Optimal full estimation of qubit mixed states

E. Bagan,[1] M. A. Ballester,[2] R. D. Gill,[2, 3] A. Monras,[1] and R. Muñoz-Tapia[1]

[1] *Grup de Física Teòrica & IFAE, Facultat de Ciències, Edifici Cn,*
Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona) Spain
[2] *Department of Mathematics, University of Utrecht, Box 80010, 3508 TA Utrecht, The Netherlands*
[3] *EURANDOM, P.O. Box 513-5600 MB Eindhoven, The Netherlands*

(Dated: December 19, 2005)

We obtain the optimal scheme for estimating unknown qubit mixed states when an arbitrary number N of identically prepared copies is available. We discuss the case of states in the whole Bloch sphere as well as the restricted situation where these states are known to lie on the equatorial plane. For the former case we obtain that the optimal measurement does not depend on the prior probability distribution provided it is isotropic. Although the equatorial-plane case does not have this property for arbitrary N, we give a prior-independent scheme which becomes optimal in the asymptotic limit of large N. We compute the maximum mean fidelity in this asymptotic regime for the two cases. We show that within the pointwise estimation approach these limits can be obtained in a rather easy and rapid way. This derivation is based on heuristic arguments that are made rigorous by using van Trees inequalities. The interrelation between the estimation of the purity and the direction of the state is also discussed. In the general case we show that they correspond to independent estimations whereas for the equatorial-plane states this is only true asymptotically.

PACS numbers: 03.67.Hk, 03.65.Ta

I. INTRODUCTION

Two-state systems or qubits are the building blocks of many applications in Quantum Information. Although they are commonly assumed to be in pure states, in real situations they are not. State preparation, processing, quantum channels, etc. are inevitably imperfect, which means that any quantum system is, in fact, in a mixed state. The accurate estimation of the parameters that characterize qubit mixed states is therefore of utmost relevance for practical applications. The aim of this work is to find the optimal (most accurate) scheme to perform this task.

So far, most of the work in state estimation has focused on pure qubit states [1–3] and fewer quantitative results have been obtained for qubit mixed states [4–9]. One obvious reason for this is the greater complexity of the estimation procedure. Whereas pure states are fully characterized by just two parameters —those specifying a point on the surface of the Bloch sphere, i.e., a unit vector— for a mixed state an additional parameter is required to specify its purity, by which we mean the distance from the center of the Bloch sphere to the point that represents the state. This brings a theoretical subtlety: we will need to identify a uniform prior distribution for the purity. In contrast to the pure-state case where there is a "natural" uniform probability distribution —the invariant measure on the 2-sphere—, for mixed states there is no unique choice. A uniform distribution must be isotropic (invariant under rotations of the Bloch sphere), but the purity, which is itself invariant, can be distributed according to a whole class of functions [10, 11], depending on several criteria. Despite this ambiguity, our results turn out to be rather general and, in particular, they do not depend on the specific choice of an isotropic purity prior.

In this paper, we assume that we have N identically prepared systems upon which we can perform generalized measurements. From their outcomes we can infer the value of the parameters that characterize the state of the systems. The quality or accuracy of the estimation is quantified by the fidelity (to be defined in the next section). The average of the fidelity over the prior and the outcome distribution provides a useful summary parameter of the overall quality of the estimation scheme. This problem was partially addressed in [5]. Here we present an alternative formulation that enables us to apply the approach to new, practically relevant situations and find many explicit results.

To be more specific, we will study two types of situation: that of estimating an *à priori* completely unknown qubit state and that of estimating a state that is known to lie on an equatorial plane of the Bloch sphere. We call the former the 3D case (or just 3D for short), as the state can be represented by any point in the 3-dimensional Bloch sphere. By the same logic, we call the latter 2D. The 2D case is useful because in many applications quantum states can be parametrized by the purity and a phase; e.g., linearly polarized photons. The 2D case also exhibits some remarkable theoretical features. For instance, we will show that while for 3D states the optimal measurement is essentially unique, independently of the isotropic prior, this is not so for 2D states, though this feature is recovered in the asymptotic limit of large N.

We will first address the problem from a Bayesian point of view, which will provide explicit results for any finite N. We will also take a steep dive into the asymptotic regime of the estimation schemes. It is clear that unknown states can only be estimated with perfect accuracy in the limit $N \to \infty$. The rate at which this perfect determination limit is achieved as N increases is a very informative parameter. It is useful, e.g., to compare different estimation schemes. If two schemes have the same rate, we say that they are (asymptotically) equivalent. The asymptotic behavior is also a central notion in statistics, where there exists a wealth of results and very powerful techniques [12, 13].

Within the statistical framework, one looks for measurements and estimators which have good behaviour for any fixed signal state. It turns out, under regularity conditions, that the maximum likelihood estimator is asymptotically optimal whatever the true signal state. The mean square error of the estimator gives a measure of the quality of the scheme. This error can be related to the fidelity through the Fisher information matrix, thus providing a connection with the Bayesian approach. In this context, the prior distribution plays a very minor role. In contrast, within the Bayesian approach the prior distribution does play a significant role because, as mentioned above, one is interested in obtaining an estimation that is optimal on *average*.

Here we present in a fairly comprehensive way the application of the two approaches to the asymptotic behavior of qubit mixed state estimation. We will see that both yield the same results. This fact has important consequences. It tells us that the asymptotic behaviour of the optimal mean fidelity only depends on the prior as an average of the optimal *pointwise* (i.e., for a fixed state) fidelities. Second, the Bayesian approach provides an explicit scheme that attains the pointwise bounds. It is worth pointing out that for some restricted schemes and some priors this might not be the case. For instance, it is known that a scheme based on fixed local measurements with the Bures prior distribution [14] does not approach unity at a rate $1/N$ [8], as a pointwise approach would indicate. Even more surprising, in this situation the Bayesian and the Maximum Likelihood estimation give different asymptotic average fidelities [8], in contrast to the common lore that both estimators should be asymptotically equivalent, pointwise. The non-equivalences here do all have simple explanations. Pointwise, everything *is* asymptotically equivalent and does converge at rate $1/N$. However, the convergence is not uniform or the integrated coefficient of $1/N$ diverges.

This paper is organised as follows. In the next section we introduce the notation and main concepts that will be used throughout this work. In Sec. III we obtain the optimal estimation protocol for any number of copies of the state in both the 3D and the 2D cases. In Secs. IV and V we compute the asymptotic expression of the fidelity from both the Bayesian and the pointwise approaches, respectively. The derivation of the latter is done through a rather self-contained presentation since some of the techniques may not be so well known among physicists. In Sec. VI we summarise our main results. We have relegated many technical details to the appendices for the benefit of readers not interested in technicalities

II. PRELIMINARIES

Consider an ensemble of N identically prepared states $[\rho(\vec{r})]^{\otimes N}$, where $\rho(\vec{r})$ is a density matrix with Bloch representation given by

$$\rho(\vec{r}) = \frac{\mathbb{1} + \vec{r}\cdot\vec{\sigma}}{2}. \tag{2.1}$$

Here $\vec{\sigma} = (\sigma^x, \sigma^y, \sigma^z)$, where σ^a, $a = x, y, z$, are the usual Pauli matrices and $\vec{r}$ is a point in the Bloch sphere $\{\vec{r} : |\vec{r}| \leq 1\}$. We will drop $\vec{r}$ and write simply ρ where no ambiguity arises.

A measurement on $\rho^{\otimes N}$ is represented by a Positive Operator Valued Measure (POVM). It is defined by a set $O = \{O_\chi\}$ of positive operators such that

$$\sum_\chi O_\chi = \mathbb{1}, \tag{2.2}$$

where χ refers to the various outcomes that can occur. It can be a discrete or a continuous variable.

In order to estimate ρ we proceed as follows. We first perform a measurement on $\rho^{\otimes N}$, from which we obtain an outcome χ. Based on χ, an estimate for ρ can be guessed: ρ_χ. Its quality is quantified by the fidelity, defined as [14]

$$f(\vec{r}, \vec{R}_\chi) = \left(\mathrm{tr}\,\sqrt{\sqrt{\rho_\chi}\rho\sqrt{\rho_\chi}}\,\right)^2, \tag{2.3}$$

which determines the maximum distinguishability between ρ and ρ_χ that can be achieved by any measurement [15]. For qubits, Eq. (2.3) reads

$$f(\vec{r}, \vec{R}_\chi) = \frac{1 + \vec{r}\cdot\vec{R}_\chi + \sqrt{1-r^2}\sqrt{1-R_\chi^2}}{2}, \tag{2.4}$$

where $\vec{r}$ and $\vec{R}_\chi$ are the Bloch vectors of the states ρ and ρ_χ respectively, $r = |\vec{r}|$ and $R = |\vec{R}|$.

In the Bayesian approach the overall performance of the estimation procedure is quantified by the average fidelity F, hereafter fidelity in short. It is the average of (2.3) over the prior probability distribution, which we denote $d\rho$, and over all possible outcomes χ of a given measurement, namely

$$F = \sum_\chi \int d\rho\, f(\vec{r}, \vec{R}_\chi) p(\chi|\vec{r}), \tag{2.5}$$

where $p(\chi|\vec{r})$ is the conditional probability of obtaining outcome χ given that the signal state has Bloch vector $\vec{r}$. These probabilities are determined by the expectation values of the positive operators O_χ, i.e., $p(\chi|\vec{r}) = \mathrm{tr}\,[O_\chi \rho]$. Our aim is to maximize (2.5).

For a given measurement O, there always exists an optimal guess or estimator. To prove this, we first introduce the four dimensional Euclidean vector

$$\mathbf{r} = (r^0, r^x, r^y, r^z) = (r^0, \vec{r}) = (\sqrt{1-r^2}, \vec{r}). \tag{2.6}$$

Note that $\mathbf{r}\cdot\mathbf{r}' = r^0 r'^0 + \vec{r}\cdot\vec{r}'$ and $|\mathbf{r}| = \sqrt{\mathbf{r}\cdot\mathbf{r}} = 1$. With this, the average fidelity reads

$$F = \sum_\chi \int d\rho \, \frac{1+\mathbf{r}\cdot\mathbf{R}_\chi}{2} p(\chi|\vec{r}), \tag{2.7}$$

where $\mathbf{R}_\chi = (R^0_\chi, \vec{R}_\chi)$ is defined in analogy to (2.6). A straightforward use of the Schwarz inequality gives an upper bound of F that is saturated with the choice

$$\mathbf{R}_\chi = \frac{\mathbf{V}_\chi}{|\mathbf{V}_\chi|}; \quad \mathbf{V}_\chi \equiv (V^0_\chi, \vec{V}_\chi) \equiv \int d\rho \, \mathbf{r} \, p(\chi|\vec{r}), \tag{2.8}$$

Using (2.8), the maximum fidelity is

$$F = \frac{1}{2}\left(1 + \sum_\chi |\mathbf{V}_\chi|\right) \equiv \frac{1}{2}(1+\Delta). \tag{2.9}$$

Since the guess (2.8) satisfies $|\mathbf{R}_\chi| = 1$ and its first component is non-negative, it *always* gives a physical state. In fact (2.8) is the best state that can be inferred and (2.9) is the maximum fidelity that can be obtained given O and the prior $d\rho$.

In the analysis below, it will prove very convenient to block-diagonalize $\rho^{\otimes N}$ by writing it in the basis of the SU(2) invariant subspaces of $(\mathbf{\frac{1}{2}})^{\otimes N}$ [we use boldfaced integers and half-integers to denote the irreducible representations of SU(2)], which are also invariant under the action of the symmetric group S_N (See App. A and also [4, 5] for details). In contrast with pure states, for which $\rho^{\otimes N}$ has projection only in the symmetric $(N+1)$-dimensional subspace of $\mathbf{J} \equiv \mathbf{\frac{N}{2}}$, for mixed states $\rho^{\otimes N}$ has also components in all the lower-dimensional invariant subspaces, which, furthermore, occur with multiplicity, n_j, greater than one. We thus write

$$\rho^{\otimes N} = \bigoplus_{j=0,1/2}^{N/2} n_j \rho_{Nj}, \tag{2.10}$$

where the lower limit in the direct sum is 0 for even N and 1/2 for odd N,

$$n_j = \binom{N}{N/2-j} \frac{2j+1}{N/2+j+1} \tag{2.11}$$

and

$$\rho_{Nj} = \left(\frac{1-r^2}{4}\right)^{N/2-j} \rho_j, \tag{2.12}$$

with

$$\begin{aligned}\rho_j = & \sum_{m=-j}^{j} \left(\frac{1-r}{2}\right)^{j-m} \left(\frac{1+r}{2}\right)^{j+m} \times \\ & U(\vec{n})|jm\rangle\langle jm|U^\dagger(\vec{n}).\end{aligned} \tag{2.13}$$

Throughout this paper $U(\vec{n})$ denotes the SU(2) unitary representation of the rotation $\mathcal{R}(\vec{n})$ that takes the unit vector $\vec{z}$ (pointing along the z-axis) into $\vec{n} \equiv \vec{r}/r$ on the Bloch sphere. Recall that

$$\langle jm|U(\vec{n})|jm'\rangle = \mathfrak{D}^{(j)}_{mm'}(\vec{n}) \tag{2.14}$$

defines the standard Wigner matrices [16]. Notice that ρ_j are not proper density matrices, since $\operatorname{tr}\rho_j \neq 1$.

For 2D states, the Bloch vector $\vec{r}$ of the state ρ lies on the equatorial xy-plane of the Bloch sphere, i.e., $\vec{r} = r(\cos\theta, \sin\theta, 0)$. We are still entitled to use the decomposition of $\rho^{\otimes N}$ above, but now we write

$$\begin{aligned}\rho_j = & \sum_{m=-j}^{j} \left(\frac{1-r}{2}\right)^{j-m} \left(\frac{1+r}{2}\right)^{j+m} \times \\ & U(\theta)U(\vec{x})|jm\rangle\langle jm|U^\dagger(\vec{x})U^\dagger(\theta),\end{aligned} \tag{2.15}$$

where $\vec{x}$ is the unit vector pointing along the x-axis and $U(\theta)$ is a unitary representation of a rotation of angle θ around the z-axis. Note that $U(\vec{x})|jm\rangle$ is an eigenstate of $\vec{x}\cdot\vec{J}$ (i.e., of the projection of the total spin operator $\vec{J}$ along the x-axis), since $U(\vec{x})$ takes $\vec{z}$ into $\vec{x}$ (i.e., is a rotation of angle $\pi/2$ around the y-axis). Hence, the Bloch vectors of the whole set of states $\{U(\theta)[U(\vec{x})|jm\rangle]\}$ lie on the xy-plane, as they should, and θ is the angle between $\vec{r}$ and the x-axis.

In the basis $|jm\rangle$ the transformation $U(\theta)$ is diagonal, and substituting (2.15) in (2.12) we obtain

$$\rho_{Nj} = \sum_{m,m'} \mathrm{e}^{i(m-m')\theta} \rho^j_{mm'} |jm\rangle\langle jm'|, \tag{2.16}$$

where

$$\begin{aligned}\rho^j_{mm'} = & \sum_{m''} \mathrm{d}^{(j)}_{mm''}(\pi/2)\mathrm{d}^{(j)}_{m'm''}(\pi/2) \\ & \times \left(\frac{1-r}{2}\right)^{N/2-m''} \left(\frac{1+r}{2}\right)^{N/2+m''}\end{aligned} \tag{2.17}$$

and $\mathrm{d}^{(j)}_{mm'}$ are the (real) reduced Wigner matrices [16].

III. FINITE NUMBER OF COPIES. BAYESIAN ESTIMATOR

In this section we obtain the optimal POVM and closed expressions of the fidelity for any number of copies of the signal state. Although the 3D and 2D cases look similar, we will show that there are remarkable differences between them.

A. 3D states

As mentioned in the introduction, we consider N identical copies of a quantum state which is chosen according to an isotropic prior distribution

$$d\rho = w(r)\, dr\, dn, \tag{3.1}$$

where dn is the invariant measure on the 2-sphere

$$dn = \frac{d(\cos\theta)\, d\phi}{4\pi} \tag{3.2}$$

and $w(r)$ is normalized such that $\int_0^1 dr\, w(r) = 1$.

Let us start by computing the optimal POVM. We first notice that because of the block-diagonal form of $\rho^{\otimes N}$ in (2.10) we may just consider also block-diagonal POVMs, of the form

$$O_\chi = \bigoplus_{j=0}^{J} n_j O_{\chi j}, \quad \text{such that} \quad \sum_\chi O_{\chi j} = \mathbb{1}_j, \tag{3.3}$$

with no loss of generality. Indeed, for any given POVM $\{O_\chi\}$, we can always construct a new one, $\{\tilde{O}_{\chi j\alpha}\}$, through

$$\tilde{O}_{\chi j\alpha} = \mathbb{1}_{j\alpha} O_\chi \mathbb{1}_{j\alpha}, \tag{3.4}$$

where $\mathbb{1}_{j\alpha}$ is the identity in the $\mathbf{j}$-representation subspace and α ($1 \le \alpha \le n_j$) labels the different occurances of $\mathbf{j}$ in the Clebsch-Gordan series of $(\frac{1}{2})^{\otimes N}$. If F ($\tilde{F}$) stands for the maximum fidelity that can be attained using $\{O_\chi\}$ ($\{\tilde{O}_{\chi j\alpha}\}$), we have $F \le \tilde{F}$. This is readily seen by noticing that the probability $p(\chi|\vec{r}) = \mathrm{tr}\,[\rho^{\otimes N} O_\chi]$ is the marginal of $p(\chi j\alpha|\vec{r}) = \mathrm{tr}\,[\rho^{\otimes N}\tilde{O}_{\chi j\alpha}]$, i.e., $p(\chi|\vec{r}) = \sum_{j\alpha} p(\chi j\alpha|\vec{r})$, and no marginal can be more informative than the initial probability distribution. Moreover, because of (2.10), if $\{\tilde{O}_{\chi j\alpha}\}$ is to be optimal, we may obviously replace $\tilde{O}_{\chi j1}$, $\tilde{O}_{\chi j2}$, …, $\tilde{O}_{\chi j n_j}$ by, say, $\tilde{O}_{\chi j1}$, $\tilde{O}_{\chi j1}$, …, $\tilde{O}_{\chi j1}$ without changing the fidelity, which leads us to (3.3).

It is important to note that (3.4) allows us to view j and α as the outcome of the measurement $\{\mathbb{1}_{j\alpha}\}$. Therefore, in Eq. (2.9) we will have $n_j|\mathbf{V}_{\chi j}|$ instead of $|\mathbf{V}_\chi|$, and an additional summation over j. Hence, our goal is to maximize $|\mathbf{V}_{\chi j}|$ for all pairs (χ, j), where

$$\mathbf{V}_{\chi j} = \int d\rho\, \mathbf{r}\, \mathrm{tr}\,(\rho^{\otimes N} O_{\chi j}). \tag{3.5}$$

The j outcomes give information about the decomposition of $\rho^{\otimes N}$ as a direct sum of SU(2) irreducible components. This, in turn, encodes information about r. For instance, if $r = 1$ (pure state), the probability of obtaining the outcome $j = N/2$ is unity. For our purposes, all the information concerning the purity of ρ comes from this source, as we now demonstrate.

Since $V^0_{\chi j}$ is invariant under rotations, whereas $\vec{V}_{\chi j}$ transforms as a 3-vector, we may apply to $\mathbf{V}_{\chi j}$ the rotation $\mathcal{R}^{-1}(\vec{n}_{\chi j}) = \mathcal{R}^\top(\vec{n}_{\chi j})$, where $\vec{n}_{\chi j} = \vec{V}_{\chi j}/|\vec{V}_{\chi j}|$, and obtain $\mathbf{V'}_{\chi j}$, such that its x- and y-components vanish, i.e., $V'^x_{\chi j} = V'^y_{\chi j} = 0$ and

$$\begin{aligned} V'^z_{\chi j} &= \int d\rho\, \left[\mathcal{R}^\top(\vec{n}_{\chi j})\vec{r}\right]^z \mathrm{tr}\,(\rho^{\otimes N} O_{\chi j}) \\ &= \int d\rho\, r\cos\theta\, \mathrm{tr}\,\left(\rho^{\otimes N}\Omega_{\chi j}\right), \end{aligned} \tag{3.6}$$

$$V'^0_{\chi j} = \int d\rho\, \sqrt{1-r^2}\,\mathrm{tr}\,(\rho^{\otimes N}\Omega_{\chi j}), \tag{3.7}$$

where we have defined

$$\Omega_{\chi j} \equiv U^\dagger(\vec{n}_{\chi j})\, O_{\chi j} U(\vec{n}_{\chi j}), \tag{3.8}$$

we have used that $d\rho$ is rotationally invariant, and we have written $\vec{r} = r\vec{n}$ in spherical coordinates, i.e., $\vec{n} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$. Therefore, $|\mathbf{V}_{\chi j}| = |\mathbf{V'}_{\chi j}|$, and the maximum fidelity can be computed using $\mathbf{V'}^j_\chi$ instead of $\mathbf{V}^j_\chi$. Hereafter, we drop the primes and write

$$\Delta_{\rm 3D} = \sum_{\chi j} n_j |\mathbf{V}_{\chi j}| = \sum_{\chi j} n_j \sqrt{(V^0_{\chi j})^2 + (V^z_{\chi j})^2}, \tag{3.9}$$

where $V^0_{\chi j}$, $V^z_{\chi j}$ are given by (3.6) and (3.7).

Using Eqs. (2.12–2.14) and recalling that $\cos\theta = \mathfrak{D}^{(1)}_{00}(\vec{n})$, we have

$$\begin{aligned} V^z_{\chi j} &= \int_0^1 drw(r)\, r \sum_{mm'm''} \rho_{jm}\, [\Omega_{\chi j}]_{m''m'} \\ &\quad\times \int dn\, \mathfrak{D}^{(1)}_{00}(\vec{n})\mathfrak{D}^{(j)}_{m'm}(\vec{n})\mathfrak{D}^{(j)*}_{m''m}(\vec{n}), \end{aligned} \tag{3.10}$$

$$\begin{aligned} V^0_{\chi j} &= \int_0^1 drw(r)\, \sqrt{1-r^2} \sum_{mm'm''} \rho_{jm}\, [\Omega_{\chi j}]_{m''m'} \\ &\quad\times \int dn\, \mathfrak{D}^{(j)}_{m'm}(\vec{n})\mathfrak{D}^{(j)*}_{m''m}(\vec{n}), \end{aligned} \tag{3.11}$$

where the sum over the indexes m, m', m'' runs from $-j$ to j, and we have defined

$$\rho_{jm} = \left(\frac{1-r^2}{4}\right)^{J-j} \left(\frac{1-r}{2}\right)^{j-m} \left(\frac{1+r}{2}\right)^{j+m}. \tag{3.12}$$

The orthogonality relations of the irreducible representations of SU(2) (Eqs. (4.6.1) and (4.6.2) on Page 62 of Ref. [16]) enable us to write

$$V^z_{\chi j} = \int_0^1 dr \frac{w(r)\, r}{j(j+1)d_j} \sum_{mm'} mm'\rho_{jm}\, [\Omega_{\chi j}]_{m'm'}\, , \tag{3.13}$$

$$V^0_{\chi j} = \int_0^1 dr \frac{w(r)\, \sqrt{1-r^2}}{d_j} \sum_{mm'} \rho_{jm}\, [\Omega_{\chi j}]_{m'm'}\, , \tag{3.14}$$

where $d_j = 2j+1$ is the dimension of the representation $\mathbf{j}$ of SU(2). We readily see that the z- and 0-components of $\mathbf{V}_{\chi j}$ are bounded by

$$|V^z_{\chi j}| \le \frac{\mathrm{tr}\,\Omega_{\chi j}}{d_j} \frac{\max_{m'}|m'|}{j(j+1)} \left| \int_0^1 drw(r)\, r \sum_m m\rho_{jm} \right|, \tag{3.15}$$

$$|V^0_{\chi j}| = \frac{\mathrm{tr}\,\Omega_{\chi j}}{d_j} \int_0^1 drw(r)\, \sqrt{1-r^2} \sum_m \rho_{jm}. \tag{3.16}$$

Note that all the χ dependence has been factored out and $\Delta_{\rm 3D}$ takes the form

$$\Delta_{\rm 3D} \le \sum_j n_j \left(\frac{\sum_\chi \mathrm{tr}\,\Omega_{\chi j}}{d_j}\right) \sqrt{(v^0_j)^2 + (v^z_j)^2}\ , \tag{3.17}$$

where v_j^0 and v_j^z can be easily worked out from (3.15) and (3.16) to be

$$v_j^z \;=\; \int_0^1 dr \frac{w(r)r}{j+1} \sum_{m=-j}^{j} m\rho_{jm}, \tag{3.18}$$

$$v_j^0 \;=\; \int_0^1 dr\, w(r)\sqrt{1-r^2} \sum_{m=-j}^{j} \rho_{jm}. \tag{3.19}$$

Eq. (3.8) clearly implies that the factor in parentheses in (3.17) is unity. Notice that the χ dependence has entirely disappeared in the final bound of the fidelity.

Inequality (3.17) is saturated iff the only non-vanishing term of the sum over m' in (3.13) corresponds to the maximum value of $|m'|$, namely, j. This implies that $[\Omega_{\chi j}]_{m'm'} \propto \delta_{m'j}$ (or the trivial symmetric choice $\delta_{m'-j}$). An obvious choice that satisfies this condition —and is independent of χ— is

$$\Omega_j = d_j |jj\rangle\langle jj|. \tag{3.20}$$

The operator Ω_j is a seed of a continuous covariant POVM, i.e.,

$$O_{\vec{\mu}\, j} = U(\vec{\mu})\Omega_j U^\dagger(\vec{\mu}), \tag{3.21}$$

where $\vec{\mu}$ plays the role of χ. It can be easily verified that, $\int d\mu\, O_{\vec{\mu}\, j} = \mathbb{1}_j$ [1], where $d\mu$ (as dn) is the invariant measure over the 2-sphere. This proves that the bound is attainable. POVMs with a finite number of outcomes can also be obtained using the results in [17].

Having obtained the optimal POVM, Eq. (3.21), it is straightforward to compute the conditional probabilities

$$\mathrm{tr}\,\left(\rho^{\otimes N} O_{\vec{\mu} j}\right) = d_j \left(\frac{1-r^2}{4}\right)^{J-j} \left(\frac{1+\vec{r}\cdot\vec{\mu}}{2}\right)^{2j}, \tag{3.22}$$

which will be needed in Sec. V. One can check that

$$\sum_j n_j \int d\mu\, \mathrm{tr}\,\left(\rho^{\otimes N} O_{\vec{\mu} j}\right) = 1, \tag{3.23}$$

as it should be. The corresponding guesses can be worked out from (3.5) by simply substituting $\vec{\mu}$ for χ. One can also verify that the angular integration indeed yields the two terms (3.18) and (3.19).

In summary, the fidelity of any optimal POVM can be written as

$$\Delta_{\mathrm{3D}} = \sum_j^J n_j \sqrt{(v_j^0)^2 + (v_j^z)^2}. \tag{3.24}$$

This equation along with (3.18) and (3.19), provide a general expression of the maximum fidelity for any given prior distribution $w(r)$. Unless an explicit expression for $w(r)$ is given, this is as far as we can get. In App. C we present closed expressions of the fidelity for arbitrary N using the Bures prior. In the asymptotic limit $N \to \infty$ however one can derive a compact formula for the fidelity in terms of the mean value of r: $\langle r\rangle = \int_0^1 dr\, w(r)\, r$. This will be done in Secs. IV and V.

Several comments are in order here. Within an optimal scheme, the purity estimator,

$$R_{\chi j} = \frac{|\vec{V}_{\chi j}|}{|\mathbf{V}_{\chi j}|} = \frac{|v_j^z|}{\sqrt{{v_j^0}^2 + {v_j^z}^2}} \equiv R_j, \tag{3.25}$$

only depends on j and comes *solely* from the measurement represented by the POVM $\{\mathbb{1}_{j\alpha}\}$ [18]. All dependence on any other kind of outcome, generically referred to as χ [e.g., $\vec{\mu}$ in Eq. (3.21)], has disappeared. This is expected from symmetry grounds: the parameter r does not change under SU(2) transformations and the optimal purity guess must thus be a function of j/N, as the only SU(2)-invariant quantity in this problem is precisely j. Furthermore, since this measurement ($\{\mathbb{1}_{j\alpha}\}$) does not alter (on average) the estimation of the orientation $\vec{n} = \vec{r}/r$ of the signal state, the optimal estimation in the sense of average fidelity of (*a priori*) isotropically distributed mixed states breaks into two independent estimations: that of the purity r and that of the orientation $\vec{n}$ in the Bloch sphere. Notice finally that after this measurement, the rest of the protocol, which involves the POVM (3.21) for a fixed j (or any version of it with a finite number of outcomes), is identical to the optimal protocol for estimating a pure state $|\vec{n}\rangle$ given $2j$ identical copies of it [2].

B. 2D states

In the situation we are about to consider, $\mathbf{V}_{\chi j}$, defined by (3.5), still determines the maximum fidelity through Eq. (2.9), but $d\rho$ is

$$d\rho = w(r)\, dr \frac{d\theta}{2\pi} \tag{3.26}$$

with $\int_0^1 dr\, w(r) = 1$. Since $\vec{r}$ is a 2-dimensional vector, we can use a complex notation and write $\vec{r} \to r\mathrm{e}^{i\theta}$. In this notation $\vec{V}_{\chi j}$ and $\vec{R}_{\chi j}$ also become complex numbers. More specifically,

$$V_{\chi j}^0 = \sum_m \int_0^1 dr\, w(r)\sqrt{1-r^2}\rho_{mm}^j O_{mm}^{\chi j}, \tag{3.27}$$

where we have raised the outcome labels χ and j in $O_{mm}^{\chi j}$ [or in $\rho_{mm'}^j$, Eq. (2.17)] to avoid a confusing proliferation of subindexes; the latter will label matrix elements, e.g., $O_{mm'}^{\chi j} = \langle jm|O_{\chi j}|jm'\rangle$. Similarly, we have

$$\begin{aligned} |\vec{V}_{\chi j}| \;&=\; \left| \int d\rho\, r \sum_{mm'} \mathrm{e}^{i(m-m'+1)\theta} \rho_{mm'}^j O_{m'm}^{\chi j} \right| \\ &\leq\; \int_0^1 dr\, w(r)\, r \sum_m \rho_{m m+1}^j \left| O_{m+1\, m}^{\chi j} \right|, \end{aligned} \tag{3.28}$$

where we have used that $\rho^j_{m\,m+1} \geq 0$ for all r. The equality in (3.28) is attained by choosing the phase of $O^{\chi j}_{m+1\,m}$ to be independent of m.

The positivity of $O_{\chi j}$ implies that

$$|O^{\chi j}_{m+1\,m}| \leq \sqrt{O^{\chi j}_{mm}}\sqrt{O^{\chi j}_{m+1\,m+1}}. \tag{3.29}$$

By choosing $|O^{\chi j}_{m+1\,m}|$ to take its maximum value in (3.29) we ensure that $|\vec{V}_{\chi j}|$ will also be maximal. So far, the optimization of $V^0_{\chi j}$ and $|\vec{V}_{\chi j}|$ can be carried out independently of one another, since the choices we have to make in order to saturate the bounds in (3.28) and (3.29) do not affect $V^0_{\chi j}$. However, we will have to check that they are compatible with the POVM condition $\sum_\chi O_{\chi j} = \mathbb{1}^j$. We will verify this by giving an explicit POVM that meets all the above conditions.

We now replace $O_{\chi j}$ by its covariant version $\tilde{O}_{\chi j \phi}$, defined in (D3) —in Appendix D we show that this change does not affect the average fidelity— and take the seed (positive) operator $\Omega_{\chi j}$ in (D1) to be given by $\Omega_{\chi j} = |u_{\chi j}\rangle\langle u_{\chi j}|$ (i.e., to be rank one), where

$$|u_{\chi j}\rangle = \sum_m u^{\chi j}_m |j, m\rangle. \tag{3.30}$$

The components $u^{\chi j}_m$ are taken to be real and must satify

$$\sum_\chi \left(u^{\chi j}_m\right)^2 = \sum_\chi O^{\chi j}_{mm} = 1, \tag{3.31}$$

as follows from

$$\tilde{O}^{\chi j \phi}_{mm'} = \mathrm{e}^{i(m-m')\phi} u^{\chi j}_m u^{\chi j}_{m'}. \tag{3.32}$$

It is important to realize that the vanishing of the off-diagonal elements in $\sum_\chi \int_0^{2\pi} d\phi/(2\pi)\, \tilde{O}^{\chi j\phi}_{mm'} = \mathbb{1}^j$ does not require further conditions on $u^{\chi j}_m$. Moreover,

$$\begin{aligned} \tilde{O}^{\chi j\phi}_{m+1\,m} &= \mathrm{e}^{i\phi} u^{\chi j}_{m+1} u^{\chi j}_m \\ &= \mathrm{e}^{i\phi}\sqrt{\tilde{O}^{\chi j\phi}_{mm}}\sqrt{\tilde{O}^{\chi j\phi}_{m+1\,m+1}}, \end{aligned} \tag{3.33}$$

hence, this choice saturates both (3.28) and (3.29).

Collecting all the pieces and defining $\Delta_{\mathrm{2D}} = \sum_j n_j \Delta^{\mathrm{2D}}_j$ [recall that $F = (1+\Delta)/2)$], we see that the maximum fidelity is given by the maximum value of

$$\begin{aligned} \Delta^{2D}_j &= \sum_\chi \left\{ \left[\sum_m \alpha^j_m (u^{\chi j}_m)^2 \right]^2 \right. \\ &+ \left. \left(\sum_m \beta^j_m u^{\chi j}_m u^{\chi j}_{m+1} \right)^2 \right\}^{1/2}, \end{aligned} \tag{3.34}$$

where $u^{\chi j}_m$ is constrained by (3.31) and α^j_m and β^j_m can be read off from (3.27) and (3.28) respectively:

$$\alpha^j_m = \int_0^1 dr\, w(r)\sqrt{1-r^2}\rho^j_{mm} \tag{3.35}$$

$$\beta^j_m = \int_0^1 dr\, w(r)\, r\, \rho^j_{mm+1}, \tag{3.36}$$

With no loss of generality we can take the index χ in (3.34) to be integer and its maximum value to be less or equal than the number of distinct values of α^j_m in (3.35). The symmetry relation $\mathrm{d}^{(j)}_{mm'} = \mathrm{d}^{(j)}_{-m'\,-m}$ further implies that $\chi \leq [d_j/2]$, where $[\ldots]$ stands for integer part. With all the above, maximizing Δ_{2D}, which can be done for each j independently, becomes a straightforward task.

The results of the 3D case may lead us to believe that the optimal POVM will be independent of the prior $w(r)$. The inspection of the low N cases gives further support to this belief. For $j \leq 5/2$ ($N \leq 5$) one can show that the optimal POVM is given by

$$u^j_m = 1 \tag{3.37}$$

for *any* prior $w(r)$, where we have dropped the index χ because it only takes one value here.[1] However, one can check that for $j \geq 3$ the choice (3.37) is *not* optimal for some priors. Take for instance $N = 6$ and consider a prior of the form $w(r) = (2r/\delta^2)\Theta(\delta - r)$, where $\Theta(x)$ is the step function [i.e., $\Theta(x) = 1$ for $x \geq 0$ and $\Theta(x) = 0$ otherwise] and δ is a positive number. If δ is sufficiently small, one can Taylor-expand Δ_3 about $\delta = 0$ and easily obtain the optimal solution at leading order, which does not turn out to be of the form (3.37). A straightforward computation yields $(\Delta^{\mathrm{opt}}_3 - \Delta^{\mathrm{Eq.(3.37)}}_3)/\Delta^{\mathrm{opt}}_3 = A\delta^4 + \mathcal{O}(\delta^5)$, where A is a constant that can be computed analytically ($A \approx 1.0 \times 10^{-3}$).

In spite of this unexpected dependence on the prior in the 2D case, there are, however, two features in the example above that are completely general: (a) the difference $\Delta^{\mathrm{opt}}_j - \Delta^{\mathrm{Eq.(3.37)}}_j$ is always very small, and (b) Δ^{opt}_j is actually different from $\Delta^{\mathrm{Eq.(3.37)}}_j$ only for priors that are very peaked about $r = 0$. There is a further, very important property: the POVM defined by (3.37) is asymptotically optimal (the proof is given in Appendix H). Hence, for practical purposes, the best one can do is to stick to the choice (3.37), for all j and m, regardless the prior knowledge one may have of ρ. Though this choice does not guarantee optimality for small N, it does guarantee that the corresponding fidelity will differ from the maximum one by a tiny amount (typically less than only one part in a thousand) and, furthermore, that this difference will decrease to zero as $N \to \infty$.

The asymptotically optimal choice (3.37) amounts to replacing $O_{\chi j}$ by

$$\tilde{O}_{\phi j} = U(\phi)\Omega_j U^\dagger(\phi), \tag{3.38}$$

where $\Omega_j = |u_j\rangle\langle u_j|$, $|u_j\rangle = \sum_m |jm\rangle$, and [hereafter we drop the superindex "Eq. (3.37)" in Δ, Δ_j, etc.]

$$\Delta_{\mathrm{2D}} = \sum_j n_j \sqrt{\left(v^0_j\right)^2 + \left(v^x_j\right)^2}, \tag{3.39}$$

[1] There are also degenerate solutions of the form $u^{\chi j}_m = \lambda_{\chi j}$ for all m, and with $\sum_\chi \lambda_{\chi j} = 1$

where

$$v^0_j=\sum_m \alpha^j_m,\qquad v^x_j=\sum_m \beta^j_m, \tag{3.40}$$

and the analogy with (3.24) is apparent.

We next recall (3.35), which involves $\operatorname{tr}\rho_{Nj}$. Since the trace is invariant under rotations, v^0_j can be straightforwardly computed using (2.12) and (2.13). No such simplification exists for v^x_j, as far as we are aware. Proceeding this way we have

$$\begin{aligned} v^0_j &= 2\sum_{m=-j}^{j}\int_0^1 dr\, w(r)\left(\frac{1-r}{2}\right)^{J-m+\frac12} \\ &\times\left(\frac{1+r}{2}\right)^{J+m+\frac12}, \\ v^x_j &= \sum_{m=-j}^{j} c^j_m \int_0^1 dr\, r\, w(r)\left(\frac{1-r}{2}\right)^{J-m} \\ &\times\left(\frac{1+r}{2}\right)^{J+m}, \end{aligned} \tag{3.41}$$

where the coefficients c^j_m are given by

$$c^j_m=\sum_{m'=-j}^{j-1}\mathrm{d}^{(j)}_{m'm}(\pi/2)\mathrm{d}^{(j)}_{m'+1\,m}(\pi/2), \tag{3.42}$$

as can be read off from (2.17). The sum over m in v^0_j can be easily performed, since it is just the sum of a geometric series, and yields

$$\begin{aligned} v^0_j &= 2\int_0^1 dr\frac{w(r)}{r}\left\{\left(\frac{1-r}{2}\right)^{J-j+\frac12}\right. \\ &\times\left.\left(\frac{1+r}{2}\right)^{J+j+\frac32}-(r\to -r)\right\}. \end{aligned} \tag{3.43}$$

The sum over m in v^x_j, however, is non trivial because of the coefficients c^j_m and no simple closed formula can be found but in the asymptotic limit $N\to\infty$.

IV. ASYMPTOTICS: BAYESIAN APPROACH

In this section we calculate the asymptotic (large N) expressions of the fidelities obtained in the previous sections using the Bayesian approach. For 2D they are summarized in (3.39), with the definitions (3.41), (3.42) and the relation (3.43). For 3D the maximum fidelity is given by (3.24), which involves the definitions (3.18) and (3.19). We here present a detailed computation only for 2D. The 3D case can be computed in a similar way and we just point out the main differences with 2D. For simplicity we consider an even number of copies $N=2n$, thus $J=n$.

We start by noticing that the coefficients c^j_m, defined in Eq. (3.42), satisfy $c^j_{-m}=-c^j_m$ (which implies $c^j_0=0$) and, hence,

$$\begin{aligned} v^x_j &= \sum_{m=1}^{j} c^j_m\int_0^1 dr\, r\, w(r)\left\{\left(\frac{1-r}{2}\right)^{n-m}\right. \\ &\times\left.\left(\frac{1+r}{2}\right)^{n+m}-(r\to -r)\right\}. \end{aligned} \tag{4.1}$$

We further note that the dominant contribution to the sum in v^x_j comes from the region where m is close to its maximum value j. We can thus replace c^j_m by the first terms of its "Taylor expansion" about $m=j$. It turns out that only the first two terms, $c^j_m\approx a_j+b_j(m-j)$, contribute at the order we are interested in. The coefficients a_j and b_j are computed in Appendix E. After substituting Eq. (E6) in (4.1) the sum over m gives:

$$\begin{aligned} v^x_j=\int_0^1 dr\frac{w(r)}{r}&\left\{\left(r-\frac{1}{4j}\right)\left(\frac{1-r}{2}\right)^{n-j}\right. \\ &\times\left.\left(\frac{1+r}{2}\right)^{n+j+1}-(r\to -r)\right\}, \end{aligned} \tag{4.2}$$

where we have dropped terms that fall off exponentially as n goes to infinity. It is convenient to combine v^0_j and v^x_j with the binomial in n_j [see Eq. (2.11)] and define $\bar v^0_j$ and $\bar v^x_j$ as

$$\bar v^0_j=\binom{2n}{n-j}v^0_j,\quad \bar v^x_j=\binom{2n}{n-j}v^x_j. \tag{4.3}$$

With this, Eq. (3.39) becomes

$$\Delta_{2\mathrm{D}}=\sum_j\frac{d_j}{n+j+1}\sqrt{\left(\bar v^0_j\right)^2+\left(\bar v^x_j\right)^2}. \tag{4.4}$$

Our goal is to compute the asymptotic behaviour of the above sum. We do so by first computing the leading order contribution: $\lim_{n\to\infty}\Delta$. We, of course, expect this to be unity, as the optimal guess must certainly lead to a perfect estimation given infinitely many copies. The calculation thus provides a consistency check of the approach and, moreover, the leading order expression of $\bar v^0_j$ and $\bar v^x_j$, which will be later used to compute the next-to-leading order contribution.

At leading order in $1/n$, we are entitled to use the well known result

$$\binom{2n}{k}q^k(1-q)^{2n-k}\approx\frac{\exp\left\{-n\frac{\left(\frac{k}{2n}-q\right)^2}{q(1-q)}\right\}}{2\sqrt{\pi nq(1-q)}}, \tag{4.5}$$

which holds for large n. In our case $k=n-j$ and $q=(1-r)/2$. Furthermore, we can approximate the gaussian in (4.5) by the Dirac delta function $\delta(k-2nq)=\delta(nr-$

$j) = \delta(r - j/n)/n$. After a straightforward calculation we end up with

$$\begin{aligned} \bar{v}_j^0 &= \frac{1}{2n}\frac{w(s)}{s}(1+s)\sqrt{1-s^2} + o(1/n), \\ \bar{v}_j^x &= \frac{1}{2n} w(s)(1+s) + o(1/n), \end{aligned} \tag{4.6}$$

where $s = j/n$.

Recalling the derivation of Eq. (3.39), we see that the optimal guess for the purity only depends on j and is given by

$$R_j = \frac{|v_j^x|}{\sqrt{(v_j^0)^2 + (v_j^x)^2}} = \frac{\bar{v}_j^x}{\sqrt{(\bar{v}_j^0)^2 + (\bar{v}_j^x)^2}}, \tag{4.7}$$

in full analogy with (3.25). [The optimal guess for θ is given by ϕ, Eq. (3.38).] One readily obtains

$$R_j = \frac{j}{n} + o(1), \tag{4.8}$$

as expected. Similarly, it also follows from (4.6) that

$$\sqrt{(\bar{v}_j^0)^2 + (\bar{v}_j^x)^2} = \frac{1}{2n}(1+s)\frac{w(s)}{s} + o(1/n). \tag{4.9}$$

At leading order the sum over j in (4.4) can be replaced by $n\int_0^1 ds$, and $d_j/(n+j+1) \approx 2j/(n+j) = 2s/(1+s)$. Hence, at leading order

$$\Delta_{\rm 2D} = \int_0^1 ds\, w(s) = 1, \tag{4.10}$$

and, as it should be, $\lim_{N\to\infty} F = 1$ for *any* prior.

We are now ready to compute the fidelity to next-to-leading order. The calculation can be greatly simplified by noticing that

$$\Delta_{\rm 2D} \geq \sum_{j=0}^{n} \frac{d_j}{n+j+1}\left(\bar{v}_j^0\sqrt{1-\xi_j^2} + \bar{v}_j^x \xi_j\right), \tag{4.11}$$

for all ξ_j such that $0 < \xi_j < 1$ [this is, in reverse, the same argument that took us from (2.7) to (2.9)]. The bound is saturated iff

$$(\sqrt{1-\xi_j^2}, \xi_j) \propto (\bar{v}_j^0, \bar{v}_j^x) \tag{4.12}$$

for all j, namely, iff $\xi_j = R_j$. With the leading order choice $\xi_j = j/n$, Eq. (4.11) provides a tight bound at order $o(1/n)$. At next-to-leading order we thus have

$$\Delta_{\rm 2D} = \sum_{j=0}^{n} \frac{d_j}{n+j+1}\left(\bar{v}_j^0\sqrt{1-\frac{j^2}{n^2}} + \bar{v}_j^x \frac{j}{n}\right), \tag{4.13}$$

where we have "linearized" the square root in (4.4), hence overcoming in a very simple way the most demanding part of the calculation. We can now use the techniques in Appendix F to evaluate the asymptotic value of this sum. We obtain

$$\Delta_{\rm 2D} = \left(1 - \frac{1}{2n}\right)\int_0^1 dr\, w(r) + o(1/n), \tag{4.14}$$

which implies

$$F^{2D} = 1 - \frac{1}{2N} + o(1/N), \tag{4.15}$$

independently of the prior $w(r)$. This result agrees with the bound derived from the pointwise approach in the next section.

The very same approach we have outlined can be applied to 3D states, we just have to replace v_j^x by v_j^z [see Sec. III A and Eqs. (C1), (C2) and (C3)]. To next to leading order we have (see Appendix F for details)

$$\Delta_{\rm 3D} = \int_0^1 dr\ w(r)\left(1 - \frac{3+2r}{4n}\right). \tag{4.16}$$

Recalling that $n = N/2$, the asymptotic fidelity reads

$$F^{\rm 3D} = 1 - \frac{3 + 2\langle r\rangle}{4N} + o(1/N), \tag{4.17}$$

where $\langle r\rangle$ stands for the mean purity over its prior distribution, namely

$$\langle r\rangle \equiv \int_0^1 dr\, w(r)\, r. \tag{4.18}$$

Particularizing (4.17) to the Bures distribution, Eq. (C4), we have

$$F^{\rm 3D}_{\rm Bures} = 1 - \left(\frac{3}{4} + \frac{4}{3\pi}\right)\frac{1}{N} + o(1/N). \tag{4.19}$$

V. ASYMPTOTICS: POINTWISE APPROACH

In the Bayesian approach, described in the previous sections, both the measurement strategy and the estimator (or guess) —i.e., the estimation scheme— are so chosen as to maximize the average fidelity with respect to a given prior distribution for any N. In contrast, in the so called pointwise approach, to which this section is devoted, one's goal is to asymptotically optimize the performance of a scheme at *each fixed point*, $\boldsymbol{\theta}_0$, in parameter space (In this section we will denote the parameters that specify the states by $\boldsymbol{\theta}$ and the guesses by $\hat{\boldsymbol{\theta}}$, as is standard in statistics).

The aim of this section is to present a bound on the quadratic cost, the so called quantum Cramér-Rao bound (QCRB), and its relation to the fidelity. The QCRB is a matrix inequality which is in general non-attainable. However there is a related bound that one can expect to be saturated asymptotically: the Holevo bound. A scheme that attains this bound is asymptotically optimal from the pointwise perspective.

The pointwise approach relies on the fact that for large N only quadratic cost functions become relevant. By appropriate algebraic manipulations and averaging over the prior distribution one can compare this approach with the Bayesian one in the asymptotic limit. It is proved rigorously in [19] that the averaged Holevo bound leads to an asymptotic upper bound to the globally optimal fidelity for "smooth" qubit estimation problems, and for "smooth" pure state estimation problems. (We have a lucky coincidence for qubits, and for pure states, that fidelity can be expressed as a quadratic form in the estimation error of certain parameters of the state.) One can expect this bound to be asymptotically valid in general, but no rigorous proof has been given yet.

As to whether or not the averaged Holevo bound is asymptotically saturated: there exist very good heuristic arguments that this should be true, but no rigorous proof. (Unpublished work of M. Hayashi: for large N the estimation problem can be approximated, around a point obtained by a preliminary rough estimate, by a Gaussian state estimation problem, for which the Holevo bound is attained by an appropriate generalized heterodyne measurement).

In Sec. III A we derived the optimal global scheme for 3D states and showed that it is the same for any isotropic prior distribution. From the previous considerations we expect it also to be asymptotically optimal in the pointwise sense. We will show that this is indeed the case, since the optimal fidelity does coincide asymptotically with the averaged Holevo bound.

For 2D states the situation is more complex. Recall that the scheme defined by (3.37) is not optimal for arbitrary N and general isotropic priors. Nevertheless, Eq. (4.15) also coincides with the averaged Holevo bound. This comes close to a proof of the asymptotic optimality of the scheme. A rigorous proof (see Appendix H) can be derived from the van Trees inequality [20] (the same inequality is used to get the more general results in [19]). Thus our approximate solution (3.37) is asymptotically optimal both from the global and from the pointwise points of view.

Both the 3D and the 2D cases confirm the conjectures that the averaged Holevo bound is a sharp asymptotic bound for fidelity, and that the global optimal scheme is also asymptotically optimal in the pointwise sense. Global asymptotic optimality does not depend on the prior or on non-local features of the figure-of-merit.

Before stating the main results, we need to introduce a bit of notation. Let ρ be a density matrix parametrized by $\boldsymbol{\theta} \equiv (\theta_1, \theta_2, \ldots, \theta_p) \in \Theta \subset \mathbb{R}^p$, where p is the number of parameters.[2] Just as in the previous sections, let us assume we perform a generalized measurement O on an arbitrary state $\rho(\boldsymbol{\theta})$. Recall that such measurement is represented by a POVM $O = \{O_\chi\}$, where $\chi \in \Omega$ labels the various outcomes. Let $\hat{\boldsymbol{\theta}}_\chi$ be the estimate (or guess) of $\boldsymbol{\theta}$ based on the outcome χ, i.e., $\hat{\boldsymbol{\theta}}$ is a mapping from the outcome set Ω to the parameter space Θ:

$$\begin{aligned} \hat{\boldsymbol{\theta}} : \Omega &\rightarrow \Theta \\ \chi &\mapsto \hat{\boldsymbol{\theta}}_\chi. \end{aligned} \tag{5.1}$$

A natural way of quantifying the performance of an estimator $\hat{\boldsymbol{\theta}}$ and a measurement O at a point $\boldsymbol{\theta}_0$ is provided by the mean square error matrix (MSE) defined by the matrix elements

$$\begin{aligned} V_{\alpha\beta}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}) &\equiv \mathbb{E}_{\boldsymbol{\theta}_0}[(\hat{\theta}_\alpha - \theta_{0\alpha})(\hat{\theta}_\beta - \theta_{0\beta})] \\ &= \sum_{\chi\in\Omega} p(\chi|\boldsymbol{\theta}_0)\ (\hat{\theta}_{\chi\alpha} - \theta_{0\alpha})(\hat{\theta}_{\chi\beta} - \theta_{0\beta}), \end{aligned} \tag{5.2}$$

where the dependence on O is understood to simplify the notation and, naturally, $p(\chi|\boldsymbol{\theta}_0) = \mathrm{tr}\,[\rho(\boldsymbol{\theta}_0)O_\chi]$. In the remaining sections of the paper $\mathbb{E}_{\boldsymbol{\theta}_0}[f]$ stands for the expectation value of f with respect to the probability distribution $p(\chi|\boldsymbol{\theta}_0)$.

An estimator is said to be *locally unbiased* (LU) at $\boldsymbol{\theta}_0$ if

$$\mathbb{E}_{\boldsymbol{\theta}_0}[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta}_0, \qquad \partial_\alpha \mathbb{E}_{\boldsymbol{\theta}}[\hat{\theta}_\beta]\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \delta_{\alpha\beta}, \tag{5.3}$$

where ∂_α is shorthand for $\partial/\partial\theta_\alpha$. Intuitively, these conditions mean that, on average, the estimator is close to the truth in a small neighborhood of $\boldsymbol{\theta}_0$. When these conditions are satisfied for all possible values of $\boldsymbol{\theta}_0$, the estimator is said to be *uniformly unbiased*, or, simply, *unbiased*. LU estimators play a fundamental role in the pointwise approach.

The Fisher information matrix (FI) is defined as

$$\begin{aligned} I_{\alpha\beta}(\boldsymbol{\theta}) &\equiv \mathbb{E}_{\boldsymbol{\theta}}\left[\partial_\alpha \ln p(\chi|\boldsymbol{\theta})\ \partial_\beta \ln p(\chi|\boldsymbol{\theta})\right] \\ &= \sum_{\chi\in\Omega} \frac{\partial_\alpha p(\chi|\boldsymbol{\theta})\ \partial_\beta p(\chi|\boldsymbol{\theta})}{p(\chi|\boldsymbol{\theta})}. \end{aligned} \tag{5.4}$$

Note that the FI depends on a specific measurement O, through the probabilities $p(\chi|\boldsymbol{\theta})$.

With the above few definitions we can already give a first important result: The Cramér-Rao bound (CRB). It states that the MSE of an estimator $\hat{\boldsymbol{\theta}}$ LU at $\boldsymbol{\theta}_0$ is lower bounded by the inverse of the FI, namely,

$$V(\boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}) \geq I(\boldsymbol{\theta}_0)^{-1}. \tag{5.5}$$

In spite of its fundamental character, the CRB has the drawback that the bound it provides refers to a particular measurement, not necessarily optimal. To go around this difficulty, some new definitions are required

The symmetric logarithmic derivative (SLD), denoted by $\lambda_\alpha(\boldsymbol{\theta})$ (recall that $\alpha = 1, 2, \ldots, p$), is defined as the self-adjoint matrix that satisfies

$$\partial_\alpha \rho(\boldsymbol{\theta}) = \frac{\rho(\boldsymbol{\theta})\lambda_\alpha(\boldsymbol{\theta}) + \lambda_\alpha(\boldsymbol{\theta})\rho(\boldsymbol{\theta})}{2}. \tag{5.6}$$

[2] In the 3D case $p = 3$, $\boldsymbol{\theta} = (r, \theta, \phi)$ and $\Theta = [0, 1] \times [0, \pi] \times [0, 2\pi)$. In the 2D case $p = 2$, $\boldsymbol{\theta} = (r, \theta)$ and $\Theta = [0, 1] \times [0, 2\pi)$.

The SLDs for the 2D and 3D cases (2D and 3D models in pointwise terminology) are given in Appendix G. With this we can now define the quantum Fisher information matrix (QFI) as

$$H_{\alpha\beta}(\boldsymbol{\theta}) = \mathrm{Re}\,\mathrm{tr}\,\rho(\boldsymbol{\theta})\lambda_\alpha(\boldsymbol{\theta})\lambda_\beta(\boldsymbol{\theta}). \tag{5.7}$$

E.g., for the two models studied in this paper the QFIs are

$$H_{3\mathrm{D}}=\begin{pmatrix} \dfrac{1}{1-r^2} & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2\sin^2\theta \end{pmatrix}; H_{2D}=\begin{pmatrix} \dfrac{1}{1-r^2} & 0 \\ 0 & r^2 \end{pmatrix}. \tag{5.8}$$

The second important result of this section, due to Braunstein and Caves [21], states that for a given model all FIs are bounded from above by the QFI, i.e.,

$$I(\boldsymbol{\theta}_0) \leq H(\boldsymbol{\theta}_0) \quad \text{for all } \{O_\chi\}, \tag{5.9}$$

from which it immediately follows the QCRB:

$$V(\boldsymbol{\theta}_0,\hat{\boldsymbol{\theta}}) \geq H(\boldsymbol{\theta}_0)^{-1} \quad \text{for all } \{O_\chi\}. \tag{5.10}$$

Although these bounds are measurement-independent —they depend only on the signal states and the geometric properties of the space they belong to— they have the drawback of not being always attainable.

We have seen above that $H(\boldsymbol{\theta}_0)$ provides information on how small the variance of an estimator can be at $\boldsymbol{\theta}_0$. There is still another remarkable property of the QFI that we will need below: its direct relation to the fidelity [14]. Indeed, from its definition [see Eq. (2.3)],

$$f(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = \left(\mathrm{tr}\sqrt{\sqrt{\rho(\boldsymbol{\theta}_1)}\rho(\boldsymbol{\theta}_2)\sqrt{\rho(\boldsymbol{\theta}_1)}}\right)^2, \tag{5.11}$$

one obtains

$$f(\boldsymbol{\theta}_0,\boldsymbol{\theta}_0+\delta\boldsymbol{\theta}) = 1-\frac{1}{4}H_{\alpha\beta}(\boldsymbol{\theta}_0)\delta\theta_\alpha\delta\theta_\beta+\ldots, \tag{5.12}$$

where the components of $\delta\boldsymbol{\theta}$ are assumed to be small (neighboring states). Given a scheme, characterized by $(\{O_\chi\},\hat{\boldsymbol{\theta}})$, the average of the fidelity over all possible outcomes is

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\theta}_0} f(\boldsymbol{\theta},\hat{\boldsymbol{\theta}}) &= \sum_{\chi\in\Omega}\mathrm{tr}\,[\rho(\boldsymbol{\theta}_0)O_\chi]\; f(\boldsymbol{\theta}_0,\hat{\boldsymbol{\theta}}_\chi) \\ &= 1-\frac{1}{4}\,\mathrm{Tr}\,H(\boldsymbol{\theta}_0)V(\boldsymbol{\theta}_0,\hat{\boldsymbol{\theta}})+\ldots. \end{aligned}\tag{5.13}$$

Our aim is, therefore, to minimize the cost

$$\mathrm{tr}\,H(\boldsymbol{\theta}_0)V(\boldsymbol{\theta}_0,\hat{\boldsymbol{\theta}}). \tag{5.14}$$

An optimal measurement, O_{opt}, is thus the one that minimizes (5.14).

The formalism and results presented so far are completely general and apply to any model, i.e., to any family of states $\rho(\boldsymbol{\theta})$. We now need to introduce the so called N-copy model. It is defined by the set of density matrices $\rho^N(\boldsymbol{\theta})$ of the form

$$\rho^N(\boldsymbol{\theta}) = [\rho(\boldsymbol{\theta})]^{\otimes N}. \tag{5.15}$$

The "original" family, $\rho(\boldsymbol{\theta})$, is sometimes referred to as the single-copy quantum model. Naturally, we can talk about the variance or MSE of an estimation of the N-copy model, which we denote by $V^N(\boldsymbol{\theta}_0,\hat{\boldsymbol{\theta}})$. It is not hard to convince oneself that the cost Eq. (5.14) of the optimal scheme necessarily scales as $1/N$, for large enough N.[3] It is well-known in classical statistics [22] that under some regularity conditions the maximum likelihood (ML) estimator is asymptotically unbiased at $\boldsymbol{\theta}_0$ and its MSE is equal to $I^N(\boldsymbol{\theta}_0)^{-1}$, i.e., the ML estimator achieves the CRB asymptotically. It follows that for an optimal measurement $\mathrm{tr}\,H(\boldsymbol{\theta}_0)I^N(\boldsymbol{\theta}_0)^{-1}$ provides an attainable bound to the cost and it will scale as $1/N$ asymptotically. This lower bound on (5.14) can be expressed as

$$\frac{\mathrm{tr}\,H(\boldsymbol{\theta}_0)\bar{I}^N(\boldsymbol{\theta}_0)^{-1}}{N}+o(1/N), \tag{5.16}$$

where $\bar{I}^N = I^N/N$ is called the normalized FI. Likewise, for the asymptotic fidelity we have

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\theta}_0} f^N(\boldsymbol{\theta}_0,\hat{\boldsymbol{\theta}}_{\mathrm{ML}}) &= 1-\frac{\mathrm{tr}\,H(\boldsymbol{\theta}_0)\bar{I}^N(\boldsymbol{\theta}_0)^{-1}}{4N} \\ &\quad +o(1/N).\end{aligned}\tag{5.17}$$

which means that our optimization problem amounts to finding a measurement O_{opt} that minimizes $\mathrm{tr}\,H(\boldsymbol{\theta}_0)\bar{I}^N(\boldsymbol{\theta}_0)^{-1}$. We next present a powerful measurement-independent bound to this expression; the so called Holevo bound.

Let G be a positive semi-definite matrix and

$$C^N_{\boldsymbol{\theta}_0}(G) = \min_{\substack{\{(O \text{ on } \rho^N,\hat{\boldsymbol{\theta}})\} \\ \text{LU at } \boldsymbol{\theta}_0}} \mathrm{tr}\,GV^N(\boldsymbol{\theta}_0,\hat{\boldsymbol{\theta}}), \tag{5.18}$$

where the minimization is over all pairs $(O,\hat{\boldsymbol{\theta}})$ of measurements on $\rho^N(\boldsymbol{\theta})$ and estimators for which the latter is LU at $\boldsymbol{\theta}_0$ (the unbiasedness of an estimator depends on the measurement through its outcome probability distribution). Eq. (5.18) is relevant to the problem we are dealing with because its right hand side can be shown to give the $1/N$ term in (5.16) and (5.17) if $G = H(\boldsymbol{\theta}_0)$. In Ref. [1] Holevo proved the following bound:

$$C^1_{\boldsymbol{\theta}_0}(G) \;\geq\; C^H_{\boldsymbol{\theta}_0}(G), \tag{5.19}$$

[3] Just consider a scheme consisting of N identical measurements on each copy $\rho(\boldsymbol{\theta})$. By definition the cost of the optimal scheme is less than or equal to the cost of the former, which obviously scales as $1/N$. This sets a bound on the cost of the latter that also scales as $1/N$.

where

$$C^H_{\boldsymbol{\theta}_0}(G) \;=\; \min_{\boldsymbol{X}\in\Xi_{\boldsymbol{\theta}_0}} \Big\{ \operatorname{tr} G \operatorname{Re} Z[\boldsymbol{X}] \\ \;+\; \operatorname{tr} \left|\sqrt{G} \operatorname{Im} Z[\boldsymbol{X}]\sqrt{G}\right| \Big\}. \qquad (5.20)$$

In this expression $\boldsymbol{X} = (X_1, X_2, \ldots, X_p)$ are hermitian matrices satisfying the following relations

$$\operatorname{tr} \rho(\boldsymbol{\theta}_0) X_\alpha \;=\; 0, \qquad (5.21)$$

$$\operatorname{tr} \partial_\alpha \rho(\boldsymbol{\theta}_0) X_\beta \;=\; \delta_{\alpha\beta}. \qquad (5.22)$$

The minimization in (5.20) is over the set $\Xi_{\boldsymbol{\theta}_0}$ of all such $\boldsymbol{X}$. Finally, $Z[\boldsymbol{X}]$ is the $p \times p$ matrix whose elements are given by

$$Z_{\alpha\beta}[\boldsymbol{X}] = \operatorname{tr} \rho(\boldsymbol{\theta}_0) X_\alpha X_\beta. \qquad (5.23)$$

Although the Holevo bound (5.19) is not attainable but for a few simple exceptions, unpublished work by M. Hayashi shows that it is *asymptotically* attainable, i.e.,

$$\lim_{N\to\infty} N C^N_{\boldsymbol{\theta}_0}(G) = C^H_{\boldsymbol{\theta}_0}(G), \qquad (5.24)$$

as previously mentioned in this section. It is important to point out here that practical use of Hayashi's construction would require a two-step measurement in order to saturate the bound. This is necessary because the optimal measurement and LU estimator at $\boldsymbol{\theta}_0$ depend themselves on $\boldsymbol{\theta}_0$, which we do not know beforehand. To overcome this difficulty, one takes an asymptotically vanishing fraction of copies, say $\sqrt{N}$, and makes an initial estimate of the parameter $\hat{\boldsymbol{\theta}}_{\rm ini}$. Then, on the remaining copies one performs the measurement that is optimal at $\hat{\boldsymbol{\theta}}_{\rm ini}$. Therefore, (5.17) and (5.24) lead us to expect that the optimal asymptotic fidelity is given by

$$\mathbb{E}_{\boldsymbol{\theta}_0} f^N(\boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}_{\rm ML}) = 1 - \frac{1}{4N} C^H_{\boldsymbol{\theta}_0}[H(\boldsymbol{\theta}_0)] + o(1/N). \quad (5.25)$$

We next apply these results to the 3D and 2D models.

1. Holevo bound for the 3D case

In this case $p = 3$ and it is not hard to show (see Appendix G) that there is only one "vector" of matrices $\boldsymbol{X} = (X_r, X_\theta, X_\phi)$ in $\Xi_{\boldsymbol{\theta}}$ and no minimization is thus required in (5.20). The Holevo bound is straightforwardly computed to be

$$C^H[H(\boldsymbol{\theta}_0)] = 3 + 2r, \qquad (5.26)$$

and (5.17) becomes

$$\mathbb{E}_{\boldsymbol{\theta}_0} f^N(\boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}_{\rm ML}) = 1 - \frac{3+2r}{4N} + o(1/N). \qquad (5.27)$$

Furthermore, we expect this result to hold regardless on whether the ML estimator or the optimal guess is used. This implies that for a "well behaved" prior, one should have (4.17) by simply averaging (5.27), and we re-obtain the result of the the preceding section, which was computed using the Bayesian approach, with much less effort. Eq. (5.27) was also obtained by Matsumoto and Hayashi [12] with an estimation strategy similar to the one developed in Section III A.

2. Holevo bound for the 2D case

In the 2D model the SLDs satisfy

$$\operatorname{Im} \operatorname{tr} \rho(\boldsymbol{\theta}_0)\, \lambda_\alpha(\boldsymbol{\theta}_0)\, \lambda_\beta(\boldsymbol{\theta}_0) = 0. \qquad (5.28)$$

It is not difficult to check that in this situation the QCRB is asymptotically attainable,[4] i.e.,

$$C^H_{\boldsymbol{\theta}_0}(G) \;=\; \operatorname{tr} G H(\boldsymbol{\theta}_0)^{-1}. \qquad (5.29)$$

Indeed, the choice $X_\alpha = \sum_\beta H^{-1}_{\alpha\beta}(\boldsymbol{\theta}_0)\lambda_\beta(\boldsymbol{\theta}_0)$ achieves this. Hence $C^H_{\boldsymbol{\theta}_0}[H(\boldsymbol{\theta}_0)] = 2$ and

$$\mathbb{E}_{\boldsymbol{\theta}_0} f^N(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_{\rm ML}) = 1 - \frac{1}{2N} + o(1/N), \qquad (5.30)$$

from which (4.15) follows for "well behaved" priors. This strongly supports the claim that the 2D measurement scheme defined by Eq. (3.37) is indeed asymptotically optimal. The Appendix H contains the rigorous proof.

VI. CONCLUSIONS

We have presented a detailed analysis of the optimal estimation of qubit mixed states given a number N of identical copies. Our results apply to arbitrary N, finite or asymptotically large.

For general states (3D) we have obtained that the structure of the optimal measurement is based on the decomposition of the signal states in irreducible blocks under the action of the symmetric group. The scheme is essentially unique, valid for *any* isotropic prior distribution and *any* number of copies. This optimal scheme has the nice property that it can be regarded as two independent protocols performed sequentially: that for estimating the purity r of the state and that for estimating its orientation $\vec{n}$ in the Bloch sphere. It turns out that the estimation of the purity only exploits rotationally invariant properties of the signal states, and a measurement of the Casimir operator $\vec{J}^2 = j(j+1)\sum_\alpha \mathbb{1}_{j\alpha}$ is optimal. In other words, the estimate of r only depends on j, which characterizes the SU(2) invariant subspaces. This should

[4] A theorem by Matsumoto [12] states that the QCRB is asymptotically attainable if and only if (5.28) holds.

not come as a surprise since the purity itself is rotationally invariant and so are the priors considered here. The estimation of the orientation is formally equivalent to pure state estimation with $2j$ copies. As an illustration of our procedure, we have obtained closed expressions of the fidelity for the particularly important Bures prior. Results for other priors can be easily obtained with the techniques presented here.

In 2D, if one wants to do precisely optimal estimation for any N, there is a subtle interplay between the estimation of the purity and the estimation of the phase and they are no longer independent, although they are *asymptotically* so. Also contrasting with 3D is that the structure of the optimal POVM depends on the prior. The roots of this unconventional behavior lie in the different group structure of 2D states. Here the relevant group is U(1) [instead of $SU(2)$] and j is not the only invariant; the magnetic number m is also invariant under U(1). Actually, the interplay purity-phase can be traced back to this symmetry property. In spite of these difficulties, we have reduced the problem of obtaining the optimal POVM for any isotropic prior to a rather trivial maximization problem [recall Eq. (3.34)]. We have also obtained a prior independent POVM that is indistinguishable from the optimal one for any practical purposes. Furthermore, it separates purity and phase estimation exactly for all N and is asymptotically optimal.

The asymptotic behaviour of the estimation procedure has also been a central issue of our work. The asymptotic fidelity in 3D has the simple form $F = 1-(3+2\langle r\rangle)/(4N)$, where $\langle r\rangle$ is the mean purity with respect to the prior. This result is proved here for isotropic priors within our Bayesian approach. It is worth emphasizing that so far the asymptotic expression was only known for the particular case of the Bures prior [8]. In 2D, the asymptotic fidelity computed with the fixed POVM described above is simply $F = 1 - 1/(2N)$, independently of the prior.

We have studied the asymptotic behavior also from the pointwise approach, which is far more common among statisticians. The main advantage of the pointwise approach over the Bayesian one is that it provides bounds on the asymptotic mean square error (as well as on any other quadratic loss function) that can be easily computed. These bounds correspond, by second order expansion of the figure-of-merit, to bounds on the average fidelity which can be shown to be rigorous in many cases ([19]), including those studied in this paper. The drawback of the approach is that though one can heuristically expect these bounds to be asymptotically sharp, and one can propose two-stage measurement schemes which can be hoped to do the job, a lot of hard work is needed in each case to prove that they can be achieved. In contrast with the 3D case where all the results we have worked out from the Bayesian approach are rigorous, the optimality in the asymptotic regime of the 2D estimation scheme defined by (3.37) or (3.38) required some further work. Here we used the pointwise approach to fill the gap. The application of the van Trees inequality [20] to 2D in Appendix H yields the asymptotic bound on the fidelity in a particularly elegant and straightforward way. In turn, this bound provides the optimality proof.

Altogether, the fact that the results obtained from the pointwise approach coincide with those derived from the Bayesian framework give further strong support for the heuristic principle that the averaged lower bound from the pointwise approach is an asymptotically sharp lower bound for the global approach; and moreover that the chosen prior distribution and to a lesser extent, figure-of-merit, has asymptotically little impact on the behaviour of the solution.

There are two extensions of our work that can be readily addressed. Here, we have considered the full estimation of a qubit mixed state, however for some applications only partial knowledge of the state, such as its purity or its orientation, may be required. The techniques developed in this work can be easily adapted to these situations (see [18] and [23]). A second line of work concerns the use of more realistic measurements, in particular those that can be implemented with current technology. In this work we have considered the most general measurements allowed by Quantum Mechanics. They yield the maximum theoretical accuracy that can possibly be achieved, and thus provide a bound (and a measuring rod) for the accuracy of any other estimation scheme. However, they involve joint operations on the whole sample of states that in general are difficult to implement in a laboratory. It is thus of great practical relevance to study schemes based on local von Neumann measurements. Preliminary results, were presented in [8]. There, it was found that, for some tomographic schemes, the rate at which the fidelity approaches unity for a Bures prior distribution is $1 - F \sim 1/N^{3/4}$, i.e., there is a *qualitative* difference with the optimal measurements. Present work in progress suggests that by using classical communication the precision rate can be similar to the optimal collective scheme $1 - F \sim 1/N$, but the coefficient of the $1/N$ term is strictly larger than the optimal one, and corresponds to the result from the pointwise approach obtained in [3].

Acknowledgments

We thank M. Baig for his collaboration at early stages of this work. We acknowledge financial support from Spanish Ministry of Science and Technology project BFM2002-02588, CIRIT project SGR-00185, Netherlands Organization for Scientific Research NWO project 613.003.047, the European Community projects QUPRODIS contract no. IST-2001-38877 and RESQ contract no IST-2001-37559. E. B. thanks EURANDOM for hospitality.

APPENDIX A: BLOCK-DIAGONAL FORM OF $\rho^{\otimes N}$

One may use the symmetric group S_N to write $\rho^{\otimes N}$ in the block-diagonal form (2.10), much in the same way as it is used to obtain the SU(2) Clebsch-Gordan decomposition

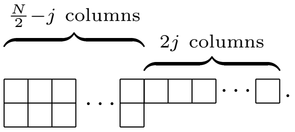
$$\left(\tfrac{\mathbf{1}}{\mathbf{2}}\right)^{\otimes N} = \bigoplus_{j=0,1/2}^{J} n_j \mathbf{j} \quad (J = N/2) \tag{A1}$$

(the multiplicity, n_j, is computed in Appendix B). However, at variance with the SU(2) case, where all Young frames have a single row, we here must also consider those with two rows, because

$$\det\rho = \frac{1-r^2}{4} \tag{A2}$$

(instead of unity). Hence, each two-box column of a frame contributes a multiplicative factor $\det\rho$.

With this observation, one can easily obtain the expression of the blocks ρ_{Nj} as follows. A generic Young frame with N boxes has the shape



(A3)

Each of the $N/2 - j$ double columns gives a factor $\det\rho$. The remaining $2j$ single columns correspond to a fully symmetric tensor on which SU(2) acts irreducibly. In the basis of the irreducible subspace of the representation $\mathbf{j}$, this tensor can be written as the matrix which we denote by ρ_j. Hence

$$\rho_{Nj} = \left(\frac{1-r^2}{4}\right)^{J-j} \rho_j. \tag{A4}$$

We now note that for $\vec{r} = r\vec{z}$ the matrices $\rho^{\otimes N}$, ρ_{Nj} and ρ_j, are all them diagonal and can thus be obtained without much effort. The result is

$$\rho_j = \sum_{m=-j}^{j} \left(\frac{1-r}{2}\right)^{j-m} \left(\frac{1+r}{2}\right)^{j+m} |jm\rangle\langle jm|. \tag{A5}$$

For arbitrary $\vec{r}$ covariance implies

$$\begin{aligned}\rho_j \;=\; &\sum_{m=-j}^{j} \left(\frac{1-r}{2}\right)^{j-m} \left(\frac{1+r}{2}\right)^{j+m} \times \\ &U(\vec{n})|jm\rangle\langle jm|U^\dagger(\vec{n}).\end{aligned} \tag{A6}$$

Notice that, in spite of what the notation might suggest, the matrices ρ_j are not proper density matrices, as $\operatorname{tr}\rho_j \neq 1$.

APPENDIX B: THE MULTIPLICITY OF THE REPRESENTATION j

Using Young tableaux techniques, there is a simple way to compute the multiplicity n_j, (2.11), with which the representation $\mathbf{j}$ shows up in the Clebsch-Gordan decomposition of $(\frac{\mathbf{1}}{\mathbf{2}})^{\otimes N}$ (this tensor product is denoted by $\Box^{\otimes N}$ in the present context).

The Young frame in (A3) can be denoted by $\lambda = [\lambda_1, \lambda_2] = [N/2 + j, N/2 - j]$ (this is a standard notation where λ_k is the number of boxes in the k-th row of the frame). This very same frame (A3) is equivalent to a single row of $2j$ boxes, i.e., to $[2j]$, which denotes the representation $\mathbf{j}$ of SU(2).

The recipe for computing SU(2) Clebsch-Gordan decompositions [24] applied to $\Box^{\otimes N}$ amounts to the following. First label N boxes each with an integer number from 1 to N. Then, starting with box number one and proceeding sequentially, build (and keep account of) all possible Young tableaux such that (i) they have at most two rows and (ii) the full sequence of integers formed by reading right to left in the first row and then in the second is *admissible*.[5] The number of occurrences of (A3) is precisely n_j. But the very same recipe gives us all *standard* Young tableaux[6] of shape $\lambda = [N/2 + j, N/2 - j]$. Hence n_j equals the number, f_λ, of such tableaux.

Recalling the Frobenius determinantal formula [25],

$$f_\lambda = N! \left\| \frac{1}{\lambda_k - k + l} \right\|, \tag{B1}$$

we get

$$n_j = N! \begin{vmatrix} \frac{1}{N/2+j} & \frac{1}{N/2+j+1} \\ \frac{1}{N/2-j-1} & \frac{1}{N/2-j} \end{vmatrix}. \tag{B2}$$

This determinant is readily seen to give (2.11).

APPENDIX C: CLOSED EXPRESSION OF THE FIDELITY USING A BURES PRIOR IN 3D

The explicit expressions of the coefficients v_j^0, v_j^z [Eqs. (3.19) and (3.18)] are

$$v_j^0 = 2\int_{-1}^{1} dr \frac{w(r)}{r} \left(\frac{1-r^2}{4}\right)^{J-j+\frac{1}{2}} \left(\frac{1+r}{2}\right)^{d_j} \tag{C1}$$

and

$$v_j^z = \eta_j - \nu_j, \tag{C2}$$

[5] A sequence of integers $p, q, r \ldots$ is admissible if at any point in the sequence at least as many 1's have occurred as 2's, at least as many 2's have occurred as 3's, etc.

[6] A Young tableaux is said to be *standard* if its labels increase from left to right along the files and from top to bottom along the columns.

with

$$\eta_j = \frac{d_j}{j+1}\int_{-1}^{1} dr \frac{w(r)}{r}\left(\frac{1-r^2}{4}\right)^{J-j}\left(\frac{1+r}{2}\right)^{d_j+1},$$
$$\nu_j = \int_{-1}^{1} dr \frac{w(r)}{r}\left(\frac{1-r^2}{4}\right)^{J-j}\left(\frac{1+r}{2}\right)^{d_j}. \qquad \text{(C3)}$$

To obtain these expressions we have recalled (2.12) and (2.13) and defined $w(r) = w(-r)$ for $-1 \le r < 0$ to extend the r-integration to the interval $[-1, 1]$.

Consider now the Bures prior [14], which is commonly regarded as the natural uniform distribution in the Bloch sphere, since it follows from the metric induced by the fidelity [9, 10]. It is given by

$$d\rho = \frac{4}{\pi}\frac{r^2 dr}{\sqrt{1-r^2}}\, dn, \qquad \text{(C4)}$$

which implies $w(r) = (4/\pi) r^2 (1-r^2)^{-1/2}$. In this case the integration in (C1) and (C3) can be performed analytically. For simplicity, we will consider an even number of copies $N = 2n$ ($J = n$). By making extensive use of

$$\int_{-1}^{1} \frac{r}{2}\left(\frac{1-r}{2}\right)^{\alpha-1}\left(\frac{1+r}{2}\right)^{\beta-1} = \frac{\beta-\alpha}{\beta+\alpha} B(\alpha,\beta), \qquad \text{(C5)}$$

where

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \qquad \text{(C6)}$$

is the standard Euler Beta function, we obtain

$$v_j^0 = \frac{8 d_j}{\pi(2n+3)} B(n-j+1, n+j+2). \qquad \text{(C7)}$$

Similarly,

$$\eta_j = \frac{8 d_j}{\pi(2n+3)} B(n-j+\tfrac{1}{2}, n+j+\tfrac{5}{2}), \qquad \text{(C8)}$$

and

$$\nu_j = \frac{4 d_j}{\pi(2n+2)} B(n-j+\tfrac{1}{2}, n+j+\tfrac{3}{2}), \qquad \text{(C9)}$$

which lead to

$$v_j^z = \frac{8 d_j j}{\pi}\frac{\Gamma(n-j+\frac{1}{2})\Gamma(n+j+\frac{3}{2})}{\Gamma(2n+4)}. \qquad \text{(C10)}$$

Putting the various pieces together we finally obtain the closed expression:

$$\Delta = \frac{4}{\pi}\sum_{j=0}^{n}\frac{2(2j+1)^2}{(2n+3)(2n+2)(2n+1)}$$
$$\times\sqrt{1+\left[\frac{j}{n+j+1}\frac{\Gamma(n-j+\frac{1}{2})\Gamma(n+j+\frac{3}{2})}{\Gamma(n-j+1)\Gamma(n+j+1)}\right]^2}. \text{(C11)}$$

APPENDIX D: COVARIANT POVMS FOR 2D STATES

For the sake of completeness, in this appendix we give a simple proof specialized to the 2D case of a more general result concerning the optimality of covariant (continuous) POVMs [1]. More precisely, we wish to prove that for any given POVM, $\{O_\chi\}$, there is always a covariant (continuous) one, with elements

$$\tilde{O}_{\chi\phi} = U(\phi)\Omega_\chi U^\dagger(\phi), \qquad \text{(D1)}$$

which gives the same average fidelity for a suitable positive operator Ω_χ. The proof goes as follows.

In the 2D case the average fidelity can be written as (in this section the integration limits 0 and 2π are understood)

$$F = \sum_\chi \int \frac{d\theta}{2\pi} f(\theta_\chi - \theta, R_\chi)\, \mathrm{tr}\,[\rho(\theta) O_\chi], \qquad \text{(D2)}$$

where θ_χ (θ) is the angle between $\vec{R}_\chi$ ($\vec{r}$) and the x-axis, and we denote the fidelity by $f(\theta_\chi - \theta, R_\chi)$ to emphasize the fact that in 2D it is a function of the difference of these two angles. Note also that we drop the explicit dependence on r which does not play any role in the proof. Thus, e.g., we denote the mixed state $\rho(\vec{r})$ simply as $\rho(\theta)$. Proving our statement amounts to proving that the POVM with elements and associated guess given by

$$\tilde{O}_{\chi\phi} = U(\phi-\theta_\chi) O_\chi U^\dagger(\phi-\theta_\chi) \xrightarrow{\text{guess}} \phi, R_\chi \qquad \text{(D3)}$$

gives the same fidelity as $\{O_\chi\}$. Note that (D3) defines Ω_χ in (D1) through

$$\begin{aligned}\tilde{O}_{\chi\phi} &= U(\phi)\left[U^\dagger(\theta_\chi) O_\chi U(\theta_\chi)\right] U^\dagger(\phi) \\ &\equiv U(\phi)\Omega_\chi U^\dagger(\phi). \qquad \text{(D4)}\end{aligned}$$

In formulæ we wish to prove that $F = \tilde{F}$, where

$$\tilde{F} = \sum_\chi \int \frac{d\phi}{2\pi}\frac{d\theta}{2\pi} f(\phi-\theta, R_\chi)\, \mathrm{tr}\,[\rho(\theta)\tilde{O}_{\chi\phi}] \qquad \text{(D5)}$$

is the fidelity we obtain with $\{\tilde{O}_{\chi\phi}\}$. We also have to prove that $\{\tilde{O}_{\chi\phi}\}$ in (D3) is indeed a POVM, namely, that

$$\sum_\chi \int \frac{d\phi}{2\pi}\tilde{O}_{\chi\phi} = \mathbb{1} \qquad \text{(D6)}$$

Let us start by proving (D6). We simply change variables $\phi \to \phi' = \phi - \theta_\chi$ and use the invariance of the U(1) Haar measure, which in this case is the trivial identity $\int_0^{2\pi} d\phi\, g(\phi) = \int_0^{2\pi} d\phi\, g(\phi+\alpha)$ satisfied by any periodic function g of period 2π. We have

$$\begin{aligned}\sum_\chi \int \frac{d\phi}{2\pi}\tilde{O}_{\chi\phi} &= \int \frac{d\phi'}{2\pi} U(\phi') \sum_\chi O_\chi\, U^\dagger(\phi') \\ &= \int \frac{d\phi'}{2\pi} U(\phi') U^\dagger(\phi') = \mathbb{1}. \qquad \text{(D7)}\end{aligned}$$

We use the same logic to prove that $F=\tilde{F}$:

$$\begin{aligned}\tilde{F} &= \sum_{\chi}\int\frac{d\theta}{2\pi}\frac{d\phi}{2\pi}f(\phi-\theta,R_{\chi})\\ &\times\ \mathrm{tr}\left[\rho(\theta)U(\phi-\theta_{\chi})O_{\chi}U^{\dagger}(\phi-\theta_{\chi})\right]\\ &= \sum_{\chi}\int\frac{d\theta}{2\pi}\frac{d\phi}{2\pi}f(\phi+\theta_{\chi}-\theta,R_{\chi})\\ &\times\ \mathrm{tr}\left[\rho(\theta)U(\phi)O_{\chi}U^{\dagger}(\phi)\right].\end{aligned}\tag{D8}$$

We now use that $U^{\dagger}(\phi)\rho(\theta)U(\phi)=\rho(\theta-\phi)$ and make the change of variable $\theta\to\theta-\phi$ to obtain $\tilde{F}=F$.

If $R_{\chi}=R$ for all χ (this is the case if the estimation of r is entirely based on j, as in the last part of Section III B), we can replace the POVM elements $\tilde{O}_{\chi\phi}$ by

$$\tilde{O}_{\phi}=\sum_{\chi}\tilde{O}_{\chi\phi}\xrightarrow{\text{guess}}\phi.\tag{D9}$$

This is equivalent to

$$\tilde{O}_{\phi}=U(\phi)\Omega U^{\dagger}(\phi),\tag{D10}$$

where the positive operator Ω can be expressed in terms of Ω_{χ} in (D1) simply as $\Omega=\sum_{\chi}\Omega_{\chi}$. The proof that achieves the same fidelity is straightforward and it amounts to pulling the sum over χ into or out of the trace in Eqs. (D5) and (D8), which we are entitled to do because we are assuming that R_{χ} is now independent of χ.

Using the results in Ref. [17], it is easy to show that for any given covariant (continuous) POVM with elements given by (D1) there is always a POVM with a *finite* number of elements $\bar{O}_{\phi_a}=U(\phi_a)\Omega U^{\dagger}(\phi_a)$, $a=0,1,2,\ldots M-1$, which achieves the same fidelity for a suitably large M. The angles ϕ_a can be chosen to be $\phi_a=2\pi a/M$, $a=0,1,2,\ldots M-1$.

APPENDIX E: COMPUTATION OF THE COEFFICIENTS a_j AND b_j

In this Appendix we give an approximation to c^j_m, defined in Eq. (3.42), of the form $c^j_m\approx a_j+b_j(m-j)$ valid for $m\approx j$ large enough.

Recalling the Wigner formula

$$\begin{aligned}&\mathrm{d}^{(j)}_{mm'}(\theta)=\sqrt{(j+m)!(j-m)!(j+m')!(j-m')!}\\ &\times\sum_{i=0}^{2j+1}\frac{(-1)^i\left(\cos\frac{\theta}{2}\right)^{2j+m'-m-2i}\left(-\sin\frac{\theta}{2}\right)^{m-m'+2i}}{(j-m-i)!(j+m'-i)!(i+m-m')!i!},\end{aligned}\tag{E1}$$

we obtain

$$\begin{aligned}c^j_j&=\frac{1}{2^{2j}}\sum_{m=-j}^{j}\binom{2j}{j-m}\sqrt{\frac{j-m}{j+m+1}},\\ c^j_{j-1}&=\sum_{m=-j}^{j}\binom{2j}{j-m}\sqrt{\frac{j-m}{j+m+1}}\frac{m(1+m)}{2^{2j-1}j}.\end{aligned}\tag{E2}$$

We note that the two coefficients c^j_j and c^j_{j-1} are binomial sums modulated by smooth functions of m in a neighborhood of $m=0$. More precisely,

$$c^j_k=\sum_{m=-j}^{j}\binom{2j}{j-m}\frac{1}{2^{2j}}\varphi_k(m),\tag{E3}$$

where $\varphi_k(m)$, which can be read off from (E2) for $k=j$, $j-1$, can be Taylor-expanded at $m=0$. For large j this expansion is

$$\begin{aligned}\varphi_j(m)&=1-\frac{1}{2j}-\frac{m}{j}+\frac{m^2}{2j^2}+\mathcal{O}(j^{-3/2}),\\ \varphi_{j-1}(m)&=\frac{2m}{j}+\left(\frac{2}{j}-\frac{3}{j^2}\right)m^2\\ &-\frac{2m^3}{j^2}+\frac{m^4}{j^3}+\mathcal{O}(j^{-3/2}).\end{aligned}\tag{E4}$$

Here the power counting is done by noticing that m is order $\sqrt{j}$, since the sum

$$S_q=\sum_{m=-j}^{j}\binom{2j}{j-m}\frac{m^q}{2^{2j}}\tag{E5}$$

is $\mathcal{O}(j^{q/2})$ for q even and vanishes for q odd, as is well known. In particular, we have $S_0=1$, $S_2=j/2$, $S_4=j(3j-1)/4$.

With all this information we obtain $c^j_j=1-1/(4j)$, $c^j_{j-1}=1-3/(4j)$, and finally have

$$\begin{aligned}c^j_m&=c^j_j+\left(c^j_j-c^j_{j-1}\right)(m-j)+\mathcal{O}[(m-j)^2]\\ &=\frac{1}{2}\left(1-\frac{1}{2j}\right)+\frac{m}{2j}+\mathcal{O}[(m-j)^2].\end{aligned}\tag{E6}$$

APPENDIX F: EXPLICIT COMPUTATION OF THE ASYMPTOTIC FIDELITY

Here we present with some detail the procedure we have used to evaluate the sum of (4.13) in the large $N=2n$ limit. We first focus on 2D states and later comment on the main differences with 3D.

In the two cases, we write n_j as the right hand side of the identity

$$\frac{d_j}{n+j+1}\binom{2n}{n-j}=\binom{2n}{n-j}-\binom{2n}{n+j+1}.\tag{F1}$$

1. The 2D case

After plugging Eqs. (3.43) and (4.2) into Eq. (4.13), we have

$$\Delta_{\mathrm{2D}}=\sum_{j=0}^{n}\left[\binom{2n}{n-j}-\binom{2n}{n+j+1}\right]$$

$$\times\left\{2\sqrt{1-\frac{j^2}{n^2}}\int_0^1 dr\frac{w(r)}{r}\right.$$
$$\times\left[\left(\frac{1-r}{2}\right)^{n-j+\frac{1}{2}}\left(\frac{1+r}{2}\right)^{n+j+\frac{3}{2}}-(r\to -r)\right]$$
$$+\int_0^1 dr\frac{w(r)}{r}\left[\frac{1}{n}\left(rj-\frac{1}{4}\right)\right.$$
$$\left.\left.\times\left(\frac{1-r}{2}\right)^{n-j}\left(\frac{1+r}{2}\right)^{n+j+1}-(r\to -r)\right]\right\}. \tag{F2}$$

We next multiply the powers of $(1\pm r)/2$ that are explicitly given in this equation by the first binomial. Likewise, we multiply those denoted by $(r\to -r)$ by the second binomial. In the resulting expressions, we next change the summation indexes according to $n-j=k$ and $n+j+1=k$, respectively, and do similar changes in the remaining crossed terms. After some algebra, we have

$$\Delta_{\rm 2D}=\frac{1}{n}\int_0^1 dr\frac{w(r)}{r}\left\{\frac{1+r}{2}\right.$$
$$\times\left[\sum_{k=0}^{n}B_k(r)\Phi_k(r)+\sum_{k=n+1}^{2n}B_k(r)\Phi_{k-1}(r)\right]$$
$$\left.-\frac{1-r}{2}[r\to -r]\right\}, \tag{F3}$$

where $B_k(r)$ is defined by

$$B_k(r)=\binom{2n}{k}\left(\frac{1-r}{2}\right)^{k}\left(\frac{1+r}{2}\right)^{2n-k} \tag{F4}$$

and

$$\Phi_k(r)=\sqrt{k(2n-k)(1-r^2)}+(n-k)r-\frac{1}{4}. \tag{F5}$$

Since the coefficients $B_k(r)$ are the terms of a binomial series, for large n only those for which $k\approx n(1-r)\le n$ (or equivalently $2n-k\ge n$) give a significant contribution to the fidelity, whereas the rest fall off exponentially with n. This enables us to expand the factor $\sqrt{(k-1)(2n-k+1)}$ in $\Phi_{k-1}(r)$ as a power series in $1/k$ and $1/(2n-k)$ and obtain the relation

$$\Phi_{k-1}(r)=\Phi_k(r)+\frac{\sqrt{1-r^2}}{2}$$
$$\times\left(\sqrt{\frac{k}{2n-k}}-\sqrt{\frac{2n-k}{k}}\right)+r+o(1/n) \tag{F6}$$

which we use in the second sum of (F3). We further define $\Psi_k(r)=\Phi_{k-1}(r)-\Phi_k(r)+o(1/n)$. It satisfies $\Psi_k(r)=-\Psi_{2n-k}(-r)$, as can be read off from (F6).

The leading contributions come from the terms that contain $\Phi_k(r)$, and the corresponding term in $[r\to -r]$. They combine into a single sum from $k=0$ to $k=2n$. The rest of the terms [those proportional to $\Psi_k(r)$ and $\Psi_k(-r)$] are subleading and can be simplified using the change of indexes $k\to 2n-k$. The result can be cast as

$$\Delta_{\rm 2D}=\frac{1}{n}\int_0^1 dr\frac{w(r)}{r}\left\{\left[\frac{1+r}{2}\sum_{k=0}^{2n}B_k(r)\Phi_k(r)\right.\right.$$
$$\left.\left.+\frac{1-r}{2}\sum_{k=0}^{n-1}B_k(r)\Psi_k(r)\right]-[r\to -r]\right\}. \tag{F7}$$

We readily see that the first sum (as well as the corresponding one obtained by the substitution $r\to -r$) is a binomial sum modulated by the function $\Phi_k(r)$, analogous to (E3) in Appendix E, and can be computed along the same line. This sum is peaked at $k\approx n(1-r)$, as we have already mentioned, which suggests expanding $\Phi_k(r)$ in powers of $k-n(1-r)$. More precisely, one can check that

$$\Phi_k(r)=n-\frac{1}{4}-\frac{[k-n(1-r)]^2}{2n(1-r^2)}+o(1/n) \tag{F8}$$

[the power counting is simply $k-n(1-r)=\mathcal{O}(\sqrt{n})$]. Recalling that the lowest moments, $S_q(r)=\sum_{k=0}^{2n}B_k(r)[k-n(1-r)]^q$, of the binomial series given by (F4) are $S_0(r)=1$, $S_2(r)=(n/2)(1-r^2)$ we obtain

$$\sum_{k=0}^{2n}B_k(r)\Phi_k(r)=n-1/2+o(1/n). \tag{F9}$$

To evaluate the second sum in (F7) we use again the approximation $B_k(r)\to\delta[k-n(1-r)]$ [see Eq. (4.5) and the comments below it], along with the substitution $\sum_{k=0}^{n-1}\to n\int_0^1 ds$, where $s=k/n$. This yields

$$\sum_{k=0}^{n-1}B_k(r)\Psi_k(r)=\mathcal{O}(1/n). \tag{F10}$$

The counterpart of (F10) in the term denoted by $[r\to -r]$, Eq. (F7), gives no contribution since $\delta[k-n(1+r)]$ lies outside the s-integration range. Collecting the various pieces we finally obtain

$$\Delta_{\rm 2D}=\left(1-\frac{1}{2n}\right)\int_0^1 dr\,w(r)+o(1/n). \tag{F11}$$

2. The 3D case

The 3D case is quite similar. Our starting point is now Eqs. (C1) and (C2). We proceed as above to obtain

$$v_j^z\;=\;\int_{-1}^1 dr\frac{w(r)}{r}\left\{\left[\frac{(2j+1)r-1}{2(j+1)}\right]\left(\frac{1-r}{2}\right)^{n-j}\right.$$
$$\left.\times\left(\frac{1+r}{2}\right)^{n+j+1}-(r\to -r)\right\} \tag{F12}$$

and a similar expression for v_j^0. From them we compute $\Delta_{\rm 3D}$ to be

$$\Delta_{\rm 3D}=\frac{1}{n}\sum_{j=0}^{n}\left[\binom{2n}{n-j}-\binom{2n}{n+j+1}\right]\int_{-1}^{1}\!dr\frac{w(r)}{r}\times\left\{\left[\sqrt{(n^2-j^2)(1-r^2)}+\frac{(2j+1)jr-j}{2(j+1)}\right]\times\left(\frac{1-r}{2}\right)^{n-j}\left(\frac{1+r}{2}\right)^{n+j+1}-[r\to -r]\right\}. \tag{F13}$$

This expression can be cast in the form of (F3), where now

$$\Phi_k(r) = \sqrt{k(2n-k)(1-r^2)}+(n-k)r - \frac{1}{2}\frac{\sqrt{n^2-k(2n-k)}}{\sqrt{n^2-k(2n-k)}+1}. \tag{F14}$$

One can check that $\Psi_k(r)$ is again defined by (F6) and $\Delta_{\rm 3D}$ can thus be expressed in the form (F7). The first sum is again Taylor-expanded about $k=n(1-r)$. Using the moments of the binomial series defined by (F4) and keeping only the relevant order we obtain

$$\sum_{k=0}^{2n}B_k(r)\Phi_k(r)=n-\frac{3+2|r|}{4}+\mathcal{O}(1/n). \tag{F15}$$

Note that we cannot drop the absolute value since the integral over r extends to the interval $[-1,1]$ [see, e.g. Eqs. (F12) and (F13)].

To evaluate the second sum in (F7) we proceed as in the previous 2D case, and find that (F10) still holds. Finally, we obtain

$$\Delta_{\rm 3D}=\int_0^1 dr\, w(r)\left(1-\frac{3+2r}{4n}\right)+o(1/n). \tag{F16}$$

APPENDIX G: SLDS AND $C^H[H(\boldsymbol{\theta}_0)]$ FOR THE 3D MODEL

The SLDs of the 3D model can be calculated to be

$$\lambda_r=\frac{1}{1+r}\frac{\mathbb{1}+\vec{n}\cdot\vec{\sigma}}{2}-\frac{1}{1-r}\frac{\mathbb{1}-\vec{n}\cdot\vec{\sigma}}{2}, \tag{G1}$$

$$\lambda_\theta=r\,\partial_\theta\vec{n}\cdot\vec{\sigma}, \tag{G2}$$

$$\lambda_\phi=r\,\partial_\phi\vec{n}\cdot\vec{\sigma}. \tag{G3}$$

[In this appendix we drop the arguments $\boldsymbol{\theta}=(r,\theta,\phi)$ and $\boldsymbol{\theta}_0$ wherever no confusion arises.] The two SLD of the 2D model, λ_r and λ_θ, are obtained by simply setting $\theta=\pi/2$ and then replacing ϕ by θ in the above expressions.

To compute $C^H(H)$ we first need $\boldsymbol{X}=(X_r,X_\theta,X_\phi)$, which are completely fixed by the conditions

$$X_\alpha = X_\alpha^\dagger, \tag{G4}$$

$$\operatorname{tr}\rho X_\alpha = 0 \tag{G5}$$

$$\operatorname{tr}\partial_\alpha\rho X_\beta = \delta_{\alpha\beta} \tag{G6}$$

Hermiticity, Eq. (G4), requires

$$X_\alpha=a_\alpha\mathbb{1}+\vec{b}_\alpha\cdot\vec{\sigma},\quad \alpha=r,\theta,\phi. \tag{G7}$$

The conditions (G5) yield

$$a_\alpha+\vec{b}_\alpha\cdot\vec{n}=0, \tag{G8}$$

and conditions (G6) give

$$\vec{b}_r = \vec{n}, \tag{G9}$$

$$\vec{b}_\theta = \frac{1}{r}\,\partial_\theta\vec{n}, \tag{G10}$$

$$\vec{b}_\phi = \frac{1}{r\sin^2\theta}\,\partial_\phi\vec{n}. \tag{G11}$$

These together with (G8) imply $a_r=-r$, $a_\theta=0$, and $a_\phi=0$. Hence, the only set of matrices satisfying (G4-G6) is

$$\begin{aligned}X_r &= -r\,\mathbb{1}+\vec{n}\cdot\vec{\sigma},\\ X_\theta &= \frac{1}{r}\,\partial_\theta\vec{n}\cdot\vec{\sigma},\\ X_\phi &= \frac{1}{r\sin^2\theta}\,\partial_\phi\vec{n}\cdot\vec{\sigma}.\end{aligned} \tag{G12}$$

To compute the Holevo bound we only need to take traces of the form $\operatorname{tr}\rho X_\alpha X_\beta$. A straightforward calculation gives

$$\operatorname{Re}Z[\boldsymbol{X}] = H_{\rm 3D}^{-1}, \tag{G13}$$

$$\operatorname{Im}Z[\boldsymbol{X}] = \begin{pmatrix}0 & 0 & 0\\ 0 & 0 & \dfrac{1}{r\sin\theta}\\ 0 & -\dfrac{1}{r\sin\theta} & 0\end{pmatrix}. \tag{G14}$$

Therefore

$$\operatorname{tr}H_{\rm 3D}H_{\rm 3D}^{-1} = 3, \tag{G15}$$

$$\operatorname{tr}\left|\sqrt{H_{\rm 3D}}\operatorname{Im}Z[\boldsymbol{X}]\sqrt{H_{\rm 3D}}\right| = 2r, \tag{G16}$$

and we obtain (5.26).

APPENDIX H: VAN TREES ASYMPTOTIC BOUND FOR 2D STATES

Let $\boldsymbol{\theta}$ be the column vector of the two real parameters r and θ of Sec. III B, which we use to parametrize the states on the equatorial plane of the Bloch sphere. Define $\boldsymbol{\psi}(\boldsymbol{\theta})=\frac{1}{2}\mathbf{r}(\boldsymbol{\theta})$ where $\mathbf{r}$ is the four-dimensional real vector (of length 1) introduced in Sec. II. By (2.4) we can now write

$$1-f(\boldsymbol{\theta}_0,\hat{\boldsymbol{\theta}})=\|\boldsymbol{\psi}(\boldsymbol{\theta}_0)-\boldsymbol{\psi}(\hat{\boldsymbol{\theta}})\|^2 \tag{H1}$$

showing that one minus the fidelity is the squared L_2 cost function for estimating $\boldsymbol{\psi}$. Taking the two states close to one another, and comparing with (5.12) shows that

$$\boldsymbol{\psi}'^{\top}\boldsymbol{\psi}' = \frac{1}{4}H \tag{H2}$$

where $\boldsymbol{\psi}'(\boldsymbol{\theta})$ denotes the 4×2 matrix of partial derivatives of $\boldsymbol{\psi}$ with respect to components of $\boldsymbol{\theta}$ and H is the QFI.

Let $\bar{I}^N = I^N/N$ denote the normalized FI for $\boldsymbol{\theta}$ based on an arbitrary collective measurement on the N copies, and let $\hat{\boldsymbol{\theta}}$ denote an arbitrary estimator of $\boldsymbol{\theta}$ based on that measurement. By $\mathbb{E}_w$ we denote averaging over $\boldsymbol{\theta}$ with respect to a prior probability density w over the equatorial plane. Then the van Trees inequality [20] states that, for any given matrix function $C(\boldsymbol{\theta})$ of size $\dim(\boldsymbol{\psi}) \times \dim(\boldsymbol{\theta})$, and under certain smoothness conditions on the probability distribution of the outcome of the measurements and on the prior w,

$$N\mathbb{E}_w\|\boldsymbol{\psi}(\boldsymbol{\theta}_0) - \boldsymbol{\psi}(\hat{\boldsymbol{\theta}})\|^2 \;\geq\; \frac{(\mathbb{E}_w \operatorname{tr} C\boldsymbol{\psi}'^{\top})^2}{\mathbb{E}_w \operatorname{tr} C\bar{I}^N C^{\top} + \frac{1}{N}\mathbb{E}_w \frac{(wC)'^{\top}(wC)'}{w^2}}, \tag{H3}$$

where by $(wC)'$ we denote the column vector of the same length as $\boldsymbol{\psi}$, with row elements $\sum_\beta \partial_\beta[w(\boldsymbol{\theta})C_{i\,\beta}(\boldsymbol{\theta})]$. By the Helstrom information inequality (5.9) we may bound $\bar{I}^N$ in the denominator by H (of the single-copy model). Without the "$1/N$" term in the denominator, the optimal choice of C would be $C = \boldsymbol{\psi}' H^{-1}$. Making this choice anyway gives

$$N\mathbb{E}_w\|\boldsymbol{\psi}(\boldsymbol{\theta}_0) - \boldsymbol{\psi}(\hat{\boldsymbol{\theta}})\|^2 \;\geq\; \frac{(\mathbb{E}_w \operatorname{tr} \boldsymbol{\psi}' H^{-1}\boldsymbol{\psi}'^{\top})^2}{\mathbb{E}_w \operatorname{tr} \boldsymbol{\psi}' H^{-1} H H^{-1}\boldsymbol{\psi}'^{\top} + \frac{1}{N}\mathbb{E}_w \frac{(wC)'^{\top}(wC)'}{w^2}}. \tag{H4}$$

Hence, provided the second term in the denominator is finite, by further substituting $\boldsymbol{\psi}'^{\top}\boldsymbol{\psi}' = \frac{1}{4}H$ and letting N converge to infinity, we obtain

$$\liminf_{N\to\infty} N\mathbb{E}_w\mathbb{E}_{\boldsymbol{\theta}}(1 - f(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})) \;\geq\; \frac{1}{2}. \tag{H5}$$

The van Trees inequality requires some modest smoothness of the probability density of the measurement outcomes as function of $\boldsymbol{\theta}$, which are satisfied in our case since the density matrix $\rho^{\otimes N}(\boldsymbol{\theta})$ is a smooth function of $\boldsymbol{\theta}$. It requires smoothness of the prior density w and also that this density converges to zero at the boundary of its support. This last property does not hold for the priors in which we are interested. However, for a given prior w and for given $\epsilon > 0$ one can construct a prior w_ϵ which is zero outside a circle of radius strictly smaller than 1, which converges smoothly to zero at the boundary of its support, and which is everywhere smaller than $(1+\epsilon)w$. The modification of w can simultaneously be done ensuring that the second term in the denominator of (H4) is finite. Since

$$\mathbb{E}_w\mathbb{E}_{\boldsymbol{\theta}}(1 - f(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})) \;\geq\; \frac{\mathbb{E}_{w_\epsilon}\mathbb{E}_{\boldsymbol{\theta}}(1 - f(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}))}{1+\epsilon} \tag{H6}$$

we can first derive (H5) with w replaced by w_ϵ, then let $\epsilon \to 0$, resulting in (H5) with the original w in place.

[1] A. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland Publishing, Amsterdam, 1982).

[2] K. R. Jones, Phys. Rev. A **50**, 3682 (1994); S. Massar and S. Popescu, Phys. Rev. Lett. **74**, 1259 (1995); Z. Hradil, Phys. Rev. A **55**, 1561(R) (1997); J. I. Latorre, P. Pascual and R. Tarrach, Phys. Rev. Lett. **81**, 1351 (1998); D. G. Fisher, S. H. Kienle and M. Freyberger, Phys. Rev. A **61**, 032306 (2000); Th. Hannemann *et al.*, Phys. Rev. A **65**, 050303 (2002); E. Bagan, M. Baig and R. Munoz-Tapia, Phys. Rev. Lett. **89**, 277904 (2002); F. Embacher and H. Narnhofer, Ann. of Phys (N.Y.) **311**, 220 (2004). E. Bagan, A. Monras and R. Munoz-Tapia, Phys. Rev. A **71**, 062318 (2005).

[3] R. D. Gill and S. Massar, Phys. Rev. A **61**, 042312 (2000).

[4] J. I. Cirac, A. K. Ekert and C. Macchiavello, Phys. Rev. Lett. **82**, 4344 (1999).

[5] G. Vidal *et al.*, Phys. Rev. A **60**, 010304 (1999).

[6] D G. Fischer and M Freyberger, Phys. Lett. A **273**, 293 (2000); H Mack, D G. Fischer and M Freyberger, Phys. Rev. A **62**, 042301 (2000).

[7] M. Keyl and R. F. Werner, Phys. Rev. A **64**, 52311 (2001).

[8] E. Bagan, M. Baig, R. Munoz-Tapia, and A. Rodriguez, Phys. Rev. A **69**, 010304 (2004).

[9] K. Zyczkowski and H. J. Sommers, Phys. Rev. A **71**, 032313 (2005).

[10] D. Petz and C. Sudar, J. Math. Phys. **37**, 2662 (1996).

[11] K. Zyczkowski and H. J. Sommers, J. Phys. A **34**, 7111 (2001); *ibid.* **37**, 8457, 2004.

[12] K. Matsumoto and M. Hayashi, (2004), quant-ph/0411073.

[13] K. Matsumoto *Uhlmann's parallelism in quantum estimation theory*, (1997), quant-ph/9711027.

[14] M. Hübner, Phys. Lett. A **163**, 239 (1992); R. Josza, J. Mod. Opt. **41**, 2315 (1994).

[15] C. A. Fuchs, PhD Dissertation, University of New Mexico, (1995) quant-ph/9601020.

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

[17] E. Bagan, M. Baig and R. Munoz-Tapia, Phys. Rev. A **64**, 022305 (2001); Phys. Rev. Lett. **87**, 257903 (2001).

[18] E. Bagan, M. A. Ballester, R. Munoz-Tapia and

O. Romero-Isart, Phys. Rev. Lett. **95**, 110504 (2005) and *e-print* `quant-ph/0505083`.

[19] R. D. Gill. *Asymptotic information bounds in quantum statistics. e-print* `math.ST/0512443`.

[20] R. D. Gill and B. Y. Levit, Bernouilli **1**, 59 (1995).

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

[22] P. J. Bickel and K. A. Doksum, *Mathematical Statistics. Basic Ideas and Selected Topics.* (Prentice Hall, New Jersey, 2001).

[23] E. Bagan et al., proceedings of the Erato Conference EQIS05, Tokyo, Japan 24-30 August 2005.

[24] S. Eidelman et al., Review of Particle Physics, Phys. Lett. B **592**, 1 (2004).

[25] M. Hamermesh, *Group Theory and its Applications to Physical Problems.* (Addison-Wesley, Massachusetts, 1962).