for example, use this to construct a POVM to discriminate between the different types of pairs from a subset of the pairs and then use it to discriminate the remaining pairs. As only two states are involved, the purpose-appropriate qubit POVM will be used for the pair states  $|\psi_0\rangle \otimes |\psi_0\rangle$  and  $|\psi_1\rangle \otimes |\psi_1\rangle$  since their distinguishability is larger than that of the single-qubit states. For instance, there are the POVMs for unambiguous discrimination [16] or for extracting the accessible information [17].

The ensemble we are looking at is described by the density matrix

$$\rho = p_0|\psi_0\rangle\langle\psi_0| \otimes |\psi_0\rangle\langle\psi_0| + p_1|\psi_1\rangle\langle\psi_1| \otimes |\psi_1\rangle\langle\psi_1|. \quad (3)$$

The density matrix and the knowledge that it is composed of two pure two-qubit product states is the only information to which we have access. We will now show that this is sufficient to find the states and probabilities.

To begin with, we note that  $|\psi_0\rangle \otimes |\psi_0\rangle$  and  $|\psi_1\rangle \otimes |\psi_1\rangle$  span a two-dimensional subspace in the four-dimensional ket space of the qubit pair; the subspace is analogous to the Bloch sphere of a qubit. In the corresponding Bloch ball, the separable mixed states are located on a line that connects the points for  $|\psi_0\rangle \otimes |\psi_0\rangle$  and  $|\psi_1\rangle \otimes |\psi_1\rangle$  on the sphere; see [18–20] and, in particular, Fig. 2.2 in [20]. Once we learn  $\rho$  from the data, its range is the Bloch sphere and hence we know the line of separable states and their endpoints.

More specifically,  $\rho$  has support in the three-dimensional symmetric subspace of two qubits, the triplet sector, and is of rank two. That means that there is a direction  $|\xi\rangle$  in the triplet sector that is orthogonal to  $\rho$ , i.e.,  $\rho|\xi\rangle = 0$ . Set

$$|\xi\rangle = c_{00}|00\rangle + c_{01}(|01\rangle + |10\rangle) + c_{11}|11\rangle, \quad (4)$$

and let's see what being orthogonal to  $|\psi_0\rangle \otimes |\psi_0\rangle$  implies. The orthogonality condition is

$$(a_0^*)^2 c_{00} + 2a_0^* a_1^* c_{01} + (a_1^*)^2 c_{11} = 0. \quad (5)$$

Dividing both by  $(a_1^*)^2$  and setting  $z = a_0^*/a_1^*$ , we find

$$c_{11} + 2c_{01}z + c_{00}z^2 = 0. \quad (6)$$

An identical equation holds for  $|\psi_1\rangle \otimes |\psi_1\rangle$  except that, in that case,  $z$  is  $b_0^*/b_1^*$ . The above equation has two solutions for  $z$ , one corresponding to  $|\psi_0\rangle$  and the other to  $|\psi_1\rangle$ . That means that if we know  $|\xi\rangle$ , then we know both  $|\psi_0\rangle$  and  $|\psi_1\rangle$ . That suggests that one way to proceed is to find a way to determine  $|\xi\rangle$ . This can be done by finding  $\rho$  by performing state tomography and then finding the solution to  $\rho|\xi\rangle = 0$ . One can also find explicit formulas for  $|\psi_0\rangle$  and  $|\psi_1\rangle$  in terms of  $\rho$ , which we shall proceed to do.

Note that the argument we just used can be easily generalized. For example, suppose that our sequence consists of three rather than two qubit states, and we receive not pairs but trios, where all the states within a given trio are guaranteed to be the same. The overall state of each trio lies in the symmetric subspace of the space of three qubits (the space with total spin 3/2, if we regard each qubit as a spin-1/2 object) which is four-dimensional, while the density matrix for the trios is

of rank three. That means there is a ket in the symmetric subspace that is annihilated by all three of the states, and consequently, by the trio density matrix. In analogy to the two-state case, this ket will lead to a cubic equation whose solutions will yield the three states. Clearly, the argument can be carried further to  $N$  states, which would require receiving identical  $N$  triplets.

It is also possible to generalize this procedure to two states in a  $d$ -dimensional space [21]. Initially you conduct single-qubit tomography to find the single-particle density matrix, which will be of rank two. You then find its two eigenstates, call them  $|0\rangle_d$  and  $|1\rangle_d$ . The states we are trying to find lie in the two-dimensional subspace spanned by  $|0\rangle_d$  and  $|1\rangle_d$ , and the problem is reduced to one of effective qubits. One can then apply the reasoning in the previous paragraphs, but substituting  $|0\rangle_d$  for  $|0\rangle$  and  $|1\rangle_d$  for  $|1\rangle$ .

Let us now return to the case of two qubit states and find explicit formulas for the states and probabilities. If the Bloch vectors for  $|\psi_0\rangle$  and  $|\psi_1\rangle$  are  $\mathbf{a}$  and  $\mathbf{b}$ , respectively, then the two-qubit pair density matrix for  $|\psi_0\rangle$  is

$$\begin{aligned}\rho_0 &= \frac{1}{2}(I_d + \mathbf{a} \cdot \boldsymbol{\sigma}) \otimes \frac{1}{2}(I_d + \mathbf{a} \cdot \boldsymbol{\sigma}) \\ &= \frac{1}{4}(I_4 + \mathbf{a} \cdot (\boldsymbol{\sigma}^{(1)} + \boldsymbol{\sigma}^{(2)}) + \boldsymbol{\sigma}^{(1)} \cdot \mathbf{a} \mathbf{a} \cdot \boldsymbol{\sigma}^{(2)}),\end{aligned}\quad (7)$$

and similarly for  $|\psi_1\rangle$  but with  $\mathbf{a}$  replaced by  $\mathbf{b}$ . Here,  $I_d$  is the  $d \times d$  identity matrix,  $\boldsymbol{\sigma}$  is the generic vector of Pauli matrices,  $\boldsymbol{\sigma}^{(1)} = \boldsymbol{\sigma} \otimes I_2$  is that for qubit 1, and  $\boldsymbol{\sigma}^{(2)} = I_2 \otimes \boldsymbol{\sigma}$  is that for qubit 2. The ensemble pair density matrix is

$$\begin{aligned}\rho &= p_0 \rho_0 + p_1 \rho_1 \\ &= \frac{1}{4}(I_4 + \mathbf{s} \cdot (\boldsymbol{\sigma}^{(1)} + \boldsymbol{\sigma}^{(2)}) + \boldsymbol{\sigma}^{(1)} \cdot \mathbf{C} \cdot \boldsymbol{\sigma}^{(2)}),\end{aligned}\quad (8)$$

where

$$\begin{aligned}\mathbf{s} &= \langle \boldsymbol{\sigma}^{(1)} \rangle = \langle \boldsymbol{\sigma}^{(2)} \rangle = p_0 \mathbf{a} + p_1 \mathbf{b}, \\ \mathbf{C} &= \langle \boldsymbol{\sigma}^{(1)} \boldsymbol{\sigma}^{(2)} \rangle = p_0 \mathbf{a} \mathbf{a} + p_1 \mathbf{b} \mathbf{b}.\end{aligned}\quad (9)$$

Note that  $\mathbf{a} \mathbf{a}$  is the dyad with matrix elements  $a_j a_k$ , and similarly for  $\mathbf{b} \mathbf{b}$ , and the dyad  $\mathbf{C}$  has the matrix elements  $C_{jk} = \langle \boldsymbol{\sigma}_j^{(1)} \boldsymbol{\sigma}_k^{(2)} \rangle = \langle \boldsymbol{\sigma}_j \otimes \boldsymbol{\sigma}_k \rangle = p_0 a_j a_k + p_1 b_j b_k$ ; as is characteristic for mixed triplet states,  $\mathbf{C}$  is symmetric,  $C_{jk} = C_{kj}$ , and has unit trace,  $\sum_j C_{jj} = 1$ .

Full tomography of the qubit pairs will provide us with the vector  $\mathbf{s}$  and the dyad  $\mathbf{C}$ , and we demonstrate now how knowledge of these quantities can be converted into knowledge of  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $p_0$ , and  $p_1$ . In the first step, we find

$$\mathbf{C} - \mathbf{s} \mathbf{s} = p_0 p_1 (\mathbf{a} - \mathbf{b})(\mathbf{a} - \mathbf{b}); \quad (10)$$

if  $\mathbf{C} - \mathbf{s} \mathbf{s} = 0$ , the source emits only one state and we are done. Otherwise,  $\mathbf{s}^2 < 1$  and  $(\mathbf{C} - \mathbf{s} \mathbf{s})/(1 - \mathbf{s}^2)$  projects on the direction of  $\mathbf{a} - \mathbf{b} \neq 0$ . We remove the component parallel to  $\mathbf{a} - \mathbf{b}$  from  $\mathbf{s}$  and obtain

$$\mathbf{s}' = \frac{\mathbf{s} - \mathbf{C} \cdot \mathbf{s}}{1 - \mathbf{s}^2} = \frac{1}{2}(\mathbf{a} + \mathbf{b}). \quad (11)$$

Then,  $(p_0 - p_1)^2$  and  $\mathbf{a} - \mathbf{b}$  are available from

$$\frac{(\mathbf{s} - \mathbf{s}')^2}{1 - \mathbf{s}'^2} = (p_0 - p_1)^2, \quad \frac{\mathbf{s} - \mathbf{s}'}{p_0 - p_1} = \frac{1}{2}(\mathbf{a} - \mathbf{b}), \quad (12)$$

and we arrive at

$$\mathbf{a} = \frac{\mathbf{s} - 2p_1 \mathbf{s}'}{p_0 - p_1}, \quad \mathbf{b} = \frac{2p_0 \mathbf{s}' - \mathbf{s}}{p_0 - p_1}, \quad (13)$$

provided that  $p_0 \neq p_1$ . If  $p_0 = p_1 = \frac{1}{2}$ , when  $\mathbf{s}' = \mathbf{s}$ , we identify  $\mathbf{a} - \mathbf{b}$  as the eigenvector of  $\mathbf{C} - \mathbf{s} \mathbf{s}$  with the eigenvalue  $\frac{1}{2}(1 - \mathbf{a} \cdot \mathbf{b})$ .

In the following sections we shall discuss two schemes for the state tomography that provides data from which one can estimate  $\rho$  and thus  $\mathbf{s}$  and  $\mathbf{C}$ . Section III deals with the symmetric informationally complete POVM (SIC POVM) in the triplet sector; see [22], for example, for properties of SIC POVMs. The high symmetry of the SIC POVM is attractive and facilitates the analysis, but we do not know how to implement the SIC POVM in the laboratory. By contrast, the tetrahedron POVM of Sec. IV can definitely be realized with existing technology.

### III. SIC POVM

#### A. Measurement

In the triplet sector, the symmetric subspace of two qubits, we expand the kets in the basis used in Eq. (4),

$$|v\rangle = a_0 |00\rangle + a_1 |11\rangle + a_2 \frac{|01\rangle + |10\rangle}{\sqrt{2}} \hat{=} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix}. \quad (14)$$

The SIC POVM for these kets is analogous to the standard SIC POVM for qutrits [23,24]. The nine POVM elements, or probability operators, are proportional to one-dimensional projectors,

$$\Pi_j = \frac{1}{3} |v_j\rangle \langle v_j| \quad \text{for } j = 1, 2, \dots, 9, \quad (15)$$

with

$$\begin{aligned}(|v_1\rangle &|v_2\rangle \dots |v_8\rangle |v_9\rangle) \\ &\hat{=} \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & -1 & 1 & 0 & -\omega & 1 & 0 & -1 & \omega \\ 1 & 0 & -1 & 1 & 0 & -\omega & \omega & 0 & -1 \\ -1 & 1 & 0 & -\omega & 1 & 0 & -1 & \omega & 0 \end{pmatrix},\end{aligned}\quad (16)$$

where  $\omega = e^{i2\pi/3}$  is the basic cubit root of unity. The sum of the elements is the projector on the triplet sector,

$$\sum_{j=1}^9 \Pi_j = \frac{1}{4}(3I_4 + \boldsymbol{\sigma}^{(1)} \cdot \boldsymbol{\sigma}^{(2)}) = I_{\text{tp}}. \quad (17)$$

After supplementing the triplet SIC POVM with the projector on the singlet

$$\Pi_0 = \frac{1}{4}(I_4 - \boldsymbol{\sigma}^{(1)} \cdot \boldsymbol{\sigma}^{(2)}) = I_{\text{sg}}, \quad (18)$$

we have a proper POVM for the qubit pair. Note that the rank of  $I_{\text{tp}}$  is three and that of  $I_{\text{sg}}$  is one.

Since

$$\begin{aligned}\text{Tr}(\Pi_j \Pi_k) &= \frac{1}{36} + \frac{1}{12} \delta_{jk}, \\ \text{Tr}(\Pi_j (12\Pi_k - I_{\text{tp}})) &= \delta_{jk},\end{aligned}\quad (19)$$

for  $j, k = 1, 2, \dots, 9$ , we reconstruct any  $\rho$  in the triplet sector from its SIC POVM probabilities

$$q_j = \text{Tr}(\Pi_j \rho), \quad (20)$$

in accordance with

$$\rho = \sum_{j=1}^9 q_j (12\Pi_j - I_{\text{tp}}). \quad (21)$$

In addition to being nonnegative and having unit sum, the nine probabilities  $q_j$  are subject to constraints that follow from  $\rho \geq 0$ . In particular, there is the purity constraint  $\frac{1}{3} \leq \text{Tr}(\rho^2) \leq 1$ , that is,

$$\frac{1}{9} \leq \sum_{j=1}^9 q_j^2 \leq \frac{1}{6}. \quad (22)$$

Moreover, in the present context of Eq. (8), we note that these separable rank-two triplet states make up a five-dimensional nonconvex set in the eight-dimensional triplet sector and that restricts the permissible probabilities stringently. We do not know, however, how to state these restrictions as explicit constraints obeyed by the  $q_j$ s.

### B. Simulated data

For different choices of  $|\psi_0\rangle$ ,  $|\psi_1\rangle$ ,  $p_0$ , and  $p_1$ , the SIC POVM was used to produce simulated measurement data; in particular, it gave us a probability distribution from which we then sampled. The data are the counts  $n_1, n_2, \dots, n_9$  for the nine different outcomes of the simulated measurement. Two different methods were then used to find the density matrix from the data, linear inversion, and maximum likelihood. We first used linear inversion to produce an empirical density matrix from the data, that is, we took the relative frequencies as estimates of the probabilities

$$\rho^{(\text{LI})} = \sum_{j=1}^9 \frac{n_j}{N} (12\Pi_j - I_{\text{tp}}) \quad \text{with} \quad N = \sum_{j=1}^9 n_j. \quad (23)$$

With this in hand, we examined two methods to find the states and probabilities.

For the first, we found the eigenvector of the empirical  $3 \times 3$  density matrix with the smallest eigenvalue and identified it with  $|\xi\rangle$ . This ket was used to create the quadratic equation in Eq. (6) and find the states. Once we know the states and the empirical density matrix, it is straightforward to find the probabilities. For the second method, we used the empirical density matrix to find the vector  $s$  and the dyad  $C$ , which then allowed us to find the vectors  $a, b$ , and their probabilities.

In both cases, we then computed the fidelities of the states from the simulation with the original states that were used to produce the data, and these were plotted versus the number of pairs received. The results for two cases are shown in Fig. 1. We used Eq. (6) for the plots in Fig. 1(a), while those in Fig. 1(b) resulted from Eq. (13).

The rate of convergence depends on the overlap of  $|\psi_0\rangle$  and  $|\psi_1\rangle$ ; in particular, the larger the overlap the slower the

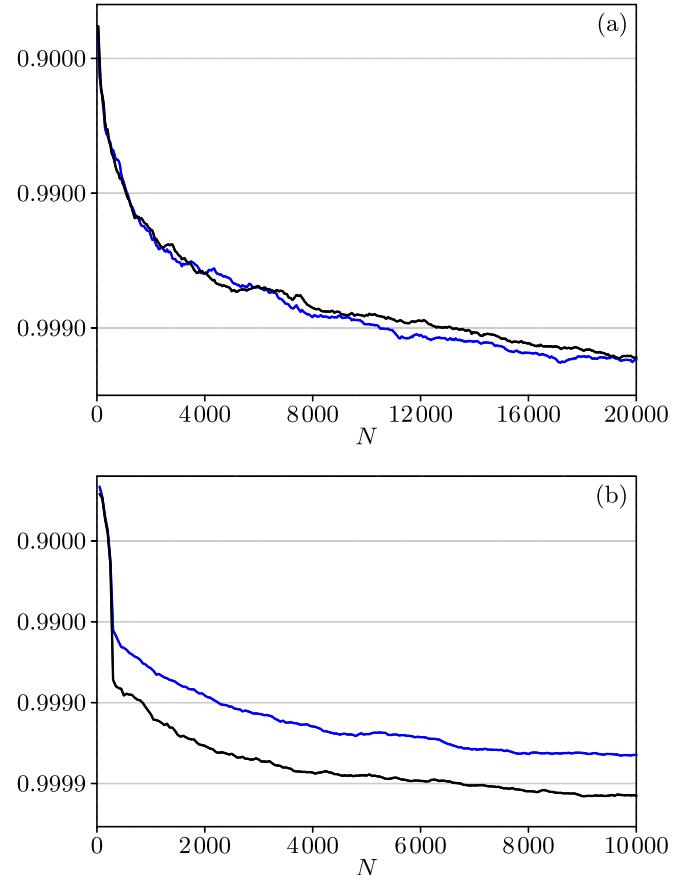


FIG. 1. Results from simulated data for the SIC POVM averaged over 50 runs. The graphs show the fidelities of the states estimated by linear inversion versus the number of detected pairs. The black curves graph  $|\langle\psi_0|\psi_0^{(\text{LI})}\rangle|^2$  and the blue curves graph  $|\langle\psi_1|\psi_1^{(\text{LI})}\rangle|^2$ . Plot (a) is for the states  $|\psi_0\rangle = |0\rangle$  and  $|\psi_1\rangle = (1/\sqrt{2})(|0\rangle + |1\rangle)$  and the probabilities  $p_0 = p_1 = 1/2$ . In the simulation, we detect up to 20 000 pairs and infer the states by using Eq. (6). Plot (b) is for the states  $|\psi_0\rangle = |0\rangle$  and  $|\psi_1\rangle = |1\rangle$  and the probabilities  $p_0 = 0.75$  and  $p_1 = 0.25$ . We learn the states from Eq. (13) for up to 10 000 detected pairs and observe that the state with the higher probability converges faster.

convergence. In the case in which one uses the ket  $|\xi\rangle$  to find the two states, a large overlap between  $|\psi_0\rangle$  and  $|\psi_1\rangle$  will lead to a  $3 \times 3$  density matrix  $\rho$  with one large and one small eigenvalue as well as the eigenvalue 0 corresponding to the eigenvector  $|\xi\rangle$ , and that can lead to problems. The ket  $|\xi\rangle$  satisfies  $\rho|\xi\rangle = 0$ , but the estimated density matrix  $\rho^{(\text{LI})}$  is not in the five-dimensional space of physical  $\rho$ s since the relative frequencies  $n_j/N$  do not obey the constraints that restrict the probabilities  $q_j$ . Rather than the exact eigenvalue 0 for  $|\xi\rangle$ ,  $\rho^{(\text{LI})}$  has an eigenvalue  $\approx 0$  that can be positive or negative (while  $\rho^{(\text{LI})}$  is hermitian, nothing ensures  $\rho^{(\text{LI})} \geq 0$ ). If, then, the overlap of  $|\psi_0\rangle$  and  $|\psi_1\rangle$  is very large,  $\rho$  has a small positive eigenvalue and it can be difficult to distinguish the corresponding approximate eigenvalue of  $\rho^{(\text{LI})}$  from the near-zero eigenvalue of  $|\xi\rangle$ .

For the maximum likelihood method we parametrize the density matrix suitably. The two states are expressed as

$|\psi_j\rangle = \cos(\theta_j)|0\rangle + e^{i\phi_j} \sin(\theta_j)|1\rangle$  for  $j = 0, 1$ , so that

$$\begin{aligned} a_x + ia_y &= \sin(2\theta_1)e^{i\phi_1}, & a_z &= \cos(2\theta_1), \\ b_x + ib_y &= \sin(2\theta_2)e^{i\phi_2}, & b_z &= \cos(2\theta_2), \end{aligned} \quad (24)$$

and the probabilities in terms of an angle  $\alpha$ ,  $p_0 = \cos(\alpha)^2$  and  $p_1 = \sin(\alpha)^2$ . Then, the two-qubit density matrix of Eq. (8) is a function of  $\theta = (\theta_0, \phi_0, \theta_1, \phi_1, \alpha)$ , the five parameters that specify the qubit states and the probabilities,  $\rho(\theta)$ . We use this density matrix to find the nine probabilities of Eq. (20) as functions of  $\theta$ , which enter the likelihood function [25]

$$L(\theta) = \prod_{j=1}^9 q_j(\theta)^{n_j} \quad (25)$$

for the given simulated data. For the purpose of finding the maximum of  $L(\theta)$ , we use a covariance matrix adaptation evolution strategy (CMA-ES), adapted from [26], to find the minimum of  $-N^{-1} \log(L(\theta))$ . Once we do this, we know the maximum likelihood estimates of both states and both probabilities, which tell us  $\rho^{(ML)}$ , and no further processing is necessary. The parametrization in terms of the five parameters ensures that every  $\rho$  in the competition is in the five-dimensional set of permissible  $\rho$ s. Put differently, the constraints mentioned after Eq. (22) are obeyed by construction. By contrast, the unconstrained maximization of  $\log(L) = \sum_j n_j \log(q_j)$  yields the linear inversion estimates  $q_j = n_j/N$  with  $\rho^{(LI)}$  outside of the set of permissible  $\rho$ s; in this sense, linear inversion is unconstrained likelihood maximization. Whereas  $\rho^{(ML)}$  is assuredly in the set of separable rank-two triplet density matrices,  $\rho^{(LI)}$  is only guaranteed to be a hermitian, unit-trace  $3 \times 3$  matrix; it is usually of rank three and has a roughly 50% chance of having a negative eigenvalue.

Nevertheless, linear inversion yields reasonable results as demonstrated by the plots in Fig. 1. This is so because, although  $\rho^{(LI)}$  is not in the five-dimensional physical set, it is very close to the actual  $\rho$ , as measured by a proper distance in the eight-dimensional convex space of hermitian unit-trace  $\rho$ s, when  $N$  is sufficiently large.

We observe that the maximum likelihood method produced higher fidelities than the linear inversion method for a given number  $N$  of data of simulated detection events. Here are two examples comparing the results of linear inversion and maximum likelihood. For the choice  $\theta = (\frac{\pi}{6}, \frac{\pi}{4}, \frac{2\pi}{3}, \frac{\pi}{2}, \frac{\pi}{6})$ , we simulated 750 experiments with  $N = 1000$  each. Figure 2(a) shows the sum of errors (one minus the fidelity) for  $|\psi_0\rangle$  and  $|\psi_1\rangle$  expressed in parts per million. We find that the maximum likelihood method consistently outperforms the linear inversion method, with average errors of 10546 ppm and 11852 ppm, respectively.

The parameters  $(\frac{\pi}{12}, \frac{\pi}{4}, \frac{5\pi}{12}, \frac{\pi}{2}, \frac{\pi}{3})$  are used in the example of Fig. 2(b). In this case, 1000 simulated experiments each with  $N = 1000$  were conducted. The average error for linear inversion was 15760 ppm and for maximum likelihood 9283 ppm.

In conclusion, the maximum likelihood method is more computationally demanding than the linear inversion method, due to the maximization, but produces better results. It also directly finds the two unknown states because, once the opti-

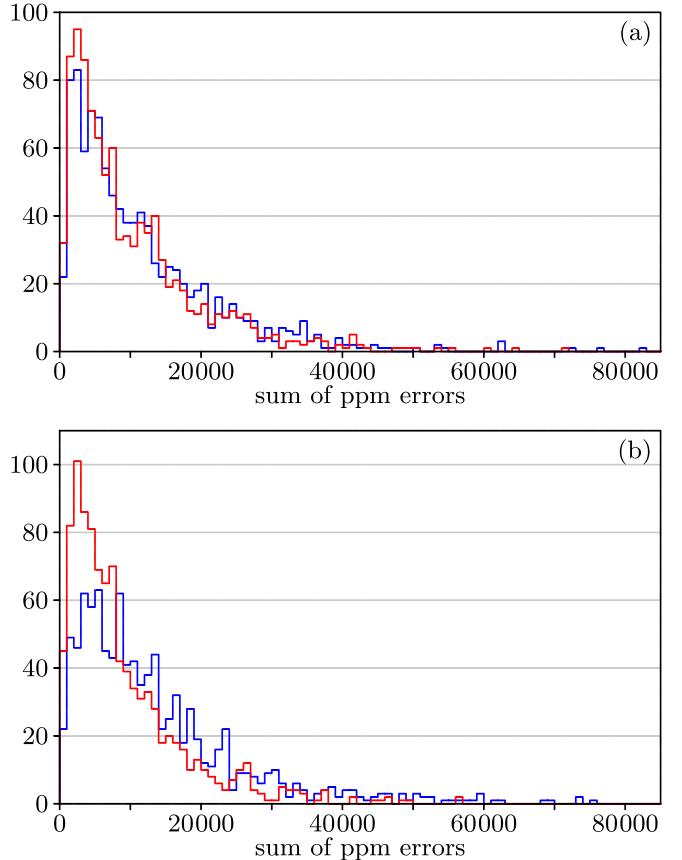


FIG. 2. The abscissa is the sum of ppm errors for the estimated  $|\psi_0\rangle$  and  $|\psi_1\rangle$  and the histograms show the frequencies with which these errors occur in each bin of 1000 abscissa units. The blue histogram is for the linear inversion method and the red histogram for the maximum likelihood method. We report results of simulated experiments (a) for the parameters  $(\theta_0, \phi_0, \theta_1, \phi_1, \alpha) = (\frac{\pi}{6}, \frac{\pi}{4}, \frac{2\pi}{3}, \frac{\pi}{2}, \frac{\pi}{6})$  and (b) for the parameters  $(\frac{\pi}{12}, \frac{\pi}{4}, \frac{5\pi}{12}, \frac{\pi}{2}, \frac{\pi}{3})$ . Observe that, in both plots, the frequency of small errors is considerably larger for the maximum likelihood method.

mal parameters are known, then so are the states, whereas the linear inversion first yields a density matrix, and then that has to be diagonalized to find the states. The simulations of Figs. 1 and 2 confirm that these two point estimators are consistent, that is,  $\rho^{(LI)} \rightarrow \rho$  and  $\rho^{(ML)} \rightarrow \rho$  as  $N \rightarrow \infty$ ; a quantitative statement about this is provided in Sec. IV B in the context of the tetrahedron POVM, see Eqs. (47) and (48), with additional details in the Appendix.

#### IV. TETRAHEDRON POVM

##### A. Measurement

A tomography experiment that could be performed with existing technology is sketched in Fig. 3. The elements of the single-qubit tetrahedron POVM are

$$\Pi_j^{(4)} = \frac{1}{4}(I_2 + \mathbf{t}_j \cdot \boldsymbol{\sigma}) \quad \text{with} \quad j = 1, 2, 3, 4; \quad (26)$$

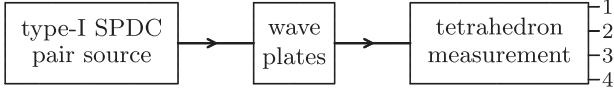


FIG. 3. A pump laser (not indicated), pulsed or continuous, illuminates a crystal for type-I spontaneous parametric down conversion (SPDC), which acts as a source of pairs of photons that propagate in the same direction and have the same polarization, both vertically polarized, say. A set of wave plates is used to change the polarization from vertical to any other kind. We switch at random between the two settings of the wave plates that correspond to the polarizations specified by the Bloch vectors  $\mathbf{a}$  and  $\mathbf{b}$  with the respective probabilities  $p_0$  and  $p_1$  for the next pair. The photons are detected by a tetrahedron measurement [27–30], the SIC POVM for single qubits, where we either register both photons in one exit port (four cases) or get a coincidence between two different exit ports (six cases).

the four tetrahedron vectors have the properties

$$\mathbf{t}_j \cdot \mathbf{t}_k = \frac{4}{3} \delta_{jk} - \frac{1}{3}, \quad \sum_{k=1}^4 \mathbf{t}_k = 0, \quad \sum_{k=1}^4 \mathbf{t}_k \mathbf{t}_k = \frac{4}{3} \mathbf{1}, \quad (27)$$

where  $\mathbf{1}$  is the unit dyad. Accordingly, the four POVM elements for detecting both photons at the same exit are

$$\begin{aligned} \Pi_k^{(s)} &= \Pi_k^{(4)} \otimes \Pi_k^{(4)} \\ &= \frac{1}{16} (I_4 + \mathbf{t}_k \cdot (\boldsymbol{\sigma}^{(1)} + \boldsymbol{\sigma}^{(2)}) + \boldsymbol{\sigma}^{(1)} \cdot \mathbf{t}_k \mathbf{t}_k \cdot \boldsymbol{\sigma}^{(2)}) \end{aligned} \quad (28)$$

for  $k = 1, 2, 3, 4$  and the six elements for the coincidence counts are

$$\begin{aligned} \Pi_{jk}^{(c)} &= \Pi_j^{(4)} \otimes \Pi_k^{(4)} + \Pi_k^{(4)} \otimes \Pi_j^{(4)} \\ &= \frac{1}{16} (2I_4 + (\mathbf{t}_j + \mathbf{t}_k) \cdot (\boldsymbol{\sigma}^{(1)} + \boldsymbol{\sigma}^{(2)}) \\ &\quad + \boldsymbol{\sigma}^{(1)} \cdot (\mathbf{t}_j \mathbf{t}_k + \mathbf{t}_k \mathbf{t}_j) \cdot \boldsymbol{\sigma}^{(2)}) \end{aligned} \quad (29)$$

for  $1 \leq j < k \leq 4$ . Clearly, the SIC POVM of Eqs. (15) to (18) is not of the kind realized by the elements in Eqs. (28) and (29) or any other product POVM. In particular, we have nonzero probabilities for qubit pairs in the singlet state

$$\text{Tr}(\Pi_k^{(s)} I_{\text{sg}}) = 0, \quad \text{Tr}(\Pi_{jk}^{(c)} I_{\text{sg}}) = \frac{1}{6}. \quad (30)$$

The ten-outcome measurement with the POVM elements in Eqs. (28) and (29) is tomographically complete in the space of density operators that are convex sums of the singlet state and any mixed state in the triplet sector, that is,

$$\rho = \frac{1}{4} (I_4 + \mathbf{s} \cdot (\boldsymbol{\sigma}^{(1)} + \boldsymbol{\sigma}^{(2)}) + \boldsymbol{\sigma}^{(1)} \cdot \mathbf{C} \cdot \boldsymbol{\sigma}^{(2)}), \quad (31)$$

with  $C_{jk} = C_{kj}$ . The singlet in Eq. (18) is of this kind and so is, of course, the state emitted by the source with  $\mathbf{s}$  and  $\mathbf{C}$  of the particular forms in Eq. (9). We recall the single-qubit identity

$$\text{Tr}(\Pi_j^{(4)} R_k^{(4)}) = \delta_{jk} \quad \text{with} \quad R_k^{(4)} = \frac{1}{2} (I_2 + 3\mathbf{t}_k \cdot \boldsymbol{\sigma}), \quad (32)$$

and recognize that the reconstruction operators  $R_k^{(s)}$  and  $R_{jk}^{(c)}$  that are defined by

$$\begin{aligned} \text{Tr}(\Pi_k^{(s)} R_{k'}^{(s)}) &= \delta_{kk'}, \quad \text{Tr}(\Pi_k^{(s)} R_{j'k'}^{(c)}) = 0, \\ \text{Tr}(\Pi_{jk}^{(c)} R_{k'}^{(s)}) &= 0, \quad \text{Tr}(\Pi_{jk}^{(c)} R_{j'k'}^{(c)}) = \delta_{jj'} \delta_{kk'}, \end{aligned} \quad (33)$$

are

$$\begin{aligned} R_k^{(s)} &= R_k^{(4)} \otimes R_k^{(4)} \\ &= \frac{1}{4} (I_4 + 3\mathbf{t}_k \cdot (\boldsymbol{\sigma}^{(1)} + \boldsymbol{\sigma}^{(2)}) + 9\boldsymbol{\sigma}^{(1)} \cdot \mathbf{t}_k \mathbf{t}_k \cdot \boldsymbol{\sigma}^{(2)}), \\ R_{jk}^{(c)} &= \frac{1}{2} R_j^{(4)} \otimes R_k^{(4)} + \frac{1}{2} R_k^{(4)} \otimes R_j^{(4)} \\ &= \frac{1}{8} (2I_4 + 3(\mathbf{t}_j + \mathbf{t}_k) \cdot (\boldsymbol{\sigma}^{(1)} + \boldsymbol{\sigma}^{(2)}) \\ &\quad + 9\boldsymbol{\sigma}^{(1)} \cdot (\mathbf{t}_j \mathbf{t}_k + \mathbf{t}_k \mathbf{t}_j) \cdot \boldsymbol{\sigma}^{(2)}). \end{aligned} \quad (34)$$

Accordingly,

$$\rho = \sum_{k=1}^4 q_k^{(s)} R_k^{(s)} + \sum_{j < k} q_{jk}^{(c)} R_{jk}^{(c)}, \quad (35)$$

reconstructs  $\rho$  from the probabilities

$$q_k^{(s)} = \text{Tr}(\Pi_k^{(s)} \rho), \quad q_{jk}^{(c)} = \text{Tr}(\Pi_{jk}^{(c)} \rho), \quad (36)$$

which means

$$\begin{aligned} \mathbf{s} &= 3 \sum_{k=1}^4 q_k^{(s)} \mathbf{t}_k + \frac{3}{2} \sum_{j < k} q_{jk}^{(c)} (\mathbf{t}_j + \mathbf{t}_k), \\ \mathbf{C} &= 9 \sum_{k=1}^4 q_k^{(s)} \mathbf{t}_k \mathbf{t}_k + \frac{9}{2} \sum_{j < k} q_{jk}^{(c)} (\mathbf{t}_j \mathbf{t}_k + \mathbf{t}_k \mathbf{t}_j), \end{aligned} \quad (37)$$

for the Bloch vector  $\mathbf{s}$  and the dyad  $\mathbf{C}$ . For the  $\mathbf{s}$  and  $\mathbf{C}$  in Eq. (9), the sum rules

$$\sum_{k=1}^4 q_k^{(s)} = \frac{1}{3}, \quad \sum_{j=1}^3 \sum_{k=j+1}^4 q_{jk}^{(c)} = \frac{2}{3} \quad (38)$$

apply.

The data are the counts  $n_k^{(s)}$  and  $n_{jk}^{(c)}$  of the ten different outcomes with the total count of detected pairs

$$N = \sum_{k=1}^4 n_k^{(s)} + \sum_{j < k} n_{jk}^{(c)}. \quad (39)$$

The law of large numbers states that the relative frequencies  $n_k^{(s)}/N$  and  $n_{jk}^{(c)}/N$  approximate the respective probabilities when  $N \gg 1$  and, therefore, we get an estimate for  $\rho$  by replacing the probabilities in Eq. (35) by the relative frequencies

$$\rho \cong \rho^{(\text{LI})} = \frac{1}{N} \sum_{k=1}^4 n_k^{(s)} R_k^{(s)} + \frac{1}{N} \sum_{j < k} n_{jk}^{(c)} R_{jk}^{(c)}, \quad (40)$$

the analog of Eq. (23). As discussed above, this linear inversion is problematic because the relative frequencies do not obey the constraints that apply to the probabilities, such as the sum rules in Eq. (38). For instance, while  $\text{Tr}(I_{\text{sg}} \rho) = 0$  for the actual  $\rho$ , we have

$$\text{Tr}(I_{\text{sg}} \rho^{(\text{LI})}) = -\frac{2}{N} \sum_{k=1}^4 n_k^{(s)} + \frac{1}{N} \sum_{j < k} n_{jk}^{(c)}, \quad (41)$$

which is almost always nonzero and negative half the time. While we can improve matters a bit by removing the singlet

component from  $\rho^{(LI)}$ , the resulting mixed triplet state is almost always *not* a rank-two separable state and can have a negative eigenvalue in the triplet sector.

### B. Plausible states

Rather than merely finding the point estimators of the linear inversion method, the maximum likelihood method, or yet other methods, we identify the plausible region in the parameter space [31, Sec. 4.5.2], the set of all separable rank-two triplet states that are supported by the data. Every  $\rho$  in the plausible region is an acceptable point estimator; additional criteria, beyond what the data tell us, would be needed for selecting a particular one. While the linear inversion estimator  $\rho^{(LI)}$  is usually improper, the maximum likelihood estimator  $\rho^{(ML)}$  is always plausible because the plausible region is one of the optimal error regions, which happen to be regions around the maximum likelihood estimator [32].

If we denote the prior probability element of the vicinity of  $\rho(\theta)$  by  $(d\theta)$ , the posterior probability element is

$$(d\theta)_{\text{post}} = \frac{(d\theta)L(\theta)}{\int(d\theta')L(\theta')}. \quad (42)$$

The data give evidence in favor of  $\rho(\theta)$  if  $(d\theta)_{\text{post}} > (d\theta)$  and evidence against  $\rho(\theta)$  if  $(d\theta)_{\text{post}} < (d\theta)$ ; the data are neutral when  $(d\theta)_{\text{post}} = (d\theta)$ . The plausible region is composed of all  $\rho(\theta)$ s with evidence in their favor, that is,

the density matrix  $\rho(\theta)$  is plausible if

$$L(\theta) > \int(d\theta')L(\theta') \quad (43)$$

and only then.

This is an application of the principle of evidence [31, Sec. 4.2]; another application to quantum data is reported in [33].

For the given data, we find the maximum likelihood estimator  $\rho^{(ML)} = \rho(\theta^{(ML)})$  and then the number

$$\lambda_{\text{pl}} = \frac{1}{L(\theta^{(ML)})} \int(d\theta)L(\theta) < 1, \quad (44)$$

so that  $\rho(\theta)$  is plausible if

$$\lambda(\theta) = \frac{L(\theta)}{L(\theta^{(ML)})} > \lambda_{\text{pl}}. \quad (45)$$

The data give strong evidence if the prior content (“size”) of the plausible region is small and its posterior content (“credibility”) is large; these are

$$\begin{aligned} \text{size: } s_{\text{pl}} &= \int(d\theta)\chi(\lambda(\theta) > \lambda_{\text{pl}}), \\ \text{credibility: } c_{\text{pl}} &= \int(d\theta)_{\text{post}}\chi(\lambda(\theta) > \lambda_{\text{pl}}), \end{aligned} \quad (46)$$

where  $\chi(A) = 1$  if the statement A is true and  $\chi(A) = 0$  if it is false.

When there are many data so that the total count  $N$  of detection events is large and the law of large numbers (central limit theorem) is applicable, the  $N$  dependence of  $\lambda_{\text{pl}}$ ,  $s_{\text{pl}}$ , and

$c_{\text{pl}}$  is given by [34, Sec. 7.4]

$$\lambda_{\text{pl}} \propto N^{-5/2}, \quad (47)$$

with the proportionality factor depending on the relative frequencies and

$$\begin{aligned} s_{\text{pl}} &\cong \lambda_{\text{pl}} \frac{\log(1/\lambda_{\text{pl}})^{5/2}}{(5/2)!} \propto N^{-5/2} \log(N)^{5/2}, \\ 1 - c_{\text{pl}} &\cong (1 - c_{\text{pl}})_{\text{c.l.t.}} = \int_0^{\lambda_{\text{pl}}} d\lambda \frac{\log(1/\lambda)^{3/2}}{(3/2)!} \\ &= s_{\text{pl}} \left( \frac{5}{2} \log(1/\lambda_{\text{pl}})^{-1} + \frac{15}{4} \log(1/\lambda_{\text{pl}})^{-2} \right) \\ &\quad + \text{erfc}(\log(1/\lambda_{\text{pl}})^{1/2}) \\ &\cong \lambda_{\text{pl}} \frac{\log(1/\lambda_{\text{pl}})^{3/2}}{(3/2)!} \propto N^{-5/2} \log(N)^{3/2}. \end{aligned} \quad (48)$$

Accordingly, the size of the plausible region shrinks with growing  $N$  and so does the gap between its credibility and unity. When there are many data, the plausible region is very small and has very large credibility (very small prior and very large posterior probability).

The five-dimensional integrals in Eqs. (42) to (46) are evaluated with Monte Carlo methods as follows [34, Sec. 8.2]. We draw a large sample of  $\theta$ s from the prior distribution and store the corresponding  $\lambda(\theta)$  values:  $\lambda^{(1)}, \lambda^{(2)}, \dots, \lambda^{(m)}, \dots, \lambda^{(M)}$ . Then,

$$\begin{aligned} \lambda_{\text{pl}} &\cong \frac{1}{M} \sum_{m=1}^M \lambda^{(m)}, \\ s_{\text{pl}} &\cong \frac{1}{M} \sum_{m=1}^M \chi(\lambda^{(m)} > \lambda_{\text{pl}}), \\ c_{\text{pl}} &\cong \frac{1}{M\lambda_{\text{pl}}} \sum_{m=1}^M \lambda^{(m)} \chi(\lambda^{(m)} > \lambda_{\text{pl}}) \end{aligned} \quad (49)$$

are approximate values for  $\lambda_{\text{pl}}$ ,  $s_{\text{pl}}$ , and  $c_{\text{pl}}$  with a sampling error  $\propto 1/\sqrt{M}$ . As it is usually CPU-cheap to draw from the prior distribution, we can easily have samples that are so large that the sampling error is of no concern.

### C. Simulated data

We simulated five runs of the experiment of Fig. 3 for the Bloch vectors specified by

$$\begin{aligned} \begin{pmatrix} \mathbf{a} \cdot \mathbf{t}_1 \\ \mathbf{a} \cdot \mathbf{t}_2 \\ \mathbf{a} \cdot \mathbf{t}_3 \\ \mathbf{a} \cdot \mathbf{t}_4 \end{pmatrix} &= \frac{1}{\sqrt{27}} \begin{pmatrix} -3 \\ -1 \\ -1 \\ 5 \end{pmatrix}, \\ \begin{pmatrix} \mathbf{b} \cdot \mathbf{t}_1 \\ \mathbf{b} \cdot \mathbf{t}_2 \\ \mathbf{b} \cdot \mathbf{t}_3 \\ \mathbf{b} \cdot \mathbf{t}_4 \end{pmatrix} &= \frac{1}{\sqrt{105}} \begin{pmatrix} -3 \\ 9 \\ 1 \\ -7 \end{pmatrix}, \end{aligned} \quad (50)$$

and the probabilities  $p_0 = 0.37$ ,  $p_1 = 0.63$ . Figure 4 shows  $\lambda_{\text{pl}}$ ,  $s_{\text{pl}}$ , and  $c_{\text{pl}}$  for  $N = 100, 200, 300, \dots, 5000$  for the five runs. There is very little variation between the runs.

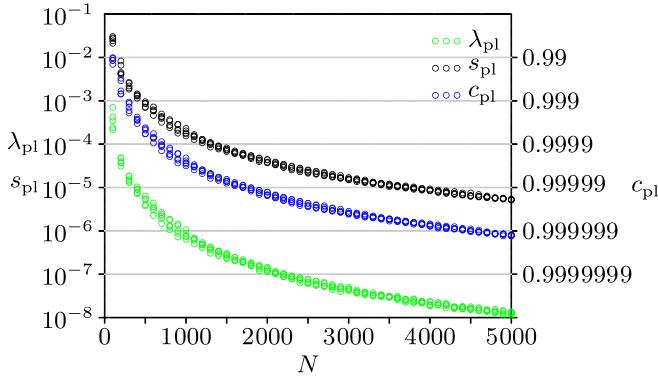


FIG. 4. The values of  $\lambda_{\text{pl}}$ ,  $s_{\text{pl}}$ , and  $c_{\text{pl}}$  as a function of  $N$  for five simulated experiments with up to 5000 detected pairs each. We observe that there is little variation between the experiments and that a few thousand detected pairs is enough to ensure a small size and large credibility of the plausible region.

The plausible region has a credibility that exceeds 0.9999 for  $N = 1000$ , 0.99999 for  $N = 2000$ , and 0.999999 for  $N = 5000$ , while the size is less than  $10^{-3}$ ,  $10^{-4}$ , and  $10^{-5}$ , respectively. It follows that a few thousand detected pairs are quite enough to acquire very strong evidence in favor of a tiny subset of the separable rank-two triplet states and against all others. We verified that the  $\rho(\theta)$  used for the simulation is in the plausible region for  $N = 100, 200, \dots, 5000$ , as it is expected to be: For  $N = 100$  already, the credibility is 0.99 so that the true state has an outside probability of only 1%. In passing, we note that the Monte Carlo integration requires ridiculously large samples for  $N > 5000$  because the peak of the likelihood function around the maximum likelihood estimator is then extremely narrow.

We used a sample with  $M = 2 \times 10^9$  entries for the Monte Carlo integration. It was drawn from the prior probability distribution that has the Bloch vectors  $\mathbf{a}$  and  $\mathbf{b}$  independently uniformly (isotropically) distributed on the unit sphere and the probability parameter  $\alpha$  uniformly distributed between 0 and  $\pi/2$ . This prior correctly reflects our complete ignorance about  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $p_0$ , and  $p_1$  before we take data, whereby we recall Wootters's insight [35] that the Jeffreys prior [36] is most natural for the probabilities.

Let us exploit the simulated data for a check of the large- $N$  approximations in Eqs. (47) and (48). Figure 5 summarizes the data of the five simulated experiments. Figure 5(a) shows the values of  $N^{5/2}\lambda_{\text{pl}}$ ; 5(b) those of  $N^{5/2}\log(N)^{-5/2}s_{\text{pl}}$ ; 5(c) those of  $N^{5/2}\log(N)^{-3/2}(1 - c_{\text{pl}})$ ; and 5(d) those of  $(1 - c_{\text{pl}})/(1 - c_{\text{pl}})_{\text{c.l.t.}}$ . It appears that a few thousand detected pairs are sufficiently many to reach the asymptotic regime, where the ratio in Fig. 5(d) is unity, and Figs. 5(a) to 5(c) show fluctuations around an  $N$ -independent value. We observe that the ratio in Fig. 5(d) is least sensitive to the fluctuations in the relative frequencies.

#### D. Imperfections in real-life experiments

The above discussion of the tetrahedron POVM assumes an ideal realization, whereas an actual experiment will have imperfections. In particular, there are deviations from the per-

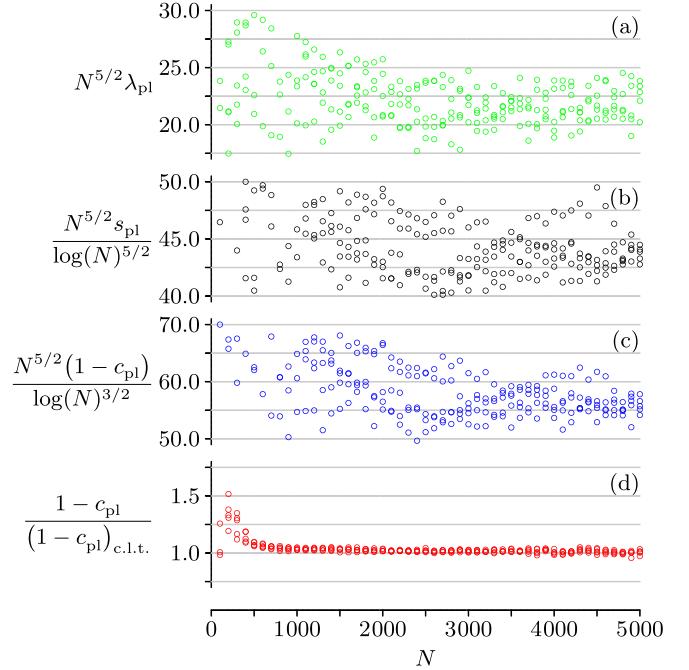


FIG. 5. An illustration of the large- $N$  approximations in Eqs. (47) and (48). For  $N = 100, 200, 300, \dots, 5000$ , panel (a) shows the values of  $N^{5/2}\lambda_{\text{pl}}$ ; panel (b) those of  $N^{5/2}\log(N)^{-5/2}s_{\text{pl}}$ ; panel (c) those of  $N^{5/2}\log(N)^{-3/2}(1 - c_{\text{pl}})$ ; and panel (d) those of  $(1 - c_{\text{pl}})/(1 - c_{\text{pl}})_{\text{c.l.t.}}$ . As in Fig. 4, there is one circle for each of the five simulated runs of the experiment of Fig. 3.

fect tetrahedron geometry, the detectors at the four outputs have nonideal detection efficiencies, and the optical elements between the source and the detectors will absorb or deflect a small fraction of the photons. As a consequence, there will be cases where only one of the two photons is detected or both escape detection. The ten POVM elements of Eqs. (28) and (29) are then modified accordingly and supplemented by five additional elements that account for the single-photon detection events and the null event. It is well known how to do all that (see, for example, [33,37]) and we shall address these matters in due course, namely when experimental data will be available for evaluation.

## V. CONCLUSION

To conclude, we considered the following problem. One receives a sequence of qubits, where each qubit is in one of two unknown states, and the goal is to determine the states. Solving this requires some extra information. We showed that providing additional quantum information, in the form of an additional copy of the state, so that one is receiving a sequence of pairs of states instead of individual states, allows one to determine the unknown quantum states as well as their probabilities of occurrence.

Proper quantum state tomography provides the data from which one can learn the unknown states and probabilities. We analyzed the SIC POVM and showed that the linear inversion method and the maximum likelihood method can be used, the former with caution. We also proposed a tomography experiment that can be realized with available technology and

demonstrated that a few thousand detected pairs are enough to locate the states and probabilities within a very small region with very high probability.

### ACKNOWLEDGMENTS

M.H. would like to thank Sarah Croke for a useful conversation. This research was supported by NSF Grant No. FET-2106447. This project is supported by the National Research Foundation, Singapore through the National Quantum Office, hosted in A\*STAR, under its Centre for Quantum Technologies Funding Initiative (Grant No. S24Q2d0009).

### APPENDIX A: SIC POVM AND LINEAR INVERSION

When the qubit pairs are measured by the SIC POVM of Eqs. (15) to (18), there are no counts associated with  $\Pi_0$  so that

$$L(\mathbf{n}|\mathbf{q}) = \frac{N!}{\prod_{j=1}^9 n_j!} \prod_{k=1}^9 q_k^{n_k}. \quad (\text{A1})$$

is the likelihood of the data  $\mathbf{n} = (n_1, n_2, \dots, n_9)$ , the list of detection event counts, given the list of detection probabilities  $\mathbf{q} = (q_1, q_2, \dots, q_9)$ ; the combinatorial factor is that for a prechosen number  $N$  of detected pairs. The expected value of a function of the data

$$\mathbb{E}(f(\mathbf{n})) = \sum_{\mathbf{n}} f(\mathbf{n}) L(\mathbf{n}|\mathbf{q}), \quad (\text{A2})$$

considers all thinkable data. The generating function

$$g(\mathbf{x}) = \mathbb{E}\left(\prod_{j=1}^9 x_j^{n_j}\right) = \left(\sum_{j=1}^9 x_j q_j\right)^N \quad (\text{A3})$$

provides the expected values of powers of the counts  $n_j$ , in particular,

$$\begin{aligned} \mathbb{E}(n_k) &= x_k \frac{\partial}{\partial x_k} g(\mathbf{x}) \Big|_{\text{all } x_j=1} = N q_k, \\ \mathbb{E}(n_j n_k) &= N(N-1) q_j q_k + \delta_{jk} N q_k. \end{aligned} \quad (\text{A4})$$

It follows that the linear inversion estimator in Eq. (23) is unbiased,  $\mathbb{E}(\rho^{(\text{LI})}) = \rho$ . The Hilbert–Schmidt norm of the error  $\Delta\rho = \rho^{(\text{LI})} - \rho$  is given by

$$\|\Delta\rho\|^2 = \text{Tr}((\Delta\rho)^2) = 12 \sum_{j=1}^9 \left(\frac{n_j}{N} - q_j\right)^2; \quad (\text{A5})$$

as a consequence of the purity constraint in Eq. (22), the expected value

$$\mathbb{E}(\|\Delta\rho\|^2) = \frac{12}{N} \left(1 - \sum_{j=1}^9 q_j^2\right) \quad (\text{A6})$$

is bounded by  $10/N \leq \mathbb{E}(\|\Delta\rho\|^2) \leq 32/(3N)$ .

We want to use the empirical density matrix  $\rho^{(\text{LI})}$  to find the triplet ket  $|\xi\rangle$  orthogonal to  $|\psi_0\rangle \otimes |\psi_0\rangle$  and  $|\psi_1\rangle \otimes |\psi_1\rangle$ .

We recall that

$$\rho^{(\text{LI})} = \sum_{j=0}^2 r_j |u_j\rangle \langle u_j| \quad \text{with} \quad r_0 \leq r_1 \leq r_2 \quad (\text{A7})$$

is hermitian, but it has rank three almost always and the eigenvalue  $r_0$  can be negative in this spectral decomposition. We have that  $\rho|\xi\rangle = 0$  and want to find a lower bound on the overlap between  $|\xi\rangle$  and  $|u_0\rangle$  under the assumption that  $\|\Delta\rho\| = \epsilon \ll 1$ , which we can ensure by choosing  $N$  large enough. The eigenvalues of  $\rho^{(\text{LI})}$  are then  $\epsilon$ -close to those of  $\rho$ ,

$$\begin{aligned} r_0 &\cong 0, \\ r_1 \Big\} &\cong \frac{1}{2} \mp \frac{1}{2} \sqrt{(p_0 - p_1)^2 + 4p_0 p_1 |\langle \psi_0 | \psi_1 \rangle|^4}. \end{aligned} \quad (\text{A8})$$

We take for granted that  $\epsilon$  is small enough that we can distinguish between  $r_0 \cong 0$  and  $r_1 > 0$ .

Now,

$$|\langle \xi | \Delta\rho | \xi \rangle| = \left| \sum_{j=0}^2 r_j |\langle u_j | \xi \rangle|^2 \right| \leq \epsilon, \quad (\text{A9})$$

where

$$|\langle u_0 | \xi \rangle|^2 = 1 - \sum_{j=1}^2 |\langle u_j | \xi \rangle|^2. \quad (\text{A10})$$

Since

$$\sum_{j=1}^2 |\langle u_j | \xi \rangle|^2 \leq \frac{1}{r_1} \sum_{j=1}^2 r_j |\langle u_j | \xi \rangle|^2 \leq \frac{\epsilon}{r_1} \quad (\text{A11})$$

if  $r_0 \geq 0$ , it follows that

$$|\langle u_0 | \xi \rangle|^2 \geq 1 - \frac{\epsilon}{r_1} \quad (\text{A12})$$

in this case. If  $r_0 < 0$ , then

$$\begin{aligned} \epsilon &\geq -|r_0| |\langle u_0 | \xi \rangle|^2 + \sum_{j=1}^2 r_j |\langle u_j | \xi \rangle|^2 \\ &\geq -|r_0| |\langle u_0 | \xi \rangle|^2 + r_1 (1 - |\langle u_0 | \xi \rangle|^2), \end{aligned} \quad (\text{A13})$$

so

$$|\langle u_0 | \xi \rangle|^2 \geq \frac{r_1 - \epsilon}{r_1 + |r_0|}. \quad (\text{A14})$$

Note that  $\langle u_0 | \Delta\rho | u_0 \rangle = r_0 - \langle u_0 | \rho | u_0 \rangle$  so that  $\epsilon \geq |r_0| + \langle u_0 | \rho | u_0 \rangle \geq |r_0|$  in this case. Therefore,

$$|\langle u_0 | \xi \rangle|^2 \geq 1 - \frac{|r_0| + \epsilon}{r_1 + |r_0|} \geq 1 - \frac{2\epsilon}{r_1 + |r_0|}, \quad (\text{A15})$$

and the second term is of order  $\epsilon$ .

What we see from this is that if  $r_1$  is small (recall that  $|r_0|$  is assumed to be substantially smaller) then a very small value of  $\epsilon$  will be required for  $|\langle u_0 | \xi \rangle|$  to be close to one. We can identify  $\epsilon$  with the square root of the expected value in

Eq. (A6), so this gives us an estimate for how many qubit pairs we will have to detect to obtain a good estimate of  $|\xi\rangle$ .

## APPENDIX B: PLAUSIBLE REGION WHEN $N \gg 1$

The consideration of all thinkable data in Eqs. (A1) to (A6) is appropriate when planning the experiment. Once the experiment has been performed, we draw inference from the actual data in conjunction with what we knew before we had the data. This prior knowledge is reflected in the prior probability element

$$(d\theta) = d\theta_0 d\phi_0 d\theta_1 d\phi_1 d\alpha w(\theta), \quad (B1)$$

where  $w(\theta)$  is the prior probability density in this parametrization. For the list  $\mathbf{q}$  of the ten probabilities of the tetrahedron POVM in Eq. (36) and the corresponding list  $\mathbf{n}$  of counts, the

fractional likelihood is

$$\frac{L(\mathbf{n}|\mathbf{q}(\theta))}{L(\mathbf{n}|\mathbf{q}(\theta^{(ML)}))} = \prod_j (q_j(\theta)/q_j(\theta^{(ML)}))^{n_j} \cong \exp\left(-\frac{N}{2} \sum_{j,k=1}^5 \varepsilon_j Q_{jk}(\theta^{(ML)}) \varepsilon_k\right), \quad (B2)$$

where  $\theta - \theta^{(ML)} = (\varepsilon_1, \dots, \varepsilon_5)$  and  $Q_{jk}(\theta^{(ML)})$  is a matrix element of a symmetric positive  $5 \times 5$  matrix; the approximation is valid when  $N$  is so large that we can invoke the central limit theorem. Under these circumstances, only the vicinity of  $\theta^{(ML)}$  contributes substantially to the integrals in Eqs. (44) and (46), and correspondingly, the prior probability element can be replaced by

$$(d\theta) \cong d\varepsilon_1 \cdots d\varepsilon_5 w(\theta^{(ML)}). \quad (B3)$$

The large- $N$  approximations in Eqs. (47) and (48) follow [34].

- 
- [1] E. Aïmeur, G. Brassard, and S. Gambs, *Mach. Learn.* **90**, 261 (2013).
- [2] S. Lloyd, M. Mohseni, and P. Rebentrost, [arXiv:1307.0411](https://arxiv.org/abs/1307.0411).
- [3] N. Wiebe, A. Kapoor, and K. Svore, *Quantum Inf. Comput.* **15**, 0318 (2015).
- [4] I. Kerenidis, J. Landman, A. Luongo, and A. Prakash, *Adv. Neural Inf. Process.* **32** (2019).
- [5] M. Schuld, I. Sinayskiy, and F. Petruccione, *Contemp. Phys.* **56**, 172 (2015).
- [6] A. Bisio, G. Chiribella, G. M. D'Ariano, S. Facchini, and P. Perinotti, *Phys. Rev. A* **81**, 032324 (2010).
- [7] A. Bisio, G. M. D'Ariano, S. Facchini, P. Perinotti, and M. Sedlak, *Phys. Lett. A* **375**, 3425 (2011).
- [8] M. Guță and W. Kotlowski, *New J. Phys.* **12**, 123032 (2010).
- [9] G. Sentís, J. Calsamiglia, R. Muñoz-Tapia, and E. Bagan, *Sci. Rep.* **2**, 708 (2012).
- [10] M. Fanizza, A. Mari, and V. Giovannetti, *IEEE Trans. Inf. Theory* **65**, 5931 (2019).
- [11] G. Sentís, A. Monras, R. Muñoz-Tapia, J. Calsamiglia, and E. Bagan, *Phys. Rev. X* **9**, 041029 (2019).
- [12] M. Hillary, *J. Phys. A* **53**, 425302 (2020).
- [13] A. Peres, *Quantum Theory: Concepts and Methods* (Kluwer, Alphen aan den Rijn, The Netherlands, 1995), Exercise 2.1.
- [14] Y. Zhan and M. G. A. Paris, *Int. J. Softw. Inform.* **8**, 241 (2014).
- [15] A. Anshu and S. Arunachalam, *Nat. Rev. Phys.* **6**, 59 (2024).
- [16] A. Chefles, *Phys. Lett. A* **239**, 339 (1998).
- [17] A. Keil, *Int. J. Quantum Inf.* **22**, 2440008 (2024).
- [18] A. Sanpera, R. Tarrach, and G. Vidal, *Phys. Rev. A* **58**, 826 (1998).
- [19] B.-G. Englert and N. Metwally, *J. Mod. Opt.* **47**, 2221 (2000).
- [20] B.-G. Englert and N. Metwally, Kinematics of qubit pairs, in *Mathematics of Quantum Computation*, edited by G. Chen and R. K. Brylinski (CRC Press, Boca Raton, FL, 2002), Chap. 2, pp. 25–75.
- [21] We would like to thank an anonymous referee for pointing this out.
- [22] T. Heinosaari and M. Ziman, *The Mathematical Language of Quantum Theory* (Cambridge University Press, Cambridge, England, 2012).
- [23] H. B. Dang, K. Blanchfield, I. Bengtsson, and D. M. Appleby, *Quantum Inf. Process.* **12**, 3449 (2013).
- [24] B. Stacey, *Mathematics* **4**, 36 (2016).
- [25] We ignore the combinatorial factors that have no bearing on what follows.
- [26] N. Hansen, [arXiv:1604.00772](https://arxiv.org/abs/1604.00772).
- [27] J. Řeháček, B.-G. Englert, and D. Kaszlikowski, *Phys. Rev. A* **70**, 052321 (2004).
- [28] B.-G. Englert, K. M. Tin, C. G. Goh, and H. K. Ng, *Laser Phys.* **15**, 7 (2005).
- [29] A. Ling, K. P. Soh, A. Lamas-Linares, and C. Kurtsiefer, *J. Mod. Opt.* **53**, 1523 (2006).
- [30] A. Ling, K. P. Soh, A. Lamas-Linares, and C. Kurtsiefer, *Phys. Rev. A* **74**, 022309 (2006).
- [31] M. Evans, *Measuring Statistical Evidence Using Relative Belief* (CRC Press, Boca Raton, FL, 2015).
- [32] J. Shang, H. K. Ng, A. Sehrawat, X. Li, and B.-G. Englert, *New J. Phys.* **15**, 123026 (2013).
- [33] Y. Gu, W. Li, M. Evans, and B.-G. Englert, *Phys. Rev. A* **99**, 022112 (2019).
- [34] B.-G. Englert, *Lectures on Quantum State Estimation* (World Scientific, Singapore, 2025).
- [35] W. K. Wootters, *Phys. Rev. D* **23**, 357 (1981).
- [36] H. Jeffreys, *Proc. R. Soc. London A* **186**, 453 (1946).
- [37] J. Y. Sim, J. Shang, H. K. Ng, and B.-G. Englert, *Phys. Rev. A* **100**, 022333 (2019).