Comprehensive analysis of quantum pure-state estimation for two-level systems

E. Bagan, A. Monras, and R. Muñoz-Tapia
Grup de Física Teòrica & IFAE, Facultat de Ciències, Edifici Cn, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain

(Received 22 December 2004; revised manuscript received 22 February 2005; published 15 June 2005)

Given N identical copies of the state of a quantum two-level system, we analyze its optimal estimation. We consider two situations: general pure states and (pure) states restricted to lie on the equator of the Bloch sphere. We perform a complete and comprehensive analysis of the optimal schemes based on local measurements, and give results (optimal measurements, maximum fidelity, etc.) for arbitrary N, not necessarily large, within the Bayesian framework. We also make a comparative analysis of the asymptotic limit of these results with those derived from a (pointwise) Cramér-Rao type of approach. We give explicit schemes based on local measurements and no classical communication that saturate the fidelity bounds of the most general collective schemes.

DOI: 10.1103/PhysRevA.71.062318 PACS number(s): 03.67.Hk, 03.65.Wj, 89.70.+c

I. INTRODUCTION

In quantum information, state characterization and measurements are not just *some* important topics. They are at the very core of the theory. Any unknown quantum state can only be unveiled by means of measurements, which provide the information that, in turn, is used to obtain an approximation of the state. This process is known as state estimation. Since this information is always gained at the expense of destroying the state, the estimation procedure requires a sample of N identical copies of it. Usually, we have a limited number of such copies and, therefore, it is very important to devise strategies with optimal performance in a variety of practical situations.

Here we address the issue of estimating the most elementary quantum state, that of a two-level system (throughout this paper we will refer to such systems as qubits for short). We consider two relevant cases: estimation of a completely unknown state, i.e., one given by an arbitrary point on the Bloch sphere; and estimation of restricted states that are known to lie on the equator of the Bloch sphere. The latter is also interesting because it is formally equivalent to phase estimation (these states are also equivalent to the so-called rebits [1]). We refer to these two situations as three- (3D) and two-dimensional (2D) cases, respectively.

Over the last few years, it has been recognized that a joint measurement on N copies (also known as collective measurement) is more efficient than N individual (or local) measurements on each copy separately [2,3]. This and other issues have been studied in various contexts [4–17], but for individual measurements not many analytical results have been obtained [18–24]. The most powerful involve sophisticated estimation theory technology that is not so widely known among physicists. Moreover, they mainly apply in the asymptotic limit (when N is large), and in a "pointwise" fashion, in the sense that no average over the prior probability distribution of unknown states is considered [25,26]. In contrast, we follow a "global" Bayesian approach that enables us to find constructively the optimal measurements, and the corresponding explicit expressions of the fidelity for finite N (not necessarily large). In particular, we present optimal schemes based on fixed (tomographic) and adaptive individual measurements in both 2D and 3D. This completes the results summarized in [22] by some of the authors, and provides further technical details that were skipped there. We also show that for large N these results are in agreement with those obtained using a Cramér-Rao type of approach. We show that there are nonadaptive schemes based on individual measurements that saturate the optimal collective bound. We consider also adaptive local schemes, usually called local operations and classical communication (LOCC) schemes. Within our framework, we give a proof that the simplest version of these LOCC schemes—the Gill-Massar scheme [19]—indeed saturates the collective bound. For more sophisticated LOCC schemes we prove that two-outcome measurements (von Neumann's) perform better than generalized measurements with a larger number of outcomes. In addition to these results, we include in Appendix A a comprehensive and unified rederivation of the optimal collective-measurement scheme in 2D and 3D. In summary, this paper is a complete and self-contained review on estimation of qubit pure states.

The paper is organized as follows. The basic concepts and notation are introduced in the next section. In Sec. III we recall the bounds on the fidelity of the optimal collective measurements. In Sec. IV we discuss several local measurement schemes, with and without classical communication. Asymptotic values of the fidelity are computed in Sec. V. A summary of results and our main conclusions are presented in Sec. VI, and four technical appendixes end the paper.

II. CONCEPTS AND NOTATION

Assume that we are given an ensemble of N identical copies of an unknown qubit state, which we denote by $|\vec{n}\rangle$, where $\vec{n}$ is the unit vector on the Bloch sphere that satisfies

$$|\vec{n}\rangle\langle\vec{n}| = \frac{1+\vec{n}\cdot\vec{\sigma}}{2}, \tag{2.1}$$

and $\vec{\sigma}=(\sigma_x,\sigma_y,\sigma_z)$ are the usual Pauli matrices. After performing a measurement (collective or local) on the N copies of $|\vec{n}\rangle$, one obtains an outcome χ. Based on χ, an estimate

 ©2005 The American Physical Society

$|\vec{M}(\chi)\rangle$ for the unknown state is guessed. To quantify how well $|\vec{M}(\chi)\rangle$ approximates the unknown state $|\vec{n}\rangle$ we use the fidelity, defined as the overlap

$$f_n(\chi) \equiv |\langle \vec{n}|\vec{M}(\chi)\rangle|^2 = \frac{1+\vec{n}\cdot\vec{M}(\chi)}{2}. \qquad (2.2)$$

Notice that $\ell_\chi(\vec{n}) \equiv 1 - f_n(\chi)$ can be regarded as a loss function. The quality of an estimation scheme is usually quantified by means of the averaged fidelity, hereafter fidelity in short, over the initial probability and all possible outcomes,

$$F \equiv \langle f\rangle = \sum_\chi \int dn\, f_n(\chi) p_n(\chi), \qquad (2.3)$$

where dn is the prior probability distribution, and $p_n(\chi)$ is the probability of getting outcome χ given that the unknown state is $|\vec{n}\rangle$. Our aim is to maximize F.

Any measurement (collective or local) is described by a positive-operator-valued measure (POVM) on the N copies. Mathematically, a POVM is a set of positive operators $\{O(\chi)\}$ (each one of them associated to an outcome) that satisfy the condition

$$\sum_\chi O(\chi) = \mathbb{1}. \qquad (2.4)$$

The probability $p_n(\chi)$ is given in terms of these operators by

$$p_n(\chi) = \mathrm{tr}[\rho_n O(\chi)], \qquad (2.5)$$

where ρ_n is the quantum state of the N identical copies, i.e., $\rho_n = (|\vec{n}\rangle\langle\vec{n}|)^{\otimes N}$.

In Eq. (2.3) there are only two elements that require optimization: the guess and the POVM, which enter this equation through Eqs. (2.2) and (2.5), respectively. The optimal guess (OG) can be obtained rather trivially. The Schwarz inequality shows that the choice

$$\vec{M}(\chi) = \frac{\vec{V}(\chi)}{|\vec{V}(\chi)|}, \qquad (2.6)$$

where

$$\vec{V}(\chi) = \int dn\, \vec{n}\, p_n(\chi), \qquad (2.7)$$

maximizes the value of the fidelity. If we define Δ through the relation

$$F \equiv \frac{1}{2}(1+\Delta), \qquad (2.8)$$

we have for the OG [22]

$$\Delta = \sum_\chi |\vec{V}(\chi)| \qquad (2.9)$$

(in the strict sense we should write F^{OG} and Δ^{OG}, but to simplify the notation, we drop the superscript when no confusion arises). Equation (2.6) gives the best state that can be inferred, and Eqs. (2.8) and (2.9) give the maximum fidelity that can be achieved for *any* prior probability and *any* measurement scheme specified by the conditional probabilities $p_n(\chi)$. The goal of this paper is to compute the maximum value of Eq. (2.8) and to obtain the optimal measurements, especially the local ones, in various interesting situations.

Before concluding this section, let us substantiate the assertion that our approach is Bayesian. For a given measurement, we view each outcome χ as an evidence. In Bayesian estimation [27], one computes the posterior probability distribution $p_\chi(\vec{n})$ that the state on which we have performed a measurement is $|\vec{n}\rangle$ given the evidence χ, and minimizes the averaged loss function

$$L_\chi \equiv \int dn\, \ell_\chi(\vec{n}) p_\chi(\vec{n}). \qquad (2.10)$$

Using Bayes' law one can straightforwardly check that the solution is the OG, $\vec{M}(\chi)$, in Eq. (2.6). In other words, the OG is the Bayesian estimator for the problem we are considering in this paper.

III. COLLECTIVE MEASUREMENTS

In this section we recall some known results concerning estimation based on collective measurements (collective schemes). Their fidelities provide absolute upper bounds for any estimation scheme. In the next sections we will view them as points of reference to which we will compare different local schemes in order to assess their performance.

As mentioned in the Introduction, a 2D state corresponds to a point that is known to lie on the equator of the Bloch sphere. If we take it to be on the xy plane, such a state $|\vec{n}\rangle$ has $\vec{n} = (\cos\theta, \sin\theta, 0)$. If no other information is available, the prior probability distribution has to be isotropic, that is, $dn = d\theta/(2\pi)$. Notice that we can write

$$\rho_n \equiv \rho(\theta) = U(\theta)\rho_0 U^\dagger(\theta), \qquad (3.1)$$

where ρ_0 is a fiducial state of angular momentum $J \equiv N/2$ and maximal magnetic quantum number $m = J$ along any fixed direction on the equator of the Bloch sphere. In particular, ρ_0 can be chosen to point along the x axis,

$$\rho_0 = |JJ\rangle_x\, {}_x\langle JJ|. \qquad (3.2)$$

In Eq. (3.1), $U(\theta)$ is the unitary representation of a rotation around the z axis. The group of such unitary matrices is isomorphic to U (1): the group of phases. One can show (See Appendix A) that

$$\Delta \leqslant \frac{1}{2^N}\sum_{m=-J}^{J} \binom{N}{J+m} \sqrt{\frac{J-m}{J+m+1}} \equiv \Delta^{\mathrm{COL}}, \qquad (3.3)$$

where Δ is defined in Eq. (2.8). This bound is tight, since there is a POVM for which it is attained. The simplest choice is given by the following set of rank-1 operators:

$$O(\phi) = U(\phi)|B\rangle\langle B|U^\dagger(\phi), \qquad (3.4)$$

where

$$|B\rangle = \sum_{m=-J}^{J} |J,m\rangle. \qquad (3.5)$$

In the asymptotic limit we have

$$\Delta^{\mathrm{COL}} = 1 - \frac{1}{2N} + \cdots, \tag{3.6}$$

and the maximum fidelity is

$$F^{\mathrm{COL}} = 1 - \frac{1}{4N} + \cdots. \tag{3.7}$$

Likewise, a 3D state $|\vec{n}\rangle$ corresponds to a general point on the Bloch sphere. We can write

$$\rho_n \equiv \rho(\vec{n}) = U(\vec{n})\rho_0 U^\dagger(\vec{n}), \tag{3.8}$$

where for convenience ρ_0 is now chosen to point along the z axis, i.e.,

$$\rho_0 = |JJ\rangle\langle JJ|, \tag{3.9}$$

and $U(\vec{n})$ is the unitary representation [i.e., the element of SU(2)] of the rotation that brings $\vec{z}$ into $\vec{n}$ (a rotation around the vector $\vec{z}\times\vec{n}$). In this case, dn is the invariant measure on the two-sphere, e.g.,

$$dn = \frac{d(\cos\theta)d\phi}{4\pi}, \tag{3.10}$$

where θ and ϕ are the standard azimuthal and polar angles. We have the upper bound (see Appendix A)

$$\Delta \leqslant \frac{J}{J+1}. \tag{3.11}$$

Recalling that $J \equiv N/2$ we finally get [3]

$$F^{\mathrm{COL}} = \frac{N+1}{N+2}. \tag{3.12}$$

As for the 2D case, there is a POVM that saturates all the inequalities. It can be chosen to be

$$O(\vec{m}) = d_J U(\vec{m})|JJ\rangle\langle JJ|U^\dagger(\vec{m}), \tag{3.13}$$

where $d_j=2j+1$ is the dimension of the invariant Hilbert space corresponding to the representation $\mathbf{j}$ of SU(2). One can easily check that the POVM condition $\int dn\, O(\vec{n})=\mathbb{1}$ is satisfied.

For large N, the fidelity behaves as

$$F^{\mathrm{COL}} = 1 - \frac{1}{N} + \cdots. \tag{3.14}$$

IV. LOCAL MEASUREMENTS

Collective measurements, although very interesting from the theoretical point of view, are difficult to implement in practice. Far more interesting for experimentalists are individual von Neumann measurements [21,28]. This paper is mainly concerned with them, to which we will loosely refer as "local measurements," for short.

Local measurements on qubits all consist in measuring the observable $\vec{m}\cdot\vec{\sigma}$, where $\vec{m}$ is a unit Bloch vector characterizing the measurement (in a spin system, e.g., $\vec{m}$ is the orientation of a Stern-Gerlach apparatus), and are represented by the two projectors

$$O(\pm\vec{m}) = \frac{1 \pm \vec{m}\cdot\vec{\sigma}}{2}. \tag{4.1}$$

In a general frame we must also allow for classical communication, i.e., the possibility of adapting the orientation of the measuring devices depending on previous outcomes [20–22].

In the next sections we study quantitatively several schemes in ascending order of optimality: from the most basic tomography inspired schemes [29,30] to the most general individual measurement procedure with classical communication [22]. Our aim is to investigate how good these local measurements are as compared to the collective ones for *any* value of N. We would like to stress that in this context few analytical results are known [19,23,26]. Our results here complement and extend the analysis carried out by some of the authors in [22].

A. Fixed measurements

Let us start with the most basic scheme for reconstructing a qubit: local measurements along two fixed orthogonal directions (say, x and y) in the equator of the Bloch sphere for 2D states, or along three fixed orthogonal directions (say, x, y, and z) for 3D states. This kind of scheme is often called tomography [28,29,31].

Before presenting the OG analysis for this specific scheme, we will briefly consider a much simpler estimator, inspired by the law of large numbers, that we will call the "frequentist" guess (FG). It is interesting for two reasons. First, its fidelity provides a lower bound for what one would call "acceptable" schemes. Second, it has the nice property of being directly (and easily) obtained from the observed frequencies without further processing.

Consider $N=2\mathcal{N}(3\mathcal{N})$ copies of the state $|\vec{n}\rangle$. After $\mathcal{N}$ measurements of the projection of $\vec{\sigma}$ along each one of the directions $\vec{e}_i$ (i.e., of $\vec{e}_i\cdot\vec{\sigma}$), $i=x, y, (z)$, we obtain a set of outcomes +1 and −1 with frequencies $\mathcal{N}\alpha_i$ and $\mathcal{N}(1-\alpha_i)$, respectively. This occurs with probability

$$p_n(\alpha) = \prod_{i=x,y,(z)} \binom{\mathcal{N}}{\mathcal{N}\alpha_i} \left(\frac{1+n_i}{2}\right)^{\mathcal{N}\alpha_i} \left(\frac{1-n_i}{2}\right)^{\mathcal{N}(1-\alpha_i)}, \tag{4.2}$$

where n_i are the projections of the vector $\vec{n}$ along each direction, $n_i \equiv \vec{n}\cdot\vec{e}_i$, and we have used the shorthand notation $\alpha=\{\alpha_i\}$. The combinatorial factor takes into account all the possible orderings of the outcomes and the remaining factors are the quantum probabilities, i.e., the appropriate powers of $\mathrm{tr}[|\vec{n}\rangle\langle\vec{n}|O(\pm\vec{e}_i)]$.

Quantum mechanics tells us that $\vec{n}$ is the expectation value of $\vec{\sigma}$, i.e., $\langle\vec{n}|\vec{\sigma}|\vec{n}\rangle=\vec{n}$. Provided N is large, the law of large numbers enables us to approximate this mean by the sample mean, obtained from the relative frequencies of the outcomes of the N measurements. This yields the estimator

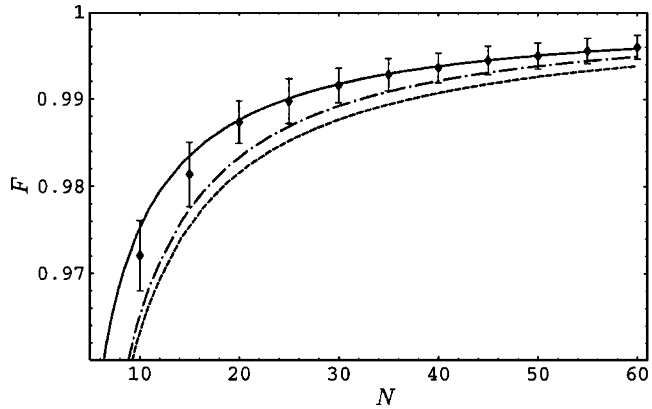

FIG. 1. Average fidelities in terms of the number of copies in the 2D case for the optimal collective measurement (solid line), tomographic OG (dot-dashed line), tomographic FG (dashed line), and a simulation of the greedy scheme (dots).

$$M_i^{\mathrm{FG}}(\alpha) = \frac{2\alpha_i - 1}{\sqrt{\sum_j (2\alpha_j - 1)^2}}. \tag{4.3}$$

Notice the normalization factor which ensures that $|\vec{M}^{\mathrm{FG}}| = 1$; hence $\vec{M}^{\mathrm{FG}}$ always corresponds to a physical pure state [30]. In this case

$$\Delta^{\mathrm{FG}} = \sum_\alpha \int dn\, \vec{n} \cdot \vec{M}^{\mathrm{FG}}(\alpha) p_n(\alpha), \tag{4.4}$$

where $p_n(\alpha)$ is defined in Eq. (4.2). The behavior of the corresponding fidelity F^{FG} for 2D and N up to 60 is represented by a dashed line in Fig. 1 (similar plots can be obtained for 3D). The fidelity of any acceptable estimation scheme is expected to lie between this line and the solid one, corresponding to the collective bound discussed in the previous section.

Next, we turn to the Bayesian approach. The OG for this measurement scheme is $\vec{M}^{\mathrm{OG}}(\alpha) = \vec{V}(\alpha)/|\vec{V}(\alpha)|$, where

$$\vec{V}(\alpha) = \int dn\, \vec{n}\, p_n(\alpha). \tag{4.5}$$

From Eqs. (2.8) and (2.9), the optimal fidelity reads $F = [1 + \sum_\alpha V(\alpha)]/2$. Closed expressions of the OG fidelity for the lowest values of N can be derived from Eq. (4.5) using

$$\int dn\, n_{i_1} n_{i_2} \cdots n_{i_q} = \frac{1}{K_q} \delta_{\{i_1 i_2} \delta_{i_3 i_4} \cdots \delta_{i_{q-1} i_q\}}, \tag{4.6}$$

where the normalization factor is $K_q = q!!$ in 2D and $K_q = (q+1)!!$ in 3D, the indexes in curly brackets are fully symmetrized, e.g., $\delta_{\{i_1 i_2 \delta_{i_3 i_4}\}} = \delta_{i_1 i_2}\delta_{i_3 i_4} + \delta_{i_1 i_3}\delta_{i_2 i_4} + \delta_{i_1 i_4}\delta_{i_2 i_3}$, and, obviously, the integral (4.6) vanishes for q odd. These results follow from very general symmetry arguments. For larger values of N the expressions become rather involved and we have resorted to a numerical calculation.

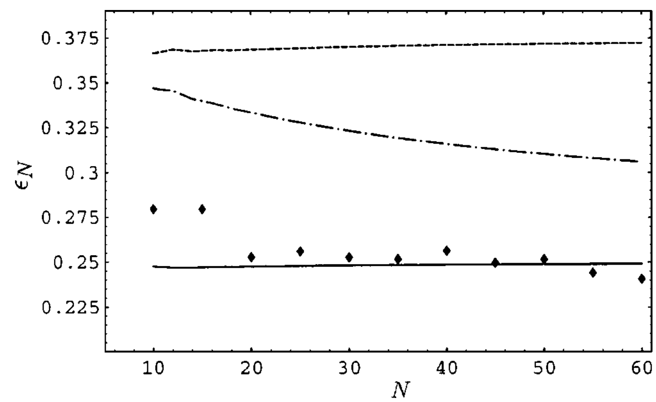

FIG. 2. Scaled error $\epsilon_N = N(1-F)$ in the 2D case for collective (solid line), OG (dot-dashed line), FG (dashed line), and greedy (diamonds) schemes.

The 2D case is included in Fig. 1, where the average fidelity for the OG guess and for N in the range 10–60 is shown. The fidelity approaches unity as N increases, and the OG (dot-dashed line) always s better than the FG, as it should. Notice that to make the graphs more easily readable we have interpolated between integer points.

At this point, it is convenient to define the scaled error function

$$\epsilon_N = N(1-F) \tag{4.7}$$

and the limit

$$\epsilon = \lim_{N\to\infty} \epsilon_N, \tag{4.8}$$

which gives the first-order coefficient of the fidelity in the large-N expansion, $F = 1 - \epsilon/N + \cdots$ (the asymptotic behavior will be properly discussed in Sec. V). Figure 2 shows ϵ_N as a function of N for 2D states. One readily sees that the FG gives $\epsilon^{\mathrm{FG}} \approx 3/8$ [22], while for collective measurements one has $\epsilon^{\mathrm{COL}} \approx 1/4$, in agreement with Eq. (3.7). The stability of the curves $\epsilon_N^{\mathrm{COL}}$ and ϵ_N^{FG} shows that the fidelity is well approximated by $F = 1 - \epsilon/N$ for such small values of N as those in the figure. This asymptotic regime is not yet achieved by the OG; however, we will show in Sec. V that the OG gives $\epsilon^{\mathrm{OG}} = 1/4$, thus matching the collective bound (3.7) for large N.

Figure 3 shows the scaled error in the 3D case. Again, one readily sees that the OG performs better than the FG, but it does not appear to be as good as in the 2D case.

In previous paragraphs we have presented the most basic scheme, i.e., that with a minimal number of orientations of the measuring devices and without exploiting classical communication. A next step in complexity is to consider a more general set of fixed directions $\{m_k\}$. It is intuitively clear that, assuming some sort of approximate isotropy, the more directions are taken into account, the better the estimation procedure will be. For instance, in 2D we may consider a set of directions given by the angles $\theta_k = k\pi/N$, where $k = 1, \ldots, N$. The set of outcomes χ can be expressed as an N-digit binary number $\chi = i_N i_{N-1} \cdots i_2 i_1$, where $i_k (=0,1)$ and

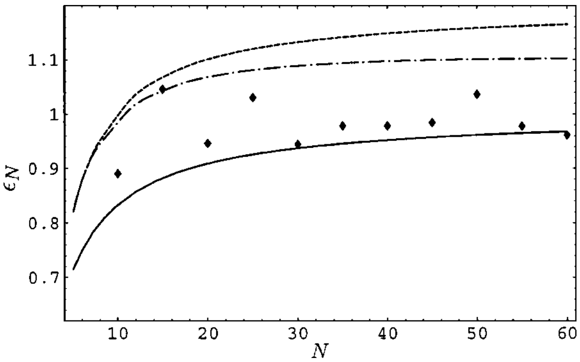

FIG. 3. Scaled error $\epsilon_N = N(1-F)$ in the 3D case for collective (solid line), tomographic OG (dot-dashed line), tomographic FG (dashed line), and greedy (diamonds) schemes.

$$\Delta^{\rm ISO} = \sum_{\chi=00\cdots 0}^{2^N-1} \left| \int dn\, \vec{n} \prod_{k=1}^{N} \frac{1+(-1)^{i_k}\vec{n}\cdot\vec{m}_k}{2} \right|. \qquad (4.9)$$

Analytical results for low N can be obtained using Eq. (4.6). For large N, numerical computations show that this "isotropic strategy" is indeed better than tomography (see [32] for explicit results), but the improvement is not very significant.

For 3D, however, one expects a completely different scenario. One can readily see from Fig. 3 that the tomographic OG line (dot-dashed) does *not* approach the collective one (solid). Indeed, it will be proved analytically in Sec. V and in Appendix B 2 that $\epsilon^{\rm OG} = 13/12 > \epsilon^{\rm COL} = 1$. Intuitively, more general fixed measurement schemes—such as a 3D version of the isotropic one we have just discussed—should perform significantly better.

There is a difficulty in implementing the isotropic scheme for 3D states since the notion of isotropic distribution of directions is not uniquely defined, which contrasts with the 2D case. A particularly interesting scheme that encapsulates this notion (at least for large enough N) and enables us to perform analytical computations consists of measurements along a set of random directions (RDs). With the same notation as in Eq. (4.9), we have

$$\Delta^{\rm RD} = \sum_{\chi=0\cdots 0}^{2^N-1} \int \prod_{k=1}^{N} dm_k \left| \int dn\, \vec{n} \frac{1+(-1)^{i_k}\vec{n}\cdot\vec{m}_k}{2} \right|. \qquad (4.10)$$

In Fig. 4 we show the scaled error ϵ_N obtained from numerical simulations for rather large N. One readily sees the improvement of the RD scheme over the tomographic OG. We will show in Sec. V that the former indeed attains the collective bound asymptotically. The RD scheme thus exemplifies that classical communication is not required for asymptotic optimality. A fit gives a value $\epsilon^{\rm RD} = 1.002 \pm 0.008$, which provides a numerical check of the analytical results of Sec. V below.

B. Adaptive measurements

In this subsection we discuss schemes that make use of classical communication. By that we mean schemes for

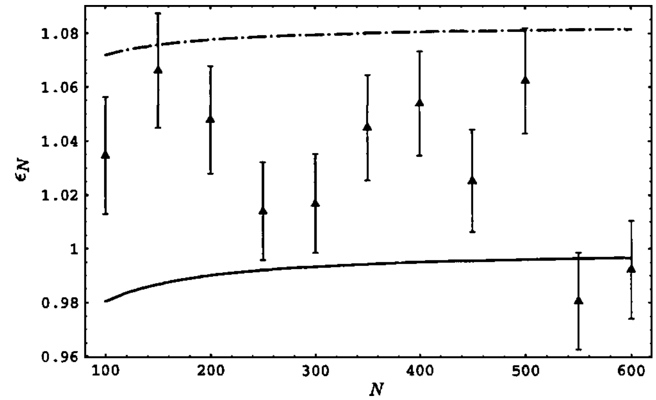

FIG. 4. Scaled error ϵ_N of the RD scheme (triangles) as compared to the OG (dot-dashed line) and the optimal collective scheme (solid line).

which we choose the local measurements according to the information gathered from previous outcomes. In principle, these schemes should be more efficient than those considered so far since a new resource is available. They can be understood from the Bayesian inference paradigm (see for instance [27] for a comprehensive treatise on Bayesian inference), where the probability distribution of the unknown parameters (in our case $\vec{n}$) is updated using the "evidence" collected after each measurement. Starting with a uniform distribution, the posterior probability distributions become more and more peaked around the true value. The convergence can be speeded up by choosing the measurement that best performs for the updated probability distribution *at each step*. This is essentially the point of view in [20,21]. Though in line with this spirit, the schemes considered in this section do not explicitly update their priors. Our approach aims to achieving optimality by directly solving the conditions of maximum averaged fidelity under some simplifying assumptions.

1. One-step adaptive

We first review a method put forward by Gill and Massar [19], which we call "one-step adaptive." It uses classical communication only once, which makes it very simple from both theoretical and practical perspectives.

The basic idea of the method is to split the measurements in two stages. In the first one, a small number of copies is used to obtain a rough estimate $\vec{M}_0$ of the state. In the second stage the remaining copies are used to refine the estimate by measuring on a plane orthogonal to $\vec{M}_0$. This strategy has a clear motivation from the information theory point of view. A measurement can be regarded as a query that one makes to a system. The most informative queries are those for which the prior probabilities of each outcome are the same. Measurements on the orthogonal plane to $\vec{M}_0$ have this feature.

To be more concrete, suppose we are given N copies of an unknown qubit state. Let N_0 stand for the number of copies used in the first stage and let $\bar{N} = N - N_0$ stand for the rest. In the 2D (3D) case, one measures $N_0/2(N_0/3)$ copies along two (three) fixed orthogonal directions and infers the guess $\vec{M}_0$. In the second stage, one measures $\bar{N}(\bar{N}/2)$ along $\vec{u}(\vec{u},\vec{v})$, which are chosen so that $\{\vec{u},\vec{v},\vec{M}_0\}$ is an orthonormal basis, and infers the final guess $\vec{M}$. The method turns out to be efficient if the number of copies used in each of the two stages is carefully chosen. Our numerical analysis reveals that the optimal choice, i.e., the one that yields the largest fidelity, corresponds to $N_0 \sim \sqrt{N}$. For other choices the scheme can even be less efficient than some fixed measurement schemes. For the benefit of the reader, we present a detailed discussion of the method and a derivation of the asymptotic limit in Appendix C. This method has also the interesting property that in spite of its simplicity, it suffices to show in a very straightforward way that local measurements attain the collective bounds in 2D and 3D.

2. Greedy scheme

We now move forward to more sophisticated schemes and discuss one that exploits classical communication much more efficiently. The idea behind it is to maximize the average fidelity at each single measurement step. It is called "greedy" because it does not take into account the total number of available copies; instead, it treats each copy as if it were the last one.

We first need to introduce some notation. Recall that the set of outcomes χ can be expressed as an N-digit binary number $\chi = i_N i_{N-1}\cdots i_2 i_1$ ($i_k = 0$, 1). Since we allow the kth measurement to depend on the list of previous outcomes $i_{k-1}i_{k-2}\cdots i_2 i_1 \equiv \chi_k$ (note that $\chi=\chi_N$), we have $\vec{m}(\chi_k)$ instead of $\vec{m}_k$. This is a compact notation where the length k of the string χ_k denotes the number of copies upon with we have already measured. The orthogonality of the von Neumann measurements is imposed by the constraint

$$\vec{m}(1\chi_{k-1}) = -\vec{m}(0\chi_{k-1}), \tag{4.11}$$

where $1\chi_{k-1}$ is the list of length k obtained by prepending 1 to the list χ_{k-1}, and similarly for $0\chi_{k-1}$. In general, the number of independent vectors for a given N is $(\sum_{k=1}^{N} 2^k)/2 = 2^N - 1$. For example, if $N=2$ there are three independent directions, which can be chosen as $\vec{m}(0)$, $\vec{m}(00)$, $\vec{m}(01)$, and the other three are obtained using Eq. (4.11). Since the first measurement can be chosen at will, this number is reduced to $2^N - 2$.

The general expression of the conditional probability thus reads

$$p_n(\chi) = \prod_{k=1}^{N} \frac{1+\vec{n}\cdot\vec{m}(\chi_k)}{2}, \tag{4.12}$$

and, as discussed in Sec. II, the OG gives

$$\Delta = \sum_{\chi=00\cdots 0}^{2^N-1} \left| \int dn\, \vec{n}\, p_n(\chi) \right|. \tag{4.13}$$

We could in principle attempt to maximize this expression with respect to *all* the independent variables, i.e., all independent $\{\vec{m}(\chi_k)\}$. However, the maximization process very quickly becomes extremely difficult. In the greedy scheme one takes a more modest approach: one maximizes at each step k. This enables us to find a compact algorithm for computing the fidelity, as we discuss below. Furthermore, we show in Appendix B that in this situation the optimal local measurement at each step is indeed of von Neumann type, i.e., any other POVM will perform worse.

Let us concentrate on the last step N of the greedy scheme. Suppose we have optimized the previous $N-1$ measurements and have obtained a string of outcomes χ_{N-1}. To ease the notation, let us denote the direction of the last measurement by $\vec{m}_N$, namely, $\vec{m}_N \equiv \vec{m}(0\chi_{N-1}) = -\vec{m}(1\chi_{N-1})$. We then need to maximize

$$d(\chi_N) = |\vec{V}(0\chi_{N-1})| + |\vec{V}(1\chi_{N-1})|. \tag{4.14}$$

Here

$$\vec{V}(i_N\chi_{N-1}) = \int dn\, \vec{n}\, p_n(\chi_{N-1}) \frac{1+(-1)^{i_N}\vec{n}\cdot\vec{m}_N}{2}, \tag{4.15}$$

or, equivalently,

$$\vec{V}(i_N\chi_{N-1}) = \frac{\vec{V}(\chi_{N-1}) + (-1)^{i_N}\mathsf{A}(\chi_{N-1})\vec{m}_N}{2}, \tag{4.16}$$

where A is the real positive symmetric matrix with elements

$$\mathsf{A}_{kl}(\chi_{N-1}) = \int dn\, n_k n_l\, p_n(\chi_{N-1}). \tag{4.17}$$

Therefore

$$d(\chi_N) = \sum_{i_N=0}^{1} \frac{|\vec{V}(\chi_{N-1}) + (-1)^{i_N}\mathsf{A}(\chi_{N-1})\vec{m}_N|}{2}. \tag{4.18}$$

Notice that for 2D states and fixed $d(\chi_N)$ the points $\vec{\mu} = \mathsf{A}\vec{m}_N$ lie on an ellipse with focus at $\pm\vec{V}$ (an ellipsoid for 3D states). In addition they satisfy the normalization constraint

$$\vec{\mu}\cdot(\mathsf{A}^{-2}\vec{\mu}) = 1, \tag{4.19}$$

which also defines an ellipse (ellipsoid in 3D) centered at the origin. As usual, optimality tells us that the maximum of $d(\chi_N)$ occurs at the points of tangency of the ellipses (ellipsoids). This provides a geometrical procedure for finding the optimal direction $\vec{m}_N$ and an algorithm for computing $|\vec{V}(\chi_N)|$.

We now proceed to obtain some explicit expressions for low N. We discuss only the 3D case, as the 2D case is completely analogous (numerical results for 2D states are shown in Figs. 1 and 2).

When we only have one copy of the state, $N=1$, the Bloch vector of the measurement can be chosen in any direction, say $\vec{e}_x$, i.e., $\vec{m}(0) = -\vec{m}(1) = \vec{e}_x$. The explicit computation of the vector $\vec{V}$ in Eq. (2.7) gives

$$\vec{V}(\chi_1) = \frac{1}{6}\vec{m}(\chi_1), \tag{4.20}$$

and $F=2/3$, as expected from Eq. (3.12) [or Eq. (3.3)].

The first nontrivial case is $N=2$. The matrix $\mathsf{A}(\chi_1)$ reads

$$\mathsf{A}_{kl}(\chi_1) = \frac{1}{6}\delta_{kl}, \quad \chi_1 = 0,1, \tag{4.21}$$

i.e., $\mathsf{A}(\chi_1)$ is independent of χ_1 and proportional to the identity. The maximum of (4.18) occurs for $\vec{m}_2 \perp \vec{e}_x$, so we choose $\vec{m}_2 = \vec{e}_y$, which means $\vec{m}(00) = \vec{m}(01) = \vec{e}_y$ [notice that in general these two vectors do not need to be equal, they are only required to be orthogonal to $\vec{m}(0)$]. Because of Eq. (4.11), we also have $\vec{m}(10) = \vec{m}(11) = -\vec{e}_y$. The OG reads

$$\vec{M}^{(2)}(\chi) = \frac{\vec{m}(\chi_2) + \vec{m}(\chi_1)}{\sqrt{2}}, \tag{4.22}$$

e.g., $\vec{M}^{(2)}(01) = [\vec{m}(01) + \vec{m}(1)]/\sqrt{2} = [\vec{m}(01) - \vec{m}(0)]/\sqrt{2} = [\vec{e}_y - \vec{e}_x]/\sqrt{2}$. We also obtain

$$|\vec{V}(\chi_2)| = \frac{\sqrt{2}}{12} \tag{4.23}$$

for all χ_2, which implies

$$F^{(2)} = \frac{3 + \sqrt{2}}{6}. \tag{4.24}$$

The case $N=3$ can be computed along the same lines. One can easily see that $\vec{m}(\chi_3)$ has to be perpendicular to $\vec{m}(\chi_2)$ and $\vec{m}(\chi_1)$. This shows that, up to $N=3$, the greedy approach does not use classical communication, i.e., the directions of the measuring devices are only required to be mutually orthogonal, independently of the outcomes. The optimal guess is straightforward generalization of (4.22):

$$\vec{M}^{(3)}(\chi) = \frac{\vec{m}(\chi_3) + \vec{m}(\chi_2) + \vec{m}(\chi_1)}{\sqrt{3}}, \tag{4.25}$$

and the fidelity reads

$$F^{(3)} = \frac{3 + \sqrt{3}}{6}. \tag{4.26}$$

The above results could have been anticipated. As already mentioned, the outcomes of a measurement on the plane orthogonal to the guess have roughly the same probability and are, hence, most informative. One can regard these measurements as corresponding to mutually unbiased observables, i.e., those for which the overlap between states of different basis (related to each observable) is constant [33]. Hence, there is no redundancy in the information about the state acquired from the different observables. This point of view also allows to extend the notion of (Bloch vector) orthogonality to states in spaces of arbitrary dimension.

The case $N=4$ is even more interesting, since four mutually orthogonal vectors cannot fit onto the Bloch sphere. We expect classical communication to start playing a role here. Indeed, the Bloch vectors $\vec{m}(\chi_4)$ do depend on the outcomes of previous measurements. They can be compactly written as

$$\vec{m}(\chi_4) = \frac{(-1)^{i_4}}{\sqrt{2}} \sum_{k=1}^{2} \vec{m}(\chi_k) \times \vec{m}(\chi_3). \tag{4.27}$$

Again, one can see that the vectors $\vec{m}(\chi_4)$ are orthogonal to the guess one would have made with the first three measurements. The fidelity in this case is

$$F^{(4)} = \frac{15 + \sqrt{91}}{30}. \tag{4.28}$$

For larger N, we have computed the fidelity of the greedy scheme by numerical simulations. In Fig. 3 (Figs. 1 and 2 for 2D states) we show the results for $10 \leqslant N \leqslant 60$ (diamonds). Notice that the greedy scheme is indeed better than fixed measurement schemes and approaches the collective bound (solid line) very fast.

Actually, the greedy scheme is the best we can use if the number of copies that will be available is not known *a priori*; obviously, the best one can do in these circumstances is to optimize at each step. However, if N is known, we have extra information that some efficient schemes could exploit to increase the fidelity. We next show that this is indeed the case.

3. General LOCC scheme

In the most general LOCC scheme one is allowed to optimize over all the Bloch vectors $\{m(\chi_k)\}$, thus taking into account the whole history of outcomes. Up to $N=3$ the results are the same as for the greedy scheme: orthogonal Bloch vectors for the measurements and no classical communication required. The results (4.24) and (4.26) are, therefore, the largest fidelity that can be attained by any LOCC scheme.

The most interesting features appear at $N=4$. Here there are 14 independent vectors, which can be grouped into two independent families of seven vectors. With such a large number of vectors an analytical calculation is too involved and we have resorted partially to a numerical optimization. The solution exhibits some interesting properties. First, one obtains that $\vec{m}(\chi_1) \perp \vec{m}(\chi_2)$, for all χ_1 and χ_2, as in the $N=2$ and 3 cases. Therefore one can choose $\vec{m}(\chi_1) = (-1)^{i_1}\vec{e}_x$ and $\vec{m}(\chi_2) = (-1)^{i_2}\vec{e}_y$. Only for the third and fourth measurements has one really to take different choices in accordance to the sequence of the preceding outcomes. The Bloch vectors of the third measurement can be parametrized by a single angle α as

$$(-1)^{i_3}\vec{m}(\chi_3) = \cos\alpha\, \vec{u}_1(\chi_2) + \sin\alpha\, \vec{v}_1(\chi_2), \tag{4.29}$$

where

$$\vec{u}_1(\chi_2) = \vec{m}(\chi_1) \times \vec{m}(\chi_2),$$

$$\vec{v}_1(\chi_2) = \vec{u}_1(\chi_2) \times \vec{s}(\chi_2),$$

$$\vec{s}(\chi_2) = \frac{\vec{m}(\chi_2) + \vec{m}(\chi_1)}{\sqrt{2}}. \tag{4.30}$$

Notice that, rather unexpectedly, $\vec{m}(\chi_1)$, $\vec{m}(\chi_2)$, and $\vec{m}(\chi_3)$ are not mutually orthogonal. The optimal value of this angle is $\alpha = 0.502$. Although we cannot give any insight as to why this value is optimal, in agreement with our intuition one sees that $\vec{m}(\chi_3) \perp \vec{M}(\chi_2)$, i.e., the third measurement probes the plane orthogonal to the Bloch vector one would guess from the first two outcomes [see Eq. (4.22)]. The vectors of the fourth measurement can be parametrized by two angles β and γ as

$$(-1)^{i_4}\vec{m}(\chi_4) = \cos\gamma\,\vec{u}_2(\chi_3) + \sin\gamma\,\vec{v}_2(\chi_3), \qquad (4.31)$$

where

$$\vec{u}_2(\chi_3) = \vec{s}(\chi_2) \times \vec{m}(\chi_3),$$

$$\vec{v}_2(\chi_3) = \cos\beta\,\vec{m}(\chi_3) - \sin\beta\,\vec{s}(\chi_2). \qquad (4.32)$$

The optimal values of these angles are $\beta=0.584$, $\gamma=0.538$, and the corresponding fidelity is $F^{(4)}_{\text{general}}=0.8206$. This is just 1.5% lower than the absolute bound $5/6=0.8333$ attained with collective measurement, Eq. (3.12). Note that this value is slightly larger than the fidelity obtained with the greedy scheme, $F^{(4)}_{\text{general}} > F^{(4)}_{\text{greedy}} = (15+\sqrt{91})/30 \approx 0.8180$. The extra information consisting of the number of available copies has indeed been used to attain a larger fidelity. We conclude that for $N>3$, it pays to relax optimality at each step, and greedy schemes [20,21] are thus not optimal. We would like to remark that if, for some reason, some copies are lost or cannot be measured, the most general scheme will not be optimal, since it has been designed for a specific number of copies. We have also computed the values of the maximal LOCC fidelities for $N=5,6$: $F^{(5)}_{\text{general}}=0.8450$ and $F^{(6)}_{\text{general}}=0.8637$. Beyond $N=6$ the small differences between this and the greedy scheme become negligible.

V. LOCAL SCHEMES IN THE ASYMPTOTIC LIMIT

Any acceptable scheme, such as those considered in this paper, achieves a unit fidelity in the limit $N\to\infty$. The subleading term of the asymptotic expansion of the fidelity (typically of order N^{-1}) carries nontrivial information, as anticipated in the discussion after Eq. (4.8). It enables us, e.g., to compare different schemes independently of the number of copies. If two schemes have the same asymptotic fidelity (the same subleading term), it is justified to say that they have the same efficiency, and conversely. In addition, this term is important in statistics since it is related to the variance of certain type of estimators to which many powerful techniques apply. In this section we will compute such asymptotic fidelities. We will show that, asymptotically, classical communication is not needed to attain the absolute upper bound given by the maximum fidelity of the most general collective measurements. Some of the results that we present below were obtained by two of the authors by explicit computation in [22]. Here we will use a statistical approach that relates the Fisher information I [34] with the average fidelity F and uses the Cramér-Rao bound. This approach will greatly simplify our earlier derivations. A brief introduction for nonpractitioners is in Appendix B, where some technical details are also included.

The main theorem that we need in this section is in Eq. (B7), whose no-frill version reads

$$F(\eta) = 1 + \frac{1}{2N}\text{tr}\frac{\mathsf{H}(\eta)}{I(\eta)} + \ldots, \qquad (5.1)$$

where H is the Hessian matrix of $f_n(\chi)$ as a function of the state parameters, which we refer generically as η, and I is the Fisher information matrix, defined in Eq. (B5). The theorem applies to the "pointwise" fidelity $F(\eta)$, which is the average of $f_n(\chi)$ over all possible outcomes, Eq. (B3). It provides an upper bound that is attained by the maximum likelihood estimator (MLE), defined after Eq. (B5), and by the OG.

Let us start with 2D states. This case is rather simple because such states have just one parameter and the Fisher information is a single number. Moreover, *any* von Neumann measurement whose vector lies on the equator of the Bloch sphere performed on a 2D system has $I=1$, as can be checked by plugging the corresponding probability

$$p_\theta(\pm 1) = \frac{1 \pm \cos(\theta - \theta_m)}{2} \qquad (5.2)$$

into Eq. (B5). In this equation θ_m is the polar angle of $\vec{m}$, the direction along which the von Neumann measurement is performed, and θ is the polar angle of $\vec{n}$. Therefore, in 2D the Fisher information for a set of N measurements (identical or not) is $I^{(N)}=N$.

Since the Hessian is $\mathsf{H}=-1/2$, *any* sensible local measurement scheme on 2D states will yield

$$F(\theta) = 1 - \frac{1}{4N} + \cdots. \qquad (5.3)$$

Note that this fidelity is independent of θ, so it coincides with the average fidelity $F=\int d\theta\, F(\theta)/(2\pi) = 1-1/(4N)$. We further note that it coincides with the collective bound Eq. (3.7), as anticipated in Sec. IV A.

We next turn to the 3D case, which is more involved. The results shown in Fig. 3 hint that the tomographic scheme with the OG (dot-dashed line) does not attain the collective bound. Indeed, a straightforward calculation (details are given in Appendix B 2) gives

$$F = 1 - \frac{13}{12N} + \cdots, \qquad (5.4)$$

which is less than $F=1-1/N$, Eq. (3.14).

At this point, the question arises whether classical communication is necessary to attain the collective bound. We next show that this is not the case by considering the so-called RD scheme. Recall that in this scheme measurements are performed along RDs chosen from an isotropic distribution. This is equivalent to performing a covariant (continuous) POVM on each one of the copies separately. Here, we stick to the RD picture and regard each individual measurement as von Neumann's *and* a classical ancilla, e.g., a "roulette," that gives us the direction of the measurement. From this point of view, the outcome parameters are given by $\chi = (\xi, (u\equiv\cos\vartheta), \varphi)$, where ϑ and φ are the azimuthal and polar angles of the direction $\vec{m}(u,\varphi)$ of the measurement, and $\xi=\pm 1$ is the corresponding outcome. Using Eq. (5.1) [or Eq. (B7)] we obtain (see Appendix B 3 for details)

$$F^{\mathrm{RD}} = 1 - \frac{1}{N} + \cdots . \tag{5.5}$$

We conclude that asymptotically classical communication is not required to saturate the collective bound: a measurement scheme based on a set of RDs does the job.

VI. SUMMARY AND CONCLUSIONS

We have presented a self-contained and detailed study of several estimation schemes when a number N of identical copies of a qubit state is available. We have used the fidelity as a figure of merit and presented a general Bayesian framework which has enabled us to find the optimal schemes not only in the asymptotic limit but for *any* number of copies. We have considered two interesting situations: that of a completely unknown qubit state (3D case), and that of a qubit lying on the equator of the Bloch sphere (2D case). For completeness, we have reviewed the optimal measurements and maximum fidelities for the most general collective strategies. However, this paper focused on measurements that can be implemented in a laboratory with technology available nowadays: local von Neumann measurements. Special emphasis has been put on situations where a finite number of copies are available, an aspect of state estimation that has not been extensively covered in the literature.

In the 2D case we have shown that, quite surprisingly, the most basic tomographic scheme, i.e., measurements along two fixed orthogonal directions with adequate data processing (the OG), gives already a fidelity that is asymptotically equal to the collective bound.

For the 3D states, tomography, i.e., measurements along three fixed orthogonal directions, fails to give the asymptotic collective bound, even with the best data processing. The main reason of this failure is that the Bloch sphere is not explored thoroughly. We have considered an extension that is asymptotically isotropic: a series of von Neumann measurements along random directions. We have proved that this scheme, which does not make use of classical communication, does saturate the collective bound. This illustrates that in the large-N limit an estimation procedure based on local measurements without classical communication does perform as well as the most efficient and sophisticated collective schemes.

We have also discussed local schemes with classical communication, i.e., schemes in which the measurements are devised in such a way that they take into account previous outcomes. We have studied in detail the one-step adaptive scheme of Gill and Massar [19]. The economy of resources in this scheme may raise doubts about its efficiency. In Appendix C we give a simple proof that for large N it indeed attains the collective bounds for both 2D and 3D.

We have also studied strategies that make a more intensive use of classical communication. In the greedy scheme, optimization is performed at each measurement step [20,21]. This scheme is the best approach one can take if the actual number of available copies is not known. We have given a geometrical condition for sequentially finding the optimal measurements and have proved that they have to be of von Neumann type (see Appendix D), i.e., no general local POVM will perform better in this context. We have illustrated the performance of the method with numerical simulations and have shown that the behavior of the optimal collective scheme is reached for very low values of N. This occurs for N as low as 20 in 2D and slightly above, 45, in 3D.

In the most general scheme we see that up to $N=3$ (2 in 2D) there is no need for classical communication: the optimal measurements correspond to a set of mutually unbiased observables. For larger N, the knowledge of the actual value of N provides an extra information that translates into an increase of the fidelity. From the practical point of view, however, this difference is negligible already at the level of a few copies ($N \gtrsim 6$).

Our approach may be extended to other situations. For instance, the problem of estimating (qubit) mixed states, which is much more involved, can be tackled along the lines described here [23]. It would also be interesting to consider qudits and check whether a set of mutually unbiased observables provides the optimal local estimation scheme when the number of copies coincides with the number of independent variables that parametrize the qudit state.

ACKNOWLEDGMENTS

It is a pleasure to thank M. Baig for his collaboration at early stages of this work. We also thank R. Gill and M. Ballester for useful discussions. We acknowledge financial support from Spanish Ministry of Science and Technology Project No. BFM2002-02588, CIRIT Project No. SGR-00185, and QUPRODIS working group EEC Contract No. IST-2001-38877.

APPENDIX A

Here we give a derivation of the results collected in Sec. III. The 2D and 3D cases are treated on the same footing, which gives a homogeneous picture of the problem under consideration.

1. 2D states

For 2D states Eq. (2.9) reads

$$\Delta = \sum_{\chi} \left| \int \frac{d\theta}{2\pi} \vec{n} \, \mathrm{tr}[\rho_n O(\chi)] \right| , \tag{A1}$$

where we can write ρ_n as in Eq. (3.1). In the standard basis $|jm\rangle \equiv |m\rangle, U(\theta)$ is diagonal and we have

$$\rho(\theta) = \sum_{m\ n} e^{i(m-n)\theta} (\rho_0)_{mn} |m\rangle\langle n| , \tag{A2}$$

where ρ_0 is defined in Eq. (3.2). We can now readily compute Eq. (A1) to obtain

$$\Delta = \sum_{\chi} \left| \sum_{mn} \int \frac{d\theta}{2\pi} e^{i\theta} e^{i(m-n)\theta} (\rho_0)_{mn} [O(\chi)]_{nm} \right|$$
$$= \sum_{\chi} \left| \sum_{m=-J}^{J-1} (\rho_0)_{mm+1} [O(\chi)]_{m+1m} \right|, \tag{A3}$$

where $[O(\chi)]_{mn} \equiv \langle m|O(\chi)|n\rangle$. The following inequalities give an upper bound for Δ:

$$\Delta \leqslant \sum_{\chi} \sum_{m=-J}^{J-1} |(\rho_0)_{mm+1}[O(\chi)]_{m+1m}|$$
$$\leqslant \sum_{m=-J}^{J-1} |(\rho_0)_{mm+1}| \sum_{\chi} |[O(\chi)]_{m+1m}| \leqslant \sum_{m=-J}^{J-1} |(\rho_0)_{mm+1}|, \tag{A4}$$

where in the last step we have used that

$$\sum_{\chi} |[O(\chi)]_{m+1\ m}| \leqslant 1, \tag{A5}$$

as follows from positivity and Eq. (2.4). More precisely, positivity implies

$$[O(\chi)]_{mm}[O(\chi)]_{m+1\ m+1} \geqslant |[O(\chi)]_{m\ m+1}|^2, \tag{A6}$$

and the Schwarz inequality yields

$$\sum_{\chi} |[O(\chi)]_{m\ m+1}| \leqslant \sqrt{\sum_{\chi} |O(\chi)]_{mm}} \sqrt{\sum_{\chi} [O(\chi)]_{m+1\ m+1}} = 1. \tag{A7}$$

Recalling Eqs. (3.2) and (A4) one has [8]

$$\Delta \leqslant \frac{1}{2^N} \sum_{m=-J}^{J-1} \sqrt{\binom{N}{J+m}\binom{N}{J+m+1}}, \tag{A8}$$

where we have used that

$$|JJ\rangle_x = |\vec{x}\rangle^{\otimes N} = \frac{1}{2^J} \sum_{m=-J}^{J} \binom{N}{J+m} |Jm\rangle \tag{A9}$$

(recall that $J \equiv N/2$). The inequality (A8) can also be written as in (3.3).

We next show that there are POVMs that attain this bound. To saturate the first inequality in (A4) the phase of $[O(\chi)]_{m\ m+1}$ must be independent of m. This is ensured if this phase is a function of $m-n$. Similarly, a set of positive operators for which $|O(\chi)_{mn}|$=const for all χ, m, and n will certainly saturate the remaining inequalities in (A4). In particular, the covariant (continuous) POVM, whose elements are given by

$$[O(\phi)]_{mn} = e^{i(m-n)\phi}, \tag{A10}$$

satisfies all the requirements. Note that we have labeled the outcomes by a rotation angle ϕ, which plays the role of χ. Hence, condition (2.4) becomes $\int O(\phi) d\phi/(2\pi) = \mathbb{1}$, which certainly holds for Eq. (A10). These are rank-1 operators, and can also be written as in (3.4).

In the asymptotic limit the fidelity can be obtained in terms of the moments of a binomial distribution $\mathrm{Bin}(n,p)$ with parameters $n=N$ and $p=1/2$. We simply need to expand (3.3) in powers of m, i.e., around $\langle m\rangle = 0$, to obtain

$$\Delta^{\mathrm{COL}} = \frac{1}{2^N} \sum_m \binom{N}{J+m} \times \left[1 - \frac{2m}{N} + \left(\frac{2m^2}{N^2} - \frac{1}{N} \right) + O(1/N^{3/2}) \right] \tag{A11}$$

(notice that the sum over m is shifted by J with respect to the usual binomial distribution). The moments are well known to be $\langle 1\rangle = 1$, $\langle m\rangle = 0$, and $\langle m^2\rangle = N/4$. The latter shows that m has "dimensions" of $\sqrt{N}$, which helps to organize the expansion in powers of $1/N$. We finally obtain Eq. (3.6).

2. 3D states

It is convenient to define an operator $\Omega(\chi)$ in such a way that

$$O(\chi) = U[\vec{M}(\chi)]\Omega(\chi)U^{\dagger}[\vec{M}(\chi)], \tag{A12}$$

where $\vec{M}(\chi)$ is given by Eq. (2.6). Taking into account that Δ is rotationally invariant one obtains

$$\Delta = \sum_{\chi} \left| \int dn\, n_z \mathrm{tr}[\rho_n \Omega(\chi)] \right|. \tag{A13}$$

We readily see that $n_z = \cos\theta = \mathfrak{D}^{(1)}_{00}(\vec{n})$, where the rotation matrices $\mathfrak{D}^{(j)}_{mm'}$ are defined in the standard way, $\mathfrak{D}^{(j)}_{mm'}(\vec{n}) = \langle jm|U(\vec{n})|jm'\rangle$. We then have

$$\Delta = \sum_{\chi} \left| \sum_{mm'} \int dn\, \mathfrak{D}^{(1)}_{00}(\vec{n})(\rho_n)_{mm'} \Omega_{m'm}(\chi) \right|$$
$$= \sum_m \left[\sum_{\chi} \Omega_{mm}(\chi) \right] \int dn\, \mathfrak{D}^{(1)}_{00}(\vec{n})(\rho_n)_{mm}, \tag{A14}$$

where in the second equality we have used that

$$\int dn\, \mathfrak{D}^{(1)}_{00}(\vec{n})(\rho_n)_{mm'} = \delta_{mm'} \int dn\, \mathfrak{D}^{(1)}_{00}(\vec{n})(\rho_n)_{mm}, \tag{A15}$$

as follows from Schur's lemma after realizing that the left-hand side of Eq. (A15) commutes with $U(\theta)$, the unitary transformations defined right after Eq. (3.2). Recall that these transformations, which are a U(1) subgroup of SU(2), have only one-dimensional irreducible representations, labeled by the magnetic quantum number m, thus yielding relation (A15). In Eq. (A14) we have removed the absolute value as all terms are positive (see below). Tracing (A12) one obtains $\Sigma_{\chi} \mathrm{tr}\, \Omega(\chi) = \Sigma_{\chi} \mathrm{tr}\, O(\chi) = d_J$. Therefore

$$\Delta \leqslant d_J \max_m \int dn\, \mathfrak{D}^{(1)}_{00}(\vec{n})(\rho_n)_{mm} = \max_m \langle 10;Jm|Jm\rangle^2 = \frac{J}{J+1}, \tag{A16}$$

where in the second equality we have used that $\rho_{mm} = \mathfrak{D}^{(J)}_{mJ}(\vec{n})\mathfrak{D}^{(J)*}_{mJ}(\vec{n})$ and the well-known orthogonality relations of the SU(2) irreducible representations [35]. This is the inequality (3.11) in the main text.

Let us finally give a POVM that saturates all the inequalities. The maximum of the Clebsch-Gordan $\langle 10;Jm|Jm\rangle$ in Eq. (A16) occurs at $m=J$. Hence, to attain the bound we need to choose

$$\Omega_{mm}(\chi) = c_\chi \delta_{Jm}, \quad \sum_\chi c_\chi = d_J, \tag{A17}$$

where the coefficients c_χ are positive. This leads straightforwardly to the optimal continuous POVM in Eq. (3.13).

APPENDIX B

1. The Cramér-Rao bound: A glossary

We label the independent state parameters by the symbol η. This symbol will refer to the two angles θ, ϕ for 3D states: $\eta \equiv (\theta, \phi)$; and the polar angle θ for 2D states: $\eta \equiv \theta$.

Assume that under a sensible measurement and estimation scheme (we mean by that a scheme that leads to a perfect determination of the state when N goes to infinity, i.e., $\lim_{N\to\infty} F=1$) the estimated state is close to the signal state, that is, their respective parameters $\hat{\eta}(\chi)$ and η differ by a small amount. In this section, a caret will always refer to estimated parameters, the fidelity $f_n(\chi)$, Eq. (2.2), will be denoted by $f_\eta(\hat{\eta})$, and similarly, the probability $p_n(\chi)$ will be written as $p_\eta(\chi)$. Note that the guessed parameters $\hat{\eta}(\chi)$ are based on a particular outcome χ. This dependence will be implicitly understood when no confusion arises.

The fidelity can be approximated by the first terms of its series expansion

$$f_\eta(\hat{\eta}) \approx 1 + \frac{1}{2} \left.\frac{\partial^2 f}{\partial \hat{\eta}_i \, \partial \hat{\eta}_j}\right|_{\hat{\eta}=\eta} (\hat{\eta}_i - \eta_i)(\hat{\eta}_j - \eta_j), \tag{B1}$$

where we have used that $f_\eta(\eta)=1$ and $\partial f_\eta / \partial \hat{\eta}|_{\hat{\eta}=\eta}=0$. Averaging over all possible outcomes, we have

$$F(\eta) \approx 1 + \frac{1}{2}\mathrm{tr}[\mathsf{H}(\eta)\mathsf{V}(\eta)], \tag{B2}$$

where

$$F(\eta) \equiv \sum_\chi p_\eta(\chi) f_\eta[\hat{\eta}(\chi)], \tag{B3}$$

is the "pointwise" fidelity, $\mathsf{H}(\eta)$ is the Hessian matrix of $f_\eta(\hat{\eta})$ at $\hat{\eta}=\eta$, and $\mathsf{V}(\eta)$ is the variance matrix, with elements $\mathsf{V}_{ij}(\eta)=\Sigma_\chi p_\eta(\chi)(\hat{\eta}_i - \eta_i)(\hat{\eta}_j - \eta_j)$.

It is well known that the variance of an unbiased estimator is bounded by

$$\mathsf{V}(\eta) \geqslant \frac{1}{I(\eta)}, \tag{B4}$$

the so called Cramér-Rao bound [19,25,36], where the Fisher information matrix $I(\eta)$ is defined as

$$I_{ij}(\eta) = \sum_\chi p_\eta(\chi) \frac{\partial \ln p_\eta(\chi)}{\partial \eta_i} \frac{\partial \ln p_\eta(\chi)}{\partial \eta_j}. \tag{B5}$$

The conditional probability $p_\eta(\chi)$ regarded as a function of η is called the likelihood function $\mathcal{L}(\eta)=p_\eta(\chi)$. It is also well known that the bound (B4) is attained by the MLE [37], defined as $\hat{\eta}^{\mathrm{MLE}}=\mathrm{argmax}\,\mathcal{L}(\eta)$. Hence this bound is tight.

A link between the Fisher information and the fidelity is obtained by combining Eqs. (B2) and (B4), and noticing that $\mathsf{H}(\eta)$ is negative definite. We thus have

$$F(\eta) \leqslant 1 + \frac{1}{2}\mathrm{tr}\frac{\mathsf{H}(\eta)}{I(\eta)} \tag{B6}$$

to leading order and for any unbiased estimation scheme.

The Fisher information is additive. This means that if $p_\eta^{(2)}(\chi,\chi')=p_\eta(\chi)p'_\eta(\chi')$, which happens when we perform two measurements [say, $\{O(\chi)\}$ and $\{O'(\chi')\}$] on two identical states, the Fisher information of the combined measurement is simply $I^{(2)}(\eta)=I(\eta)+I'(\eta)$. In particular, for N *identical* measurements, we have $I^{(N)}(\eta)=NI(\eta)$.

Finally, since the OG is a better estimator, and it is asymptotically unbiased, we must have

$$F^{\mathrm{OG}}(\eta) = F^{\mathrm{MLE}}(\eta) = 1 + \frac{1}{2N}\mathrm{tr}\frac{\mathsf{H}(\eta)}{I(\eta)} \tag{B7}$$

to leading order, where the fidelities refer to an estimation scheme consisting of N identical measurements. We use Eq. (B7) to compute the asymptotic limits of the fixed measurement schemes discussed in Sec. IV A (details are given below).

2. Tomography of 3D states

Consider a scheme that consists in repeating $\mathcal{N}$ times the following: take three copies of the state and perform a measurement along $\vec{e}_x$ on the first copy, along $\vec{e}_y$ on the second copy, and along $\vec{e}_z$ on the third copy (recall that $N=3\mathcal{N}$). These three von Neumann measurements can be regarded as a single measurement with 2^3 possible outcomes labeled by $\chi=(\chi_1,\chi_2,\chi_3)$, where $\chi_j=\pm 1$. The probability of obtaining an outcome χ is

$$p_\eta(\chi) = \prod_{j=x,y,z} \frac{1}{2}(1 + \chi_j \vec{n}\cdot\vec{e}_j). \tag{B8}$$

The Fisher information matrix $I(\theta,\phi)$ of this elementary measurement is obtained by substituting Eq. (B8) in Eq. (B5). Note that the Fisher information of this scheme as a whole is $I^{(\mathcal{N})}(\theta,\phi)=\mathcal{N}I(\theta,\phi)$.

With this, we obtain

$$\mathrm{tr}\frac{\mathsf{H}(\theta,\phi)}{I(\theta,\phi)} = \frac{3}{4}\frac{1+6\cos^2\theta+\cos^4\theta-\sin^4\theta\cos 4\phi}{1+7\cos^2\theta-\sin^2\theta\cos 4\phi}. \tag{B9}$$

Integrating over the isotropic prior probability dn, Eq. (3.10), we obtain Eq. (5.4).

3. Random direction scheme

We parametrize the state by $\eta=((v\equiv\cos\theta),\phi)$. Since this strategy is isotropic, the pointwise fidelity $F(\eta)$ is independent of η, and we conveniently choose $\eta=((v=0),0)=\underline{0}$. By the same argument, no average over η will be needed: $F=F(\eta)$. The probability is given by

$$p_\eta(\chi)=\frac{1+\xi\,\vec{n}\cdot\vec{m}(u,\varphi)}{2}, \tag{B10}$$

and the Fisher information reads

$$I_{ij}(\eta)=\sum_{\xi=\pm1}\int\frac{du\,d\varphi}{4\pi}p_\eta(\chi)\frac{\partial\ln p_\eta(\chi)}{\partial\eta_i}\frac{\partial\ln p_\eta(\chi)}{\partial\eta_j}. \tag{B11}$$

The diagonal elements read

$$I_{vv}(\underline{0})=\frac{1}{8\pi}\sum_{\xi=\pm1}\int\frac{u^2du\,d\varphi}{1+\xi\sqrt{1-u^2}\cos\varphi}=\frac{1}{2}, \tag{B12}$$

$$I_{\phi\phi}(\underline{0})=\frac{1}{8\pi}\sum_{\xi=\pm1}\int\frac{(1-u^2)\sin^2\varphi\,du\,d\varphi}{1+\xi\sqrt{1-u^2}\cos\varphi}=\frac{1}{2}. \tag{B13}$$

As for the off-diagonal elements, a straightforward calculation gives $I_{v\phi}(\underline{0})=I_{\phi v}(\underline{0})=0$, as one could expect, since gaining information on v does not provide information on ϕ, and vice versa. Plugging these results and the Hessian of the fidelity, which is $\mathsf{H}_{ij}(\underline{0})=-\delta_{ij}/2$, in Eq. (B7) we finally obtain Eq. (5.5).

APPENDIX C

In this appendix we review the "one-step adaptive" (or Gill-Massar) scheme and give a straightforward and comprehensive proof that it saturates the collective bound for large N. We consider only the 3D case, as the simpler 2D case can be worked out along the same lines.

First stage. One performs $N_0=N^a(0<a<1)$ measurements with a sensible estimator, in the sense of Sec. V, and obtains an estimation $\vec{M}_0$ with a fidelity F_0:

$$F_0=\sum_{\chi_0}\int dn\frac{1+\vec{n}\cdot\vec{M}_0(\chi_0)}{2}p_n(\chi_0), \tag{C1}$$

where χ_0 stands for the list of outcomes obtained in this first stage.

Second stage. At this point we use the FG on the remaining $\overline{N}\equiv2\mathcal{N}\equiv N-N_0$ copies by measuring along *two* perpendiculars directions $\vec{u}$ and $\vec{v}$ on the plane orthogonal to $\vec{M}_0$. In this basis the final guess can be written as

$$\vec{M}=\vec{M}_0(\chi_0)\cos\omega+(\vec{u}\cos\tau+\vec{v}\sin\tau)\sin\omega. \tag{C2}$$

This parametrization ensures that $\vec{M}$ is unitary. The angles ω and τ depend on the outcomes of this second stage, which are the frequencies $\alpha_i\mathcal{N},(1-\alpha_i)\mathcal{N}$ (for $i=u,v$). The probabilities are given by $p_n(\alpha)$ in Eq. (4.2), with $n_u=\vec{n}\cdot\vec{u}$ and $n_v=\vec{n}\cdot\vec{v}$. Since we measure on the plane orthogonal to $\vec{M}_0$, the two outcomes of each measurement have roughly the same probability, $\alpha_i\approx1/2$, and they are most informative. It is convenient to define the two-dimensional vector $\vec{r}$, with components

$$r_i\equiv2\alpha_i-1,\quad i=u,v, \tag{C3}$$

which, on average, is close to $\vec{0}$. This vector gives an estimation of the projection of the signal Bloch vector $\vec{n}$ on the measurement plane (uv plane). Hence, ω is expected to be small ($\vec{M}_0\approx\vec{M}$) and we make the ansatz

$$\omega=\lambda\sqrt{r_u^2+r_v^2},\quad\tan\tau=\frac{r_v}{r_u}, \tag{C4}$$

where the positive parameter λ will be determined later.

The final fidelity for a signal state $\vec{n}$ and outcomes $(\chi_0,\vec{r})$ is

$$f_n(\chi_0,\vec{r})=\frac{1+\vec{n}\cdot\vec{M}(\chi_0,\vec{r})}{2}, \tag{C5}$$

and the average fidelity F reads

$$F=\sum_{\chi_0,\vec{r}}\int dn\frac{1+\vec{n}\cdot\vec{M}(\chi_0,\vec{r})}{2}p_n(\chi_0)p_n(\vec{r}|\chi_0). \tag{C6}$$

Notice that the probability of obtaining the outcome $\vec{r}$, namely, $p_n(\vec{r}|\chi_0)[\equiv p_n(\alpha)$ in Eq. (4.2) with $i=u,v]$, is conditioned on χ_0 through the dependence of the second-stage measurements on $\vec{M}_0(\chi_0)$.

Since we will compute different averages over χ_0, $\vec{r}=(r_u,r_v)$, and $\vec{n}$, it is convenient to introduce the following notation:

$$\langle f\rangle_0=\sum_{\chi_0}f_n(\chi_0,\vec{r})p_n(\chi_0), \tag{C7}$$

$$\langle f\rangle_r=\sum_{\vec{r}}f_n(\chi_0,\vec{r})p_n(\vec{r}|\chi_0), \tag{C8}$$

$$\langle f\rangle_n=\int dn\,f_n(\chi_0,\vec{r}), \tag{C9}$$

and similarly for averages of other functions of χ_0, $\vec{r}$, and $\vec{n}$. We will denote composite averaging by simply combining subscripts (i.e., $\langle\langle F\rangle_r\rangle_0\equiv\langle F\rangle_{r,0}$). Therefore, we write

$$F\equiv\langle f\rangle_{r,0,n}\equiv\langle f\rangle. \tag{C10}$$

Since $F=(1+\Delta)/2$, we have

$$\Delta=\langle\vec{n}\cdot\vec{M}\rangle. \tag{C11}$$

In the expansions that we perform below, we keep only the terms that contribute to the fidelity up to order $1/N$. Recalling that ω is expected to be small, it follows that

$$\cos\omega=1-\frac{\lambda^2}{2}[r_u^2+r_v^2], \tag{C12}$$

$$\sin\omega\cos\tau=\lambda r_u, \quad (C13)$$

$$\sin\omega\sin\tau=\lambda r_v, \quad (C14)$$

to leading order. Therefore, the expectation value in Eq. (C11) can be written as

$$\Delta=\left\langle\left(1-\frac{\lambda^2}{2}\langle r_u^2+r_v^2\rangle_r\right)\vec{n}\cdot\vec{M}_0\right\rangle_{0,n}+\lambda\langle\langle r_u\rangle_r n_u+\langle r_v\rangle_r n_v\rangle_{0,n}. \quad (C15)$$

Since r_u, r_v (or equivalently α_u, α_v) are binomially distributed, one readily sees that

$$\langle r_i\rangle_r=n_i, \quad (C16)$$

$$\langle r_i^2\rangle_r=n_i^2+\frac{1-n_i^2}{\mathcal{N}}. \quad (C17)$$

We further recall that $\vec{n}$ is unitary and that $\{\vec{M}_0,\vec{u},\vec{v}\}$ is an orthonormal basis, hence $n_u^2+n_v^2=1-(\vec{n}\cdot\vec{M}_0)^2$, and Eq. (C15) can be cast as

$$\Delta=\lambda+\left[1-\frac{\lambda^2}{2}\left(1+\frac{1}{\mathcal{N}}\right)\right]\langle\vec{n}\cdot\vec{M}_0\rangle_{0,n}-\lambda\langle(\vec{n}\cdot\vec{M}_0)^2\rangle_{0,n}+\frac{\lambda^2}{2}\left(1-\frac{1}{\mathcal{N}}\right)\langle(\vec{n}\cdot\vec{M}_0)^3\rangle_{0,n}. \quad (C18)$$

To compute the moments $\langle(\vec{n}\cdot\vec{M}_0)^q\rangle_{0,n}$, we consider the angle δ between $\vec{n}$ and $\vec{M}_0$, which is also expected to be small. We have

$$2F_0-1=\langle\vec{n}\cdot\vec{M}_0\rangle_{0,n}=\langle\cos\delta\rangle_{0,n}\simeq 1-\frac{\langle\delta^2\rangle_{0,n}}{2}, \quad (C19)$$

where we have used Eq. (C1). Therefore

$$\langle(\vec{n}\cdot\vec{M}_0)^q\rangle_{0,n}=\langle\cos^q\delta\rangle_{0,n}\simeq 1-\frac{q}{2}\langle\delta^2\rangle_{0,n}=1-2q(1-F_0). \quad (C20)$$

Now we plug this result back into Eq. (C18) to obtain

$$F=\langle f\rangle=1-(1-\lambda)^2(1-F_0)-\frac{1-4(1-F_0)}{2\mathcal{N}}\lambda^2. \quad (C21)$$

Since the term $(1-\lambda)^2(1-F_0)$ is always positive, the maximum fidelity is obtained with the choice $\lambda=1$, and we are left with

$$F=1-\frac{1-4(1-F_0)}{2\mathcal{N}}. \quad (C22)$$

Since the first estimation is asymptotically unbiased, $1-F_0$ vanishes for large N_0 (i.e., for large N) and

$$F\simeq 1-\frac{1}{2\mathcal{N}}. \quad (C23)$$

Recalling that $\mathcal{N}=(N-N^a)/2$, we finally have

$$F=1-\frac{1}{N}+\cdots. \quad (C24)$$

This concludes the proof.

APPENDIX D

In this appendix we prove that in the greedy scheme the optimal individual measurements on each copy are of von Neumann type. We sketch the proof for 2D states. The 3D case can be worked out along the same lines.

The history of outcomes will be denoted, as usual, by χ. Notice that here we consider general local measurements (local POVMs) with R outcomes, where R is possibly larger than 2. Therefore χ is an N-digit integer number in base R: $\chi=i_N i_{N-1}\cdots i_1(i_k=0,1,\ldots,R-1)$. As in Sec. IV B we use the notation $\chi_k=i_k i_{k-1}\cdots i_1$. A measurement on the kth copy is defined by a set of non-negative rank-1 operators $\{O(\chi_k)\}_{i_k=0}^{R-1}=\{O(i_k\chi_{k-1})\,|\,i_k=0,1,\ldots,R-1\}$, where

$$O(\chi_k)=c(\chi_k)[1+\vec{m}(\chi_k)\cdot\vec{\sigma}]. \quad (D1)$$

The non-negative constants $c(\chi_k)$ and the vectors $\vec{m}(\chi_k)$ are subject to the constraints

$$\sum_{i_k=0}^{R-1}c(\chi_k)=1, \quad (D2)$$

$$\sum_{i_k=0}^{R-1}c(\chi_k)\vec{m}(\chi_k)=0, \quad (D3)$$

$$|\vec{m}(\chi_k)|=1, \quad (D4)$$

which ensure that $O(\chi_k)\geqslant 0$ and $\sum_{i_k}O(\chi_k)=\mathbb{1}$. Note that we allow $c(\chi_k)$ to be zero, thus taking into account the possibility that each local POVM may have a different number of outcomes without letting R depend on k.

Assume we have measured all but the last copy and we wish to optimize the last measurement. To simplify the notation, let us define $r\equiv i_N$, $\vec{m}_r=\vec{m}(r\chi_{N-1})$, and $c_r=c(r\chi_{N-1})$. Then,

$$p_n(\chi)=p_n(\chi_{N-1})[c_r(1+\vec{n}\cdot\vec{m}_r)] \quad (D5)$$

and

$$\Delta=\sum_{\chi_{N-1}}\sum_r|\vec{V}(r\chi_{N-1})|=\sum_{\chi_{N-1}}d(\chi_{N-1}), \quad (D6)$$

where we have defined $d(\chi_{N-1})$ as

$$d(\chi_{N-1})\equiv\sum_r c_r\left|\int dn\,\vec{n}p_n(\chi_{N-1})(1+\vec{n}\cdot\vec{m}_r)\right|. \quad (D7)$$

We further write $\vec{V}\equiv\vec{V}(\chi_{N-1})$ and define the symmetric positive matrix

$$\mathsf{A}_{ij}\equiv\mathsf{A}_{ij}(\chi_{N-1})\equiv\int dn\,n_i n_j p_n(\chi_{N-1}). \quad (D8)$$

Equation (D7) becomes

$$d = \sum_r c_r |\vec{V} + \mathsf{A}\vec{m}_r|. \tag{D9}$$

(Hereafter the dependency on χ_{N-1} will be implicitly understood to simplify the notation.)

Our task is to maximize Eq. (D9). Introducing the Lagrange multipliers λ, $\vec{\gamma}$, and ω_r, the function we need to maximize is actually

$$L = d - \lambda\ \Lambda - \vec{\gamma}\cdot\vec{\Gamma} - \sum_r \omega_r \Omega_r, \tag{D10}$$

where the constraints

$$\Lambda = \sum_r c_r - 1, \tag{D11}$$

$$\vec{\Gamma} = \sum_r c_r \vec{m}_r, \tag{D12}$$

$$\Omega_r = \frac{\vec{m}_r^2 - 1}{2} \tag{D13}$$

can be read off from Eqs. (D2)–(D4). The factor 2 in the last expression is introduced for later convenience. Variations with respect to c_r yield

$$\frac{\delta L}{\delta c_r} = |\vec{V} + \mathsf{A}\vec{m}_r| - \vec{\gamma}\cdot\vec{m}_r - \lambda = 0. \tag{D14}$$

Notice that the points $\vec{m}_r$ that satisfy this equation define an ellipse $\mathcal{E}$ with focus at $-\mathsf{A}^{-1}\vec{V}$. Notice also that the parameter λ at the maximum is the value of d in Eq. (D9) [just multiply Eq. (D14) by c_r and sum over r taking into account the constraints $\Lambda=0$ and $\vec{\Gamma}=\vec{0}$].

Finally, consider the variations of $\vec{m}_r$ in Eq. (D10). We obtain

$$c_r\left(\mathsf{A}\frac{\vec{V} + \mathsf{A}\vec{m}_r}{|\vec{V} + \mathsf{A}\vec{m}_r|} - \vec{\gamma}\right) = \omega_r \vec{m}_r, \tag{D15}$$

which means that the vector inside the parentheses is proportional to $\vec{m}_r$. Note that the condition $\Omega_r=0$ defines a unit circle, and the orthogonal vector to this curve is $\vec{m}_r$. So we only need to prove that the orthogonal vector at point $\vec{m}_r$ of the ellipse $\mathcal{E}$ defined in Eq. (D14) has precisely this direction. This follows straightforwardly by taking variations with respect to $\vec{m}_r$ in Eq. (D14). Therefore, the solution is given by the tangency points of the ellipse $\mathcal{E}$ and the circle $\Omega_r=0$. There are only two such points; they are in opposite directions, and all constraints and maximization equations are satisfied with $c_{1,2}=1/2$. This proves that the optimal measurements in the greedy scheme are indeed von Neumann's.[1] Notice that this is a stronger statement than it looks: local measurements with a larger number of outcomes will perform worse.

[1]For the 3D case we just have to replace ellipses by ellipsoids and circles by spheres.

[1] C. M. Caves, C. A. Fuchs and P. Rungta, Found. Phys. Lett. **14**, 199 (2001); J. Batle *et al.*, Opt. Spectrosc. **94**, 700 (2003).
[2] A. Peres and W. K. Wootters, Phys. Rev. Lett. **66**, 1119 (1991).
[3] S. Massar and S. Popescu, Phys. Rev. Lett. **74**, 1259 (1995).
[4] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982).
[5] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).
[6] *Quantum State Estimation*, edited by M. Paris and J. Rehácek, Lecture Notes in Physics Vol. 649 (Springer, Berlin, 2004).
[7] S. L. Braunstein and H. J. Kimble, Phys. Rev. A **61**, 042302 (2000).
[8] R. Derka, V. Buzek, and A. K. Ekert, Phys. Rev. Lett. **80**, 1571 (1998).
[9] J. I. Latorre, P. Pascual, and R. Tarrach, Phys. Rev. Lett. **81**, 1351 (1998); S. Iblisdir and J. Roland, e-print quant-ph/0410237.
[10] N. Gisin and S. Popescu, Phys. Rev. Lett. **83**, 432 (1999); S. Massar, Phys. Rev. A **62**, 040101 (2000).
[11] E. Bagan *et al.*, Phys. Rev. Lett. **85**, 5230 (2000); Phys. Rev. A **63**, 052309 (2001); E. Bagan, M. Baig, and R. Munoz-Tapia, *ibid.* **64**, 022305 (2001).
[12] A. Peres and P. F. Scudo, Phys. Rev. Lett. **86**, 4160 (2001).
[13] A. Acin, E. Jane, and G. Vidal, Phys. Rev. A **64**, 050302(R) (2001); E. Jane, Ph.D. thesis, University of Barcelona, Barcelon, 2002 (unpublished).
[14] E. Bagan, M. Baig, and R. Munoz-Tapia, Phys. Rev. Lett. **87**, 257903 (2001); Phys. Rev. A **69**, 050303 (2004).
[15] E. Bagan, M. Baig, and R. Munoz-Tapia, Phys. Rev. A **70**, 030301 (2004).
[16] A. Peres and P. F. Scudo, Phys. Rev. Lett. **87**, 167901 (2001); J. Mod. Opt. **49**, 1235 (2002); N. H. Lindner, A. Peres, and D. R. Terno, Phys. Rev. A **68**, 042308 (2003).
[17] S. D. Bartlett, T. Rudolph, and R. W. Spekkens, Phys. Rev. Lett. **91**, 027901 (2003).
[18] K. R. Jones, Phys. Rev. A **50**, 3682 (1994).
[19] R. D. Gill and S. Massar, Phys. Rev. A **61**, 042312 (2000).
[20] D. G. Fischer, S. H. Kienle, and M. Freyberger, Phys. Rev. A **61**, 032306 (2000).
[21] Th. Hannemann *et al.*, Phys. Rev. A **65**, 050303 (2002).
[22] E. Bagan, M. Baig, and R. Munoz-Tapia, Phys. Rev. Lett. **89**, 277904 (2002).
[23] E. Bagan, M. Baig, R. Munoz-Tapia, and A. Rodriguez, Phys. Rev. A **69**, 010304 (2004).
[24] F. Embacher and H. Narnhofer, Ann. Phys. (N.Y.) **311**, 220

(2004).

[25] M. Hayashi, in *Quantum Communication, Computing and Measurement*, edited by O. Hirota, A. S. Holevo, and C. M. Caves, (Plenum Publishing, New York, 1997); M. Hotta and M. Ozawa, Phys. Rev. A **70**, 022327 (2004).

[26] K. Matsumoto, J. Phys. A **35**, 3111 (2002); K. Usami *et al.*, Phys. Rev. A **68**, 022314 (2003); M. Hayashi and K. Matsumoto, e-print quant-ph/0308150.

[27] J. M. Bernardo and A. F. M. Smith, *Bayesian Theory* (Wiley, Chichester, 2000).

[28] A. G. White *et al.*, Phys. Rev. Lett. **83**, 3103 (1999); D. F. V. James *et al.*, Phys. Rev. A **64**, 052313 (2001); R. T. Thew *et al.*, *ibid.* **66**, 012303 (2002); J. B. Altepeter *et al.*, Phys. Rev. Lett. **90**, 193601 (2003); E. Skovsen, H. Stapelfeldt, S. Juhl, and K. Molmer, *ibid.* **91**, 090406 (2003).

[29] U. Leonhardt, *Measuring the Quantum State of Light* (Cambridge University Press, Cambridge, England, 1997).

[30] Z. Hradil, Phys. Rev. A **55**, R1561 (1997); K. Banaszek, *ibid.* **59**, 4797 (1999); Z. Hradil, J. Summhammer, and H. Rauch, Phys. Lett. A **261**, 20 (1999); K. Banaszek *et al.*, Phys. Rev. A **61**, 010304 (2000); Z. Hradil *et al.*, *ibid.* **62**, 014101 (2000); J. Fiurásek and Z. Hradil, *ibid.* **63**, 020101 (2001).

[31] Y. I. Bogdanov *et al.*, Phys. Rev. A **70**, 042303 (2004); H. F. Hofmann and S. Takeuchi, *ibid.* **69**, 042108 (2004).

[32] A. Monras, Ms thesis, Universitat Autònoma de Barcelona, Barcelona, 2004. (unpublished).

[33] D. Ivanovic, J. Phys. A **14**, 3241 (1981); W. K. Wootters and B. D. Fields, Ann. Phys. (N.Y.) **191**, 363 (1989).

[34] T. Cover and J. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).

[35] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton University Press, Princeton, NJ, 1960).

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

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