Optimal experiment design for quantum state tomography: Fair, precise, and minimal tomography

J. Nunn,[*] B. J. Smith, G. Puentes, and I. A. Walmsley

Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, United Kingdom

J. S. Lundeen

National Research Council of Canada, Ottawa, Ontario K1A 0R6, Canada

Given an experimental setup and a fixed number of measurements, how should one take data to optimally reconstruct the state of a quantum system? The problem of optimal experiment design (OED) for quantum state tomography was first broached by Kosut *et al.* [R. Kosut, I. Walmsley, and H. Rabitz, e-print arXiv:quant-ph/0411093 (2004)]. Here we provide efficient numerical algorithms for finding the optimal design, and analytic results for the case of 'minimal tomography'. We also introduce the *average OED*, which is independent of the state to be reconstructed, and the *optimal design for tomography* (ODT), which minimizes tomographic bias. Monte Carlo simulations confirm the utility of our results for qubits. Finally, we adapt our approach to deal with constrained techniques such as maximum-likelihood estimation. We find that these are less amenable to optimization than cruder reconstruction methods, such as linear inversion.

I. INTRODUCTION

Quantum state tomography [1–5] is an indispensable tool in quantum information processing, being essential for the characterization of quantum sources [6–8], gates [9,10], processes [11–14], and measurements [15,16]. The reconstruction of quantum states is, however, essentially a classical problem—that of estimating the parameters of a density matrix from a data set. The classical theory of multiple parameter estimation is well developed [17–27], and so a considerable amount is known about the precision that can be achieved with quantum tomography [1,28–31].

In this paper we are concerned with designing an experiment so as to optimize this precision. Choosing the right set of measurements is of paramount importance, and the optimal measurements for tomography are now known [32–38]. But it is not always possible to implement them: More often than not, technical constraints permit only a nonideal set of measurements [39]. Given such a set, and a finite time in which to acquire data, one encounters the question "How much time should be spent on each measurement, so as to perform the best tomographic inversion?". This is the problem of optimal experiment design (OED). We will see that a judicious design can significantly improve the performance of tomographic reconstruction. Rather paradoxically, however, the OED generally *depends* on the state we wish to reconstruct, so that one cannot find the OED if one is completely ignorant of the quantum state. In this paper we introduce two alternative approaches to experiment design that do not suffer from this state dependence.

The paper is structured as follows. In Sec. II we review the theory of multiparameter estimation, and we introduce the notation to be used subsequently. In Sec. III we will show how to find the OED quickly using standard numerical techniques. We then move on in Sec. IV to derive an analytic formula for the OED, which holds in the case that the quantum state

is not overdetermined by the available data. Sections V and VI address the problem of state tomography when no prior knowledge of the true quantum state exists; that is, when one really has no idea at all what quantum state we expect to find. On the one hand, we may decide that we should design our experiment so that, on average, whatever the true state, the tomography is as precise as possible. This we call the average OED. On the other hand, we might want our tomography to be as "fair" as possible, so that the precision is as close as possible to being independent of the true state. We call this the optimal design for tomography (ODT). We find that these two optimizations are different, so that a choice should be made about whether fairness or precision is paramount in designing tomographic experiments. In Sec. VII we present the results of Monte Carlo simulations to corroborate our predictions. Finally in Sec. VIII we consider adapting our results to the case of constrained estimators, such as maximum-likelihood estimation: We conclude that our optimizations still apply for such estimators, even though the improvement is less marked than for unconstrained tomography. Section IX wraps up the paper with some concluding remarks.

II. BACKGROUND

A. Bloch representation

We suppose that we are given an ensemble of identical N-dimensional quantum systems, all of which are prepared in the same way. Quantum state tomography amounts to estimating $2N^2$ real numbers comprising the elements of ρ, the complex-valued $N \times N$ density matrix describing the ensemble. But not all of these numbers are independent of one another because ρ must always be Hermitian with unit trace and positive eigenvalues [2]. The first two conditions reduce the number of independent real parameters in the density matrix down to $N^2 - 1$. An equivalent description of the quantum system is then given by the *Bloch vector* $\boldsymbol{r}$, which is a real column vector whose $N^2 - 1$ elements uniquely

─────────
[*]j.nunn1@physics.ox.ac.uk

determine ρ, according to the relation

$$\rho = \frac{1}{N} I + r\sigma, \qquad (1)$$

where I is the $N \times N$ identity matrix. Here σ is an $N^2 - 1$-dimensional vector whose elements are $N \times N$ matrices that form an orthonormal basis for the space of traceless Hermitian operators [4,40]. That is to say,

$$\text{tr}\{\sigma_j\} = 0, \quad \text{and} \quad \text{tr}\{\sigma_j \sigma_k\} = \delta_{jk}. \qquad (2)$$

Any set of matrices satisfying these conditions will suffice for constructing σ. For $N = 2$, the three elements of σ are conventionally taken to be the familiar Pauli matrices, and then r is the standard three-dimensional vector that describes the state of a qubit in the Bloch sphere [2]. The vector r is known as the *Bloch* or *Fano* representation of ρ; its usefulness becomes clear when we consider the way it enters into the calculation of measurement statistics.

We will perform our tomographic experiment by sending multiple copies of the state ρ into our apparatus. The apparatus can be configured in one of M different ways, so that there are M different possible measurements we can make (for instance, we could measure the three Cartesian components of a spin, and $M = 3$ in this case). Let γ denote the particular measurement we are making. For each γ, there are n_γ different possible measurement outcomes (in our spin example, $n_\gamma = 2S + 1$ for all the measurements, where S is the total spin quantum number). Associated to each measurement outcome α is a so-called *positive operator-valued measure element* (POVM element) [2], which is a Hermitian, positive $N \times N$ matrix $\Pi_{\alpha\gamma}$, such that the probability $p_{\alpha\gamma}$ of obtaining the outcome α in configuration γ is given by[1]

$$p_{\alpha\gamma} = \text{tr}\{\Pi_{\alpha\gamma}\rho\}. \qquad (3)$$

The POVM elements contain the physics of our tomography setup. All noise and detector inefficiency can be incorporated into them, so that once they are fixed, the only issues we must deal with are statistical. Like the density matrix, they are Hermitian, and can also be written in the Bloch representation

$$\Pi_{\alpha\gamma} = c_{\alpha\gamma} I + a_{\alpha\gamma} \sigma, \qquad (4)$$

where $c_{\alpha\gamma}$ is a real number, and $a_{\alpha\gamma}$ is a real $(N^2 - 1)$-dimensional vector. For each measurement configuration γ, the probabilities $p_{\alpha\gamma}$ must sum to unity, and this imposes the sum constraint $\sum_{\alpha=1}^{n_\gamma} \Pi_{\alpha\gamma} = I$. In the Bloch representation, we must have, accordingly,

$$\sum_{\alpha=1}^{n_\gamma} c_{\alpha\gamma} = 1, \quad \text{and} \quad \sum_{\alpha=1}^{n_\gamma} a_{\alpha\gamma} = 0, \qquad (5)$$

for all values of γ. Now, substituting Eq. (4) into Eq. (3) we obtain the simple affine relation

$$p_{\alpha\gamma} = c_{\alpha\gamma} + a_{\alpha\gamma} r, \qquad (6)$$

which connects the measurement statistics with the Bloch vector. The probabilities can be collected into an n_{tot}-dimensional vector p, where

$$n_{\text{tot}} = \sum_{\gamma=1}^{M} n_\gamma, \qquad (7)$$

is the total number of outcomes our experiment can produce, across all measurement configurations. Then the statistics are summarized by the matrix equation

$$p = c + Ar, \qquad (8)$$

where c is a real vector whose elements are the $c_{\alpha\gamma}$, and where A is the $n_{\text{tot}} \times (N^2 - 1)$ matrix whose rows are given by the vectorized POVM elements

$$A = \begin{pmatrix} \boxed{a_{11}^\top} \\ \boxed{a_{21}^\top} \\ \vdots \\ \boxed{a_{\alpha\gamma}^\top} \\ \vdots \\ \boxed{a_{n_M M}^\top} \end{pmatrix}. \qquad (9)$$

This notation uses no more matrix elements than necessary, unlike using ρ and $\Pi_{\alpha\gamma}$ in (3). As is often the case, the notational compactness of matrices and vectors will prove invaluable in teasing out the structure of the optimizations that follow.

B. Cramer-Rao bound

Suppose that we use a total of N_{tot} copies of our quantum system (note that N_{tot} is different to n_{tot}, the latter being the total number of possible outcomes). We measure N_γ times in each of the M configurations, so that we have

$$N_{\text{tot}} = \sum_{\gamma=1}^{M} N_\gamma. \qquad (10)$$

At the end of the experiment, we are left with an N_{tot}-dimensional data vector n whose elements $n_{\alpha\gamma}$ are the number of times the outcome α was observed in configuration γ, so that $\sum_{\alpha=1}^{n_\gamma} n_{\alpha\gamma} = N_\gamma$. We must process n somehow to produce an estimate of the true state r. The probability of obtaining the data vector n is given (up to an unimportant combinatorial factor) by the likelihood function p,

$$p(n|r) = \prod_{\alpha\gamma} p_{\alpha\gamma}^{n_{\alpha\gamma}}. \qquad (11)$$

We will see in the following that the sensitivity of this likelihood function to r, through Eq. (6), determines the precision of our tomographic inversion. The OED is the set of numbers $\{N_\gamma\}$ which makes our estimate of r as precise as possible.

_____

[1]Much of our notation is taken from the paper by Kosut *et al.* [1], which first introduced the idea of OED for quantum tomography.

Our approach to finding the OED follows that presented in [1], in which we seek to minimize the *Cramer-Rao bound* [17–19]. This is a classical information-theoretic lower bound on the mean-squared error in the estimate of $\boldsymbol{r}$, which does not depend on the algorithm used to extract that estimate, as long as the estimate is *unbiased* (we will discuss certain biased estimators in Sec. VIII). The Cramer-Rao bound is defined in terms of the *Fisher information matrix*[2]

$$F = \frac{1}{N_{\text{tot}}} \langle \boldsymbol{\Delta} \boldsymbol{\Delta}^{\mathsf{T}} \rangle, \tag{12}$$

where $\langle \cdot \rangle$ indicates an ensemble average and the symbol T denotes the transpose. The vector $\boldsymbol{\Delta} = \nabla_{\boldsymbol{r}} \mathcal{L}$ is the gradient of the *log-likelihood function*

$$\mathcal{L} = \ln p\,(\boldsymbol{n}|\boldsymbol{r}) = \sum_{\alpha\gamma} n_{\alpha\gamma} \ln p_{\alpha\gamma}. \tag{13}$$

One might expect that "informative" measurements are those associated with a greater sensitivity of p to $\boldsymbol{r}$, and indeed with these definitions, a strong $\boldsymbol{r}$ dependence of p contributes to the magnitude of the Fisher information F. Substituting Eq. (6) into Eq. (13), we obtain $\boldsymbol{\Delta} = \sum_{\alpha\gamma} n_{\alpha\gamma} \boldsymbol{a}_{\alpha\gamma} / p_{\alpha\gamma}$, so that

$$F = \frac{1}{N_{\text{tot}}} \sum_{\alpha\gamma} \sum_{\beta\delta} \frac{\langle n_{\alpha\gamma} n_{\beta\delta} \rangle}{p_{\alpha\gamma} p_{\beta\delta}} \boldsymbol{a}_{\alpha\gamma} \boldsymbol{a}_{\beta\delta}^{\mathsf{T}}. \tag{14}$$

The statistics for each measurement are found from independent experiments, so the cross correlation $\langle n_{\alpha\gamma} n_{\beta\delta} \rangle$ factorizes with respect to the different measurement configurations; within the same configuration we can use the standard result for multinomial distributions, so we have

$$\langle n_{\alpha\gamma} n_{\beta\delta} \rangle = \delta_{\gamma\delta} [N_{\gamma}(N_{\gamma} - 1) p_{\alpha\gamma} p_{\beta\delta} + N_{\gamma} \delta_{\alpha\beta} p_{\alpha\gamma}]. \tag{15}$$

Only the last term contributes (the other terms vanish due to the second condition in Eq. (5) [4]). Finally, we are left with the expression

$$F = \frac{1}{N_{\text{tot}}} \sum_{\alpha\gamma} \frac{N_{\gamma}}{p_{\alpha\gamma}} \boldsymbol{a}_{\alpha\gamma} \boldsymbol{a}_{\alpha\gamma}^{\mathsf{T}}. \tag{16}$$

The inverse of this matrix is a lower bound on the covariance matrix of our reconstructed state, known as the Cramer-Rao bound. That is, if we estimate the state to be $\hat{\boldsymbol{r}}$, then the matrix $N_{\text{tot}} \text{Cov}(\boldsymbol{r}, \hat{\boldsymbol{r}}) - F^{-1}$ has positive eigenvalues. In particular, taking the trace yields the condition

$$\langle |\boldsymbol{r} - \hat{\boldsymbol{r}}|^2 \rangle \geqslant B/N_{\text{tot}}, \tag{17}$$

where

$$B = \text{tr}\{F^{-1}\}. \tag{18}$$

The mean-squared error in our reconstruction is therefore bounded by the quantity B, which we will henceforth refer to as the Cramer-Rao bound (CRB). It has a clear operational meaning as the best mean-squared error achievable by our tomography experiment (scaled by N_{tot}). Furthermore, as the

simulations in Sec. VII demonstrate, the CRB represents a tight bound—in the limit of a large number of measurements, the achieved precision (i.e., mean-squared error) for unbiased tomography is well-described by B/N_{tot}.

III. NUMERICAL OPTIMIZATION

We are now ready to give a more formal rendering of the problem at hand. Our aim in finding the OED is to discover the numbers $\{N_{\gamma}\}$ which minimize B, subject to the constraint in Eq. (10). Of course the N_{γ} must be integers since one cannot perform an experiment a fractional number of times. But the solution of this problem is intractable, being combinatorial in nature. We will instead consider what the authors of [1] called the *relaxed* problem, where we allow "fractional experiments." We define the real, positive M-dimensional vector $\boldsymbol{\lambda}$ with elements λ_{γ} such that

$$\sum_{\gamma} \lambda_{\gamma} = 1. \tag{19}$$

The λ's are the "weights" representing the experiment design, so that in the limit $N_{\text{tot}} \longrightarrow \infty$ of a large number of measurements, we have

$$N_{\gamma} \longrightarrow N_{\text{tot}} \lambda_{\gamma}. \tag{20}$$

The experiment design $\boldsymbol{\lambda}$ is only asymptotically correct, but as our simulations in Sec. VII show, the optimization of $\boldsymbol{\lambda}$ produces results which are beneficial for finite data sets with a wide range of sizes: One simply rounds the right-hand side of (20) to the nearest integer to obtain N_{γ}. The Fisher information becomes

$$F = \sum_{\alpha\gamma} \frac{\lambda_{\gamma}}{p_{\alpha\gamma}} \boldsymbol{a}_{\alpha\gamma} \boldsymbol{a}_{\alpha\gamma}^{\mathsf{T}}. \tag{21}$$

In [1], similar expressions were derived, and a numerical convex optimization routine was invoked to find the vector $\boldsymbol{\lambda}$ which minimized $B = \text{tr}(F^{-1})$, subject to the normalization constraint in Eq. (19). But this problem does not require specialized convex optimization software. The optimization can be performed quickly using standard gradient-ascent algorithms since the constraint in Eq. (19) can be incorporated using a Lagrange multiplier. We define the *cost function*

$$J = B + \eta \left(\sum_{\gamma} \lambda_{\gamma} - 1 \right), \tag{22}$$

where η is a real Lagrange multiplier that imposes the normalization constraint on $\boldsymbol{\lambda}$. The OED is the vector $\boldsymbol{\lambda}^{\text{OED}}$ such that the cost function is minimized, at which point

$$\nabla_{\boldsymbol{\lambda}} J|_{\boldsymbol{\lambda} = \boldsymbol{\lambda}^{\text{OED}}} = 0. \tag{23}$$

To proceed further, we need to differentiate B. This is most easily done by re-writing the Fisher information as a product of matrices, as follows

$$F = A^{\mathsf{T}} \Lambda P^{-1} A, \tag{24}$$

where $P = \text{diag}(\boldsymbol{p})$ and

$$\Lambda = \text{diag}([\lambda_1 \mathbb{1}_{n_1}^{\mathsf{T}}, \ldots, \lambda_{\gamma} \mathbb{1}_{n_{\gamma}}^{\mathsf{T}}, \ldots, \lambda_M \mathbb{1}_{n_M}^{\mathsf{T}}]^{\mathsf{T}}), \tag{25}$$

---

[2]We have scaled the matrix by N_{tot}, so that strictly speaking we are dealing with the Fisher information *per datum*. The use of this "normalized" matrix will avoid notational clutter elsewhere.

are both $n_{\mathrm{tot}} \times n_{\mathrm{tot}}$ diagonal matrices, and where A was defined in Eq. (9). Here we used $\mathbb{1}_m$ to denote the m-dimensional column vector whose elements are all equal to 1. With the expression in Eq. (24) in hand, it is easy to differentiate the cost function [41]. We obtain

$$\nabla_\lambda J = \eta \mathbb{1}_M - \mathrm{diag}\{\mathrm{tr}_\alpha[P^{-1}AF^{-2}A^{\mathsf{T}}]\}, \qquad (26)$$

where $(\mathrm{tr}_\alpha[X])_{\gamma,\delta} = \sum_\alpha X_{\alpha\gamma,\alpha\delta}$ indicates a partial trace over the measurement outcomes. The OED is found by minimizing the norm of $\nabla_\lambda J$, which can be done efficiently using, for example, MATLAB's lsqnonlin routine. Convergence is accelerated by providing the algorithm with an expression for the Hessian H of J—the matrix containing the gradients of $\nabla_\lambda J$

$$
\begin{aligned}
H_{\delta\gamma} &= \partial_{\lambda_\delta}\partial_{\lambda_\gamma} J \\
&= \mathrm{tr}\,[(\partial_{\lambda_\gamma}\Lambda)P^{-1}AK_\delta A^{\mathsf{T}}], \qquad (27)
\end{aligned}
$$

where

$$K_\delta = \{F, F^{-1}A^{\mathsf{T}}P^{-1}(\partial_{\lambda_\delta}\Lambda)AF^{-1}\}, \qquad (28)$$

with $\{\cdots\}$ denoting the anticommutator. The matrix $\partial_{\lambda_\gamma}\Lambda$ contains the vector $\mathbb{1}_{n_\gamma}$ along the diagonal of its γth subblock and zeros everywhere else. To proceed with the optimization, we first pick an initial "guess" for λ. We then perform a quick line-search optimization to find the value of η that minimizes J, given our guess. We then feed λ and η into the lsqnonlin routine.

In Fig. 1, we plot the OED predicted using the previous procedure for a model polarimetric experiment introduced by Kosut *et al.* [1] (see Sec. 2.4, and in particular the top panel of Fig. 3 therein). The polarization of an incident photon is analyzed by means of a polarizing beamsplitter (PBS) with a pair of detectors placed behind each of its output ports, so the two outcomes of a given measurement correspond to one of the two detectors firing. A half-wave plate and a quarter-wave plate before the analyzer are used to modify the measurement basis, so the different measurement configurations correspond to different orientations of the waveplates' optical axes. In Fig. 1, the angles of both waveplates are allowed to vary in $5°$ increments from $0°$ to $45°$ with respect to the vertical, as defined by the direction of polarization reflected by the PBS. This gives 10 configurations for each waveplate, and so a total number of measurement configurations of $n_{\mathrm{tot}} = 100$. The input state, which determines the $p_{\alpha\gamma}$, is assumed to be purely diagonal polarization. Details of how the POVM's representing the measurements are constructed can be found in [1]. After
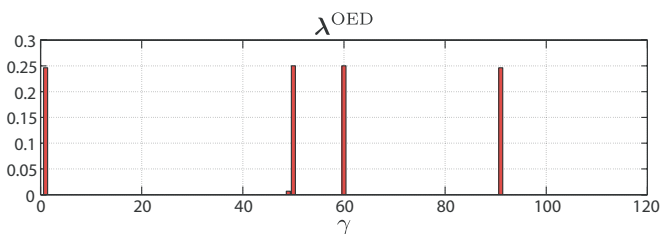

FIG. 1. (Color online) The OED for a simple polarization tomography experiment (described in [1]). This result is to be compared with the top panel of Figure 3 in [1]. The optimization we use is simple and fast, requiring no specialist software.

transforming into the Bloch representation and applying the formalism developed previously, we obtain results that stand in excellent agreement with those presented in [1]. The advantage of the method presented here is that no dedicated convex optimization packages are required. The optimization is therefore easier to implement; only standard numerical tools (any "conjugate gradients" algorithm will perform well) are required. Furthermore, the provision of the Hessian makes for very rapid convergence. In Sec. V, we will see that this numerical method can also be employed to find the *average* OED.

Here we comment that, although we are considering the problem of what design should be used, given a set of measurements, the results of this optimization can be interpreted in terms of which measurements should actually be used. That is, those measurements for which $\lambda_\gamma \sim 0$ can simply be omitted, and a reduced set of measurements should be performed: Those for which the λ's are largest [1]. Therefore, if one is able to parametrize the POVM's that can be implemented in the laboratory (as was done here using the angles of the two waveplates), one can use OED to determine what the optimal measurements are, subject to experimental constraints, in addition to finding the "weights" that should be assigned to each one. This application of OED also applies to its generalization, the average OED, presented shortly.

IV. MINIMAL OPTIMAL EXPERIMENT DESIGN

In certain circumstances, we can do better than the previous numerical optimization: we can write down an analytic expression for the OED. This is possible if we are able to invert the Fisher information matrix F "by hand." The form of Eq. (24) is suggestive. If we can take the inverse of the matrices A, P^{-1}, and Λ, we can multiply them to form F^{-1}. But in general A is not invertible. First, because only square matrices have inverses, and it may be that $n_{\mathrm{tot}} \neq N^2 - 1$, and second, because even if A is square, it must be of full rank, to be invertible. That is, none of its rows can be linearly dependent upon any of the others. But the rows of A are given by the $a_{\alpha\gamma}^{\mathsf{T}}$, which are *always* linearly dependent upon each other because of the second condition in Eq. (5)—ultimately because of the conservation of probability. This latter issue is potentially fatal, but fortunately it is possible to re-write the Fisher information in terms of full-rank matrices. To see how, note that from Eq. (5), we can re-express the last POVM element in each configuration in terms of the other elements,

$$a_{n_\gamma \gamma} = -\sum_{\alpha'=1}^{\widetilde{n}_\gamma} a_{\alpha'\gamma}, \qquad (29)$$

where $\widetilde{n}_\gamma = n_\gamma - 1$ is the number of *independent* measurement outcomes associated with configuration γ. In subsequent calculations we will always use primed indices, such as α', to enumerate these independent outcomes (i.e., omitting the last outcome with $\alpha = n_\gamma$). Substituting Eq. (29) into Eq. (21), we obtain

$$F = \sum_\gamma \lambda_\gamma \left(\sum_{\alpha'} \frac{a_{\alpha'\gamma}a_{\alpha'\gamma}^{\mathsf{T}}}{p_{\alpha'\gamma}} + \frac{1}{p_{n_\gamma \gamma}}\sum_{\alpha'\beta'}a_{\alpha'\gamma}a_{\beta'\gamma}^{\mathsf{T}} \right). \qquad (30)$$

Some redefinitions then afford a matrix decomposition for F

$$F = \widetilde{A}^{\mathsf{T}} \widetilde{\Lambda} \left(\widetilde{P}^{-1} + Q^{-1} \mathbb{I} \right) \widetilde{A}, \qquad (31)$$

where $\widetilde{A}$ is essentially the same as A, except that the last POVM element $\boldsymbol{a}_{n_\gamma \gamma}^{\mathsf{T}}$ has been removed for each configuration. The matrix $\widetilde{A}$ is therefore of size $\widetilde{n}_{\text{tot}} \times (N^2 - 1)$, where $\widetilde{n}_{\text{tot}} = \sum_\gamma \widetilde{n}_\gamma = n_{\text{tot}} - M$ is the total number of independent measurement outcomes. Similarly, the $\widetilde{n}_{\text{tot}} \times \widetilde{n}_{\text{tot}}$ diagonal matrix $\widetilde{P} = \text{diag}(\widetilde{\boldsymbol{p}})$ is formed from the vector $\widetilde{\boldsymbol{p}}$ of probabilities remaining after the last probability $p_{n_\gamma \gamma}$ for each configuration has been removed from $\boldsymbol{p}$. The $\widetilde{n}_{\text{tot}} \times \widetilde{n}_{\text{tot}}$ diagonal matrix $\widetilde{\Lambda}$ is defined accordingly as

$$\widetilde{\Lambda} = \text{diag}\left(\left[\lambda_1 \mathbb{1}_{\widetilde{n}_1}^{\mathsf{T}}, \ldots, \lambda_\gamma \mathbb{1}_{\widetilde{n}_\gamma}^{\mathsf{T}}, \ldots, \lambda_M \mathbb{1}_{\widetilde{n}_M}^{\mathsf{T}} \right]^{\mathsf{T}} \right). \qquad (32)$$

Q is the $\widetilde{n}_{\text{tot}} \times \widetilde{n}_{\text{tot}}$ "complementary" matrix to $\widetilde{P}$, which contains the probabilities we removed from $\widetilde{P}$,

$$Q = \text{diag}\left(\left[p_{n_1 1} \mathbb{1}_{\widetilde{n}_1}^{\mathsf{T}}, \ldots, p_{n_\gamma \gamma} \mathbb{1}_{\widetilde{n}_\gamma}^{\mathsf{T}}, \ldots, p_{n_M M} \mathbb{1}_{\widetilde{n}_M}^{\mathsf{T}} \right]^{\mathsf{T}} \right). \qquad (33)$$

Finally, the matrix $\mathbb{I}$ appearing in Eq. (31) is the $\widetilde{n}_{\text{tot}} \times \widetilde{n}_{\text{tot}}$ block-diagonal matrix whose γth subblock is equal to $\mathbb{1}_{\widetilde{n}_\gamma} \mathbb{1}_{\widetilde{n}_\gamma}^{\mathsf{T}}$.

If the matrix $\widetilde{A}$ is square, it should now be invertible (unless we are unlucky and it is rank deficient for some other reason, unrelated to the normalization of the POVM elements). To make $\widetilde{A}$ square, we must have that

$$\widetilde{n}_{\text{tot}} = N^2 - 1. \qquad (34)$$

That is, we require the total number of independent measurement outcomes to equal the number of independent real parameters specifying the quantum state. If $\widetilde{n}_{\text{tot}} < N^2 - 1$, tomographic inversion is not possible (this is intuitively obvious, but the Fisher information is also singular in this case). If $\widetilde{n}_{\text{tot}} > N^2 - 1$, the quantum state is overdetermined by our measurements. For this reason, we refer to the situation $\widetilde{n}_{\text{tot}} = N^2 - 1$ as "minimal tomography." There are good reasons why overdetermined tomography is advantageous since this introduces redundancy which suppresses the effects of statistical noise [39]. However, when attempting tomography of high-dimensional systems, the number of measurements required can become large, and in this case minimal tomography, involving the fewest possible number of independent measurement outcomes, is appealing.

For the case of a qubit, there are three distinct types of experiment that are minimal in this sense. We have $N = 2$, so one requires $N^2 - 1 = 3$ independent measurement outcomes. This can be achieved with (a) a single POVM measurement with four outcomes, (b) a pair of POVM measurements, one with three outcomes and one with two, and finally (c) a set of three POVM measurements, each with two possible outcomes. In each case we have $n_{\text{tot}} - M = N^2 - 1$. In case (a), the experiment design problem is trivial since there is only one measurement to make, so we just have $\boldsymbol{\lambda} \to \lambda = 1$. Case (b) is slightly more interesting; the experimentalist's time must be divided between two measurements. In what follows we will frequently treat case (c), whose experiment design is described by a three-element vector, for illustrative purposes. We refer to this case as "minimal qubit tomography with binary POVM's."

We now present an exact analytic result for the OED for minimal tomography on arbitrarily large systems (i.e., the result is not just restricted to qubits). We start by explicitly inverting F to obtain the CRB. Define

$$K = \widetilde{A}\widetilde{A}^{\mathsf{T}}, \qquad (35)$$

as the symmetric matrix whose elements are the scalar products between the POVM elements, $K_{\alpha'\gamma, \beta'\delta} = \boldsymbol{a}_{\alpha'\gamma}^{\mathsf{T}} \boldsymbol{a}_{\beta'\delta}$. Then we can write

$$\begin{aligned} B &= \text{tr}\{ \widetilde{\Lambda}^{-1} K^{-1} (\widetilde{P}^{-1} + Q^{-1} \mathbb{I})^{-1} \} \\ &= \text{tr}\{ \widetilde{\Lambda}^{-1} K^{-1} (\widetilde{P} - \mathbb{P}) \}, \end{aligned} \qquad (36)$$

where in the second line we have inverted $\widetilde{P}^{-1} + Q^{-1} \mathbb{I}$ blockwise using the rank-1 update to a matrix inverse [41], introducing the block-diagonal matrix $\mathbb{P}$ whose γth subblock is equal to $\widetilde{\boldsymbol{p}}_\gamma \widetilde{\boldsymbol{p}}_\gamma^{\mathsf{T}}$, where $\widetilde{\boldsymbol{p}}_\gamma$ is the $\widetilde{n}_\gamma$-dimensional vector of probabilities for all but the last outcome of the γth measurement.

After a little algebra, we can rewrite this result in the following way,

$$B = \sum_{\alpha'\gamma} \frac{b_{\alpha'\gamma}}{\lambda_\gamma}, \qquad (37)$$

where the $b_{\alpha'\gamma}$ are the elements of the $\widetilde{n}_{\text{tot}}$-dimensional vector

$$\boldsymbol{b} = \widetilde{\boldsymbol{p}} * (\boldsymbol{d} - D\widetilde{\boldsymbol{p}}), \qquad (38)$$

with D an $\widetilde{n}_{\text{tot}} \times \widetilde{n}_{\text{tot}}$ block-diagonal matrix whose blocks are given by the diagonal blocks of K^{-1}, and with the vector $\boldsymbol{d}$ given by

$$\boldsymbol{d} = \text{diag}(D) = \text{diag}(K^{-1}). \qquad (39)$$

Here the symbol $*$ indicates the Hadamard product (i.e., elementwise multiplication).

To find the OED, we differentiate the cost function J in Eq. (22). Armed with the expression (37) for B, we can do this analytically, and we arrive at the result (omitting an unimportant normalization factor)

$$\lambda_\gamma^{\text{OED}} = \sqrt{\sum_{\alpha'} b_{\alpha'\gamma}}. \qquad (40)$$

For the special case of binary measurements (such as "click"/"no-click" photon counters), the formula becomes especially simple since binary POVM's have only a single independent outcome each. The matrix D becomes purely diagonal, and we can write

$$\lambda_\gamma^{\text{OED:binary}} = \sqrt{d_\gamma p_\gamma (1 - p_\gamma)}, \qquad (41)$$

where we have dropped the redundant index α'.

V. AVERAGE OPTIMAL EXPERIMENT DESIGN

So far we have considered the problem of OED when the true state is known—both the optimization in Sec. III and the analytic expressions of the previous section require knowledge of $\boldsymbol{r}$ in order that the statistics $\boldsymbol{p}$ are fixed. What if we really have no information about the true state? This is surely when tomography is most useful, but we cannot then calculate the OED. In this section we will introduce the *average* OED, which does not require any knowledge of $\boldsymbol{r}$. The idea is a simple one: Where there is a dependence on $\boldsymbol{r}$, we

average over the space of all possible states, to account for our complete uncertainty about the true state. We will introduce two approaches. In the first, we average the Fisher information before minimizing the resulting CRB. In the second, we calculate the CRB analytically, and then perform the average. It seems that these two approaches produce essentially the same results, which is perhaps natural (but not obvious!). Using the first approach, one can calculate the average OED using a numerical optimization similar to the one described in Sec. III. The second approach yields an analytic result for the average OED, but of course it can only be applied to minimal tomography.

Both approaches, however, are exact *only for qubits*. The reason is that the averaging requires an integral over the space of physical states. This space is the $(N^2 - 1)$-dimensional volume such that all states $\boldsymbol{r}$ within it correspond to positive density matrices. Evaluating the boundaries of this region is nontrivial, and so, in general, the averaging cannot be performed exactly. But it is known that the space of physical states is a simply connected convex region, whose boundary lies between two concentric $(N^2 - 2)$-dimensional hyperspheres whose radii are given, respectively, by [40]

$$R_{\min} = \frac{1}{\sqrt{N(N-1)}} \quad \text{and} \quad R_{\max} = \sqrt{\frac{N-1}{N}}. \quad (42)$$

For the case of qubit tomography, with $N = 2$, these hyperspheres coincide, with $R_{\min} = R_{\max} = R_2 = 1/\sqrt{2}$, the radius of the well-known Bloch sphere on which all pure qubit states lie. For higher-dimensional systems, we will continue to approximate the space of physical states as being bounded by a hypersphere of radius R_N, where R_N must lie somewhere in between $R_{\min}$ and $R_{\max}$. If, at the end of the calculation, we find that the average OED does not depend (or depends only weakly) on our choice for R_N, then we can be confident that we have closely approximated the "true" average OED. In practice, we have found this to be the case, so that our averaging procedure—crude though it is—produces useful results.

A. Averaging the Fisher information

The definition of the Fisher information in Eq. (12) involves the evaluation of an ensemble average. For calculating the OED, we assumed that every member of our ensemble was prepared in the same state, and so we averaged over the statistical distribution of the data $\boldsymbol{n}$, assuming a given state $\boldsymbol{r}$. If we consider that, in fact, the prepared state is itself drawn from a large ensemble of possibilities, then we should extend our calculation of the expectation by averaging our result over those possibilities. With no prior information at all, we should average over the space of all physical states. The Fisher information obtained by averaging in this way is

$$\langle F \rangle = A^{\mathsf{T}} \Lambda G A, \quad (43)$$

where the elements of the diagonal matrix $G = \langle P^{-1} \rangle$ are given by

$$g_{\alpha\gamma} = \int \frac{\mathfrak{p}(\boldsymbol{r})}{p_{\alpha\gamma}(\boldsymbol{r})} dV, \quad (44)$$

with $\mathfrak{p}$ the probability distribution from which the states are drawn. In what follows we consider a uniform distribution

$\mathfrak{p} = \text{const}$, and take the integral to run over the $(N^2 - 1)$-dimensional volume enclosed by a hypersphere of radius R_N centered at the origin. For $N = 2$ the integral in Eq. (44) can be evaluated analytically to give

$$g_{\alpha\gamma} = \frac{3}{4R_2^3 |\boldsymbol{a}_{\alpha\gamma}|^3} \left\{ (c_{\alpha\gamma}^2 - R_2^2 |\boldsymbol{a}_{\alpha\gamma}|^2) \ln \left[\frac{c_{\alpha\gamma} - R_2 |\boldsymbol{a}_{\alpha\gamma}|}{c_{\alpha\gamma} + R_2 |\boldsymbol{a}_{\alpha\gamma}|} \right] \right.$$
$$\left. + 2c_{\alpha\gamma} R_2 |\boldsymbol{a}_{\alpha\gamma}| \right\}. \quad (45)$$

The generalization to higher dimensions can be found recursively,

$$g_{\alpha\gamma} = \frac{1}{|\boldsymbol{a}_{\alpha\gamma}|} \{ I_{N^2-3}(c_{\alpha\gamma}, |\boldsymbol{a}_{\alpha\gamma}|) - I_{N^2-3}(c_{\alpha\gamma}, -|\boldsymbol{a}_{\alpha\gamma}|) \}, \quad (46)$$

where the integral $I_n(a,b) = \int_0^{R_N} x^n \ln(a + bx) dx$ satisfies the recurrence relation

$$(n+1)I_n(a,b) = \frac{R_N^n}{b} \{ (a + bR_N) [\ln(a + bR_N) - 1] + a \}$$
$$+ R_N^{n+1} \frac{n}{n+1} - n \frac{a}{b} I_{n-1}(a,b), \quad (47)$$

with $I_0(a,b) = \frac{1}{b}(a + bR_N)[\ln(a + bR_N) - 1] - \frac{a}{b}[\ln(a) - 1]$. Once the elements $g_{\alpha\gamma}$ of G are known, we can apply the numerical method described in Sec. III to find the average OED, which is the experiment design $\lambda^{\langle \text{OED} \rangle}$ that minimizes the associated CRB $\langle B \rangle = \text{tr}(\langle F \rangle^{-1})$. The only difference is that the matrix P^{-1} in Eqs. (26), (27), and (28) should be replaced with G.

In Fig. 2, we show the result of such a numerical optimization, for the same set of POVM's as were used to generate Fig. 1 in Sec. III (the model is described in [1]). It is notable that the two figures look very different: The OED of Fig. 1 is optimal for a particular state (a pure state in this case), while the average OED has been constructed so as to be optimal for any state, or more precisely, optimal when we do not know which state has been prepared.

If we are interested in minimal tomography [so that Eq. (34) is satisfied], we can write down an analytic formula for the average OED. The averaged Fisher information is

$$\langle F \rangle = \widetilde{A}^{\mathsf{T}} \widetilde{\Lambda} \langle \widetilde{P}^{-1} + Q^{-1} \mathbb{I} \rangle \widetilde{A}. \quad (48)$$

Inverting this gives the associated CRB

$$\langle B \rangle = \sum_{\alpha'\gamma} \frac{\langle b_{\alpha'\gamma} \rangle}{\lambda_\gamma}, \quad (49)$$
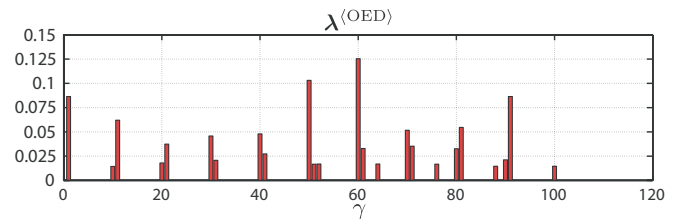

FIG. 2. (Color online) The average OED computed numerically from the averaged Fisher information. The set of POVM's is the same as those used to generate Fig. 1. The marked difference between the two experiment designs shows how ignorance of the true state changes the optimization.

where the averaging procedure yields

$$\langle \boldsymbol{b} \rangle = \widetilde{\boldsymbol{g}}^{-1} * (\boldsymbol{d} - \boldsymbol{f} * D\widetilde{\boldsymbol{g}}^{-1}), \tag{50}$$

with $\widetilde{\boldsymbol{g}}^{-1}$ the elementwise inverse of the vector $\widetilde{\boldsymbol{g}}$, whose elements are the $g_{\alpha'\gamma}$, defined in Eqs. (45) and (46). The tilde symbol indicates, as usual, that the element associated with the last outcome of each POVM has been removed. The $\widetilde{n}_{\text{tot}}$-dimensional vector $\boldsymbol{f}$ is defined as follows,

$$f_{\alpha'\gamma} = \frac{1}{\sum_\beta g_{\beta\gamma}^{-1}}, \quad \text{for all } \alpha'. \tag{51}$$

Differentiation of the appropriate cost function then yields the formula for the average OED (again omitting a normalization factor),

$$\lambda_\gamma^{\langle \text{OED} \rangle} = \sqrt{\sum_{\alpha'} \langle b_{\alpha'\gamma} \rangle}. \tag{52}$$

For the case of binary measurements, we drop the index α', and we have $f_\gamma = g_\gamma$. The formula in Eq. (50) then vanishes,

$$\langle b_\gamma \rangle = \frac{d_\gamma}{g_\gamma} \left(1 - \frac{g_\gamma}{g_\gamma} \right) = 0,$$

but the term in brackets cancels when the λ_γ are normalized, so we arrive at the simple expression

$$\lambda_\gamma^{\langle \text{OED:binary} \rangle} = \sqrt{d_\gamma / g_\gamma}. \tag{53}$$

B. Averaging the CRB directly

We now describe a second method for computing the average OED. Our approach here is to average the analytic expression in Eq. (37) for the CRB over a hypersphere—our approximation to the space of all physical states. Since the analytic expression only holds for minimal tomography, when Eq. (34) is satisfied, this method is applicable only in this situation. To distinguish this from the procedure outlined previously, where we averaged F, we will use two angle brackets to denote this type of average

$$\langle\!\langle B \rangle\!\rangle = \sum_{\alpha'\gamma} \frac{\langle\!\langle b_{\alpha'\gamma} \rangle\!\rangle}{\lambda_\gamma}. \tag{54}$$

Performing this averaging yields the result

$$\langle\!\langle \boldsymbol{b} \rangle\!\rangle = \widetilde{\boldsymbol{c}} * (\boldsymbol{d} - D\widetilde{\boldsymbol{c}}) - \langle\!\langle x^2 \rangle\!\rangle \text{diag}\,(DK), \tag{55}$$

where $\widetilde{\boldsymbol{c}}$ is generated from $\boldsymbol{c}$ by removing the last element from each configuration, and where $\langle\!\langle x^2 \rangle\!\rangle$ stands for the integral of a Cartesian coordinate x^2 over the hypersphere, divided by the hypersphere volume. For qubits, with $N = 2$, we have $\langle\!\langle x^2 \rangle\!\rangle = 1/10$, and in general, we find [42,43]

$$\langle\!\langle x^2 \rangle\!\rangle = \frac{R_N^2 (N^2 - 1)\Gamma\left(\frac{N^2-1}{2}\right)}{2(N^2 + 1)\Gamma\left(\frac{N^2+1}{2}\right)}, \tag{56}$$

where $\Gamma(.)$ is the Euler Gamma function. Differentiating the appropriate cost function leads to the following analytic formula for the average OED,

$$\lambda_\gamma^{\langle\!\langle \text{OED} \rangle\!\rangle} = \sqrt{\sum_{\alpha'} \langle\!\langle b_{\alpha'\gamma} \rangle\!\rangle}. \tag{57}$$

For the case of binary measurements, the formula reduces to

$$\lambda_\gamma^{\langle\!\langle \text{OED:binary} \rangle\!\rangle} = \sqrt{d_\gamma [c_\gamma (1 - c_\gamma) - \langle\!\langle x^2 \rangle\!\rangle |\boldsymbol{a}_\gamma|^2]}. \tag{58}$$

As a simple application of the these approaches to calculating the average OED, consider the particular case of three binary measurements on a qubit chosen so that the three independent vectors specifying the POVM's in the Bloch sphere all have the same length a. Then we have $\boldsymbol{a}_\gamma^\mathsf{T} \boldsymbol{a}_\delta = a^2 \cos(\theta_{\gamma\delta})$, where $\theta_{\gamma\delta}$ is the angle between the vectors $\boldsymbol{a}_\gamma$ and $\boldsymbol{a}_\delta$. The matrix K of scalar products can be easily inverted in this case. Suppose in addition that the measurements are symmetric in the sense that the "identity components" associated with the two outcomes of each measurement are the same, $c_\gamma = 1 - c_\gamma = 1/2$. Then Eqs. (58) and (53) both yield the same surprisingly simple result for the average optimal experiment design (up to an unimportant normalization factor),

$$\boldsymbol{\lambda}^{\langle\!\langle \text{OED:binary} \rangle\!\rangle} = \boldsymbol{\lambda}^{\langle \text{OED:binary} \rangle} = \begin{pmatrix} \sin(\theta_{23}) \\ \sin(\theta_{13}) \\ \sin(\theta_{12}) \end{pmatrix}. \tag{59}$$

Note that when the POVM's are all mutually orthogonal with $\theta_{\gamma\delta} = \pi/2$, the average OED is a uniform distribution. This is consistent with the well-known result that projective measurements chosen from the three mutually unbiased bases (MUB's [34]) for a qubit are optimal in the sense that, given a uniform distribution for the experiment design, and no prior knowledge of the state, they provide the best tomographic precision [33].

It is possible to use the result in Eq. (59) to analyze the effect of small deviations from the ideal MUB tomography, as might occur in a polarimetry experiment where the waveplates can only be oriented with limited accuracy. Suppose that one makes three measurements, such that $\theta_{23} = \theta_{13} = \pi/2$ and $\theta_{12} = \pi/2 - \vartheta$, where ϑ is a small deviation. If one uses a uniform design instead of the average OED, the average CRB $\langle\!\langle B \rangle\!\rangle$ will increase slightly, and the best achievable precision will suffer. But how large is this effect? One finds that the ratio $r = \langle\!\langle B \rangle\!\rangle_{\text{uni}} / \langle\!\langle B \rangle\!\rangle_{\text{opt}}$ of the average CRB's for the uniform and optimized designs is given by

$$r = 3\frac{(2 + \cos^2 \vartheta)}{(2 + \cos \vartheta)^2} \approx 1 + \vartheta^4/18, \tag{60}$$

where the approximation holds for $\vartheta \lesssim 1$. That is, the misalignment affects the tomographic precision only at fourth order. This shows that, while the average OED is useful when one has access to nonideal measurements, it is not particularly important in the context of small deviations from ideal measurements. If we set $\vartheta \approx \pi/2$, so that the last measurement is almost degenerate with (i.e., parallel to) one of the others, then we have $r \approx 3/2$; in this situation it is certainly worth using the optimized design.

In Fig. 3, we plot the average OED $\boldsymbol{\lambda}^{\langle\!\langle \text{OED} \rangle\!\rangle}$ predicted by Eq. (57), alongside the prediction $\boldsymbol{\lambda}^{\langle \text{OED} \rangle}$ of the method presented previously [Eq. (52)], for the case of minimal qubit
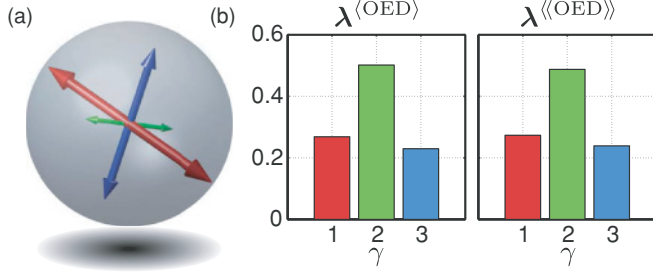
FIG. 3. (Color online) Comparison of the two methods for calculating the average OED. (a) The set of three binary POVM's used in the qubit tomography setup. Arrows represent the vectors $\boldsymbol{a}_{\alpha\gamma}$ on the Bloch sphere. (b) On the left is the average OED predicted by averaging the Fisher information [see Eq. (52)]; on the right is the corresponding result predicted by averaging the CRB [see Eq. (57)]. The two results are indistinguishable. This is typical: sometimes differences between the two methods are discernible, but generally they produce very similar experiment designs.

tomography using a randomly chosen set of *three* binary POVM's, given by

$$
\widetilde{c} = \begin{pmatrix} 0.43 \\ 0.81 \\ 0.49 \end{pmatrix}; \quad \widetilde{A} = \begin{pmatrix} -0.17 & 0.47 & 0.19 \\ -0.19 & 0.11 & 0.05 \\ -0.19 & -0.04 & -0.36 \end{pmatrix}. \quad (61)
$$

Note that the results are exact since the state space is exactly spherical for a qubit. What is notable is that the two predictions appear to coincide—although the formulas are different, the two averaging procedures seem to result in identical, or very similar, experiment designs. Conceptually, minimizing the averaged CRB is slightly different to minimizing the CRB associated to the averaged Fisher information, so it is not immediately obvious why this is so. Empirically, however, we have found that both methods generate the same results.

VI. OPTIMAL DESIGN FOR TOMOGRAPHY

The method just presented has an obvious generalization: Instead of choosing $\boldsymbol{\lambda}$ so as to minimize the averaged Cramer-Rao bound $\langle\!\langle B \rangle\!\rangle$, we could find an expression for $\langle\!\langle \delta B^2 \rangle\!\rangle$, the *variance* of B over the state space, and try to find the vector $\boldsymbol{\lambda}$ that minimizes this variance. Such an experiment design makes the tomographic reconstruction as "fair" as possible since the reconstruction should have as close as possible to a uniform precision for all states. We call this experiment design the *optimal design for tomography* (ODT). Using the expression in Eq. (37), we can write the variance of B as

$$
\langle\!\langle \delta B^2 \rangle\!\rangle = \langle\!\langle B^2 \rangle\!\rangle - \langle\!\langle B \rangle\!\rangle^2
$$
$$
= \boldsymbol{\lambda}^{-1\mathsf{T}} V \boldsymbol{\lambda}^{-1}, \quad (62)
$$

where $\boldsymbol{\lambda}^{-1}$ is the elementwise inverse of $\boldsymbol{\lambda}$, and where the real, symmetric (and therefore positive definite) $M \times M$ variance matrix V is given by

$$
V_{\gamma\delta} = \sum_{\alpha'\beta'} \langle\!\langle b_{\alpha'\gamma} b_{\beta'\delta} \rangle\!\rangle - \langle\!\langle b_{\alpha'\gamma} \rangle\!\rangle \langle\!\langle b_{\beta'\delta} \rangle\!\rangle
$$
$$
= \sum_{\alpha'\beta'} v_{\alpha'\gamma,\beta'\delta}. \quad (63)
$$

Performing the average explicitly requires a stout heart; the result is

$$
\begin{aligned}
v_{\alpha'\gamma,\beta'\delta} = {}& \langle\!\langle x^2 \rangle\!\rangle d_{\alpha'\gamma} d_{\beta'\delta} K_{\alpha'\gamma,\beta'\delta} \\
& - 2\langle\!\langle x^2 \rangle\!\rangle \sum_j c_{\alpha'\gamma} d_{\beta'\delta} W_{\alpha'\gamma,j} \widetilde{A}_{\beta'\delta,j} \\
& + 4\langle\!\langle x^2 \rangle\!\rangle \sum_j c_{\alpha'\gamma} c_{\beta'\delta} W_{\alpha'\gamma,j} W_{\beta'\delta,j} \\
& + \langle\!\langle x^4 \rangle\!\rangle \sum_j X_{\alpha'\gamma,j} X_{\beta'\delta,j} \\
& + \langle\!\langle x^2 y^2 \rangle\!\rangle \sum_{j\neq k} X_{\alpha'\gamma,j} X_{\beta'\delta,k} \\
& + 2\langle\!\langle x^2 y^2 \rangle\!\rangle \sum_{j\neq k} \widetilde{A}_{\alpha'\gamma,j} \widetilde{A}_{\beta'\delta,j} W_{\alpha'\gamma,k} W_{\beta'\delta,k} \\
& - \langle\!\langle x^2 \rangle\!\rangle \sum_{j\neq k} X_{\alpha'\gamma,j} X_{\beta'\delta,k}, \quad (64)
\end{aligned}
$$

where we defined the matrices $W = D\widetilde{A}$ and $X = \widetilde{A} * W$, and where the roman indices j,k number the Cartesian coordinates of the state space, and run from 1 to $N^2 - 1$. The average $\langle\!\langle x^2 \rangle\!\rangle$ is defined in Eq. (56) and the other averages are given by [42,43]

$$
\langle\!\langle x^2 y^2 \rangle\!\rangle = \frac{R_N^4 (N^2 - 1) \Gamma\left(\frac{N^2-1}{2}\right)}{4(N^2 + 3) \Gamma\left(\frac{N^2+3}{2}\right)},
$$
$$
\langle\!\langle x^4 \rangle\!\rangle = 3\langle\!\langle x^2 y^2 \rangle\!\rangle. \quad (65)
$$

We can now find the experiment design that minimizes Eq. (62). Setting the derivative of the cost function $J = \langle\!\langle \delta B^2 \rangle\!\rangle + \eta(\sum_\gamma \lambda_\gamma - 1)$ to zero, we obtain the condition

$$
V\boldsymbol{\lambda}^{-1} = \eta \boldsymbol{\lambda}^2, \quad (66)
$$

where $\boldsymbol{\lambda}^2$ is a column vector whose elements are given by λ_γ^2. It is not clear how to solve Eq. (66) analytically, but it is not hard to find the solution $\boldsymbol{\lambda}^{\mathrm{ODT}}$ numerically– almost any optimization algorithm will suffice.

A. ODT or average OED?

Both the ODT and the average OED are reasonable candidates for an experiment design that does not rely on prior knowledge of the true state ρ. They are both the results of a minimization, but their objective functions are different, so how do they compare? To contrast the two optimizations, we used the analytic results in Eqs. (54) and (62) to evaluate the objective functions $\langle\!\langle B \rangle\!\rangle$ and $\langle\!\langle \delta B^2 \rangle\!\rangle$ using many sets of $N^2 - 1$ randomly chosen binary POVM's, for a range of dimensionalities $N = 2$–10. Operationally, the quantities $\sqrt{\langle\!\langle B \rangle\!\rangle}$ and $\langle\!\langle \delta B^2 \rangle\!\rangle^{1/4}$ are more intuitive, having natural interpretations as distances in the Bloch representation. The former is the bound on the root-mean-squared (rms) error in the reconstructed state $\hat{\boldsymbol{r}}$—the typical value of $|\hat{\boldsymbol{r}} - \boldsymbol{r}|$; the latter is the square root of the standard deviation of the CRB, which can be thought of as the range over which the rms error $|\hat{\boldsymbol{r}} - \boldsymbol{r}|$ varies.

In Fig. 4, we show in the top two panels how the average OED and the ODT improve (i.e., reduce) the averages of
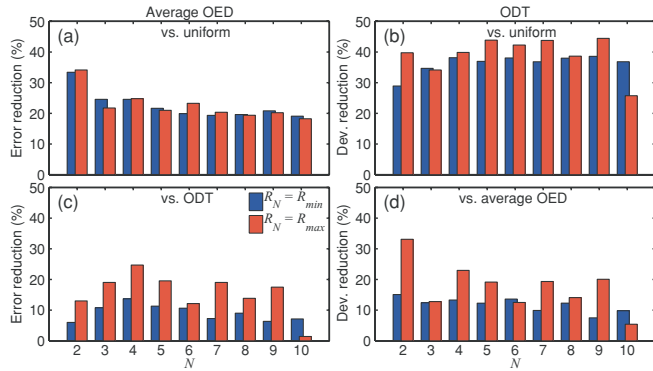
FIG. 4. (Color online) Comparison of ODT and average OED. The top two panels show how (a) average OED and (b) ODT compare with a uniform design $\lambda_\gamma = 1/M$. In part (a), the relative reduction in the average of the rms error $\sqrt{\langle\!\langle B \rangle\!\rangle}$ afforded by using $\lambda^{\langle\!\langle \text{OED} \rangle\!\rangle}$ instead of a uniform design, averaged over 3000 sets of randomly generated POVM's, is plotted against the dimension N of the quantum systems: the red bars (foreground) result from setting $R_N = R_{\max}$; the blue (behind) from setting $R_N = R_{\min}$. In part (b), we show the relative percentage reduction in the average of the "deviation" $\langle\!\langle \delta B^2 \rangle\!\rangle^{1/4}$ afforded by using λ^{ODT} instead of a uniform design. The lower two plots compare ODT and average OED with each other. In part (c), we plot the relative reduction in the average rms error arising from using the average OED instead of the ODT. In part (d), the relative reduction in the deviation given by choosing the ODT over the average OED is shown. Note that the averages $\langle\!\langle \cdot \rangle\!\rangle$ over the state space are exact for qubits, for which $R_{\min} = R_{\max}$, so the discrepancy between the red and blue bars is entirely statistical for $N = 2$; the results for higher dimensions should be interpreted with this in mind.

these quantities (averaged over 3000 sets of random POVM's) when compared against the results generated with a uniform experiment design $\lambda = \mathbb{1}_M/M$. The rms error $\sqrt{\langle\!\langle B \rangle\!\rangle}$ is typically reduced by around 20% when using $\lambda^{\langle\!\langle \text{OED} \rangle\!\rangle}$ [part (a)], and the "deviation" $\langle\!\langle \delta B^2 \rangle\!\rangle^{1/4}$ falls by $\sim$40% when using λ^{ODT} [part (b)]. Improvements at this level are certainly significant enough to motivate spending the time to calculate the appropriate optimal design.

The lower two panels show how the two designs $\lambda^{\langle\!\langle \text{OED} \rangle\!\rangle}$ and λ^{ODT} perform in terms of the other's objective function. Part (c) shows that the average OED improves the rms error by around 15% over that achieved by the ODT. Similarly part (d) shows that the ODT, on average, reduces the deviation $\langle\!\langle \delta B^2 \rangle\!\rangle^{1/4}$ by roughly the same amount with respect to the deviation produced by using the average OED.

In general, then, the two designs are clearly different, in that using one does not achieve optimality for the other's objective function: maximally fair tomography is not maximally precise, and vice versa. When designing a tomographic experiment, one should decide whether precision or fairness is the more important attribute. That said, it does, however, appear that the average OED is slightly fairer than a uniform design, and similarly that the ODT is more precise than a uniform design, so the two procedures are not entirely at loggerheads.

These results are not definitive because we cannot integrate over the true state space—only our hyperspherical approximations to it. We can conclude that these approximations are reasonable if the results do not depend strongly on the choice

made for the radius R_N. Examining Fig. 4, we see that the red and blue bars, which correspond to the minimal and maximal choices for R_N, appear to lie relatively close together, but there are some marked differences. However, note that for $N = 2$, $R_{\min} = R_{\max}$, so the differences here are purely statistical. It may be that the discrepancies between the results for $R_{\min}$ and $R_{\max}$, which are most noticeable in parts (c) and (d), can be explained largely in terms of statistical fluctuations; further investigation of this issue is probably required.

VII. MONTE CARLO SIMULATIONS

So far we presented a number of mathematical results involving approximate averages of asymptotic bounds. The reader could be forgiven for doubting the utility of these results in the laboratory. Short of performing real experiments—the ideal proving ground—we cannot do better than applying our techniques to simulated data, so we have performed some Monte Carlo simulations of minimal qubit tomography that provide some insight into when the CRB is saturated, and how much is gained by implementing an optimized experiment design.

In these simulations, we generate a set of qubit states distributed throughout the Bloch sphere, as shown in Fig. 5(a). The states are chosen so that it is easy to calculate averages over the Bloch sphere by polynomial interpolation: We use Chebyshev grids for the radial and polar coordinates and an equally spaced grid for the azimuth; averages over the sphere can then be computed accurately using Clenshaw-Curtis quadrature radially and polarwise, and Fourier interpolation around the equator [44]. Using *six* points for each coordinate yields averages that are accurate to within about 1% with reasonable computing times.

Figure 6 shows the mean-squared errors predicted by the CRB along with the errors actually achieved using simulated data equivalent to 2000 experimental runs, each consisting of $N_{\text{tot}} = 1000$ measurements. We chose a minimal set of binary POVM's randomly [shown in Fig. 5(b)], given by

$$\widetilde{c} = \begin{pmatrix} 0.42 \\ 0.41 \\ 0.49 \end{pmatrix}; \quad \widetilde{A} = \begin{pmatrix} -0.13 & 0.08 & -0.35 \\ -0.00 & 0.01 & -0.09 \\ -0.26 & -0.15 & -0.03 \end{pmatrix}. \quad (67)$$
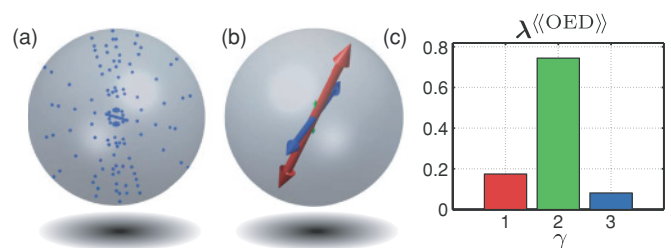

FIG. 5. (Color online) (a) The 216 sample states used to construct averages over the Bloch sphere using polynomial interpolation. The points are distributed evenly around the equator, but Chebyshev grids are used for the polar and radial distributions—this explains the clustering of points toward the poles and the center of the Bloch sphere. (b) The POVM elements used in the Monte Carlo simulations presented in the text. (c) The average OED for this POVM set.
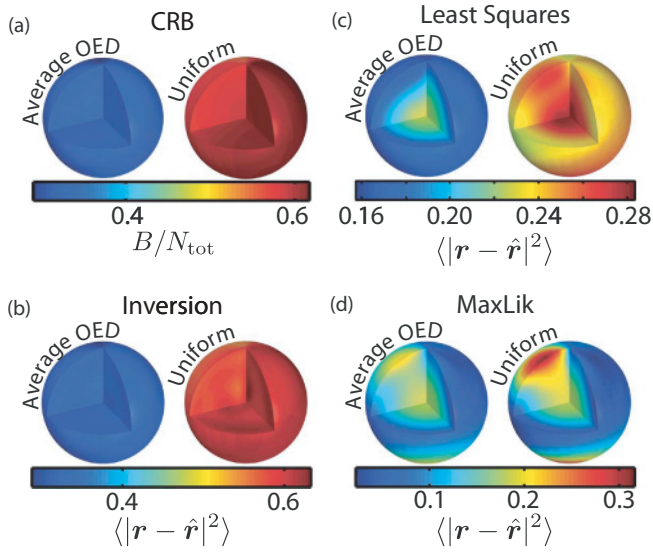
FIG. 6. (Color online) Monte Carlo results. Part (a) shows the variation of the CRB B/N_{tot} over the Bloch sphere, parts (c) and (d) show the mean-squared error $\langle |\boldsymbol{r} - \hat{\boldsymbol{r}}|^2 \rangle$ for the various reconstruction methods. The results for the average OED appear to the left; those for a uniform design to the right.

We then performed two simulations—one with a uniform experiment design, and one using the average OED predicted by Eq. (57) [shown in Fig. 5(c)]. Results are plotted for three different methods of state reconstruction, which we call *inversion*, *least squares*, and *maximum likelihood* [3,45].

A. Reconstruction methods

Inversion is the simplest: We first generate an estimate $\overline{\boldsymbol{p}}$ of the "true" statistics $\boldsymbol{p}$ using the relative frequencies of each measurement outcome in the simulated data $\{n_{\alpha\gamma}\}$, by setting $\overline{p}_{\alpha\gamma} = n_{\alpha\gamma}/N_\gamma$. We then substitute this into Eq. (8) and solve for the Bloch vector

$$\hat{\boldsymbol{r}} = A^- (\overline{\boldsymbol{p}} - \boldsymbol{c}). \tag{68}$$

Here the notation A^- stands for the Moore-Penrose pseudoinverse of A [46–48], which exists even though A is generally rectangular (and therefore has no true inverse). In MATLAB we use the backslash operator [49], which computes $\hat{\boldsymbol{r}}$ directly by Gaussian elimination. A problem with direct inversion is that sometimes the estimate $\hat{\boldsymbol{r}}$ lies outside the Bloch sphere, producing an unphysical reconstructed density matrix. In the least-squares method, we remove any of these unphysical estimates by using the "nearest" (in the least-squares sense) physical state. This is particularly easy for the qubit states we consider here, since the state space is spherical: Whenever the norm of $\hat{\boldsymbol{r}}$ exceeds $R_2 = 1/\sqrt{2}$, we renormalize it,

$$\hat{\boldsymbol{r}} \longrightarrow R_2 \frac{\hat{\boldsymbol{r}}}{|\hat{\boldsymbol{r}}|}. \tag{69}$$

The maximum-likelihood method is a more nuanced approach to the same problem [1,4,50–52]. The rationale is to try to find the physical state which is most likely to have produced the observed (simulated) data. This is the state $\hat{\boldsymbol{r}}$ that maximizes the likelihood function $p(\boldsymbol{n}|\hat{\boldsymbol{r}})$ defined in Eq. (11), or—more

conveniently—the state that renders the corresponding log-likelihood $\hat{\mathcal{L}}$ [see Eq. (13)] stationary with respect to changes in the estimated density matrix $\hat{\rho}$. A simple iterative scheme that converges on this state, while including the positivity and trace constraints on $\hat{\rho}$, has been derived by Hradil [4,51,52]. Starting with an unphysical state $\hat{\boldsymbol{r}}$ generated by inversion, we first normalize it as per the least-squares method. We then construct the corresponding density matrix $\hat{\rho}$ using Eq. (1), and then make the replacement

$$\hat{\rho} \longrightarrow \tfrac{1}{2} \left[R(\hat{\rho}) \, \hat{\rho} + \hat{\rho} \, R(\hat{\rho}) \right], \tag{70}$$

where the matrix R, which depends on $\hat{\rho}$ through the probabilities $\hat{p}_{\alpha\gamma} = \text{tr}\{\Pi_{\alpha\gamma} \hat{\rho}\}$, is given by [4,51]

$$R = \frac{1}{M} \sum_{\alpha\gamma} \frac{\overline{p}_{\alpha\gamma}}{\hat{p}_{\alpha\gamma}} \Pi_{\alpha\gamma}. \tag{71}$$

We repeat this procedure with the updated estimate, and its associated operator R, until the algorithm converges. The resulting state is guaranteed to be physical, where the initial state was not. In the cases that inversion produces a physical estimate to start with, we accept it without applying the previous procedure since it can be shown that under these circumstances inversion already produces the most likely estimate [4].

B. Results

Figure 6(a) shows that the average OED reduces the CRB significantly; the improvement over the uniform design is so large that the two plots appear almost uniform in color since they share the same color scale. The precision achieved by direct inversion, shown in part (b), matches the CRB very closely, and is accordingly improved markedly by using the average OED.

However, the precision of the other two reconstruction methods is not well described by the CRB. The least-squares method performs significantly better than inversion, while the maximum-likelihood method leaves both other methods standing, achieving mean-squared errors nearly 2 orders of magnitude smaller than inversion for some states. The reason why these methods perform better is clear: Knowledge of the boundaries of the space of physical states is utilized to improve the tomographic reconstruction. The least-squares method is too crude to take full advantage of this constraint, but the maximum-likelihood method exploits it to impressive effect. The positivity constraint on the density matrix is not present anywhere in our derivation of the CRB, and this is why an estimation method which implements this constraint is able to beat the "lower bound" on the average errors represented by the CRB. These simulations highlight the fact that the CRB as presented in this paper applies rigorously only to state tomography via direct linear inversion of the measured statistics.

Does this limit the applicability of our results relating to optimal experiment design? In the previous example, use of the average OED reduces the average rms error achieved by direct inversion over a uniform design by around 25%, while this number falls to 14% for the least-squares method, and 5% for the maximum-likelihood method. In the next

section, we will consider an alternative formulation of the CRB that includes the positivity constraint, and which therefore correctly describes the precision achieved by the maximum-likelihood method. We will see that its optimization results in very similar experiment designs to those presented previously, from which we conclude that our results remain close to optimal even when used with estimation methods that enforce positivity.

VIII. CONSTRAINED ESTIMATORS

It is well known in classical statistics that the bias and variance of an estimator are complementary. The precision of an estimate can be improved using some prior knowledge of the estimated quantities, but inevitably this biases the estimate, shifting the mean of the estimator away from the "true" value [23,53]. The least-squares and maximum-likelihood methods are biased estimation techniques because they incorporate the positivity constraint that excludes unphysical states, but at the same time they are more precise than direct inversion—which is unbiased—and they are able to beat the unbiased CRB.

It is useful to visualize the true quantum state as a point in the Bloch sphere, surrounded by a spherical "bubble" that represents isotropic statistical fluctuations. If the true state lies close to the boundary of the Bloch sphere, the bubble may extend into the region of unphysical states. A constrained estimation method such as maximum likelihood will exclude these unphysical states, which distorts the shape of the error bubble. This reduces its size, improving the tomographic precision, but it is clear that it also introduces bias, since the center of mass of the bubble no longer coincides with the true state.

The CRB can be modified to yield a bound on the precision of biased estimators, but it requires an analytic expression for the gradient of the bias as a function of the true state [25,54]. This depends on the particular estimator being used, which removes some of the generality of the expression; in the case of the maximum-likelihood method, it is not possible to write down such an expression in any case, since the estimator itself is only defined implicitly, as the fixed point of the iteration described in Eq. (70) [23]. In general, constructing CRB's for estimators biased by inequality constraints—such as positivity—is a hard problem [25], and so it is not obvious how to adapt the foregoing analysis to maximum-likelihood tomography.

In this section, we show that an alternative parametrization of the density matrix allows us to replace the inequality constraint of positivity with the *equality* constraint of unit trace. The CRB for estimators with equality constraints is a much more tractable problem, and we will derive the appropriate constrained CRB (CCRB), which correctly describes maximum-likelihood tomography.

A. The Cholesky representation

Instead of using the Bloch representation of ρ, we will use its *Cholesky decomposition* [45,50,55]. That is, we write ρ in the form

$$\rho = T^\dagger T, \tag{72}$$

where T is a unique[3] upper triangular matrix with real elements along its main diagonal. The positivity of ρ is built in to this parametrization because the eigenvalues of ρ are given by the squares of the singular values of T. Clearly the Hermiticity of ρ is also guaranteed. However the trace condition must be added separately (compare this with the Bloch representation, in which the trace condition and Hermiticity are "automatic," and positivity must be imposed separately),

$$\text{tr}\{T^\dagger T\} = 1. \tag{73}$$

The problem of quantum state tomography now reduces to finding the N^2 real numbers that parametrize the matrix T, subject to the equality constraint in Eq. (73). We define the Cholesky vector $\boldsymbol{\theta}$ as the column vector comprising these numbers, by analogy with the Bloch vector $\boldsymbol{r}$. To construct $\boldsymbol{\theta}$, we first define

$$\boldsymbol{t} = \begin{bmatrix} \textbf{vec}\,(\text{Re}\{T^\dagger\}) \\ \textbf{vec}\,(\text{Im}\{T^\dagger\}) \end{bmatrix}, \tag{74}$$

as the $(2N^2)$-dimensional column vector formed by vectorizing first the real part, and then the imaginary part of $T^\dagger$, and concatenating the two. We then eliminate all the redundant elements of $\boldsymbol{t}$, which contain zeros because of the structure of T (that is, because T is upper triangular with a real diagonal—"accidental" zeros we keep). This gives us $\boldsymbol{\theta}$. With these definitions, the trace condition becomes simply $|\boldsymbol{t}|^2 = 1$, or equivalently (since we only remove zero elements to produce $\boldsymbol{\theta}$),

$$|\boldsymbol{\theta}|^2 = 1. \tag{75}$$

That is, the Cholesky vector is of unit length, so that physical states are restricted to lying on an $(N^2 - 1)$ sphere in what we might call "Cholesky space." To derive the CCRB—the Cramer-Rao bound that incorporates this equality constraint—we need to evaluate the Fisher information using the Cholesky parametrization, which requires differentiation of the measurement statistics with respect to $\boldsymbol{\theta}$. Substituting Eq. (72) in to Eq. (3), we find

$$p_{\alpha\gamma} = \text{tr}\{T\,\Pi_{\alpha\gamma}T^\dagger\}. \tag{76}$$

In terms of $\boldsymbol{t}$, we have

$$p_{\alpha\gamma} = \boldsymbol{t}^\mathsf{T} P_{\alpha\gamma}\boldsymbol{t}, \tag{77}$$

where the $2N^2 \times 2N^2$ matrices $P_{\alpha\gamma}$ are generated from the POVM elements $\Pi_{\alpha\gamma}$ according to the relation

$$P_{\alpha\gamma} = M \times \begin{bmatrix} I \otimes \Pi_{\alpha\gamma} & \\ & I \otimes \Pi_{\alpha\gamma}^\mathsf{T} \end{bmatrix} \times M^{-1}, \tag{78}$$

with I the $N \times N$ identity matrix and M the matrix that separates real and imaginary parts, defined by [56,57]

$$M = \begin{bmatrix} 1 & 1 \\ -\text{i} & \text{i} \end{bmatrix} \otimes I_{N^2}, \tag{79}$$

---

[3]T is only unique when ρ is of full rank, which condition excludes pure states, but in such cases one can recover the uniqueness of T by adopting a consistent convention on the positions of zeros along its main diagonal.

where $I_{N^2} = I \otimes I$ is the $N^2 \times N^2$ identity matrix. Finally, we arrive at the following bilinear expression for the statistics in terms of $\boldsymbol{\theta}$

$$p_{\alpha\gamma} = \boldsymbol{\theta}^\mathsf{T} Q_{\alpha\gamma} \boldsymbol{\theta}, \tag{80}$$

where the $N^2 \times N^2$ matrices $Q_{\alpha\gamma}$ are formed from the $P_{\alpha\gamma}$ by deleting rows and columns with indices given by those of the elements of $\boldsymbol{t}$ that are deleted to produce $\boldsymbol{\theta}$. The $Q_{\alpha\gamma}$ mirror the properties of the POVM elements themselves, being real, symmetric matrices that sum to the identity operator,

$$Q_{\alpha\gamma} = Q_{\alpha\gamma}^\mathsf{T}, \quad \sum_\alpha Q_{\alpha\gamma} = I_{N^2}. \tag{81}$$

Differentiating the log-likelihood function with respect to $\boldsymbol{\theta}$, we find the Fisher information in the Cholesky representation to be

$$F = 4 \sum_{\alpha\gamma} \frac{\lambda_\gamma}{p_{\alpha\gamma}} Q_{\alpha\gamma} \boldsymbol{\theta}\boldsymbol{\theta}^\mathsf{T} Q_{\alpha\gamma}$$
$$= Z^\mathsf{T} \Lambda P^{-1} Z, \tag{82}$$

where the diagonal matrices Λ and P are as defined in Sec. III, and where the rows of the $n_{\text{tot}} \times N^2$ matrix Z are given by the row vectors

$$\boldsymbol{z}_{\alpha\gamma} = 2\boldsymbol{\theta}^\mathsf{T} Q_{\alpha\gamma}. \tag{83}$$

Note that unlike the matrix A appearing in the Bloch representation [see Eq. (24)], Z depends on the state $\boldsymbol{\theta}$ because the state enters the measurement statistics quadratically, rather than linearly, in the Cholesky representation. This more complicated dependence on the state makes further manipulations, such as averaging over the state space, more involved than in the Bloch representation. But we can find the CCRB, and we can find the associated OED numerically.

B. The constrained Cramer-Rao bound

In general, a set of n equality constraints on $\boldsymbol{\theta}$ can be written as $\boldsymbol{s}(\boldsymbol{\theta}) = 0$, where $\boldsymbol{s}$ is an n-dimensional column vector of constraint functions. Let the derivative of $\boldsymbol{s}$ with respect to $\boldsymbol{\theta}^\mathsf{T}$ be the $n \times N^2$ matrix S. The CCRB is then given by [22,26,27]

$$B_C = \text{tr}\{U(U^\mathsf{T} F U)^{-1} U^\mathsf{T}\}, \tag{84}$$

where U is the unitary matrix whose columns span the null space of S,

$$SU = 0. \tag{85}$$

Since $U^\mathsf{T} U = I_{N^2}$, the cyclic property of the trace gives

$$B_C = \text{tr}\{(U^\mathsf{T} F U)^{-1}\}. \tag{86}$$

In our case, this formula produces an extremely simple result. We have just a single equality constraint, so that $s = s = |\boldsymbol{\theta}|^2 - 1$. The matrix S then collapses to the row vector $2\boldsymbol{\theta}^\mathsf{T}$, and then U can be any matrix whose columns span the $(N^2 - 1)$-dimensional subspace that is orthogonal to $\boldsymbol{\theta}$. Now, multiplying out the product $F\boldsymbol{\theta}$ reveals that $F\boldsymbol{\theta} = 4\boldsymbol{\theta}$, so that $\boldsymbol{\theta}$ is always an eigenvector of the Fisher matrix with eigenvalue

4. Since F is Hermitian, its $N^2 - 1$ remaining eigenvectors $\boldsymbol{u}_j$ are all orthogonal to $\boldsymbol{\theta}$, and to each other, and we can write

$$F = 4\boldsymbol{\theta}\boldsymbol{\theta}^\mathsf{T} + \sum_{j=2}^{N^2} F_j \boldsymbol{u}_j \boldsymbol{u}_j^\mathsf{T}, \tag{87}$$

where the F_j are the eigenvalues of F associated with the $\boldsymbol{u}_j$. Now, we are free to take the $\boldsymbol{u}_j$ as the columns of U, and then we have

$$U^\mathsf{T} F U = 4U^\mathsf{T} \boldsymbol{\theta}\boldsymbol{\theta}^\mathsf{T} U + \sum_{j=2}^{N^2} F_j U \boldsymbol{u}_j \boldsymbol{u}_j^\mathsf{T} U,$$

$$= \begin{pmatrix} F_2 & & & \\ & F_3 & & \\ & & \ddots & \\ & & & F_{N^2} \end{pmatrix}.$$

Inverting this diagonal matrix is trivial, and so we obtain the final expression for the CCRB

$$B_C = \sum_{j=2}^{N^2} \frac{1}{F_j} = \text{tr}\{F^{-1}\} - \frac{1}{4}. \tag{88}$$

That is, up to an additive constant, the CCRB is the same as the unconstrained CRB in the Cholesky representation.

In Fig. 7, we show the results of another Monte Carlo simulation, in which we compare the mean-squared errors $\langle |\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}|^2 \rangle$ predicted by the CCRB with those produced by maximum-likelihood estimation, and by the least-squares method, for the same set of qubit states as depicted in
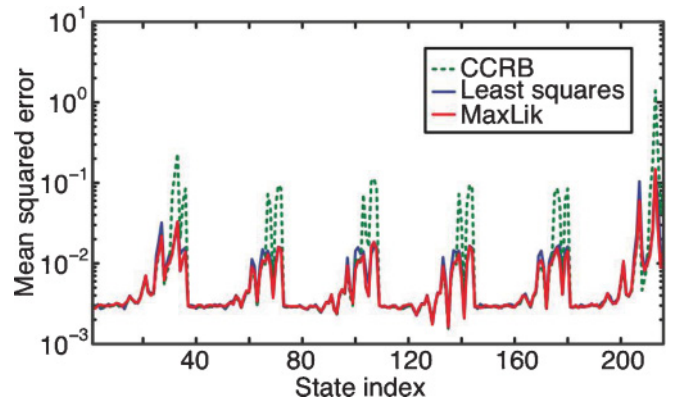

FIG. 7. (Color online) The mean-squared error $\langle |\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}|^2 \rangle$ achieved by the maximum-likelihood and least-squares methods are plotted alongside the CCRB for the same set of sample states as used to generate Fig. 6. In these Monte Carlo simulations, we chose a random minimal set of binary qubit POVM's, and a uniform experiment design. We then averaged the errors over 2000 virtual experiments, involving 10^5 measurements each—it seems that convergence on the CCRB requires larger samples than for the CRB in the Bloch representation. The precision attained by both constrained estimation methods is clearly well described by the CCRB, although it is discernible that maximum likelihood performs slightly better, as might be expected. For some states, the CCRB seems to "blow up," predicting much worse performance than actually achieved. These states are close to the pure state boundary, on which the Fisher matrix in the Cholesky representation becomes singular.

Fig. 5(a). The CCRB describes the errors achieved by both methods well, although the least-squares method is slightly less precise. It is not possible to compare the errors produced by inversion since nonpositive density matrices cannot be represented with a Cholesky vector.

For pure states, the Cholesky matrix T contains just a single nonzero diagonal element, so that the Cholesky vector $\boldsymbol{\theta}$ contains zero elements. The determinant of the outer product $\boldsymbol{\theta}\boldsymbol{\theta}^{\mathsf{T}}$ then vanishes, rendering the Fisher information singular. This explains the divergences of the CCRB for pure states; of course the precision actually achieved remains finite. These singularities would appear to be a pathology of the formalism and they should not be given a physical interpretation. Numerical failure in computing the CCRB is easily avoided by slightly shifting the pure states away from the boundary of the Bloch sphere.

C. Experiment designs for constrained estimators

The CCRB can be used to find the OED for constrained estimators. The analytic formulas from Sec. IV do not quite carry over to the Cholesky representation, but the numerical method described in Sec. III can be applied directly. Eq. (82) for the Fisher information in the Cholesky representation has precisely the same structure as Eq. (24) for F in the Bloch representation. Ignoring the constant shift of $\frac{1}{4}$ in Eq. (88), finding the OED simply requires that we minimize the trace of the inverse of F. To find the solution, we implement exactly the same numerical algorithm as for the Bloch representation, the only difference being that we replace the matrix A with Z in the Eqs. (26) through (28).

To examine the difference between these two optimizations—the first appropriate to unbiased tomography and the second designed for constrained estimators—we would like to compare the average OED's predicted by the two calculations. Unfortunately we cannot use any of the techniques described previously to evaluate the average OED in the Cholesky representation since the Fisher information defies attempts to average or invert it analytically. But it is possible to arrive at the average OED—or at least *an* average OED—by brute force. First, one calculates the OED for each of the sample states $\boldsymbol{r}_j$ shown in Fig. 5(a), so as to obtain $\boldsymbol{\lambda}^{\text{OED}}(\boldsymbol{r}_j)$. Next, one uses polynomial interpolation to numerically approximate the average of $\boldsymbol{\lambda}^{\text{OED}}$ over the Bloch sphere, yielding the design

$$\boldsymbol{\lambda}^{\langle\text{OED:brute}\rangle} = \frac{1}{\frac{4}{3}\pi R_2^3} \int \boldsymbol{\lambda}^{\text{OED}}(\boldsymbol{r})\, dV, \qquad (89)$$

which is literally the average OED. This method is only feasible for qubit states, where a direct numerical average can be performed quickly by polynomial interpolation with a few sample points. When this brute force method is implemented in the Bloch representation, we have found that the resulting distributions seem to coincide very closely with the designs $\boldsymbol{\lambda}^{\langle\text{OED}\rangle}$ and $\boldsymbol{\lambda}^{\langle\langle\text{OED}\rangle\rangle}$ introduced in Sec. V; this is why we still refer to $\boldsymbol{\lambda}^{\langle\text{OED:brute}\rangle}$ as the average OED.

We compared the average OED's for the Bloch and Cholesky representations using 1000 randomly chosen minimal sets of binary qubit POVM's. For each POVM set, we calculated the "discrepancy" $\mathcal{D}$ as the sum of the absolute difference between the two designs,

$$\mathcal{D} = \sum_\gamma \left| \lambda^{\langle\text{OED:brute}\rangle}_{\gamma,\text{Bloch}} - \lambda^{\langle\text{OED:brute}\rangle}_{\gamma,\text{Cholesky}} \right|. \qquad (90)$$

Since by definition the λ's sum to unity, one would expect an average discrepancy of order 1, when comparing two completely unrelated distributions. We found $\langle\mathcal{D}\rangle \sim 0.0390(7)$ for the average discrepancy, which shows that the optimal design for reconstruction by inversion is very close to the optimal design for maximum-likelihood tomography. Although we are not able to compare the ODT's for the two parametrizations (because the Cholesky representation is too unwieldy), this suggests that the results we derived in the Bloch representation for unbiased estimators are useful for constrained estimators too, even though the Bloch representation does not account for the constraints.

Despite this, the precision of maximum-likelihood estimation seems to improve less than the precision of inversion upon adoption of the average OED (whether derived using either the Bloch or the Cholesky representation), indicating that this reconstruction technique is less amenable to optimization generally. It seems that maximum-likelihood estimation is able to recruit the positivity constraint to counteract the statistical errors that a poor experiment design exacerbates. After our foray into the Cholesky representation, we are able to conclude that the smaller improvements in maximum-likelihood tomography are "intrinsic" to the method, and are not due to the neglect of the positivity constraint when optimizing the design in the Bloch representation. Of course, if precision matters, it is always better to use an optimized design over a uniform one.

IX. CONCLUSION

We revisited the problem of optimal experiment design for quantum state tomography, first introduced in [1]. We showed that specialist convex optimization software is not necessary to calculate the OED, and further that analytic results exist for minimal tomography, in which the measurements do not overdetermine the state. The reliance of the OED on knowledge of the true state is rather paradoxical, but we have explored a number of averaging methods that remove this state dependence. The averaging relies on approximating the state space as hyperspherical (an exact procedure for qubits), but the results are generally insensitive to the radius chosen for the hypersphere, indicating that this approximation is reasonably robust.

We introduced the average OED, which optimizes the average precision, or the "precision on average," depending on the details of the method—the resulting designs appear to be the same. We also considered the ODT, which seeks to render a tomographic experiment maximally fair. These optimizations are different, so the experimenter should make a choice about which property, precision, or fairness is most important when designing a tomographic setup. Finally, we confirmed that the formalism correctly describes the achieved precision for unbiased state estimation techniques, but that constrained estimators appear to perform better than predicted. However, a new formulation of the problem allows the treatment of

these constrained estimators, and shows that designs optimal for unbiased estimators are also very close to optimal for constrained methods.

Although this paper contains a rather large number of results, it seems that the various methods we proposed for computing the average OED are effectively interchangeable. That is, numerical evidence suggests that $\lambda^{\langle\mathrm{OED}\rangle} \approx \lambda^{\langle\!\langle\mathrm{OED}\rangle\!\rangle} \approx \lambda_{\mathrm{Cholesky}}^{\langle\mathrm{OED:brute}\rangle}$. Of the methods presented, probably the most useful is the numerical method for finding the average OED described in Sec. V A. This method is fast, and works for any number of measurements on a system of arbitrary dimension.[4] The analytic result for the average OED in the case of minimal tomography with binary measurements in Eq. (58), and the particularly simple special case in Eq. (59), may also prove useful.

————

[4]Details of the simple (though not pretty) MATLAB codes used are available from JN on request.

Further work would help to justify some of the claims we have made, particularly regarding the applicability of our averaging method in higher dimensions. An interesting possibility is to apply these techniques in examining the effectiveness of a detector; perhaps they can be used to design better measurements under laboratory constraints. In any case, as quantum tomography becomes increasingly indispensable, we hope that our results will be of assistance in real experiments.

[1] R. Kosut, I. Walmsley, and H. Rabitz, e-print arXiv:quant-ph/0411093 (2004).

[2] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, England, 2004).

[3] Z. Hradil, D. Mogilevtsev, and J. Rehacek, Phys. Rev. Lett. **96**, 230401 (2006).

[4] *Quantum State Estimation*, edited by M. Paris and J. Rehacek, Lecture Notes in Physics, Vol. 649 (Springer, Berlin, 2004).

[5] J. Řeháček, B.-G. Englert, and D. Kaszlikowski, Phys. Rev. A **70**, 052321 (2004).

[6] G. Puentes *et al.*, Phys. Rev. Lett. **102**, 080404 (2009).

[7] A. I. Lvovsky and M. G. Raymer, Rev. Mod. Phys. **81**, 299 (2009).

[8] S. R. Huisman, N. Jain, S. A. Babichev, F. Vewinger, A. N. Zhang, S. H. Youn, and A. I. Lvovsky, Opt. Lett. **34**, 2739 (2009).

[9] J. L. O'Brien, G. J. Pryde, A. Gilchrist, D. F. V. James, N. K. Langford, T. C. Ralph, and A. G. White, Phys. Rev. Lett. **93**, 080502 (2004).

[10] M. Riebe, K. Kim, P. Schindler, T. Monz, P. O. Schmidt, T. K. Korber, W. Hansel, H. Haffner, C. F. Roos, and R. Blatt, Phys. Rev. Lett. **97**, 220407 (2006).

[11] M. Branderhorst, I. Walmsley, and R. Kosut, in Proceeding of the European Conference on Lasers and Electro-Optics, 2007 and the International Quantum Electronics Conference. CLEOE IQEC 2007, (unpublished), p. 1.

[12] K. Modi and E. Sudarshan, e-print arXiv:0904.4663 (2009).

[13] C. Yu and H. Song, EPL **86**, 40007 (2009).

[14] M. Mohseni, A. T. Rezakhani, and D. A. Lidar, Phys. Rev. A **77**, 032322 (2008).

[15] J. S. Lundeen, A. Feito, H. Coldenstrodt-Ronge, K. L. Pregnell, C. Silberhorn, T. C. Ralph, J. Eisert, M. B. Plenio, and I. A. Walmsley, Nature Physics **5**, 27 (2008).

[16] K. Audenaert and S. Scheel, New J. Phys. **11**, 023028 (2009).

[17] H. Cramer, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, NJ, 1946).

[18] C. Radhakrishna Rao, Bull. Calcutta Math. Soc. **37**, 81 (1945).

[19] Y. Bard, *Nonlinear Parameter Estimation* (Academic Press, New York, 1974).

[20] E. Ollila, V. Koivunen, and J. Eriksson, in *Proceedings of the 5th IEEE Sensor Array and Multichannel Signal Processing Workshop, 2008* (IEEE, New York, 2008), pp. 414–418.

[21] A. Van den Bos, IEEE Transactions on Signal Processing 42 (1994).

[22] A. Jagannatham and B. Rao, IEEE Signal Process. Lett. **11**, 875 (2004).

[23] J. Fessler, IEEE Transactions on Image Processing **5**, 493 (1996).

[24] T. Marzetta, IEEE Transactions on Signal Processing **41**, 2247 (1993).

[25] C. Matson and A. Haji, J. Opt. Soc. Am. A **23**, 2702 (2006).

[26] P. Stoica and B. Ng, IEEE Signal Process. Lett. **5**, 177 (1998).

[27] J. Gorman and A. Hero, IEEE Trans. Inf. Theory **36**, 1285 (1990).

[28] S. Braunstein and C. Caves, Phys. Rev. Lett. **72**, 3439 (1994).

[29] K. C. Young, M. Sarovar, R. Kosut, and K. B. Whaley, Phys. Rev. A **79**, 062301 (2009).

[30] R. Demkowicz-Dobrzanski *et al.*, Phys. Rev. A **80**, 013825 (2009).

[31] U. Dorner, R. Demkowicz-Dobrzanski, B. J. Smith, J. S. Lundeen, W. Wasilewski, K. Banaszek, and I. A. Walmsley, Phys. Rev. Lett. **102**, 040403 (2009).

[32] P. Kurzynski, W. Kaszub, and M. Czechlewski, e-print arXiv:0905.1723 (2009).

[33] W. Wootters and B. Fields, Ann. Phys. (NY) 191 (1989).

[34] T. Paterek, B. Dakic, and C. Brukner, Phys. Rev. A **79**, 012109 (2009).

[35] S. Brierley and S. Weigert, Phys. Rev. A **79**, 052316 (2009).

[36] S. Weigert and M. Wilkinson, Phys. Rev. A **78**, 020303(R) (2008).

[37] R. Adamson and A. Steinberg, e-print arXiv:0808.0944 (2008).

[38] R. Vianna, M. Yang, A. Delgado, and C. Saavedra, e-print arXiv:0806.0391 (2008).

[39] M. de Burgh, N. Langford, A. Doherty, and A. Gilchrist, Phys. Rev. A **78**, 052122 (2008).

[40] J. Harriman, Phys. Rev. A **17**, 1249 (1978).

[41] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook* (Technical University of Denmark, 2008).

[42] G. Folland, Am. Math. Mon. **108**, 446 (2001).

[43] M. Abramowitz and I. Stegun, New York (1965).

[44] L. Trefethen, *Spectral Methods in MATLAB* (Society for Industrial Mathematics, Philadelphia, 2000).

[45] M. S. Kaznady and D. F. V. James, Phys. Rev. A **79**, 022109 (2009).

[46] A. Albert, *Regression and the Moore-Penrose Pseudoinverse.* (Academic Press, New York, 1972).

[47] G. Golub and W. Kahan, Journal of the Society for Industrial and Applied Mathematics: Series B, Numerical Analysis **2**, 205 (1965).

[48] T. Greville, SIAM Review **2**, 15 (1960).

[49] G. Recktenwald, *Numerical Methods with MATLAB: Implementations and Applications* (Prentice Hall, 2000).

[50] K. Banaszek, G. M. D'Ariano, M. G. A. Paris, and M. F. Sacchi, Phys. Rev. A **61**, 010304(R) (1999).

[51] Z. Hradil, J. Rehácek, J. Fiurásek, and M. JEZEK, Lect. Notes Phys. **649**, 59 (2004).

[52] A. I. Lvovsky, J. Opt. B: Quantum and Semiclassical Optics **6**, S556 (2004).

[53] A. Hero III, J. Fessler, and M. Usman, IEEE Transactions on Signal Processing **44**, 2026 (1996).

[54] J. Fessler and A. Hero, in *Proceedings of the 36th Midwest Symposium on Circuits and Systems, 1993* (IEEE, New York, 1993), pp. 253–256.

[55] L. Trefethen and D. Bau, *Numerical Linear Algebra* (Society for Industrial Mathematics, Philadelphia, 1997).

[56] A. Van Den Bos, IEEE Proceedings-Vision, Image and Signal Processing **141**, 380 (1994).

[57] D. Brandwood, in *IEEE Proceedings H Microwaves, Optics and Antennas* (IEEE, New York, 1983), Vol. 130, pp. 11–16.