Joint Quantum State and Measurement Tomography with Incomplete Measurements

Adam C. Keith,[1,2,*] Charles H. Baldwin,[1] Scott Glancy,[1,†] and E. Knill[1,3]

[1] *Applied and Computational Mathematics Division, National Institute of Standards and Technology, Boulder, Colorado, 80305, USA*

[2] *Department of Physics, University of Colorado, Boulder, Colorado, 80309, USA*

[3] *Center for Theory of Quantum Matter, University of Colorado, Boulder, Colorado 80309, USA*

(Dated: March 22, 2018)

Estimation of quantum states and measurements is crucial for the implementation of quantum information protocols. The standard method for each is quantum tomography. However, quantum tomography suffers from systematic errors caused by imperfect knowledge of the system. We present a procedure to simultaneously characterize quantum states and measurements that mitigates systematic errors by use of a single high-fidelity state preparation and a limited set of high-fidelity unitary operations. Such states and operations are typical of many state-of-the-art systems. For this situation we design a set of experiments and an optimization algorithm that alternates between maximizing the likelihood with respect to the states and measurements to produce estimates of each. In some cases, the procedure does not enable unique estimation of the states. For these cases, we show how one may identify a set of density matrices compatible with the measurements and use a semi-definite program to place bounds on the state's expectation values. We demonstrate the procedure on data from a simulated experiment with two trapped ions.

PACS numbers: 03.65.Wj, 03.67.-a, 37.10.Ty,

I. INTRODUCTION

Recent experiments have demonstrated high-fidelity unitary operations in various platforms for quantum information processing; for examples see Refs. [1–6]. Even in the most advanced systems, some operations are harder to accomplish and have significantly lower fidelity than others. For example, in the systems reported in Refs. [1–6] single-qubit gates are accomplished with significantly higher fidelity than two-qubit gates. A natural question is then, how can we use the high-fidelity operations to diagnose other parts of the quantum system? In this work, we propose such a procedure that uses a single high-fidelity state initialization and a limited set of unitary operations to diagnose other state preparations and measurement operators.

The standard method to diagnose state preparations and measurements is quantum tomography (QT). Quantum state tomography (QST) is a procedure to estimate an unknown quantum state from experimental data. When the measurements are informationally complete, the resulting data can be used to estimate the corresponding density matrix [7, 8], and we say that the state is "identifiable." Quantum detector tomography (QDT) is a procedure to estimate an unknown quantum measurement operator [9, 10]. When the unknown measurement is applied to an informationally-complete set of already known quantum states, the resulting data can be used to create an estimate of the measurement operator, and the measurements are identifiable.

A drawback to standard QST and QDT is that they require well-known measurements and state preparations respectively. However, it is difficult to produce such measurements and state preparations when some processes have significantly lower fidelity than others. Attempting standard QT in this case typically results in estimates with systematic errors.

To combat systematic errors, we adapt standard QST and QDT to the situation where a single state preparation and a limited set of unitary operations (for example, single-qubit rotations) have significantly higher fidelity than other processes and measurements. For this situation, we develop a procedure to estimate other states and the measurement operators simultaneously with an alternating maximum likelihood estimation (MLE) algorithm. When the measurements are not informationally complete the state is non-identifiable. However, it is "set identifiable," meaning that we can specify a set of density matrices that are compatible with the measurements. This set can be used to upper- and lower- bound important quantities like fidelity of the state preparations or expectation values of other observables. The estimates also provide information about the quantum processes that produced the unknown states.

Systems with a limited set of high-fidelity processes are common in state-of-the art quantum information experiments. Our original motivation comes from trapped-ion experiments where single-qubit gates have been demonstrated with high-fidelity, but entangling gates have lower fidelities. The procedure described here was implemented in trapped-ion experiments described in Ref. [1, 11]. We return to the trapped-ion example throughout this paper to illustrate our procedure.

QT with unknown states and measurements has been considered in previous work [12, 13]. Our procedure dif-

* adam.keith@colorado.edu

† sglancy@nist.gov

fers from these works by using alternating MLE, which yields estimates consistent with all the data collected. Our procedure also resembles other variants of QT, such as self-consistent [14, 15] and gate-set tomography (GST) [16]. These methods treat the quantum system as a "black-box" about which nothing (or almost nothing) is known. Notably, in GST one has access to a collection of unknown quantum processes or "gates," and one creates a series experiments, applying the gates in different orders. By performing the proper experiments, one can find a full estimate of all the gates simultaneously (up to an unobservable gauge) [16]. The method has been successful in experiments for single qubit systems [17–19]. In GST, the goal is to learn everything about the system. In our procedure, we assume more about the system and ask less about the outcome, thereby requiring fewer experiments. While this requires stronger pre-experiment knowledge about the system than is available in GST, we believe that it is well suited to a situation that occurs in many state-of-the-art experiments and complements these other proposals.

We begin by defining and describing the standard methods for QST and QDT in Sec. II. We also discuss and physically motivate some additional assumptions about the measurement that greatly simplify our procedure. Then, in Sec. III, we introduce the experiments required for our procedure and discuss the application to the example trapped-ion system. Next, we discuss our numerical technique to extract estimates of the states and measurements from the experimental data in Sec. IV. In Sec. V, we show how to upper- and lower-bound important expectation values if the states are only set-identifiable. We then discuss how to estimate the uncertainties Sec. VI. We summarize our procedure and discuss future directions in Sec. VII. We also include three appendices that describe our software implementation (App. A), a method for coarse-graining measurement outcomes (App. B), and the stopping criteria for iterations of the likelihood maximization (App. C).

II. GENERAL MODEL

Our procedure is rooted in QST and QDT, so we begin with a brief description of the standard version of each. We focus on quantum systems that are described with finite, d-dimensional Hilbert spaces. QST is a procedure to estimate the unknown $d \times d$ density matrix ρ, that describes the state of a quantum system. To estimate ρ, we prepare many identical copies of the quantum state and apply a known quantum measurement to each. In the following, a sequence of identical state preparations and measurements is called an "experiment", while a particular state preparation and measurement in an experiment is referred to as a "trial." A quantum measurement is associated with the measurement operators F_b of a POVM $\{F_b\}_b$, where $F_b \geq 0$ and $\sum_b F_b = \mathbb{1}$. The probability of outcome b in a trial is given by the Born rule $p_b = \mathrm{Tr}\,(F_b\rho)$. In some instances, several different POVMs are used for QST, in these cases we perform separate experiments for each POVM. The QST formalism assumes that uncertainty about the measurement operators of the POVMs is negligible. If the measurement operators from all experiments span the bounded operators on the Hilbert space, we call the set of POVMs "informationally complete" (IC). If the state's probabilities for the outcomes of a set of IC POVMs with specified measurement operators are exactly known, one has all the information necessary to exactly reconstruct the unknown density matrix [7, 8]. In an experimental implementation, the probabilities are never exactly known, because one only runs a finite number of trials. One therefore numerically estimates the quantum state from the measured frequencies of the outcomes via techniques such as MLE [20, 21].

In QDT, the goal is to estimate the measurement operators associated with an unknown quantum measurement. To estimate the measurement operators, we prepare many identical copies of members of a family of known quantum states $\{\rho_j\}_j$. We then perform experiments where we prepare many copies of one member of this family and apply the unknown quantum measurements. For a given trial in one of these experiments the probability of each outcome is given by the Born rule $p_{j,b} = \mathrm{Tr}\,(F_b\rho_j)$. In the QDT formalism, we assume the uncertainty about the states $\{\rho_j\}_j$ is negligible. We define an IC set of states for QDT as a set that spans the bounded operators on the Hilbert space, which is analogous to IC POVMs in QST. The probabilities from each outcome along with the exact description of an IC set of states allow for unique reconstruction of each measurement operator. As with QST, in practice we cannot determine the probability of each outcome due to a finite number of copies of the unknown state. Thus, the measurement operators are numerically estimated from the frequencies of the outcomes via techniques such as MLE [10].

In this work, we use aspects of QST and QDT to create a hybrid procedure that is useful when one has access to a limited set of high-fidelity unitary operations. We designate such a set of unitary operators with indices as $\boldsymbol{\mathcal{U}} = \{\mathcal{U}_i | i = 1, \ldots, r\}$, where $\mathcal{U}_i[\rho] = U_i \rho U_i^\dagger$ and U_0 is the identity, $U_0 = \mathbb{1}$. We additionally assume that we can reliably prepare a known state ρ_0. Then, by applying the high-fidelity processes to this state, we generate a family of known states $\boldsymbol{\rho} = \{\rho_i = \mathcal{U}_i[\rho_0]\}_i$. There could be repetitions in this family. In addition to these known states, other processes generate a family of unknown states $\boldsymbol{\sigma} = \{\sigma_j | j = 1, \ldots, s\}$ to be characterized. All states are read-out by an unknown quantum measurement described by the POVM $\boldsymbol{F} = \{F_b | b = 1, \ldots, M\}$. To measure in different bases, we apply one of the high-fidelity processes prior to the measurement, thereby creating the measurement operators $F_{i,b} = \mathcal{U}_i^\dagger[F_b] = U_i^\dagger F_b U_i$, where $F_{0,b} = F_b$ by definition of U_0.

To significantly simplify our procedure, we add an ad-

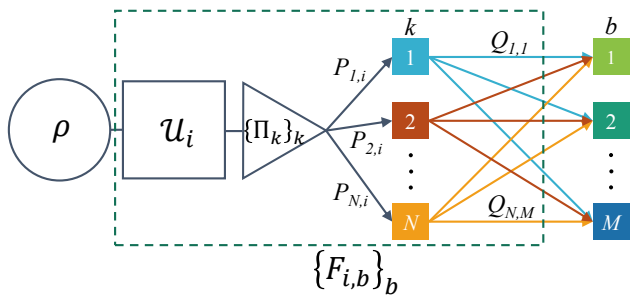

FIG. 1. Schematic describing the measurement model for a single engineered measurement $\{F_{i,b}\}_b$. A density matrix ρ is rotated by the unitary operator U_i, which determines the basis in which ρ is measured. In the measurement, first the underlying POVM $\{\Pi_k\}_k$ is applied, and result k is obtained with probability $P_{k,i}$. Then, a Markov process acts to give random outcome b with probability $Q_{k,b}$. In the experiment we only perceive the outcome b. Ellipses and dotted arrows show that there may be an arbitrary finite number of underlying measurement operators Π_k and values of b.

ditional assumption that the measurement operators F_b are unknown mixtures of the operators of an underlying POVM $\mathbf{\Pi} = \{\Pi_k | k = 1, \ldots, N\}$, whose members are well known. To avoid numerical instabilities, we assume that no two measurement operators in $\mathbf{\Pi}$ are equal. The resulting measurement is depicted in Fig. 1. We model the measurement process by the application of the underlying POVM $\mathbf{\Pi}$ yielding hidden outcome k, followed by an unknown Markov process, after which outcome b is observed. The Markov process is described by the transition matrix $Q_{k,b}$. (If the outcome b is a continuous variable, our procedure requires discretization.) This situation is common in many experiments, where physics constrains the underlying quantum measurement process, but subsequent incoherent effects make identification of the measurement operators corresponding to outcomes difficult. Previous techniques have further constrained the model by assuming a form for the Markov process. For example, in some ion trap experiments, the observed outcomes are assumed to be a mixture of Poissonians given the hidden outcomes [22, 23]. By not constraining the Markov process by a model, we avoid systematic errors from model mismatch.

The measurement operator F_b can now be expressed in terms of the Markov process and the underlying measurement operators as

$$F_b = \sum_k Q_{k,b} \Pi_k. \tag{1}$$

The probability of observing b given state ρ is $p_b = \mathrm{Tr}\,(F_b \rho)$. The POVM $\boldsymbol{F}$ is referred to as the "bare POVM." As mentioned above, we also apply the high-fidelity processes $\boldsymbol{\mathcal{U}}$ prior to the measurement to create "engineered POVMs" whose members are

$$F_{i,b} = \sum_k Q_{k,b} \Pi_{i,k}, \tag{2}$$

where $\Pi_{i,k} = \mathcal{U}_i^\dagger[\Pi_k]$. Because we designated $U_0 = \mathbb{1}$, $\{F_{i,b}\}_{i,b}$ is the complete family of measurement operators, bare and engineered.

In the theory described below, the underlying POVM $\mathbf{\Pi}$ may be any POVM, but we specifically discuss POVMs whose measurement operators are orthogonal subspace projectors. This further assumption was originally motivated by trapped-ion experiments where the underlying POVM consists of projectors onto orthogonal internal states of each ion. This model is a useful description because the quantization axis and measurements are aligned with high accuracy (further details are given in the next section). The situation also occurs in other experiments with similarly high accuracy alignment. If the assumption does not hold naturally, it can be easily enforced if the operations take place in a rotating frame. Then the measurements can be decohered by randomizing the phase (or time) between the operations and measurements for each trial. Then, averaged over many trials any coherence will be lost and the underlying measurement operators are orthogonal subspace projectors.

III. EXPERIMENTAL MEASUREMENT PROCEDURE

Our procedure consists of two experimental steps followed by numerical processing of the data. In this section, we describe the experimental steps and introduce an example application to trapped-ion systems to illuminate the discussion. A schematic for the measurement and numerical estimation procedure is in Fig. 2.

The first experimental step is related to standard QDT of the bare POVM. We call the experiments done during this step the "reference" experiments since they will be used to initialize the algorithm described in the next section. There is one reference experiment for each known pure state ρ_i in $\boldsymbol{\rho}$. The ith reference experiment, indexed i, consist of n trials. In each trial, we prepare ρ_i and apply the bare POVM. The probability for outcome b in a given trial is,

$$p_{i,b}^{(0)} = \mathrm{Tr}\,(F_b \rho_i) = \sum_k Q_{k,b} \mathrm{Tr}(\Pi_k \rho_i). \tag{3}$$

The measurement outcomes are sampled from the distribution given by Eq. (3). We assume the outcomes of the trials are independent and identically distributed. The outcomes of all trials from all reference experiments are collected into a matrix $H^{(0)}$ such that row i is the relative frequency histogram for experiment i. Specifically, the matrix element $H_{i,b}^{(0)}$ is the number of times outcome b is observed for experiment i divided by n. We call this matrix the "histogram matrix" in the following.

The second experimental step is similar to QST, where the goal is to extract information about the unknown quantum states $\boldsymbol{\sigma}$. We call the experiments done during this step the "probing experiments," because they are designed to probe the unknown states. There is one prob-

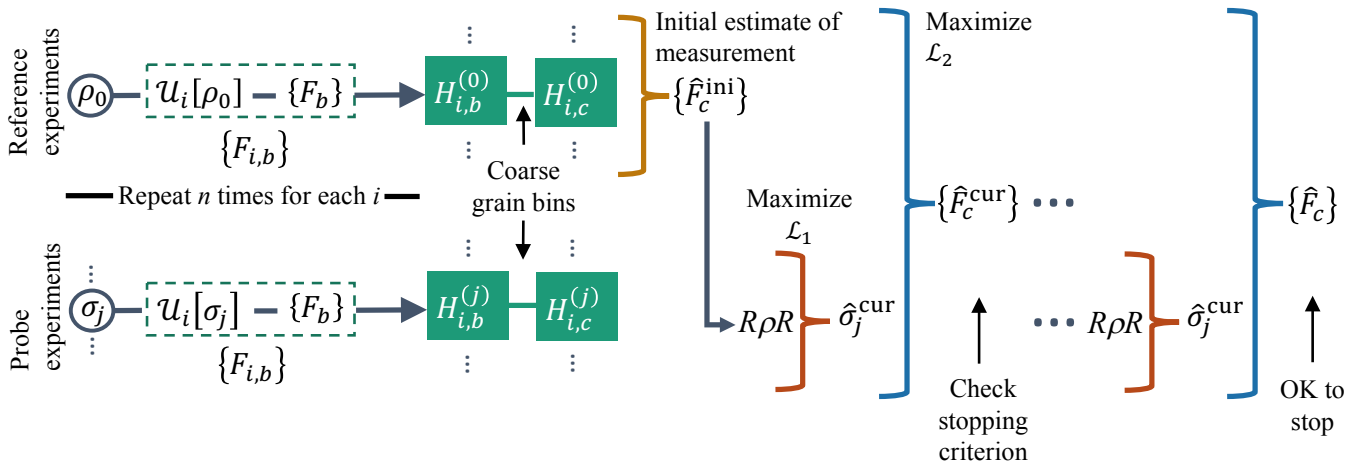

FIG. 2. Schematic showing measurement and estimation procedure. Reference experiments begin with the preparation of the known state ρ_0, and probe experiments begin with unknown states from $\boldsymbol{\sigma} = \{\sigma_j | j = 1, \ldots, s\}$. During each experiment the states are transformed by a process in the set $\boldsymbol{\mathcal{U}} = \{\mathcal{U}_i | i = 0 \ldots r\}$ and then measured with the POVM $\boldsymbol{F} = \{F_b | b = 1, \ldots, M\}$, giving outcome b. For simplicity, we assume every known process acts on each unknown state, but this is not necessary. After n trials of each experiment, histograms $\{H^{(j)}_{b,i}\}_{b,i}$ for each state are recorded. These histograms are then coarse-grained to produce new histograms $\{H^{(j)}_{c,i}\}_{c,i}$ based on a training data set composed of 10 % of the trials randomly selected without replacement from each reference experiment. From the reference experiments' data, we calculate an initial estimate for the POVM $\hat{\boldsymbol{F}}^{\text{ini}}$. We divide the likelihood maximization into two concave subproblems: maximization of $\mathcal{L}_1$ with respect to $\boldsymbol{\sigma}$ and maximization of $\mathcal{L}_2$ with respect to $\boldsymbol{F}$. We alternate between using the $R\rho R$ algorithm to maximize $\mathcal{L}_1$ and a standard nonlinear optimizer to maximize $\mathcal{L}_2$. At each iteration we find maximizing parameters $\hat{\boldsymbol{\sigma}}^{\text{cur}}$ and $\hat{\boldsymbol{F}}^{\text{cur}}$, and we check a stopping criterion. When the stopping criterion signals that we can stop iterations, we have the numerical maximum likelihood estimates $\hat{\boldsymbol{\sigma}}$ and $\hat{\boldsymbol{F}}$.

ing experiment for each engineered POVM and unknown state. The probing experiment, indexed by i, j, consists of n trials, where in each trial, we prepare σ_j and apply the engineered POVM indexed by i. The probability of getting outcome b in a given trial is

$$p^{(j)}_{i,b} = \text{Tr}(\mathcal{U}_i^\dagger[F_b]\sigma_j) = \sum_k Q_{k,b}\text{Tr}(\Pi_{i,k}\sigma_j). \tag{4}$$

The outcomes of the trials from the probing experiments on state σ_j are collected into the histogram matrix $H^{(j)}$.

In general, the probability of outcome b in a trial of a reference or probing experiment has the probability distribution,

$$p^{(j)}_{i,b} = \text{Tr}(\mathcal{U}_i^\dagger[F_b]\tau_j) = \sum_k Q_{k,b}\text{Tr}(\Pi_{i,k}\tau_j), \tag{5}$$

where $\boldsymbol{\tau} = (\tau_j)_{j=0}^s = (\rho_0, \sigma_1, \ldots, \sigma_s)$. For $j = 0$, Eq. (5) reduces to Eq. (3) and for $j > 0$ it reduces to Eq. (4).

In this discussion we have made a few simplifying assumptions about the measurement record, such as that every experiment contains the same number of trials and that every state τ_j is subjected to the same set of measurements. These assumptions are not necessary for our procedure and are made here only to simplify the mathematical notation.

As an example of the complete procedure we consider the task of diagnosing the internal state of two trapped ions, similar to Ref. [1]. Each ion is treated as a single qubit with computational basis $|\uparrow\rangle$, called the "bright state," and $|\downarrow\rangle$, called the "dark state." Optical pumping is used to initialize each ion in the bright state, $\rho_0 = |\uparrow\uparrow\rangle\langle\uparrow\uparrow|$, with high fidelity.

As in Ref. [1], we consider the high-fidelity processes to be single-qubit rotations applied to both ions. A single-qubit rotation is defined by

$$U(\theta, \phi) = \exp\left[-i\tfrac{\theta}{2}(\sigma_x \cos\phi + \sigma_y \sin\phi)\right], \tag{6}$$

where σ_x and σ_y are Pauli operators. Collective rotations $U(\theta, \phi)^{\otimes 2}$ can be accomplished with, e.g., microwaves applied uniformly across the ion trap. For the example discussed below, we choose the subset $\boldsymbol{U} = \{U(0,0)^{\otimes 2}, U(\frac{\pi}{2}, 0)^{\otimes 2}, U(\pi, 0)^{\otimes 2}, U(\frac{\pi}{2}, \frac{\pi}{2})^{\otimes 2}\}$ as the set of high-fidelity processes. Since two-qubit entangling operations have significantly lower fidelities, a natural task is to diagnose entangled states created by these operations.

Measurement of the ions is accomplished by stimulating a cycling transition between the bright state and another internal state (outside of the computational basis). This transition is well aligned with the quantization axis and highly detuned from the dark state, which implies observing fluorescence is a projective measurement of an ion in the bright state. The ions are close together in the trap, so the measurement cannot distinguish which ion is fluorescing. The corresponding POVM consists of subspace projectors, $\boldsymbol{\Pi} = \{\Pi_0 = |\downarrow\downarrow\rangle\langle\downarrow\downarrow|, \Pi_1 = |\downarrow\uparrow\rangle\langle\downarrow\uparrow| + |\uparrow\downarrow\rangle\langle\uparrow\downarrow|, \Pi_2 = |\uparrow\uparrow\rangle\langle\uparrow\uparrow|\}$. However, we can-

not observe the outcome of this POVM directly in any experiment since the number of photons in the fluorescence signal is distributed according to counting statistics as well as other detection errors such as dark counts, repumping to the dark state, or detector inefficiency. These effects act as the Markov process described above, so the observed outcomes are associated with the POVM, $\boldsymbol{F}$.

Simulated histograms of the above states are shown in Fig. 3. With the processes considered, we cannot create an IC set of engineered POVMs because the POVM $\boldsymbol{\Pi}$ (and therefore also $\boldsymbol{F}$) cannot distinguish the two single-ion-bright states and all of the rotations in $\boldsymbol{\mathcal{U}}$ act equally on both qubits. However, the measurements are sufficient to identify a Bell state fidelity, as is discussed in Sec. V.

IV. MAXIMUM LIKELIHOOD ESTIMATION TECHNIQUE

We now present a maximum likelihood estimation (MLE) technique that produces numerical estimates of the measurement operators and the unknown states. Like all MLE techniques, if the true operators and states are not on the boundary, the technique is optimal in that its variance asymptotically approaches the Cramér-Rao lower bound [24]. Our algorithm starts by coarse graining the outcome space b, which condenses the data to speed up the numerical processing. We then create an initial estimate of the transition matrix Q. Finally, we search for the estimates of Q and $\boldsymbol{\sigma}$ that maximize the total log-likelihood function of all the data. This is accomplished by an alternating approach. While this is not proven to yield the maximally likely operators and states in all instances, we have found no counterexamples and believe that the maximum likelihood solution is typically found.

The first step is to coarse grain the outcome space. In many cases, the set of possible values of the outcomes b is large, that is $b = 1, \ldots, M$ where $M \gg 1$. The collected histograms and transition matrix are then also large, which is a main contributor to the computational cost of likelihood maximization. Coarse graining the outcome space shrinks the histograms and transition matrix, thereby reducing the complexity. The coarse graining is accomplished by creating a pre-specified number $G < M$ of contiguous bins on the set of possible outcomes. Each bin is defined by its edges in the outcome space. We index the bins, which correspond to new coarse-grained outcomes, by $c = 1, \ldots, G$. Each outcome b is then reassigned to a coarse-grained outcome c based on the bin edges. The bin edges are constructed according to a heuristic that minimizes the information loss between the original outcomes and the coarse grained outcomes (details are in Appendix B). We randomly select without replacement 10 % of the trials from the reference experiments as a training set with which we choose the bin edges. The training set is then excluded from further analysis. (To simplify notation, in the results presented below, we simulated reference experiments with $(10/9)n$ trials so that the number of trials for reference and probing experiments is equal after coarse graining, although the technique can be used with unequal numbers of trials.) The coarse-grained outcomes then determine a new set of histogram matrices $H^{(j)}_{i,c}$, measurement operators F_c, and transition matrix $Q_{k,c}$. For the remainder of the paper we discuss the coarse-grained version of each.

The next step is to use the reference histogram matrix $H^{(0)}$ to derive an initial estimate of the transition matrix Q, which we later use to initialize the likelihood maximization. In the limit of infinite trials $H^{(0)}$ is equal to the probability distribution from Eq. (3) but with a finite number of trials $H^{(0)}_{i,c} \approx \sum_k Q_{k,c}\mathrm{Tr}(\Pi_k \rho_i)$. We can re-express this relation in matrix form, $H^{(0)} \approx PQ$, where Q has elements $Q_{k,c}$, and P has elements $P_{i,k} = \mathrm{Tr}(\Pi_k \rho_i)$. We refer to the elements $P_{i,k}$ as "populations," since they are the probabilities that state i is in subspace k. The population matrix P is known *a priori*, because the states are prepared with the high-fidelity processes. Therefore, we can derive an initial estimate for Q, which describes the bare measurement, by applying the left pseudo-inverse of P,

$$\hat{Q}^{\mathrm{ini}} = \left(P^\top P\right)^{-1} P^\top H^{(0)}. \tag{7}$$

The left pseudo-inverse exists when $\mathrm{rank}(P) = N$, that is, the number of subspace projections. This requires that the states span the subspace that is spanned by the underlying POVM. We only consider situations where this condition is met. The initial estimate of the transition matrix is also used to calculate an initial estimate of the observed measurement operators, $\hat{F}^{\mathrm{ini}}_c = \sum_k \hat{Q}^{\mathrm{ini}}_{k,c}\Pi_k$.

The final step, is to determine estimates of the states and measurement operators that maximize the total likelihood function L. To derive L, we first need the probability of observing a given histogram matrix $H^{(j)}$, which is the product of the probabilities of observing each outcome from Eq. (5) and is given by

$$\mathrm{Prob}(H^{(j)}|\boldsymbol{\tau}, \boldsymbol{F}, \boldsymbol{\mathcal{U}}) = \prod_{i,c} \mathrm{Tr}\left(\mathcal{U}_i^\dagger [F_c] \tau_j\right)^{H^{(j)}_{i,c}}. \tag{8}$$

The probability of obtaining the histogram matrix from all of the experiments given the unknown and the known parameters is then

$$\mathrm{Prob}(\boldsymbol{H}|\boldsymbol{\tau}, \boldsymbol{F}, \boldsymbol{\mathcal{U}}) = \prod_{j,i,c} \mathrm{Tr}\left(\mathcal{U}_i^\dagger [F_c] \tau_j\right)^{H^{(j)}_{i,c}}, \tag{9}$$

where $\boldsymbol{H} = \{H^{(j)}\}_j$. This probability is the total likelihood function L. When we maximize L, we vary the unknown parameters $\boldsymbol{\sigma}$ and $\boldsymbol{F}$, which are the parameters of the statistical model, and keep the known parameters ρ_0, $\boldsymbol{\mathcal{U}}$ and the histogram matrix $\boldsymbol{H}$ fixed. To emphasize the distinction between varying and fixed parameters, we write the likelihood function as $L(\boldsymbol{\sigma}, \boldsymbol{F}|\boldsymbol{H}, \boldsymbol{\mathcal{U}}, \rho_0) =$

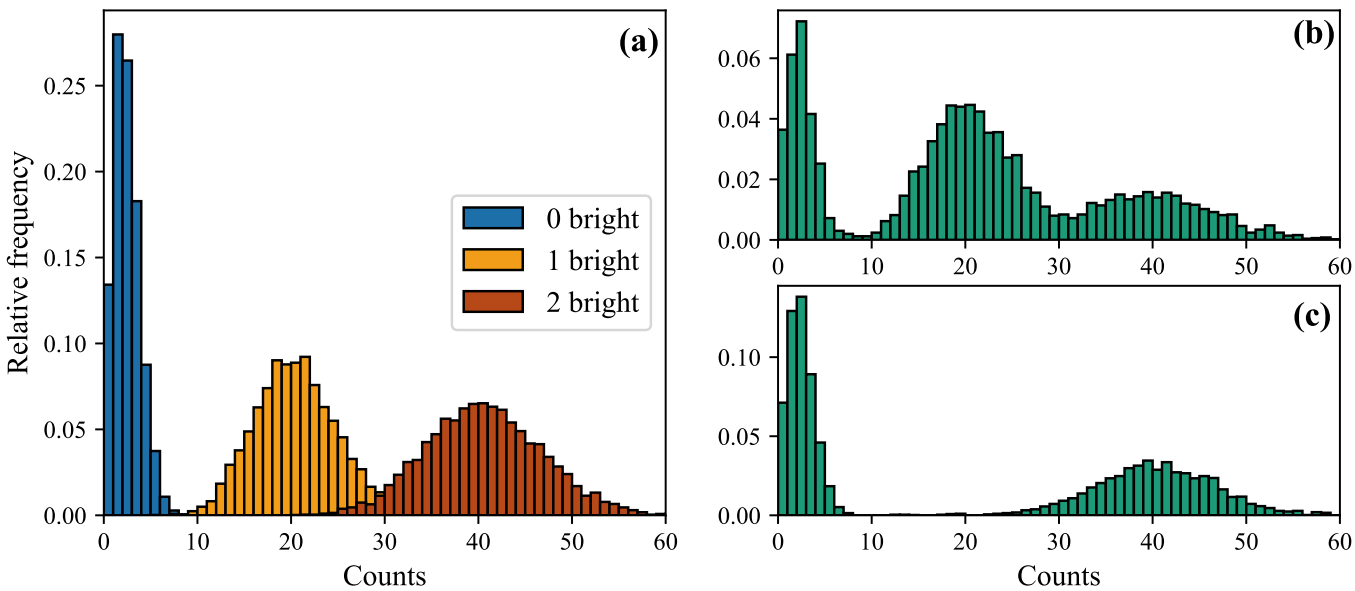

FIG. 3. Histograms from simulated measurements of two ions. An individual ion fluoresces if it is in the $|\uparrow\rangle$ (bright) state. We model the distributions of photon counts as Poissonians. In a real experiment the distributions differ from Poissonians due to processes such as repumping from the dark to the bright state. For two ions, there are three possible count distribution corresponding to the total number of ions in the bright state: zero ions bright (mean 2, due to dark counts), one ion bright (mean 20), and two ions bright (mean 40). For one ion bright, the photon detector cannot determine which of the two ions is fluorescing. (a) Relative frequency histograms for both ions in the dark state (blue), one ion on the dark state (orange), both ions in the bright state (red). (b) Relative frequency histogram for the reference experiment with $U_1 = U(\frac{\pi}{2}, 0)^{\otimes 2}$. (c) Relative frequency histogram for the probing experiment with $U_3 = U(\frac{\pi}{2}, \frac{\pi}{2})^{\otimes 2}$.

$\mathrm{Prob}(\boldsymbol{H}|\boldsymbol{\tau}, \boldsymbol{F}, \boldsymbol{\mathcal{U}})$. The estimates for the states and measurement operators that maximize L also maximize the log-likelihood function $\mathcal{L} = \log(L)$. As is standard practice, we maximize $\mathcal{L}$ instead of L because $\mathcal{L}$ has convenient concavity properties, and it avoids numerical issues with extremely small values. The log-likelihood is

$$\mathcal{L}(\boldsymbol{\sigma}, \boldsymbol{F}|\boldsymbol{H}, \boldsymbol{\mathcal{U}}, \rho_0) = \sum_{i,j,c} H^{(j)}_{i,c} \log \mathrm{Tr}\,(\mathcal{U}_i^\dagger[F_c]\sigma_j). \quad (10)$$

In this case, the log-likelihood is not a concave function, which makes it difficult to determine the global maximum value. However, $\mathcal{L}$ is separately concave in $\boldsymbol{\sigma}$ and $\boldsymbol{F}$. This can be confirmed by computing the log-likelihood's second derivatives with respect to $\boldsymbol{\sigma}$ and $\boldsymbol{F}$ and noting that they are negative semidefinite.

We present an iterative technique to find the maximum of $\mathcal{L}$. Since the log-likelihood is separately concave in $\boldsymbol{\sigma}$ and $\boldsymbol{F}$, optimizing over one while holding the other fixed is a concave optimization problem whose local maxima are global maxima. So to maximize $\mathcal{L}$ over both σ and $\boldsymbol{F}$ jointly, we alternate between two subproblems: (1) concave optimization over $\boldsymbol{\sigma}$ with $\boldsymbol{F}$ fixed, and (2) concave optimization over $\boldsymbol{F}$ with $\boldsymbol{\sigma}$ fixed (details are given below). Here, we describe the subproblem optimizations in the case of a single unknown state, $\boldsymbol{\sigma} = \{\sigma\}$ (we drop the boldface notation for the unknown states since the family has a single member), which requires a single set of probing experiments with histogram matrix $H^{(1)}$.

The first subproblem is maximization of the log-likelihood with respect to the unknown density matrix σ keeping the measurement operators fixed at their current estimate $\hat{\boldsymbol{F}}^{\mathrm{cur}}$. For the initial step, we fix the measurement as $\hat{\boldsymbol{F}}^{\mathrm{cur}} = \hat{\boldsymbol{F}}^{\mathrm{ini}}$. For this subproblem the objective function is

$$\mathcal{L}_1(\sigma|H^{(1)}, \boldsymbol{\mathcal{U}}, \hat{\boldsymbol{F}}^{\mathrm{cur}}) = \sum_{c,i} H^{(1)}_{c,i} \log \mathrm{Tr}\,(\mathcal{U}_i^\dagger[\hat{F}_c^{\mathrm{cur}}]\sigma). \quad (11)$$

Note that $\mathcal{L}_1$ only uses the histogram matrix $H^{(1)}$ from the probing experiments because the reference experiments are independent of σ. We numerically search for the density matrix σ that maximizes $\mathcal{L}_1$ by solving

$$\begin{aligned} \underset{\sigma}{\text{maximize:}} \quad & \mathcal{L}_1(\sigma|H^{(1)}, \boldsymbol{\mathcal{U}}, \hat{\boldsymbol{F}}^{\mathrm{cur}}), \\ \text{subject to:} \quad & \mathrm{Tr}\,\sigma = 1, \\ & \sigma \succeq 0, \end{aligned} \quad (12)$$

where $\sigma \succeq 0$ means that σ is a positive semidefinite matrix. To accomplish this, we use the $R\rho R$ algorithm [21, 25], initialized with $\sigma^{(1)} = \hat{\sigma}^{\mathrm{cur}}$ ($\hat{\sigma}^{\mathrm{cur}} = \mathbb{1}/d$ for the first iteration). At the kth iteration of the algorithm, the state $\sigma^{(k)}$ is updated to $\sigma^{(k+1)} = \mathcal{N}(R(\sigma^{(k)})\sigma^{(k)}R(\sigma^{(k)}))$, where $\mathcal{N}$ indicates normalization, and R is

$$R(\sigma) = \sum_{i,c} \frac{H^{(1)}_{i,c}\hat{F}^{\mathrm{cur}}_{i,c}}{\mathrm{Tr}\,(\sigma\hat{F}^{\mathrm{cur}}_{i,c})}. \quad (13)$$

This ensures that at each iteration the estimate is physical and the likelihood is non-decreasing [21]. The algorithm is run until the stopping condition derived in Ref. [26] is met. For multiple unknown states, $\mathcal{L}_1$ is the sum of the log-likelihoods for each state $\mathcal{L}_1 = \sum_j \mathcal{L}_1(\sigma_j)$, where $\mathcal{L}_1(\sigma_j)$ has the form of Eq. (11). To maximize $\mathcal{L}_1$ in this case, we calculate R for each σ_j and iteratively update each σ_j individually.

The second subproblem is maximization of the log-likelihood with respect to the measurement operators, $\boldsymbol{F}$, with σ fixed at its current estimate $\hat{\sigma}^{\text{cur}}$ returned by the most recent use of the $R\rho R$ algorithm. The objective function for this subproblem is

$$\mathcal{L}_2(\boldsymbol{F}|\boldsymbol{H},\boldsymbol{\mathcal{U}},\hat{\boldsymbol{\tau}}^{\text{cur}}) = \sum_{c,i,j} H_{c,i}^{(j)} \log \text{Tr}\,(\mathcal{U}_i^\dagger[F_c]\hat{\tau}_j^{\text{cur}}), \quad (14)$$

where $\hat{\boldsymbol{\tau}}^{\text{cur}} = (\rho_0, \hat{\sigma}^{\text{cur}})$. Because the measurement operators are constrained according to Eq. (2), we re-write the objective function as

$$\mathcal{L}_2(Q|\boldsymbol{H},\hat{\boldsymbol{P}}^{\text{cur}}) = \sum_{c,i,j} H_{c,i}^{(j)} \log\left(\sum_k Q_{k,c}\hat{P}_{k,i,j}^{\text{cur}}\right), \quad (15)$$

where $\hat{\boldsymbol{P}}^{\text{cur}} = \{\text{Tr}\,(\mathcal{U}_i^\dagger[\Pi_k]\hat{\tau}_j^{\text{cur}})\}_{k,i,j}$. We use a standard nonlinear multi-variable optimizer to solve the following problem:

$$\begin{aligned} \underset{Q}{\text{maximize:}} \quad & \mathcal{L}_2(Q|\boldsymbol{H},\hat{\boldsymbol{P}}^{\text{cur}}), \\ \text{subject to:} \quad & \sum_c Q_{c,k} = 1,\ \forall\, k, \\ & 0 \le Q_{c,k} \le 1,\ \forall\, k, c. \end{aligned} \quad (16)$$

The program is initialized with the current estimate of the transition matrix $\hat{Q}^{\text{cur}}$ ($\hat{Q}^{\text{ini}}$ for the first iteration) and runs until a pre-specified stopping tolerance is reached. All histogram matrices are considered in the optimization, which ensures that the estimated transition matrix $\hat{Q}$ is consistent with both the reference and probing experiments described in the previous section.

The assumption that the measurement operators are linear combinations of an underlying POVM significantly simplifies the program in Eq. (16). Without this assumption, we would require semi-definite constraints to ensure that the measurement is physical. With the assumption, the linear constraints ensure that the estimated transition matrix $\hat{Q}$ is a probability distribution and thus the corresponding observed measurement operators $\hat{\boldsymbol{F}}$ are physical.

We alternate between these two subproblems until a gradient-based stopping criterion is met (see Appendix C), yielding estimates $\hat{\sigma}$ and $\hat{\boldsymbol{F}}$. While there is no guarantee that these estimates correspond to a global maximum [27], in numerical experiments we find the algorithm converges to estimates that are close to the true parameters.

There may be states other than $\hat{\sigma}$ that produce the same maximum value of the likelihood function. This occurs when the high-fidelity processes do not generate an IC set of POVMs, that is, $\{\hat{F}_{i,c}\}_{i,c}$ does not span the space of bounded operators. This is the case in the trapped-ion example given at the end of Sec. III, where the POVM cannot determine which ion is in the bright state. In this case, $\hat{\sigma}$ is an element of the "set of MLE states" where each element of the set produces the same maximum value of the likelihood function. Similarly, there may be transition matrices other than $\hat{Q}$ that yield the same maximum value of the likelihood function. This occurs when the family of known input states do not span the subspace of the underlying POVM. However, as mentioned previously in this section, we do not consider this case because it would cause other complications with our initialization and would complicate the estimation of expectation values described in the next section.

V. ESTIMATION OF EXPECTATION VALUES

To gain information about the unknown states, we can estimate the expectation value $\langle O\rangle = \text{Tr}(O\sigma)$ of any observable O. We can calculate the observable's expectation value according to the point estimate from the likelihood maximization as $\langle\hat{O}\rangle = \text{Tr}\,(O\hat{\sigma})$. However, if the set of POVMs is not IC then it is possible that the set of MLE states contains more than one element, and each element may have a different expectation value with respect to O. This leads to a set of possible expectation values for the unknown state, which are all equally likely. Nevertheless, we can still learn about the unknown states by determining bounds on the expectation value. This is accomplished by searching for the elements in the MLE set that provide the largest and smallest expectation values, which corresponds to the pair of semidefinite programs (SDPs)

$$\begin{aligned} \underset{\rho}{\text{minimize:}} \quad & \pm\,\text{Tr}\,(O\rho), \\ \text{subject to:} \quad & \text{Tr}\,\rho = 1, \\ & \rho \succeq 0, \\ & \text{Tr}\left(\hat{F}_{i,c}(\rho - \hat{\sigma})\right) = 0,\ \text{for all } i, c. \end{aligned} \quad (17)$$

The last constraint ensures that ρ is in the set of MLE states by constraining the expected probability of each outcome to be equal to that of $\hat{\sigma}$. When this is the case, the log-likelihood of any ρ obeying the constraint is equal to the log-likelihood of $\hat{\sigma}$.

As formulated above, the last line of the SDP contains $|\{F_{i,c}\}_{i,c}|$ constraints, many of which may be redundant, thereby causing increased computation time. We reduce the number of constraints by finding an orthogonal basis that spans $\{\hat{F}_{i,c}\}_{i,c}$. This is done by defining a size $|\{F_{i,c}\}_{i,c}| \times d^2$ matrix $\boldsymbol{\mathcal{F}}$, where each row is the measurement operator $\hat{F}_{i,c}$ written as a $1 \times d^2$ vector. We then calculate the singular value decomposition, $\boldsymbol{\mathcal{F}} = \boldsymbol{W}\boldsymbol{S}\boldsymbol{V}^\dagger$. The measurement operators $\{\hat{F}_{i,c}\}_{i,c}$ are

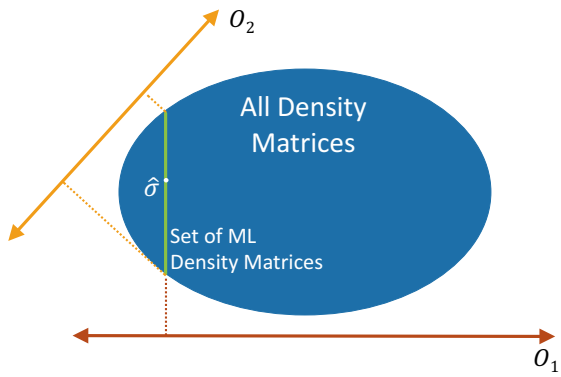

FIG. 4. Conceptual schematic of bounding expectation values. The likelihood maximization returns an estimate $\hat{\sigma}$ that is an element of the set of density matrices that maximize the likelihood (green line). Projecting from the set of maximum likelihood density matrices onto the line representing observable O_1 gives a unique expectation value. Because O_1 is in the span of the estimated measurement operators, each maximum likelihood density matrix has the same expectation value for O_1. Observable O_2 is only partially contained within the span of estimated measurement operators. The SDP provides upper and lower bounds of its expectation value.

in the span of the rows of $\boldsymbol{V}^\dagger$. To reduce the number of constraints in the SDP, we create $\boldsymbol{V}'^\dagger$ by discarding the rows of $\boldsymbol{V}^\dagger$ corresponding to the ℓ smallest singular values, where the value of ℓ is equal to d^2 minus the number of nonzero singular values of the related matrix formed by the underlying measurement operators, $\{\Pi_{i,k}\}_{i,k}$. Now, we replace the last constraint in the above SDP with $\mathrm{Tr}\,[V_k'^\dagger(\rho-\hat{\sigma})]=0$, where $V_k'^\dagger$ is the matrix formed from the kth row of $\boldsymbol{V}'^\dagger$. This reduces the number of these constraints to d^2 or fewer.

Observables can be divided into two types, identifiable and set-identifiable, shown in Fig. 4. First, identifiable observables have the same expectation value for every element in the set of MLE states. In this case the two SDPs return the same expectation value. An identifiable observable O_1 is in the span of the estimated measurement operators, namely there exist coefficients $f_{i,c}$ such that $O_1=\sum_{i,c} f_{i,c}\hat{F}_{i,c}$. The expectation values are then proportional to the estimated outcome probabilities

$$\langle\hat{O}\rangle=\mathrm{Tr}\,(O\hat{\sigma})=\sum_{i,c} f_{i,c}\,\mathrm{Tr}\,(\hat{F}_{i,c}\hat{\sigma}). \tag{18}$$

Since $\mathrm{Tr}\,(\hat{F}_{i,c}\hat{\sigma})$ has an associated empirical estimate given by $H^{(1)}_{c,i}$, we can directly calculate $\langle\hat{O}\rangle$. When $\{F_{i,c}\}_{i,c}$ are IC, they span the space of Hermitian operators, and every observable is identifiable. We do not explicitly check if a decomposition of the form used in Eq. (18) exists apart from running the SDPs.

Set-identifiable observables have a range of expectation values over the MLE states. If $\{F_{i,c}\}_{i,c}$ does not span all Hermitian operators, then there necessarily exist observables that are set-identifiable. The SDPs provide tight upper and lower bounds on the ranges of the expectation values.

Observable	$O_1=\lvert\Phi^+\rangle\langle\Phi^+\rvert$	$O_2=\lvert{\downarrow\uparrow}\rangle\langle{\downarrow\uparrow}\rvert$
True $\langle O\rangle$	0.9925	2.5×10^{-3}
M.L. $\langle O\rangle$	0.9902	2.452×10^{-3}
SDPs	(0.9902, 0.9902)	$(2.361, 2.543)\times10^{-3}$
Basic C.I.	(0.9849, 0.9943)	$(1.296, 3.451)\times10^{-3}$
Bias corrected C.I.	(0.9841, 0.9944)	$(2.522, 4.357)\times10^{-3}$

TABLE I. Numerical results from the two-ion example. The procedure is applied with measurement operators and high-fidelity unitary processes described in Sec. IV and unknown state $\sigma=0.99\lvert\Phi^+\rangle\langle\Phi^+\rvert+\frac{0.01}{4}\mathbb{1}$. "True $\langle O\rangle$" is the true expectation value for the prepared state, "M.L. $\langle O\rangle$" is the estimated expectation value with $\hat{\sigma}$, "SDPs" are the (lower, upper) bounds returned from the semidefinite programs, "Basic C.I." is the 95 % confidence interval found with the basic method and "Bias corrected C.I." is the 95 % confidence interval found with the bias corrected method.

In the two-ion example, the measurements are not informationally complete so there are both identifiable and set-identifiable observables. An example of an identifiable observable is the Bell state, $O_1=\lvert\Phi^+\rangle\langle\Phi^+\rvert$, where $\lvert\Phi^+\rangle=\frac{1}{\sqrt{2}}(\lvert{\uparrow\uparrow}\rangle+\lvert{\downarrow\downarrow}\rangle)$. Since σ is an attempted Bell state preparation, O_1 measures the fidelity. The Bell state can be written in terms of the measurement operators,

$$\begin{aligned}\lvert\Phi^+\rangle\langle\Phi^+\rvert &= \Pi_0+\Pi_2\\ &\quad+\left[U(\tfrac{\pi}{2},0)^{\otimes2}\right]^\dagger\Pi_1U(\tfrac{\pi}{2},0)^{\otimes2}\\ &\quad-\left[U(\tfrac{\pi}{2},\tfrac{\pi}{2})^{\otimes2}\right]^\dagger\Pi_1U(\tfrac{\pi}{2},\tfrac{\pi}{2})^{\otimes2}.\end{aligned} \tag{19}$$

In principle, one could calculate the fidelity by finding the probability of each measurement outcome with an operator in the expansion above. However, the SDP algorithm allows for exact calculation without knowing the expansion.

An example of a set-identifiable observable about which we can gain partial information is the probability that the 2nd ion is in the bright state, $O_2=\lvert{\downarrow\uparrow}\rangle\langle{\downarrow\uparrow}\rvert$. This operator does not have an expansion in terms of the engineered measurements, unlike the Bell state. However, we can still gain partial information about the 2nd ion bright via the SDP. Upper and lower bounds of both example observables are reported in Table I along with the true value of the expectation value and the value with respect to the point estimate.

VI. UNCERTAINTIES

To quantify the uncertainty in our procedure we use parametric bootstrap resampling. For this method, one generates samples from the estimated probability distribution, which simulates new repetitions of the experiments, and uses the results to estimate the distribution of various parameters [28]. For our case, the es-

timated probability distribution is defined by Eq. (5) with $\boldsymbol{\tau} = (\rho_0, \hat{\sigma})$ and $Q = \hat{Q}$, the estimates produced by the MLE algorithm. We numerically sample this distribution n times for each combination of state (index j) and engineered POVM (index i) to create a simulated histogram for each of the experiments described in Sec. II. The bootstrap resampling is performed with coarse-grained measurement outcomes according to the same binning rule so that the bootstrap histogram matrices have the same structure as the original binned experimental histogram matrices. We repeat the numerical technique from Sec. IV and expectation value bounding SDPs from Sec. V on the numerically generated data to produce new estimates for the measurements, states, and upper and lower bounds for $\langle O \rangle$. We repeat this entire bootstrap procedure t times, producing a distribution of estimated measurement operators, states, and bounds on observables.

We use the bootstrap distribution to estimate the uncertainty in our method by calculating approximate 95 % bootstrap confidence intervals of the bounds on observables. For identifiable observables (such as the Bell-state fidelity) the confidence interval can be calculated with standard techniques (see below). However, when reporting uncertainty for expectation values of set-identifiable observables a complication arises. In these cases, the expectation values are lower and upper bounded by the SDPs, but in each bootstrap resample different lower and upper bounds will be estimated. Therefore, there exists uncertainty from the observable being set-identifiable as well as from the bootstrap distribution. To incorporate both uncertainties, we calculate 97.5 % one sided confidence intervals for the lower and upper bounds separately and report the overlapping range, which is a 95 % confidence interval.

There are many methods to determine the bootstrap confidence intervals, but these methods make assumptions that are often not satisfied in our situation. In simulations, we see that the bootstrap distributions for the upper and lower bounds of set-identifiable observables are commonly biased or asymmetric. This is detected by observing a difference between the median of the bootstrap resamples and the estimate from the original data. Bias in maximum likelihood quantum state tomography has also been reported in Ref. [29–31]. These effects are caused by the nonlinearity of the estimator and the complicated structure of the boundary of quantum state space, which necessarily affects our lower and upper bounds on expectation values due to the nature of the SDPs. Certain techniques to construct confidence intervals, such as the percentile method [32], are sensitive to bias and asymmetry, and are therefore contraindicated. Further, the bias and boundary issues imply that the theory underlying other bootstrap confidence interval methods is not applicable. As a result, we expect systematic coverage probability errors that cannot be removed by increasing the number of bootstrap samples.

Nevertheless, for moderate confidence levels between 60 % and 95 %, intervals obtained can still be useful for descriptive purposes. To this end, we report two methods, the basic [33] and bias corrected [32] bootstraps, which both have some robustness to bias and asymmetry, though neither is designed for high dimensional estimates with boundary constraints. These methods return sometimes very different confidence intervals (c.f. in Table I for O_2). Therefore, we stress that both should be taken only as qualitative descriptions of the uncertainty and not used for further inference.

More sophisticated methods to deal with bias and asymmetry exist (e.g. the bias corrected and accelerated method in Ref. [32]). However, such methods are not designed to address the underlying problems encountered. Another possibility may be to use methods that involve a double bootstrap, but they are impractical for our procedure. Further research is required to obtain high-quality bootstrap confidence intervals for quantum tomography. Alternatively, one might adapt the confidence regions described in Ref. [34, 35], which are not based on the bootstrap.

The bootstrap procedure also allows us to perform a likelihood ratio test to determine how well our model fits the observed data. The "likelihood ratio" is the ratio of the likelihood of our null model, the estimates of $\hat{\sigma}$ and $\hat{\boldsymbol{F}}$ from the experimental data (with log-likelihood given by the iterative program discussed in Sec. IV), to the likelihood of an alternative model, the experimental frequency histogram matrix with log-likelihood function,

$$\mathcal{L}_{\mathrm{frq}}(\boldsymbol{H}) = \sum_{i,j,c} H_{i,c}^{(j)} \log H_{i,c}^{(j)}. \tag{20}$$

To perform the likelihood ratio test we compare the likelihood ratio of the original estimates to the distribution of likelihood ratios created by the bootstrap procedure. The likelihood ratio test statistic is

$$\Lambda_0 = \mathcal{L}(\hat{\sigma}, \hat{\boldsymbol{F}} | \boldsymbol{H}, \boldsymbol{\mathcal{U}}, \rho_0) - \mathcal{L}_{\mathrm{frq}}(\boldsymbol{H}). \tag{21}$$

We also compute an analogous statistic for each of the t bootstrapped data sets (each computed using its own simulated histogram matrices and respective state and measurement estimate) to generate a distribution of likelihood test statistics, $\{\Lambda_i | i = 1, \ldots, t\}$. Because the bootstrap data sets are certainly well described by the null model, their likelihood ratios should typically be comparable to the likelihood ratio of the experimental data set, which may or may not obey the model. A statistically significant difference between the experimental data set's likelihood ratio and the bootstrap data sets' is evidence that the null model does not match the experiment. To quantify evidence against the null model, we compute an empirical p-value by determining the percentile at which Λ_0 falls in the distribution from the bootstrap data sets [28].

VII. CONCLUSION

We have developed a procedure to simultaneously characterize unknown quantum measurements and states by using a limited set of high-fidelity quantum operations. The protocol requires two types of experiments, which we designated as "reference" and "probing." The reference experiments use the high-fidelity processes to estimate the unknown measurement operators, similar to QDT. The probing experiments use the high-fidelity processes to probe unknown state preparations, similar to QST. In our procedure the estimation of the measurement operators and density matrices is achieved simultaneously via alternating MLE. This means the estimates produced are consistent with both reference and probe experiments. We also introduced a method for estimating expectation values of the unknown states by two SDPs when the high-fidelity processes do not produce IC measurements.

Our procedure applies to systems where we can apply certain operations with high fidelity. These need not be universal, a small set of single-qubit unitaries suffices. A sufficient set of such operations is available in many state-of-the-art quantum information processors, for example trapped ions. Moreover, our protocol has the advantage that it is efficient to implement relative to previous proposals. This is because we make use of prior information about the measurement operators and do not seek to diagnose all parts of the quantum system. However, since our protocol is dependent on prior information about the system, it is not ideal for experiments that have not previously been diagnosed. In the absence of a well defined initial state and sufficiently many high-fidelity operations, methods such as GST may be better suited.

Our implementation and testing of this estimation procedure has focused on a few trapped ions. In this system, measurements can be modeled as classical noise following projection onto a small number of orthogonal subspaces and measurement outcomes that can be partitioned into a small number of bins with little information loss. In principle, the procedure can handle higher dimensional systems and measurements that consist of multi-dimensional histograms but further optimization is required to run these cases efficiently. There are also opportunities for expanding our procedure by applying it to systems with measurements that cannot be modeled as orthogonal projection followed by classical noise.

Appendix A: Software implementation

We implemented our procedure as a Python package, available at Ref. [36]. Instructions for installing and running the package are given in the accompanying documentation. Our package contains a version of $R\rho R$ for the first optimization subproblem, and uses `scipy.optimize` for the second optimization subproblem. The SDPs that estimate the expectation values described in Sec. V require the MATLAB API engine. The SDPs are solved with YALMIP [37], which is a package for MATLAB.

Appendix B: Coarse-graining measurement outcomes

The measurement device may have a very large number of possible outcomes $b = 1, \ldots, M$ where $M \gg 1$. This is a main contributor to the computational complexity of the log-likelihood maximization, because each outcome adds a term to the log-likelihood function, which requires more computation for finding R in the first subproblem and increases the optimization space for the second subproblem. To reduce the complexity, we coarse grain the outcome space by collecting the original outcomes b into $G < M$ bins. (The original outcomes can also be thought of as binned, so this procedure is technically a re-binning.) We assume that the ordering of the original outcomes $1, \ldots, M$ is meaningful so that it makes sense to bin consecutive outcomes for minimum information loss. The bins are then identified by the bin edges B_c in the outcome space $\{B_c | c = 0, \ldots, G\}$ where $B_0 = 0$, and $B_G = M$, so an outcome $b \in (B_{c-1}, B_c]$ is mapped to the coarse-grained outcome c. Coarse graining reduces the information about the unknown states and measurements that was captured by our experiments. For an extreme example choose $G = 2$ with $B_1 = 0$, then every outcome b is mapped to the coarse-grained outcome $c = 1$, which provides no information. To combat this problem, we construct a heuristic algorithm to chose the bin edges. We apply this algorithm to "training data," which is composed of 10 % of the trials from the reference experiments randomly sampled without replacement and set aside from all further analysis. Though we describe the algorithm in terms of one-dimensional histograms here, our software also supports multidimensional histograms.

For a given number G of coarse-grained bins, our algorithm maximizes the amount of information retained in the coarse-grained outcomes c by a heuristic based on mutual information. First, we need to identify the information about the unknown states and measurement. In our procedure, this information is contained in the histograms collected from the experiments. These histogram matrices $H^{(j)}_{i,b}$ are rectangular with vertical dimension M, the number of possible outcomes. Coarse graining the outcomes then shrinks the vertical dimension to $G < M$. For bin edges $\{B_c\}_c$ the coarse grained histogram matrix is defined as

$$H'^{(j)}_{i,c} = \sum_{b=B_{c-1}+1}^{B_c} H^{(j)}_{i,b}. \tag{B1}$$

Note that every histogram (that is, every row of the histogram matrix) is treated identically. The coarse-

graining is also applied uniformly to the transition matrix

$$Q'_{k,c} = \sum_{b=B_{c-1}+1}^{B_c} Q_{k,b}. \tag{B2}$$

Therefore, both $H'^{(j)}$ and Q' have smaller vertical dimension due to the coarse graining. In the following, we study the coarse-grained histogram matrices and transfer matrix estimated from the training data $H^{\text{train}} = H'^{(0)}$ and $\hat{Q}^{\text{train}} = Q'$ from Eq. (7).

We quantify the amount of information retained in coarse-grained histograms with the mutual information. Consider a joint probability distribution $P(k, c)$ where c are the bin indices and k are underlying outcomes. Let C and K denote the random variables with values c and k, respectively. The mutual information $I_P(K; C)$ between K and C quantifies the amount of information C has about K (or K about C). It is given by

$$I(K; C) = \sum_{k,c} P(c)P(k|c)\log_2\left(\frac{P(k|c)}{P(k)}\right), \tag{B3}$$

where $P(k|c) = P(k, c)/P(k)$ is the conditional probability distribution. Our goal for coarse graining is to maximize the mutual information, so as to approach the full information in K, achieved when $C = K$. We could calculate the mutual information between the original and the coarse-grained histograms, but what we really care about is how well the coarse-graining retains information about the underlying outcome distribution $p_k = \text{Tr}\,(\Pi_k\rho)$ discussed in Sec. II. This distribution is dependent on the state, so we choose a representative state $\rho = \rho^{\text{eq}}$ with the property that underlying outcomes have equal probability, namely $P(k) = \text{Tr}\,(\Pi_k\rho^{\text{eq}}) = 1/N$ for all k. We could estimate the coarse-grained histogram matrix H^{eq} for ρ^{eq} based on the training data's histogram, but this relation is dependent on the family of known input states $\boldsymbol{\rho}$. Instead we find it more convenient to determine H^{eq} based on the estimated transfer matrix Q^{train}, which already contains information about the histogram and the known states, such that $H^{\text{eq}}_c = \sum_k \hat{Q}^{\text{train}}_{k,c}\frac{1}{N} = P(c)$. The needed joint probability distribution is given by $P(k, c) = P(c|k)P(k) = P(c|k)/N$ with

We determine a good coarse graining by finding bin edges $\{B_c\}_c$ with high mutual information for $P(k, c)$ as defined in the previous paragraph. Searching over all possible bin edges to maximize $I(K; C)$ is impractical, so instead we use a heuristic algorithm that iteratively adds bin edges. We pre-specify the target number of bins G. Then, starting with two bins, we compute the mutual information for each possible location for the bin edge between 0 and M. The boundary location that gives the largest mutual information is then fixed as an element of what will be our final list of bin edges. To add a third bin, leaving the existing bin edge fixed, we compute the mutual information for all possible locations for the new edge. The new edge that gives the largest mutual information is added to $\{B_c\}_c$. We continue this procedure until the target number of G bins is reached. We apply the resulting bin edge rule to the remaining reference histograms not used in the training data, all probing histograms, and $\hat{Q}^{\text{ini}}$ by Eq. (B1) and (B2). An example of binning ion fluorescence data is in Fig. 5.

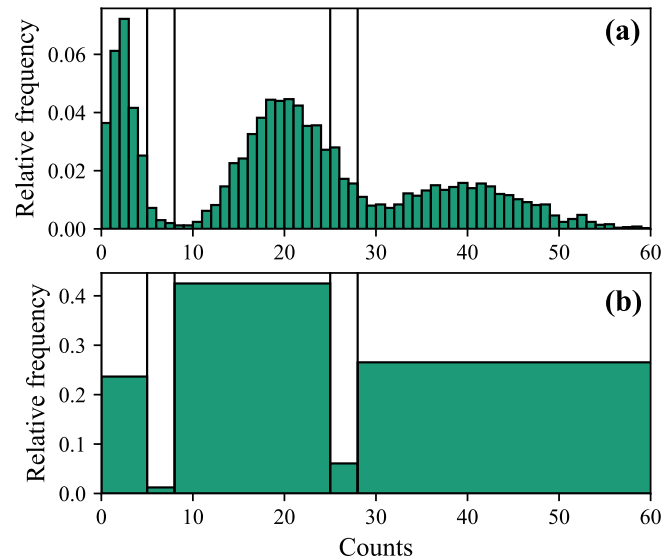

FIG. 5. Illustration of the binning procedure. (a) Relative frequency histogram for the second reference experiment, $U_1 = U(\frac{\pi}{2}, 0)^{\otimes 2}$ (b) Same histogram data but after binning. The black lines show the bin boundaries found with the heuristic algorithm.

We have observed that binning consecutive elements in the rows of $H^{(j)}$ has been effective when systems occupying a single subspace produce unimodal distributions, such as those produced by the ion measurements. Analysis of multimodal distributions may benefit from more complicated binning strategies.

Appendix C: Stopping criteria

After each optimization subproblem is run in an iteration, we bound the difference between the current log-likelihoods $\mathcal{L}_i$ ($i = 1$ and $i = 2$ for the first and second subproblems) and their respective maximum possible values. The bounds tell us how much the log-likelihoods could increase with further iterations. We stop the algorithm when both differences are individually below pre-specified thresholds T_σ and T_Q.

To find the difference bound for the first subproblem, we follow the method introduced in Ref. [26]. The bound $\hat{S}_\sigma$ is given by,

$$\mathcal{L}_1(\sigma_{\text{ML}}) - \mathcal{L}_1(\hat{\sigma}^{\text{cur}}) \leq \max\{\text{eig}[R(\hat{\sigma}^{\text{cur}})]\} - n = \hat{S}_\sigma, \tag{C1}$$

where σ_{ML} is the maximum likelihood state, and $\max\{\text{eig}[R(\hat{\sigma}^{\text{cur}})]\}$ is the maximum eigenvalue of $R(\hat{\sigma}^{\text{cur}})$, which was computed by the $R\rho R$ algorithm. We stop it-

erations of $R\rho R$ when $\hat{S}_\sigma \leq T_\sigma$.

The difference bound for the second subproblem is calculated in a similar way. Since the log-likelihood function of this subproblem is also concave, the difference between the maximum log-likelihood and the log-likelihood of $\hat{Q}^{\rm cur}$ is upper bounded by,

$$\mathcal{L}_2(Q_{\rm ML}) - \mathcal{L}_2(\hat{Q}^{\rm cur}) \leq \sum_{k,c} (\hat{Q}^{\rm cur}_{k,c} - Q^{\rm ML}_{k,c}) \frac{\partial\, \mathcal{L}_2(\hat{Q}^{\rm cur})}{\partial Q_{k,c}}, \tag{C2}$$

where $\frac{\partial\, \mathcal{L}_2(\hat{Q}^{\rm cur})}{\partial Q_{k,c}}$ are the elements of the gradient of $\mathcal{L}_2$. Since we do not know $Q_{\rm ML}$, we bound the right side of Eq. (C2) by finding the maximum value over all distributions by solving the optimization problem

$$\begin{aligned} \underset{X}{\text{maximize:}} \quad & \sum_{k,c} (\hat{Q}^{\rm cur}_{k,c} - X_{k,c}) \frac{\partial\, \mathcal{L}_2(\hat{Q}^{\rm cur})}{\partial \hat{Q}^{\rm cur}_{k,c}}, \\ \text{subject to:} \quad & \sum_c X_{k,c} = 1,\ \forall k, \\ & 0 \leq X_{k,c} \leq 1,\ \forall\, k, c. \end{aligned} \tag{C3}$$

The value $\hat{S}_Q$ returned is an upper bound on the difference between the maximum log-likelihood and the log-likelihood at the current iteration.

We compare both bounds to the pre-specified thresholds T_σ and T_Q. Default values for these thresholds in the code are currently $T_\sigma = 0.3$ and $T_Q = 0.25$, which were chosen empirically by testing on simulated data representative of our ion-trap applications. When $\hat{S}_\sigma \leq T_\sigma$ and $\hat{S}_Q \leq T_Q$ the procedure is stopped, and the final estimates are returned.

ACKNOWLEDGMENTS

This work includes contributions of the National Institute of Standards and Technology, which are not subject to U.S. copyright. The use of trade names allows the calculations to be appropriately interpreted and does not imply endorsement by the US government, nor does it imply these are necessarily the best available for the purpose used here.

[1] J. P. Gaebler, T. R. Tan, Y. Lin, Y. Wan, R. Bowler, A. C. Keith, S. Glancy, K. Coakley, E. Knill, D. Leibfried, and D. J. Wineland, Phys. Rev. Lett. **117**, 060505 (2016), arXiv:1604.00032 [quant-ph].

[2] J. M. Nichol, L. A. Orona, S. P. Harvey, S. Fallahi, G. C. Gardner, M. J. Manfra, and A. Yacoby, npj Quantum Information **3**, 3 (2017), arXiv:1608.04258 [cond-mat.mes-hall].

[3] S. Sheldon, E. Magesan, J. M. Chow, and J. M. Gambetta, Phys. Rev. A **93**, 060302 (2016), arXiv:1603.04821 [quant-ph].

[4] L. Casparis, T. W. Larsen, M. S. Olsen, F. Kuemmeth, P. Krogstrup, J. Nygård, C. M. K. D. Petersson, and Marcus, Phys. Rev. Lett. **116**, 150505 (2016), arXiv:1512.09195 [cond-mat.mes-hall].

[5] R. Barends, J. Kelly, A. Megrant, A. Veitia, D. Sank, E. Jeffrey, T. C. White, J. Mutus, A. G. Fowler, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, C. Neill, P. O'Malley, P. Roushan, A. Vainsencher, J. Wenner, A. N. Korotkov, A. N. Cleland, and J. M. Martinis, Nature **508**, 500 (2014), arXiv:1402.4848 [quant-ph].

[6] C. J. Ballance, T. P. Harty, N. M. Linke, M. A. Sepiol, and D. M. Lucas, Phys. Rev. Lett. **117**, 060504 (2016), arXiv:1512.04600 [quant-ph].

[7] E. Prugovečki, Int. J. Theor. Phys. **16**, 321 (1977).

[8] C. A. Fuchs and R. Schack, in *Quantum State Estimation*, Lecture Notes in Physics, Vol. 649, edited by M. G. A. Paris and J. Řeháček (Springer, Berlin, 2004) Chap. 5, p. 164.

[9] A. Luis and L. L. Sánchez-Soto, Phys. Rev. Lett. **83**, 3573 (1999).

[10] J. Fiurášek, Phys. Rev. A **64**, 024102 (2001), arXiv:quant-ph/0101027.

[11] Y. Lin, J. P. Gaebler, F. Reiter, T. R. Tan, R. Bowler, Y. Wan, A. Keith, E. Knill, S. Glancy, K. Coakley, A. S. Sørensen, D. Leibfried, and D. J. Wineland, Phys. Rev. Lett. **117**, 140502 (2016), arXiv:1603.03848 [quant-ph].

[12] D. Mogilevtsev, J. Řeháček, and Z. Hradil, New J. Phys. **14**, 095001 (2012).

[13] D. Mogilevtsev, A. Ignatenko, A. Maloshtan, B. Stoklasa, J. Řeháček, and Z. Hradil, New J. Phys. **15**, 025038 (2013).

[14] S. T. Merkel, J. M. Gambetta, J. A. Smolin, S. Poletto, A. D. Córcoles, B. R. Johnson, C. A. Ryan, and M. Steffen, Phys. Rev. A **87**, 062119 (2013), arXiv:1211.0322 [quant-ph].

[15] C. Stark, Phys. Rev. A **89**, 052109 (2014), arXiv:1209.5737 [quant-ph].

[16] R. Blume-Kohout, J. K. Gamble, E. Nielsen, J. Mizrahi, J. D. Sterk, and P. Maunz, "Robust, self-consistent, closed-form tomography of quantum logic gates on a trapped ion qubit," (2013), arXiv:1310.4492 [quant-ph].

[17] J. Medford, J. Beil, J. M. Taylor, S. D. Bartlett, A. C. Doherty, E. I. Rashba, D. P. DiVincenzo, H. Lu, A. C. Gossard, and C. M. Marcus, Nature Nanotechnology **8**, 654 (2013), arXiv:1302.1933 [cond-mat.mes-hall].

[18] J. P. Dehollain, J. T. Muhonen, R. Blume-Kohout, K. M. Rudinger, J. K. Gamble, E. Nielsen, A. Laucht, S. Simmons, R. Kalra, A. S. Dzurak, and A. Morello, New Journal of Physics **18**, 103018 (2016), arXiv:1606.02856 [cond-mat.mes-hall].

[19] R. Blume-Kohout, J. K. Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier, and P. Maunz, Nat. Commun. **8** (2017), arXiv:1605.07674 [quant-ph].

[20] Z. Hradil, J. Řeháček, J. Fiurášek, and M. Ježek, in

Quantum State Estimation, Lecture Notes in Physics, Vol. 649, edited by M. G. A. Paris and J. Řeháček (Springer, Berlin, 2004) Chap. 3, pp. 59–112.

[21] J. Řeháček, Z. Hradil, E. Knill, and A. I. Lvovsky, Phys. Rev. A **75**, 042108 (2007), arXiv:quant-ph/0611244.

[22] W. M. Itano, J. C. Bergquist, J. J. Bollinger, J. M. Gilligan, D. J. Heinzen, F. L. Moore, M. G. Raizen, and D. J. Wineland, Phys. Rev. A **47**, 3554 (1993).

[23] M. Acton, K.-A. Brickman, P. C. Haljan, P. J. Lee, L. Deslauriers, and C. Monroe, Quantum Info. Comput. **6**, 465 (2006), arXiv:quant-ph/0511257.

[24] J. Shao, *Mathematical Statistics*, Springer Texts in Statistics (Springer, New York, 1998) Chap. 4.

[25] D. S. Gonçalves, M. A. Gomes-Ruggiero, and C. Lavor, Quantum Info. Comput. **14**, 966 (2014), arXiv:1306.3057 [math-ph].

[26] S. Glancy, E. Knill, and M. Girard, New J. Phys. **14**, 095017 (2012), arXiv:1205.4043 [quant-ph].

[27] J. Gorski, F. Pfeuffer, and K. Klamroth, Math. Method. Oper. Res. **66**, 373 (2007).

[28] D. D. Boos, Statistical Science **18**, 168 (2003).

[29] T. Sugiyama, P. S. Turner, and M. Murao, New J. Phys. **14**, 085005 (2012), arXiv:1205.2976 [quant-ph].

[30] C. Schwemmer, L. Knips, D. Richart, H. Weinfurter, T. Moroder, M. Kleinmann, and O. Gühne, Phys. Rev. Lett. **114**, 080403 (2015), arXiv:1310.8465 [quant-ph].

[31] G. B. Silva, S. Glancy, and H. M. Vasconcelos, Phys. Rev. A **95**, 022107 (2017), arXiv:1604.00321 [quant-ph].

[32] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, Monographs on Statistical and Applied Probability (Chapman and Hall, New York, 1993).

[33] A. C. Davison and D. V. Hinkley, *Bootstrap Methods and their Application* (Cambridge University Press, New York, NY, 1997).

[34] M. Christandl and R. Renner, Phys. Rev. Lett. **109**, 120403 (2012), arXiv:1108.5329v1 [quant-ph].

[35] R. Blume-Kohout, "Robust error bars for quantum tomography," (2012), arXiv:1202.5270 [quant-ph].

[36] "Joint quantum state and measurement tomography," `https://github.com/usnistgov/state_meas_tomo`.

[37] J. Löfberg, in *In Proceedings of the CACSD Conference* (Taipei, Taiwan, 2004).