Beyond the swap test: optimal estimation of quantum state overlap

M. Fanizza,1 M. Rosati,2 M. Skotiniotis,2 J. Calsamiglia,2 and V. Giovannetti1
1NEST, Scuola Normale Superiore and Istituto Nanoscienze-CNR, I-56126 Pisa, Italy
2Física Teórica: Información i Fenòmens Quàntics, Departament de Física, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona) Spain

We study the estimation of the overlap between two Haar-random pure quantum states in a finite-dimensional Hilbert space, given $M$ and $N$ copies of them. We compute the statistics of the optimal measurement, which is a projection onto permutation-invariant subspaces, and provide lower bounds for the mean square error for both local and Bayesian estimation. In the former case, the bound is asymptotically saturable by a maximum likelihood estimator, whereas in the latter we give a simple exact formula for the optimal value. Furthermore, we introduce two LOCC strategies, relying on the estimation of one or both the states, and show that, although they are suboptimal, they outperform the commonly-used swap test. In particular, the swap test is extremely inefficient for small values of the overlap, which become exponentially more likely as the dimension increases. Finally, we show that the optimal measurement is less invasive than the swap test and study the robustness to depolarizing noise for qubit states.

Introduction.— The estimation of the overlap between two unknown quantum states is a basic primitive in quantum information processing, with applications ranging from quantum fingerprinting [1,3], entanglement estimation [4,7], communication without a shared reference frame [5,11] to quantum machine learning [12,20]. In most applications, overlap estimation (OvE) is carried out through the swap test [12,13,18]: given two systems in the state $|\psi\rangle|\phi\rangle$, the probability of projecting it on the permutation-invariant subspace is determined by the overlap between $|\psi\rangle$ and $|\phi\rangle$. By repeating this measurement on several pairs of copies one can obtain a good estimate of this probability, and hence of the overlap. It is then natural to ask whether, for the same number of copies, one could reach a larger accuracy via a collective strategy that extracts the relevant information using a joint and less-destructive measurement. In this article we answer in the positive, evaluating the ultimate precision attainable in the OvE of two pure quantum states, given a number of copies of each and assuming no prior knowledge about them.

OvE has been extensively studied in the past, as the overlap is an archetypical instance of quantum relative information [21–28]. Recently, with the advent of quantum machine learning protocols [14–20], OvE has attracted renewed interest as a fundamental primitive and its efficient implementation and generalization on near-term quantum computers are subjects of current research [12,13]. Furthermore, when the states are unlabelled, OvE constitutes in itself an instance of unsupervised quantum-classical learning problem, in a setting similar to [29]. Our aim in this paper is to construct a machine that classifies unlabelled quantum inputs according to their mutual overlap by minimizing the mean square error of the estimate.

The measurements optimizing the average information gain [8] and the minimum average error [9,10] have been derived for the case of qubits, with only numerical solutions [11] for higher dimensions. In this article, we tackle OvE in full generality, considering an arbitrary number of copies of each state (say $M$ and $N$ respectively) in any finite dimension, see Fig. 1.

We study the precision of the estimation, quantified by the mean square error, within both local and global (Bayesian) approaches. For local estimation, we provide an asymptotically achievable lower bound using the quantum Fisher information (QFI) [30,32], whereas for Bayesian estimation [31,33] we provide an exact solution using a Haar-uniform prior over the states, generalizing the results of [11]. We find that the optimal local strategies are also Bayesian-optimal asymptotically. We compare our results with the swap test and with two LOCC strategies based on estimating either one or both the species of states, see Fig. 1. Such strategies are useful in distributed scenarios where the two species are produced in distant laboratories. We show that in the limit $M + N \gg 1$, $|M − N|$ finite, when there is no preferred bounded reference frame, the global strategy displays a finite asymptotic gap with respect to the others. Moreover, we show that the optimal measurement is less invasive than the swap test and robust against local qubit noise.

Average state at fixed overlap.— The task we consider is as follows. Given $N$ and $M \geq N$ copies of two pure states $|\psi\rangle, |\phi\rangle$, provide a precise estimate $\hat{c}$
of their overlap \( c = |\langle \psi | \phi \rangle|^2 \). The pure states in question belong to an arbitrary Hilbert space \( \mathcal{H}_d \) of finite dimension \( d \), but for the moment we shall start considering the case of qubits, i.e., \( d = 2 \), since the result for arbitrary \( d \) is similar. As the states are unknown and the overlap is the only quantity we are interested in, the effective state of the system for a fixed value of \( c \) is given by

\[
\rho(c) = \int_{U \in \text{SU}(2)} dU U^\dagger \otimes (N+M) |\psi\rangle \langle \psi| U^{\otimes (N+M)},
\]

where \( U \) is an arbitrary reference-frame orientation and \( |\psi\rangle = |\psi\rangle^\otimes N \otimes |\phi\rangle^\otimes M \). We stress that \( \rho(c) \) contains the same information about the overlap as the fixed-reference state, which we may choose without loss of generality such that \( \langle \psi | = |\frac{1}{2}, \frac{1}{2}\rangle \) and \( |\phi\rangle = \sum \frac{1}{2} \delta_{\frac{1}{2}, \frac{1}{2}} (\frac{1}{2}, k) \langle \frac{1}{2}, k | \). Here, we use a basis for \( \mathcal{H}_2 \) labeled by the angular momentum \( j = \frac{1}{2} \) and its projection onto the \( z \)-axis, while \( d_{J,M}^{(N)}(\theta) \) are the matrix-elements of a Wigner small-d matrix. Note that in this basis the overlap is simply given by \( c = \sum \frac{1}{2} \langle \frac{1}{2}, k | \theta)^2 = \cos^2 \frac{1}{2} \).

Using the addition rules for angular momentum, the joint state can be written as

\[
|\Psi\rangle = \sum_{J = \text{min}}^{J = \text{max}} \sum_{M = \text{min}}^{M = \text{max}} C_J^M \frac{N}{2}, \frac{M}{2}, \frac{N}{2}, \frac{M}{2}, k \langle J, N, k | J, N, k \rangle.
\]

where \( C_J^M \frac{N}{2}, \frac{M}{2}, \frac{N}{2}, \frac{M}{2}, k \) are the Clebsch-Gordan coupling coefficients, \( J_{\text{min}} = \frac{M-N}{2} \) and \( J_{\text{max}} = \frac{M+N}{2} \). Plugging this expression in Eq. (1) and computing the average, results in a state \( \rho(c) \) that is diagonal in the total-angular-momentum basis, see the supplemental material (SM):

\[
\rho(c) = \sum_{J = \text{min}}^{J = \text{max}} p(J|c) \frac{1}{2J+1} |J\rangle \langle J|.
\]

with \( |J\rangle = \sum_{M = -J}^{M = J} |J, M\rangle \langle J, M| \) the projector on the subspace of total angular momentum \( J \)

\[
p(J|c) = \frac{(2J+1)N!M!(1-c)^M P_{J,M}^{(0,2\text{max})} \left( \frac{1+c}{1-c} \right) }{(J_{\text{max}}-J)!(J_{\text{max}}+1+J)!}
\]

where \( P_n^{(a,b)}(x) \) is the \( n \)-th degree Jacobi polynomial.

For generic \( d \), the average of \( \rho(c) \) is in general performed over all \( U \in \text{SU}(d) \). Nevertheless, and without loss of generality, we find it more convenient to first average over \( U \in \text{SU}(2) \) in the 2-dimensional subspace spanned by \( |\psi\rangle \) and \( |\phi\rangle \), and only after average over \( U \in \text{SU}(d) \). The resulting \( p(c) \) differs from Eq. (3) only in that \( |J\rangle \) are now projectors over the irreducible representations (irreps) \( U^{(J)} \) of \( \text{SU}(d) \) arising from the tensor product of two completely symmetric representations, with a proper normalization accounting for the dimension \( \chi_J(d) \) of \( U^{(J)} \). These irreps are still indexed by an (half)-integer \( J \in \left\{ \frac{M-N}{2}, \frac{M+N}{2} \right\} \). Importantly, note that the \( J \)-statistics \( p(J|c) \) is left unchanged and thus it does not depend on the dimension.

**Local estimation.**— Suppose that we perform a POVM \( \{E_k\} \) and wish to extract from the outcome \( k \) an unbiased estimator of the overlap, \( c(k), \) i.e.,

\[
\sum_k c(k) \text{Tr}[E_k \rho(c)] = c, \text{ whose mean square error is minimal.}
\]

The quantum Cramer-Rao bound \([32]\) places a lower bound on the Mean Square Error (MSE) of all unbiased quantum estimators as \( v(c) = \sum_k (c(k) - c)^2 \text{Tr}[E_k \rho(c)] \geq H(c)^{-1} \), where \( H(c) = \text{Tr}[\partial^2 \rho(c) L_c] \) is the QFI, and \( L_c \) is the symmetric logarithmic derivative (SLD), implicitly defined as \( \partial \rho(c) = \rho(c) L_c \rho(c) L_c \rho(c) \).

As \( \rho(c) \) is block-diagonal in \( J \), with each block proportional to the identity, it is trivial to see that the optimal measurement for estimating the overlap is given by the projectors onto the total-angular-momentum subspaces, i.e., \( E_j = \mathbf{1}_J \), and the QFI reduces to the Fisher information of the classical probability distribution \( p(J|c) \).

In the limit \( M + N \to \infty \) and \( M - N \ll (M + N)\sqrt{c} \), we can use an approximation of the Jacobi polynomial given in \([34]\) to obtain the following asymptotically-unbiased estimator and its associated MSE:

\[
v_{\text{op}}^\text{loc}(c) = \sum_{J = \text{min}}^{J_{\text{max}}} \left( \frac{2J}{M+N} \right)^2 \mathbf{1}_J, \quad v_{\text{op}}(c) = \frac{4c(1-c)}{M+N},
\]

where we have written the estimator as an operator \( \hat{c} = \sum_j c(j) \mathbf{1}_j \), since the optimal measurement is projective. In the SM we show that \( v_{\text{op}}^\text{loc}(c) \) coincides with \( H(c)^{-1} \) to leading order in \( \frac{1}{M+N} \) and hence the quantum Cramer-Rao bound is achievable in this limit. If instead \( M \to \infty \) and \( N \) is finite, it is clear that \( |\phi\rangle \) can be estimated perfectly and hence the optimal strategy is to project the copies of \( |\psi\rangle \) in this known direction, with resulting \( v_{\text{op}}^\text{loc}(c) = \frac{c(1-c)}{N+1} \). Finally, note that the results obtained so far do not depend on \( d \) and hence also hold for states in an infinite-dimensional Hilbert space.

**Bayesian estimation.**— We now consider the problem of estimating the value of \( c \) when the two states are chosen completely at random from the Hilbert space. This implies that \( c \) has a prior distribution induced by the Haar measure in dimension \( d \), e.g., see \([29,35]\):

\[
p(c) = \int_{U \in \text{SU}(d)} dU \delta(c - |\langle 0| U |0\rangle|^2) = (d-1)(1-c)^{d-2}.
\]

In this case we want to find the estimator \( \hat{c} \) that minimizes the average MSE: \( v_{\text{bay}} = \int d\rho(c) p(c) v(c) \). Clearly, the optimal measurement does not change, while the optimal estimator is given by the expected value of the overlap after the measurement \([33]\). As explained in the SM, using graphical calculus techniques for the recoupling theory of Clebsch-Gordan coefficients \([39]\), we obtain

\[
v_{\text{bay}} = \frac{d(d-1)(d+M+N)}{d(d+1)(d+M)(d+N)}.
\]
attainable with the optimal estimator
\[
\nu^\text{bay}_{\text{op}} = \sum_{J=0}^{M} \frac{d + J + J^2 + M + N - (M + N)^2}{(d + M)(d + N)} \mathbf{1}_J.
\]

We pause to highlight the following important observations: i) when \(M, N \gg 1\), the average of \(\nu^\text{loc}(c)\) agrees with \(\nu^\text{bay}_{\text{op}}\) to leading order in \(\frac{1}{M+N}\), implying that the local optimal estimator is also a good Bayesian estimator in this case; ii) contrarily to the local-estimation results, the average MSE of Eq. (7) is exact for all \(M, N\) and depends on \(d\) due to the prior, Eq. (6); iii) in particular, \(\nu^\text{bay}_{\text{op}}\) decays as \(d^{-2}\) if one of either \(M\) or \(N\) is kept finite, whereas it decays only as \(d^{-4}\) when \(M, N \gg 1\).

1-LOCC strategies.— While the swap test and the optimal measurement act, respectively, locally and collectively on the copies of the two states, we can also consider a family of intermediate strategies that employ 1-LOCC on \(|\psi\rangle \otimes N\) and \(|\phi\rangle \otimes M\). These strategies are relevant in the case of distributed scenarios and for the generalization of the problem to multiple species. On the other hand, at variance with the optimal measurement, these strategies require the input copies to be labelled in order to perform distinct measurements of \(\psi\) and \(\phi\).

The estimate-and-project (EP) strategy consists in estimating the species with the largest number of copies, then projecting each copy of the second species on this estimate and counting the number of successful projections. The corresponding POVM elements can be written as \(E_{V_{\text{ep}}}^\otimes = dV_{\text{ep}}(M) \otimes V \otimes N \Pi_k^\otimes \otimes N\), where \(E_{V_{\text{ep}}}^\otimes = \frac{N}{d} \sum_{0}^{N} dV_{\text{ep}}(d)(\pi_0 \otimes N) V \perp \otimes N\) is the optimal covariance measurement to estimate the direction of \(\phi\) \[\text{EB}^{\text{EE}} \text{EE}^{\text{EE}}\] and \(\Pi_k^\otimes = G_{\psi_0}(0) |0\rangle \langle 0| V \perp \otimes M\) represents \(k\) successful projections, acting on the copies of \(\psi\), averaged over all possible permutations of \(N\) systems. We take as local estimator \(\nu^\text{bay}_{\text{op}}(k) = \frac{k}{d}\).

The estimate-and-estimate (EE) strategy instead consists in estimating both species separately, then computing the overlap between the estimated directions. The corresponding POVM elements can be written as \(E_{V_{\text{ep}}}^\otimes = dV_{\text{ep}}(M) \otimes E_{V_{\text{ep}}}^\otimes(N)\), i.e., a product of two covariant measurements to estimate the directions of \(\phi\) and \(\psi\). We take as local estimator

\[
M \sim N \gg 1 \quad \frac{4 + (d-1)}{M+N} \quad \frac{2}{N} \quad \frac{2}{N} \quad 2
\]

\[
M \to \infty \quad \frac{4 + (d-1)}{M+N} \quad \frac{2}{N} \quad \frac{2}{N} \quad 2
\]

\[
\begin{array}{cccc}
\text{Local MSE} & \nu^\text{loc}(c) & \nu^\text{op}(c) & \nu^\text{bay}(c) \\
M \sim N \gg 1 & 4 \frac{(d-1)}{M+N} & \frac{2}{N} & 2 \\
M \to \infty & 4 \frac{(d-1)}{M+N} & \frac{2}{N} & 2
\end{array}
\]

\[
\begin{array}{cccc}
\text{Average MSE} & \nu^\text{bay}_{\text{op}} & \nu^\text{bay}_{\text{op}} & \nu^\text{bay}_{\text{op}} \\
M \sim N \gg 1 & 4 \frac{(d-1)}{M+N} & \frac{2}{N} & 2 \\
M \to \infty & 4 \frac{(d-1)}{M+N} & \frac{2}{N} & 2
\end{array}
\]

\[
M = N \gg 1
\]

\[
\begin{array}{cccc}
N \cdot \nu^\text{loc}(c) & 0.2 & 0.4 & 0.6 & 1.0 \\
\text{Optimal} & \text{Swap} & \text{EP} & \text{EE}
\end{array}
\]

\(c = (0|V^\dagger W|0)^2\)

In the SM we provide exact results for local and Bayesian estimation using EP and EE. Table (I) compares the performance of these strategies with the optimal one in two asymptotic limits. We find that, for both local and Bayesian estimation, the EE strategy is always worse than the optimal by a factor of 2, whereas the EP strategy attains a MSE equal to the optimal in the limit \(M \to \infty\), \(N\) finite.

\[
J = \sum_{k=0}^{N} \langle k | V | k \rangle^2 \approx \sum_{k=0}^{N} \langle k | V | k \rangle^2
\]

\[
\frac{\nu^\text{loc}(V, W)}{\nu^\text{loc}(V, W)} = \frac{(0|V^\dagger W|0)^2}{(0|V^\dagger W|0)^2}
\]

\[
\text{For} \quad c < \frac{1}{2}\quad \text{and the use of the EE, EP or optimal strategies strongly enhances the estimation in this case. These features are confirmed in general by the average MSE,}
\]

\[
\begin{array}{cccc}
0 & 0.2 & 0.4 & 0.6 & 1.0 \\
\text{Optimal} & \text{Swap} & \text{EP} & \text{EE}
\end{array}
\]
plotted in Fig. 3 as a function of $N$ for $M$ fixed and increasing $d$ (inset). We observe that the swap test is comparable with EE for $M \sim N$ and $d = 2$, but with a small increase in dimension this feature disappears. Moreover, there is in general a gap between the EP and EE strategies, the former being closer to the optimal one.

**Measurement invasiveness.** — Another relevant figure of merit for applications is the fidelity between the post-measurement state and the initial one, averaged over the measurement outcomes. This quantity is important when the states need to be reused after the post-measurement stage. The optimal measurement and the swap test are projective measurements. We assume that the post-measurement states are given by the result of such projections and hence the average post-measurement fidelity can be written as

$$F(c) = \frac{1}{U \in SU(d)} \int dU \sum_k |\langle \Psi_U | E_k | \Psi_U \rangle|^2,$$  \hspace{1cm} \text{(9)}

with $\{E_k \equiv \frac{1}{2} I_d\}$ for the optimal measurement and $\{E_k = G_{S_n}(I_2^{\otimes k} \otimes I_2^{\otimes N-k})\}$ for the swap test, where $I_2^{\otimes k}$ are the projectors on the singlet/triplet components of $H_2^{\otimes k}$. Then Eq. (9) is simply given by

$$F_{\text{op}}(c) = \sum_{J = J_{\text{min}}}^{J_{\text{max}}} p(J|c)^2, \quad F_{\text{sw}}(c) = \left(\frac{1+c^2}{2}\right)^N, \hspace{1cm} \text{(10)}$$

as shown in the SM. In Fig. 4 we plot these two quantities as a function of $c$, showing that the optimal measurement is less invasive than the swap test, especially for small overlap values. In the SM we provide general bounds on the fidelity of any instrument realizing the optimal measurement and the swap test.

**Noise-robustness.** — Finally, we consider how the optimal strategy changes when the states, which are expected to be pure, are independently subject to depolarizing noise during their production, storage or transmission, restricting to $d = 2$ for simplicity. Defining the depolarizing channel as $\Delta_r = rI + (1-r)\frac{1}{2}I_4$, with $I$ the identity channel, the overall state of the system can now be written as $\Delta_{r_M}(\psi) \otimes \Delta_{r_S}(\phi) \otimes M$.

FIG. 3. (a) Plot of the minimum average MSE $\psi_{\text{bay}}$ vs. the number of copies of one state $N$, for a fixed number of copies of the other $M = 1000$, in dimension $d = 2$, for the optimal, EP and EE strategies. (b) Plot of the minimum average MSE $\psi_{\text{bay}}$ vs. the dimension $d$, for a fixed number of copies $M = N = 1000$ for all the strategies studied.

FIG. 4. Plot of the average post-measurement fidelity with the initial state $F(c)$ vs. the true value of the overlap $c$ with a fixed and equal number of copies $M = N = 100$, for the optimal strategy and swap test.

To compute the average state we may proceed as before, with one major difference: the joint states of the $N$ and $M$ copies are now a mixture of different angular momenta, respectively $J_0$ and $J_1$, with $z$-component determined by a function of their purity $R_i = (1 + r_i)/(1 - r_i)$ [38]. For example, for the first state it holds

$$\Delta_{r_0}(\psi) \otimes N = \sum_{J_0 = 0}^{J_0} \sum_{m_0 = 0}^{m_0} p_{J_0} R_0 \psi_{J_0} \langle J_0, m_0 \rangle \langle J_0, m_0 \rangle \hspace{1cm} \text{(11)}$$

and similarly for the second one. Hence the average state of the system is a mixture of block-diagonal-states in the total-angular-momentum basis $|J_0, J_1; J, m\rangle$, with weights depending on the value of the local angular momenta $J_i$:

$$\tilde{\rho}(c) = \sum_{J_0 = 0}^{J_0} \sum_{J_1 = 0}^{J_1} \sum_{J = J_{\text{min}}}^{J_{\text{max}}} p_{J_0} p_{J_1} J_{\text{max}} \langle J_0, J_1; c \rangle \frac{1}{\chi_J(d)} \chi_J(d) \hspace{1cm} \text{(12)}$$

Using the same methods as before, we can compute the minimum average MSE attainable with $p(J|J_0, J_1; c)$, as detailed in the SM. In particular, in the limit $M, N \rightarrow \infty$ with $M/N$ finite, the average MSE at leading order is $\psi_{\text{bay}}^{\text{opt, max}} = \frac{1}{2} R_0 + \frac{1}{2} R_1$, which agrees with the previously found limit of Eq. (7) for zero-noise, $r_i = 1$. Hence the net effect of white noise is to rescale the MSE by a factor $r_i^{-2}$ for each state.

**Conclusions.** — In this article we have computed the ultimate precision attainable in estimating the overlap of two arbitrary pure quantum states, as a function of their Hilbert-space dimension and their number of copies. We showed that the commonly-used swap test is highly inefficient for small values of the overlap and also on average over Haar-distributed random states. The optimal strategy is a collective measurement on all the copies and can be implemented efficiently using the Schur transform [10, 11], although it remains experimentally challenging.

In addition, we proposed two intuitive strategies that estimate separately one or both species of states and showed that they also outperform the swap test. Finally, we showed that the optimal measurement is less invasive than the swap test and robust to white noise. The strategies we introduced provide several clear advantages over the swap test and they could...
become a standard tool for various quantum technologies, as well as providing improvements in the runtime of quantum algorithms.

Acknowledgements.—M. F. and V. G. acknowledge support from PRIN 2017 "Taming complexity with quantum strategies: a hybrid integrated photonics approach". M. R., M. S. and J. C. acknowledge support from the Spanish MINECO, project FIS2016-80681-P with the support of AEI/FEDER funds; the Generalitat de Catalunya, project CIRIT 2017-SGR-1127. M. S. also acknowledges support from the Spanish MINECO project IJC-2015-24643.

[1] H. Buhrman, R. Cleve, J. Watrous, and R. De Wolf, Phys. Rev. Lett. 87, 167902 (2001).
[2] J. N. de Beaudrap, Phys. Rev. A, 69, 022307 (2004).
[3] N. Kumar, E. Diamanti, and I. Kerenidis, Phys. Rev. A 95, 032337 (2017).
[4] A. K. Ekert, C. M. Alves, D. K. Oi, M. Horodecki, P. Horodecki, and L. C. Kwek, Phys. Rev. Lett. 88, 217901 (2002).
[5] F. Mintert, M. Ku, and A. Buchleitner, Phys. Rev. Lett., 95, 260502 (2005).
[6] S. P. Walborn, P. H. S. Ribeiro, L. Davidovich, F. Mintert, and A. Buchleitner, Phys. Rev. A, 75, (2007).
[7] S. D. Bartlett, T. Rudolph, and R. W. Spekkens, Phys. Rev. A, 70, 032321 (2004).
[8] E. Bagan, S. Iblisdir, and R. Munoz-Tapia, Phys. Rev. A, 73, 022341 (2006).
[9] N. H. Lindner, P. F. Scudo, and D. Bruß, Int. J. of Quantum Information 4, 131 (2006).
[10] N. Gisin and S. Iblisdir, European Physical Journal D 39, 321-327 (2006).
[11] L. Cincio, Y. Suhasi, A. T. Sornborger, and P. J. Coles, New J. Phys. 20, 113022 (2018).
[12] U. Chabaud, E. Diamanti, D. Markham, E. Kashefi, and A. Joux, Physical Review A 98 (6), 062318 (2018).
[13] S. Lloyd, M. Mohseni, and P. Rebentrost, arXiv preprint quant-ph/1307.0411 (2013).
[14] P. Rebentrost., M. Mohseni, and S. Lloyd, Phys. Rev. Lett., 113, 130503 (2014).
[15] A. W. Harrow, A. Hassidim, and S. Lloyd, Phys. Rev. Lett., 103, 150502 (2009).
[16] Z. Zhao, J.K. Fitzsimons, and J.F. Fitzsimons, Phys. Rev. A 99, 052331 (2019).
[17] N. Wiebe, A. Kapoor, and K. Svore, Quantum Inf. Comput. 15, 0318-0358 (2015).
[18] V. Hlavacek, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, Nature, 567, 209-212 (2019).
[19] N. Liu and P. Rebentrost, Phys. Rev. A, 97, 042315 (2018).
[20] Lindner, Netanel H., Petra F. Scudo, and Dagmar Bruss, "Quantum estimation of relative information." International Journal of Quantum Information 4.01 (2006): 131-149.
[21] A. Peres and P. F. Scudo, in Quantum Theory: Reconsideration of Foundations, edited by A. Khrennikov (Växjö Univ. Press, Växjö Univ, Sweden), p.283, e-print quantph/0201017 (2002).
**APPENDIX**

**Average state at fixed overlap**

Here we derive the form of the average state of $N$ copies of a pure state $|\psi\rangle \in \mathcal{H}_d$ and $M \geq N$ copies of a pure state $|\phi\rangle \in \mathcal{H}_d$ with $|\psi\rangle$ and $|\phi\rangle$ having fixed overlap. Without loss of generality we may write the latter of the two states as

$$|\phi\rangle = V(\theta) |\psi\rangle \equiv \sqrt{\tau} |\psi\rangle + \sqrt{1-\tau} |\psi_\perp\rangle$$

where $\tau = \cos^2 \frac{\theta}{2} \in (0,1)$, $\theta \in (0,\pi)$, $\langle \psi_\perp |\psi\rangle = 0$.

As both $|\psi\rangle$, $|\phi\rangle \in \mathcal{H}_d$ are randomly chosen the global state describing the $N + M$ qudits is given by

$$\mathcal{G}_{SU(d)}[\{(|\psi\rangle\langle \psi|) \otimes (|\phi\rangle\langle \phi|)\}^\otimes M] \equiv \int_{g \in SU(d)} dg \ U(g) \ (|\psi\rangle\langle \psi|) \otimes (|\phi\rangle\langle \phi|) \ U^\dagger(g)$$

where $U : SU(d) \to \mathcal{H}_d$ denotes a unitary representation of $SU(d)$ with $dg$ the corresponding invariant Haar measure. We can decompose the integral over $SU(d)$ in Eq. (14) as follows. First we consider the two-dimensional subspace $\mathcal{E} \equiv \text{span}\{|\psi\rangle, |\psi_\perp\rangle\} \subseteq \mathcal{H}_d$ and a corresponding unitary representation $V : SU(2) \to \mathcal{E}$. Using the invariance of the Haar measure, we first perform the group average over $SU(2)$. Afterwards, we can average over all $SU(d)$. Mathematically this implies that Eq. (14) can be written as

$$\mathcal{G}_{SU(d)}[\{(|\psi\rangle\langle \psi|) \otimes (|\phi\rangle\langle \phi|)\}^\otimes M] = \int_{h \in SU(d)} dh \ W(h) \ \left(\int_{\theta \in SU(2)} d\theta \ V(\theta) \otimes (|\phi\rangle\langle \phi|)\right) \ W^\dagger(h) \otimes (|\phi\rangle\langle \phi|)$$

$$= \int_{h \in SU(d)} \mathcal{G}_{SU(2)}[\{(|\psi\rangle\langle \psi|) \otimes (|\phi\rangle\langle \phi|)\}^\otimes M]$$

where $W : SU(d) \to \mathcal{H}_d$ and $h \theta = g$. For any compact Lie group, $G$ whose action on a vector space $\mathcal{H}$ is described by a unitary, generally reducible, representation $U : G \to \mathcal{H}$, the action of the completely positive, trace-preserving map $\mathcal{G}_G : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ can be written as

$$\mathcal{G}_G[\cdot] = \int_{g \in G} dg \ U(g) \cdot U^\dagger(g) = \sum_{\lambda} (\mathcal{D}_{M(\lambda)} \otimes \mathcal{I}_{N(\lambda)}) \circ \mathcal{P}_\lambda[\cdot]$$

where $\lambda$ labels the irreducible representations (irreps), $U(\lambda) : G \to \mathcal{H}(\lambda)$, of $U : G \to \mathcal{H}$, i.e., $U = \bigoplus_{\lambda} \alpha(\lambda) U(\lambda)$, with $\alpha(\lambda) \in \mathbb{N}_+$ denoting the multiplicity of each irrep. The map $\mathcal{P}_\lambda$ is the projector onto the irreducible subspace $\mathcal{H}(\lambda) \in \mathcal{H}$. The latter can be further decomposed into a tensor product $\mathcal{H}(\lambda) = \mathcal{M}(\lambda) \otimes \mathcal{N}(\lambda)$, of vector spaces where $\mathcal{M}(\lambda)$ are called the carrier spaces of the irreps $U(\lambda)$ and are of dimension $\chi_\lambda = \dim(U(\lambda))$ whilst $\mathcal{N}(\lambda)$ are the multiplicity spaces of the irreducible decomposition on which $G$ acts trivially and whose dimensions are $\alpha(\lambda)$. In Eq. (16), $\mathcal{I}_{N(\lambda)}$ is the identity map on $\mathcal{N}(\lambda)$, whereas, due to Schur’s lemma [27], the map $\mathcal{D}_{M(\lambda)}$ is the completely depolarizing map, i.e., $\forall A \in \mathcal{M}(\lambda)$, $\mathcal{D}_{M(\lambda)}[A] = \frac{\text{Tr}(A)}{\chi_\lambda} \mathbf{1}_{M(\lambda)}$.

Let us now restrict to $d = 2$. Using the addition rules for angular momentum on $\mathcal{H}_2^{N+M}$ we thus have

$$|\psi\rangle^\otimes N \otimes |\phi\rangle^\otimes M = \left| \frac{N}{2}, \frac{N}{2} \right\rangle \otimes \sum_{k=-\frac{M}{2}}^{\frac{M}{2}} d_{\frac{N}{2} \frac{M}{2}}^{\frac{N}{2} \frac{M}{2}} \left| \frac{M}{2}, k \right\rangle$$

$$= \sum_{J=J_{\min}}^{J_{\max}} \sum_{k=-\frac{M}{2}}^{\frac{M}{2}} C_{\frac{N}{2} \frac{N}{2} + k}^{\frac{N}{2} \frac{M}{2}} \left| \theta \right\rangle \left| J\frac{N}{2} + k \right\rangle,$$

where $C_{\frac{N}{2} \frac{N}{2} + k}^{\frac{N}{2} \frac{M}{2}} = \langle J\frac{N}{2} + k | \langle \frac{N}{2} | \frac{N}{2} \frac{M}{2} | k \rangle$ are the Clebsch-Gordan coefficients. Using Eq. (16) the first Haar-measure group average of Eq. (15) reads

$$\mathcal{G}_{SU(2)}[\{(|\psi\rangle\langle \psi|) \otimes (|\phi\rangle\langle \phi|)\}^\otimes M] = \sum_{J=J_{\min}}^{J_{\max}} p(J) c \frac{1}{2J+1}.$$
where $\mathbf{1}_f^{(2)}$ is the projector on $\mathcal{U}(SU(2))_J \otimes \mathcal{U}(S_{N+M})_J$, which are the irreps of $SU(2)$ and $S_{N+M}$ on $\mathcal{H}_d^{\otimes N+M}$, respectively. The coefficients of Eq. (18) are:

$$p(J|c) = \sum_{k=-\frac{d}{2}}^{\frac{d}{2}} \left( C^{(\frac{J}{2}+\frac{k}{2})}_{\frac{J}{2},-\frac{k}{2}} \right)^2 (2 \arccos \sqrt{c})^2$$

$$= \frac{(2J+1)(J+J_{\text{min}})!N!}{(J-J_{\text{min}})!(J_{\text{max}}-J)!(J_{\text{max}}+1+J)!} \sum_{k=-\frac{d}{2}}^{\frac{d}{2}} \left( \frac{M}{2} - k \right)! \left( \frac{N}{2} + k \right)! \left( d^{(\frac{J}{2}+\frac{k}{2})}_{\frac{J}{2},-\frac{k}{2}} \right)^2 (2 \arccos \sqrt{c})^2$$

$$= \frac{(2J+1)(J+J_{\text{min}})!N!M!}{(J-J_{\text{min}})!(J_{\text{max}}-J)!(J_{\text{max}}+1+J)!} \sum_{k=-\frac{d}{2}}^{\frac{d}{2}} (J+\frac{N}{2}+k)! \left( d^{(\frac{J}{2}+\frac{k}{2})}_{\frac{J}{2},-\frac{k}{2}} \right)^2 (1-c)^{\frac{M}{2}-k} c^{\frac{N}{2}+k}$$

and we have made use of the following expression for the Wigner small-$c$ matrix in going from the second to the third line in Eq. [19]

$$d^{(J)}_{z,z'}(\theta) = \sqrt{\frac{(J+z)!(J-z)!}{(J+z')!(J-z')}} \sin^{(z-z')} \left( \frac{\theta}{2} \right) \cos(2z') \left( \frac{\theta}{2} \right) P^2_{(z,z'+z')} \left[ \cos(\theta) \right],$$

with $P_n^{(\alpha,\beta)}(x)$ the Jacobi polynomials, defined in general as

$$P_n^{(\alpha,\beta)}(x) = \frac{\Gamma(\alpha+n+1)}{n!\Gamma(\alpha+\beta+n+1)} \sum_{m=0}^{n} \binom{n}{m} \frac{\Gamma(\alpha+\beta+n+m+1)}{\Gamma(\alpha+m+1)} \left( \frac{x-1}{2} \right)^m.$$  

For general $d$ instead we have, applying Eq. [16],

$$G_{SU(d)} \left[ G_{SU(2)} \left[ (|\psi\rangle\langle\psi|) \otimes N \otimes (|\phi\rangle\langle\phi|) \otimes M \right] \right] = \sum_{\lambda} (\mathcal{D}_{\mathcal{M}(\lambda)} \otimes \mathcal{T}_{N(\lambda)}) \circ \mathcal{P}_{\lambda} \left[ \sum_{J=J_{\text{min}}}^{J_{\text{max}}} p(J|c) \frac{\mathbf{1}_{f}^{(2)}}{2J+1} \right]$$

$$= \sum_{J=J_{\text{min}}}^{J_{\text{max}}} p(J|c) \mathbf{1}_{f}^{(d)} \chi_{J}.$$  

In the last equality we used that the support of $\mathbf{1}_f^{(2)}$ is a subspace of the support of $\mathbf{1}_f^{(d)}$, the projector on $\mathcal{U}(SU(d))_J \otimes |J\rangle$, where the latter is a one-dimensional subspace of $\mathcal{U}(S_{N+M})_J$. The proof of this fact proceeds as follows. First observe that by Schur-Weyl duality, one has the decomposition

$$\mathcal{H}_d^{\otimes N+M} = \oplus_{\lambda} \mathcal{U}(\mathcal{M}(\lambda)) \otimes \mathcal{U}(\mathcal{S}_r(\lambda)),$$

where $\mathcal{U}(\mathcal{S}_r(\lambda)) = \mathcal{M}(\lambda)$ and $\mathcal{U}(\mathcal{S}_r(\lambda)) = \mathcal{N}(\lambda)$ are the irreps of $SU(d)$ and $SU$ respectively, and $\lambda$ labels all Young diagrams with at most $d$ rows and $r$ boxes, $Y(d,r)$. In our problem $r = N + M$ and the states are symmetric under $S_N \times S_M \subset S_{N+M}$, i.e., they are supported on the tensor product of maximally symmetric subspaces of $N$ and $M$-d level systems:

$$\text{Sym}_{N,M}^{(d)} := \text{Sym}_{N}^{(d)} \otimes \text{Sym}_{M}^{(d)} \subset \mathcal{H}_d^{\otimes N+M}.$$  

This space admits a decomposition

$$\text{Sym}_{N,M}^{(d)} = \oplus_{J} \left( \mathcal{U}(SU(d))_J \otimes |J\rangle \right),$$

with $\mathcal{U}(SU(d))_J$ being representations with Young diagrams with two rows, indexed by an integer $J$, $\frac{M+N}{2} \leq J \leq \frac{M-N}{2}$, and $|J\rangle \in \mathcal{U}(S_{N+M})_J$. Second, notice that since $\mathcal{E}$ is a two-dimensional subspace of $\mathcal{H}_d$, the total state of the system before acting with $\mathcal{G}_d$ is also supported on

$$\text{Sym}_{N,M}^{(2)} := \oplus_{J_2} \left( \mathcal{U}(SU(2))_J_2 \otimes |J_2\rangle \right) \subset \text{Sym}_{N,M}^{(d)}.$$  

What is more, $\mathcal{E}^{\otimes N} \otimes \mathcal{E}^{\otimes M}$ is an invariant subspace of $(\mathcal{H}_d)^{\otimes N+M}$ under the action of the symmetric group, therefore the irreps of the symmetric group supported on $\mathcal{E}^{\otimes N} \otimes \mathcal{E}^{\otimes M}$ can also be taken as valid irreps on $(\mathcal{E}^{\otimes N})^{\otimes M+N}$ and it follows that $\mathcal{U}(SU(2))_J_2 \otimes |J_2\rangle \subset \mathcal{U}(SU(d))_J \otimes \mathcal{U}(S_{N+M})_J$. Finally, since $\text{Sym}_{N,M}^{(2)} \subset \text{Sym}_{N,M}^{(d)}$, and each of the $\mathcal{U}(SU(2))_J_2 \otimes |J_2\rangle$ and $\mathcal{U}(SU(2))_J \otimes |J\rangle$ are included in the same set $\mathcal{U}(SU(2))_J \otimes \mathcal{U}(S_{N+M})_J$, it follows that

$$\mathcal{U}(SU(2))_J \otimes |J\rangle \subset \mathcal{U}(SU(2))_J \otimes |J\rangle$$

and $J = J_2$. 
Fisher Information of \( p(J|c) \)

In this section we derive the Fisher information, Eq. (5), and the corresponding optimal estimator for the probability distribution given by Eq. (4). We begin by recalling the definition of the Fisher information

\[
H(c) = \sum_{J=J_{\min}}^{J_{\max}} \left( \frac{dp(J|c)}{dc} \right)^2 \frac{1}{p(J|c)}.
\]  

(28)

Using the identity

\[
\frac{d^n P_n^{(\alpha, \beta)}(x)}{dx^n} = \frac{(\alpha + \beta + n + m)!}{2^n (\alpha + \beta + n)!} P_{n-m}^{(\alpha + m, \beta + m)}(x),
\]  

(29)

it follows that

\[
\frac{dp(J|c)}{dc} = \frac{(2J + 1)M!N!(1 - c)^M - 2}{(J_{\max} - J)! (J + J_{\max} + 1)!} \times \left( (J - J_{\min} + 1) P^{(1, 1 - J_{\min})}_{J + J_{\min} - 1} \frac{1 + c}{1 - c} - (1 - c) M P^{(0, -2J_{\min})}_{J + J_{\min}} \frac{1 - c}{1 + c} \right).\]

(30)

For \( \frac{1 + c}{1 - c} \geq 1 \) the following asymptotic expansion for the Jacobi polynomials holds

\[
P_n^{(\alpha, \beta)}(x) \xrightarrow{n \to \infty} \frac{\sqrt{\frac{x + \sqrt{x^2 - 1}}{x - \sqrt{x^2 - 1}}}}{\sqrt{2 \pi n} (\sqrt{x - 1})^\alpha (\sqrt{x + 1})^\beta \varphi^2} n^{\frac{\alpha + \beta}{2}}.\]

(31)

Using Eq. (31) one obtains, after some algebra

\[
\left( \frac{dp(J|c)}{dc} \right)^2 \frac{1}{p(J|c)} = \frac{(2J + 1)M!N!(1 + \sqrt{c})^{2J - 1} (1 - c)^{J_{\max} - J}}{2 \sqrt{\pi} (1 - \sqrt{c})^{2J/5} (J + J_{\min} - 1) \sqrt{J + J_{\min} (J_{\max} - J)}! (J + J_{\max} + 1)!} \times \left( (J - J_{\min} + 1) \sqrt{J + J_{\min} - M \sqrt{c(J + J_{\min} - 1)}} \right)^2.
\]

(32)

Observe that in Eq. (32) one can identify the following binomial distribution:

\[
\text{Bin}(2J_{\max} + 1, J_{\max} - J, p) = \binom{2J_{\max} + 1}{J_{\max} - J} p^{J_{\max} - J} (1 - p)^{J_{\max} + J + 1},
\]

(33)

with \( p = \frac{1 - \sqrt{c}}{2} \). The mean and variance of this binomial distribution are given by

\[
\mu = (2J_{\max} + 1) \left( \frac{1 - \sqrt{c}}{2} \right)
\]

(34)

\[
\sigma^2 = \frac{(2J_{\max} + 1)}{4} (1 - c).
\]

Note that \( \mu = J_{\max} - J \) so that

\[
\langle J \rangle = \sqrt{c} \left( J_{\max} + \frac{1}{2} \right) - \frac{1}{2}.
\]

(35)

In order for the asymptotic expansion of Eq. (31) to hold we need \( J + J_{\min} \gg 2J_{\min} \), and to use it consistently we need that terms which violate this condition are suppressed. Using Eq. (35), we must have that

\[
\sqrt{c} J_{\max} \gg J_{\min}.
\]

(36)

This means that if \( M - N \) is fixed and \( M + N \to \infty \), the approximation holds for the dominant terms of \( p(J|c) \).

As we are interested in computing the QFI in the limit \( J_{\max} \to \infty \) we will perform a Taylor series expansion around the mean value \( \langle J \rangle \) of the binomial distribution. For any analytical function \( h(J) \) an expansion about the mean value is given by

\[
\sum_J p(J) h(J) = \sum_J p(J) \left( \sum_{x=0}^{\infty} (J - \langle J \rangle)^x \frac{d^x h(J)}{dJ^x} \bigg|_{J=\langle J \rangle} \right)
\]  

(37)
Writing
\[
\left( \frac{d p(J|c)}{d c} \right)^2 \frac{1}{p(J|c)} = f(M, N, c) g(J, M, N, c) \text{Bin}(2J_{\text{max}} + 1, J_{\text{max}} - J, p)
\]
(38)

where
\[
f(M, N, c) = \frac{M!N!2^{J_{\text{max}}}}{\sqrt{\pi} (1 - c)^2 c^{J_{\text{max}}}}
\]
\[
g(J, M, N, c) = \frac{(2J + 1) \left( (J - J_{\text{min}} + 1) \sqrt{J + J_{\text{min}} - M \sqrt{c(J + J_{\text{min}} - 1)}} \right)^2}{(J + J_{\text{min}} - 1) \sqrt{J + J_{\text{min}}}}
\]
(39)

the QFI explicitly reads
\[
H(c) = f(M, N, c) \sum_J g(J, M, N, c) \text{Bin}(2J_{\text{max}} + 1, J_{\text{max}} - J, p)
\]
(40)

In the limit \( J_{\text{max}} \to \infty \)
\[
f(M, N, c) = \frac{1}{2(1 - c)^2 c^\frac{1}{2}} \frac{1}{\sqrt{J_{\text{max}}}} + O\left(J_{\text{max}}^{-\frac{1}{2}}\right)
\]
(41)

whereas using Eq. (37) and the fact that \( \frac{d^m g(J, M, N, c)}{d x^m} \bigg|_{x=J} = O\left(J_{\text{max}}^{\frac{m}{2}}\right) \)
\[
\sum_J g(J, M, N, c) \text{Bin}(2J_{\text{max}} + 1, J_{\text{max}} - J, p) = g(J, M, N, c) + \frac{1}{2} c^\frac{1}{2} \frac{d^2 g(J, M, N, c)}{d J^2} \bigg|_{J=(J)} + O\left(J_{\text{max}}^{-\frac{1}{2}}\right)
\]
(42)

Multiplying Eqs. (41), (42) finally gives
\[
H(c) = \frac{J_{\text{max}}}{2c(1 - c)} + O(1) = \frac{M + N}{4c(1 - c)} + O(1).
\]
(43)

We now show that \( H(c) \) is achievable by the estimator of Eq. (5). First note that the estimator is indeed asymptotically unbiased since
\[
\langle \hat{c}_{\text{op}}^{\text{loc}} \rangle := \text{Tr} \left[ \hat{c}_{\text{op}}^{\text{loc}} \rho(c) \right] = \sum_{J=J_{\text{min}}}^{J_{\text{max}}} \left( \frac{J}{J_{\text{max}}} \right)^2 \frac{1}{J_{\text{max}}} p(J|c)
\]
(44)
\[
= \frac{2^{2J_{max}} M!N!}{c^\frac{1}{2} \sqrt{\pi} (2J_{\text{max}} + 1)! J_{\text{max}}^2} \sum_{J=J_{\text{min}}}^{J_{\text{max}}} \frac{J^2 (2J + 1)}{\sqrt{J + J_{\text{min}}}} \text{Bin}(2J_{\text{max}} + 1, J_{\text{max}} - J, p),
\]
(45)

where we have again used the asymptotic expansion of Eq. (31) and identified the binomial distribution. In the limit \( J_{\text{max}} \to \infty \) one can proceed as above, separating the \( J \)-dependence and applying Eq. (37), to obtain
\[
\frac{2^{2J_{\text{max}}}}{c^\frac{1}{2} \sqrt{\pi} (2J_{\text{max}} + 1)! J_{\text{max}}^2} \sum_{J=J_{\text{min}}}^{J_{\text{max}}} \frac{J^2 (2J + 1)}{\sqrt{J + J_{\text{min}}}} \text{Bin}(2J_{\text{max}} + 1, J_{\text{max}} - J, p) = \frac{2c^\frac{1}{2} J_{\text{max}}^\frac{1}{2}}{J_{\text{max}}^\frac{1}{2}} + O(J_{\text{max}}^\frac{1}{2}).
\]
(46)

It follows that
\[
\langle \hat{c}_{\text{op}}^{\text{loc}} \rangle = c + O(J_{\text{max}}^{-1}).
\]
(47)

A similar calculation gives for the MSE of the estimator
\[
\langle (\hat{c}_{\text{op}}^{\text{loc}} - c)^2 \rangle := \frac{2c(1 - c)}{J_{\text{max}}} + O\left(J_{\text{max}}^{-\frac{1}{2}}\right).
\]
(48)
Optimal average mean squared error

In this appendix we derive the optimal estimator and corresponding average mean squared error (AvMSE) for the case where the overlap \( c \) is a random variable with a distribution induced by the Haar-uniform measure of SU\((d)\). One obtains [33] Eq. (13)]

\[
p(c) = \int_{g \in \text{SU}(d)} dg \delta(c - |\langle \psi | U(g) | \psi \rangle|^2) = (d - 1)(1 - c)^{d-2}
\]

Following [33] the optimal estimator, \( S \), satisfies

\[
\frac{S\Gamma + \Gamma S}{2} = \eta
\]

where

\[
\Gamma \equiv \int p(c) \rho(c) \, dc
\]

\[
\eta \equiv \int c p(c) \rho(c) \, dc.
\]

and is explicitly given by

\[
S = \int_0^\infty e^{-\alpha \Gamma} \eta e^{-\alpha \Gamma} \, d\alpha
\]

Plugging Eq. (3) into Eqs. (51), (52) gives

\[
S = \sum_{J=J_{\text{min}}}^{J_{\text{max}}} \frac{\int c p(J,c) \, dc}{\int p(J,c) \, dc} \mathbf{1}^{(d)}_J
\]

\[
= \sum_{J=J_{\text{min}}}^{J_{\text{max}}} \int c p(c|J) \, dc \mathbf{1}^{(d)}_J
\]

\[
= \sum_{J=J_{\text{min}}}^{J_{\text{max}}} \langle c|J \mathbf{1}^{(d)}_J,
\]

where \( p(J,c) = p(c)p(J|c) \), \( \int p(J,c) \, dc = p(J) \) and we have used Bayes’ theorem in going from the second to the third line of Eq. [53]. Again, the optimal measurement corresponds to measuring the total irrep label \( J \). Upon a given outcome the estimator that minimizes the AvMSE is \( \hat{c}_J = \langle c|J \rangle \) where the expectation value is taken with respect to the conditional probability distribution \( p(c|J) \).

The operators \( \Gamma \) and \( \eta \) are given by

\[
\Gamma = \int_0^1 p(c) \left( \int_{g \in \text{SU}(d)} dg \left( U(g) |\psi\rangle\langle\psi | U^\dagger(g) \right) \otimes \left( U(g) V(c) |\psi\rangle V^\dagger(c) U^\dagger(g) \right) \right) \, dc
\]

\[
= \int_{g \in \text{SU}(d)} dg \left( U(g) |\psi\rangle\langle\psi | U^\dagger(g) \right) \otimes \int_0^1 dc \int_{h \in \text{SU}(d)} dh \delta(c - |\langle \psi | U(h) | \psi \rangle|^2) \left( U(g) V(c) |\psi\rangle V^\dagger(c) U^\dagger(g) \right)
\]

\[
= \int_{g \in \text{SU}(d)} dg \left( U(g) |\psi\rangle\langle\psi | U^\dagger(g) \right) \otimes \int_{h \in \text{SU}(d)} dh \left( U(g) W U(h) U(h) |\psi\rangle V^\dagger(c) U^\dagger(h) \right)
\]

\[
= \mathcal{G}_{U(d)} \left( |\psi\rangle\langle\psi | \otimes \mathcal{G}_{U(d)} \left( |\psi\rangle\langle\psi | \right) \right)
\]

\[
= \mathbf{1}^{(d)} \mathbf{1}^{(d)} \otimes \mathcal{G}_{U(d)} \left( |\psi\rangle\langle\psi | \right)
\]

\[
= \frac{1}{\chi^{(d)}_N} \sum_{J=J_{\text{min}}}^{J_{\text{max}}} \mathbf{1}^{(d)}_J,
\]

\[
(54)
\]
where we have made use of the fact that \( W_{U(h)} U(h) = V(c) \) for a unitary \( W_{U(h)} \) such that \( W_{U(h)} |\psi\rangle \langle \psi| W_{U(h)}^\dagger = |\psi\rangle \langle \psi| \) in the third equality, the invariance of the Haar measure for the fourth equality, and used the addition rules for SU(d) representations for the last equality.

To compute \( \eta \) we make use of

\[
\int_{g \in SU(2)} dg (d-1)(1 - |V(g)\frac{1}{2} (|0\rangle \langle 0|)^{\otimes d} - |V(g)\frac{1}{2} |0\rangle \langle 0| V(g)\frac{1}{2} )^{\otimes M} = \\
\int_{SU(2)} dD \frac{d-1}{d} \frac{d-1}{d} (g) D \frac{d-1}{d} \frac{d-1}{d} (g)^* D \frac{d-1}{d} \frac{d-1}{d} (g) D \frac{d-1}{d} \frac{d-1}{d} (g) |\langle N, N| = |\langle M, M|, k, k'| = \\
\frac{1}{d + M} \sum_{J = \frac{d-1}{2}}^{J = \frac{d+1}{2}} \left( C_{J,N,M} \frac{d-1}{2} \frac{d-1}{2} k, C_{J,N,M} \frac{d-1}{2} \frac{d-1}{2} k \right)^2 \left( \frac{M}{2} \frac{d-1}{2} J \frac{d-1}{2} \frac{d+1}{2} \right)
\]

with \( D_{m,n}^J (g) \) being Wigner matrices, so that

\[
\eta = \int_0^1 p(c) \left( \int_{g \in SU(d)} dg (U(g)|\psi\rangle \langle \psi| U^\dagger (g) )^{\otimes N} \otimes (U(g)V(c)|\psi\rangle \langle \psi| V^\dagger (c) U^\dagger (g) )^{\otimes M} \right) dc
\]

\[
= \int_{g \in SU(d)} dg (U(g)|\psi\rangle \langle \psi| U^\dagger (g) )^{\otimes N} \otimes \int_0^1 dc p(c) \int_{h \in SU(2)} dh \delta(c - |\langle U(h)| \psi\rangle |^2) (U(g)V(c)|\psi\rangle \langle \psi| V^\dagger (c) U^\dagger (g) )^{\otimes M}
\]

\[
= \mathcal{G}_{SU(d)} \left[ \int_{h \in SU(2)} dh (d-1)(1 - |\langle \psi| V(h) \psi\rangle |^2)^{d-2} ) |\langle \psi| V(h) \psi\rangle |^2 |\langle \psi| \psi\rangle |^N \otimes (V(h)|\psi\rangle \langle \psi| V^\dagger (h) )^{\otimes M} \right]
\]

\[
= \frac{1}{d + M} \sum_{J = J_{\min}}^{J = \frac{d}{2}} \sum_{k = -J}^{k = J} \left( C_{J,M-N} \frac{d-1}{2} \frac{d-1}{2} k, C_{J,M-N} \frac{d-1}{2} \frac{d-1}{2} k \right)^2 \frac{1}{\chi_J^{(d)}}
\]

It is now trivial to compute the optimal Bayesian estimator for a given measurement outcome \( J \):

\[
c(J) = \frac{\text{Tr}\left[ (|J\rangle \langle J|) \right]}{\text{Tr}\left[ (|\Gamma\rangle \langle \Gamma|) \right]} = \frac{\chi_J^{(d)}}{\chi_J^{(d)}(d)\chi_J^{(d)}(d)} = \frac{1}{d + M} \sum_{k = -J}^{k = J} \left( C_{J,M-N} \frac{d-1}{2} \frac{d-1}{2} k, C_{J,M-N} \frac{d-1}{2} \frac{d-1}{2} k \right)^2 \frac{1}{\chi_J^{(d)}}
\]

\[
= \frac{1}{d + M} \sum_{k = -J}^{k = J} \left( C_{J,M-N} \frac{d-1}{2} \frac{d-1}{2} k, C_{J,M-N} \frac{d-1}{2} \frac{d-1}{2} k \right)^2 = (2J + 1) \left( \frac{d + J + \frac{N + M}{2} - 1}{(d - 1)(d - 2)!} \right) \left( \frac{d + J + \frac{N + M}{2} - 2}{(d - 1)(d - 2)!} \right) - \left( \frac{d + J + \frac{N + M}{2}}{2} \right)!
\]

\[
\times \frac{N!M!(d - 1 + J + \frac{N + M}{2} - 1)!}{4(d + N)!((d + M)!(d + M)!(1 + J + \frac{N + M}{2})!)}
\]

where the term in curly brackets is the Wigner 6-j symbol. Plugging everything together the optimal AvMSE estimator for a given measurement outcome \( J \) is given by

\[
c(J) = \frac{d + J + J^2 + \frac{M + N}{2} - \left( \frac{M + N}{2} \right)^2 + MN}{(d + M)(d + N)}
\]

with its corresponding AvMSE

\[
v_{\text{op}} = (v_{\text{op}} - c)^2 = \int_0^1 p(c) c^2 - \sum_{J} p(J)c(J)^2 = \frac{(d - 1)(d + M + N)}{(d + 1)(d + M)(d + N)}
\]
1-LOCC strategies

Here we derive the MSE attainable by the estimate-and-project (EP) and estimate-and-estimate (EE) strategies. The EP strategy consists in first estimating at best one of the states using an optimal collective measurement on all of its copies, and then projecting each copy of the other state on the estimate of the first one. The optimal estimation of a random state $|\phi\rangle$, given $N$ copies of it, is provided by the covariant measurement of Ref. [36], $\{M_{AV}\}$, that produces an estimate $|\phi_{V}\rangle = V|0\rangle$, $V \in SU(d)$, with probability density

$$d\mu(V) = \frac{1}{\hat{M}}(d) |\phi_{V}\rangle |\phi\rangle^{2M} dV.$$  

Finally, we perform the projective measurement $\{ |\phi_{V}\rangle \langle \phi_{V}|, I - |\phi_{V}\rangle \langle \phi_{V}| \}$ on each copy of $|\psi\rangle$. This succeeds with probability

$$p_{V}(c) = |\langle \phi_{V}|\psi\rangle|^{2}.$$  

The overall measurement operator is then $E_{V,k}^{(ep)} = dV E_{V}^{(M)} \otimes V^{\otimes N} \Pi_{k}^{(N)} V^{\dagger \otimes N}$, as defined in the main text. Its outcome statistics for a fixed value of the overlap can be written as

$$p(k|c) = \frac{1}{\hat{M}}(d) \int_{U \in SU(d)} dU \text{Tr} \left[ \left( |0\rangle \langle 0| \otimes \Pi_{k}^{(N)} \right) U^{\otimes(M+N)} \left( \left( U_{c} |0\rangle \langle 0| U_{c}^{\dagger} \right) \otimes |0\rangle \langle 0| \right) U^{\dagger \otimes(M+N)} \right],$$  

where we have set without loss of generality $|\psi\rangle = |0\rangle$, $|\phi\rangle = U_{c} |0\rangle$ thanks to the presence of the reference-frame average. Moreover, since the dependence on $V$ can be included in the reference-frame average and results in constant, we averaged over $V$ without loss of generality. We now consider both local and Bayesian estimation with this strategy.

For local estimation with EP, we employ as an estimator $v_{ep}(k) = \frac{k}{N}$, i.e., the fraction of successful projections. This estimator is in general biased, except in the limit of large $N$, and it is not necessarily optimal, yet it provides a natural guess for the overlap, given our strategy. Its MSE is

$$v_{ep}(c) = \int_{V \in SU(d)} d\mu(V) \sum_{k=0}^{N} \frac{1}{N} \binom{k}{N} \left( \frac{k}{N} - c \right)^{2} \left( c^{2} - 2cp_{V}(c) + \frac{1}{N}p_{V}(c)(1 - p_{V}(c)) + p_{V}(c)^{2} \right)$$  

$$=\frac{1}{\hat{M}}(d) \int_{V \in SU(d)} d\mu(V) \left( I_{1}(M) + \left( 1 - \frac{1}{N} \right) I_{2}(M) \right),$$  

where in the first equality we have introduced the binomial distribution $\binom{k}{N}$ for $m$ successes out of $N$ trials, with a single-trial success probability $p$, in the second one we have expanded the square, used the mean and variance of the binomial distribution and in the third one we defined the integrals

$$I_{i}(M) = \frac{1}{\hat{M}}(d) \int_{V \in SU(d)} dV |\phi_{V}\rangle |\phi\rangle^{2M} |\langle \phi_{V}|\psi\rangle|^{2i}.$$  

These can be computed by expanding the scalar products and writing the states in a collective-spin basis, obtaining the expectation value of the operator in Eq. (4), with the substitution $N \rightarrow i$:

$$I_{i}(M) = \frac{1}{\hat{M}}(d) \langle 0 \rangle^{\otimes M+i} \Phi_{SU(d)} \left[ |\phi\rangle \langle \phi| \otimes |\psi\rangle \langle \psi|^{\dagger} \right] |0\rangle^{\otimes M+i},$$  

where we have eliminated the integral over $V$ by including this rotation into the average over unknown reference-frame. The result of the group average is then given by Eqs. (19), (22) with the same substitution. Plugging in the expression of $\chi_{J}(d) = \left( \begin{array}{c} 2J + d - 1 \\ d - 1 \end{array} \right)$, Ref. [36], we have

$$I_{i}(M) = \left( \begin{array}{c} M + i + d - 1 \\ i \end{array} \right)^{-1} (1 - c)^{i} I_{i}^{(0,M-i)} \left( \frac{1}{1 - c} \right),$$  

which can be computed explicitly for $i = 1, 2$. Inserting these expressions in Eq. (63) we finally obtain

$$v_{ep}^{loc}(c) = \frac{c^{2}(dM + d^{2}M + N - 2MN - N^{2}) + c(-2M - 2dM - 3N + dN + 2MN + N^{2}) + d - 1 + 2M + N}{M(d + N)(1 + d + N)}.$$  

In the limit $M \to \infty$, $N$ constant we have
\[ v_{\text{loc}}^c(c) \sim \frac{c(1-c)}{N}, \] (68)
which coincides with the optimal strategy, corresponding to a projection on the known direction of $|\phi\rangle$. In the limit $M + N \to \infty$, $M - N$ fixed we have instead
\[ v_{\text{loc}}^c(c) \sim \frac{6c(1-c)}{(M+N)}. \] (69)
which is $3/2$ times larger than the optimal strategy.

For Bayesian estimation with EP, as in the previous section, the optimal classical estimator is given by
\[ c(k) = \int dc \; p(c|k) = \frac{\tilde{c}(k)}{p(k)}, \] with $\tilde{c}(k) := \int dc \; p(c)p(k|c)$. (70)

We start by computing the probability distribution of the outcomes, using Eq. (62):
\[ p(k) = \int dc \; p(c|k|c) \]
\[ = \frac{1}{\chi_{\mathcal{S}}(d)\mathcal{N}} \int_{U \in SU(d)} dU \; dU_c \; \text{Tr} \left[ \left(0 \right)_{0}^{\otimes M} \otimes \Pi_k^{(N)} \right] U_{\sigma}^{\otimes (M+N)} \left( \left(U_c |0\right)_{0}^{\otimes M} \right) U_{\sigma}^{\dagger \otimes (M+N)} \right] \]
\[ = \int_{U \in SU(d)} dU \; \text{Tr} \left[ \left(0 \right)_{0}^{\otimes M} \otimes \Pi_k^{(N)} \right] \left( U_{\sigma}^{\otimes (M+N)} \right) \left( U_{\sigma}^{\dagger \otimes (M+N)} \right) \] (71)
In the third equality we have performed the average over the representation $U_{\sigma}^{\otimes M}$ of SU($d$), defining $1_{\mathcal{N}}^{\text{sym}} = 1_{M=\frac{d}{2}}$ as the projector on the completely symmetric irrep of SU($d$) arising from the coupling of $N$ completely symmetric representations and employed its invariance under $U_{\sigma}$. In the fourth equality we have performed the average over $U_{\sigma}^{\otimes N}$ with a similar result. We are then left to compute the overlap of $\Pi_k^{(N)}$ with $1_{\mathcal{N}}^{\text{sym}}$. In order to do so, we recall that the latter can be written as the average of all permutation operators $\pi_\sigma, \sigma \in \Sigma_N$ of $N$-level systems. Then the symmetrization of $\Pi_k^{(N)}$ gives a trivial binomial factor and we can write
\[ p(k) = \frac{1}{\chi_{\mathcal{S}}(d)\mathcal{N}} \int_{U \in SU(d)} dU \; dU_c \; \text{Tr} \left[ \left(0 \right)_{0}^{\otimes M} \otimes \Pi_k^{(N)} \right] \left( U_{\sigma}^{\otimes (M+N)} \right) \left( U_{\sigma}^{\dagger \otimes (M+N)} \right) \] (72)
where $\tilde{c}(k)$ is the number of times the integer $j$ appears in the sequence $\tilde{i}$. The sum over $\tilde{i}$ can then be broken up into the sum over all partitions of $N - k$ systems in $d - 1$ sectors, i.e., the sum over all possible vectors $\beta$ of $d - 1$ components that add up to $N - k$ times the sum of all permutations of $N - k$ systems which are equal in groups of size $\beta_j$. The latter can be carried out immediately since the summand is invariant under permutation of the $\beta_j$:
\[ p(k) = \frac{1}{\chi_{\mathcal{S}}(d)\mathcal{N}} \int_{U \in SU(d)} dU \; dU_c \; \text{Tr} \left[ \left(0 \right)_{0}^{\otimes M} \otimes \Pi_k^{(N)} \right] \left( U_{\sigma}^{\otimes (M+N)} \right) \left( U_{\sigma}^{\dagger \otimes (M+N)} \right) \] (73)
where in the last equality we finally carried out the sum over $\beta$.

For the first moment of the distribution we get similarly,
\[ \tilde{c}(k) = \frac{1}{\chi_{\mathcal{S}}(d)\mathcal{N}} \int_{U \in SU(d)} dU \; dU_c \; \text{Tr} \left[ \left(0 \right)_{0}^{\otimes (M+1)} \otimes \Pi_k^{(N)} \right]. \]
\[ = \frac{1}{\chi_{\mathcal{S}}(d)\mathcal{N}} \int_{U \in SU(d)} dU \; dU_c \; \text{Tr} \left[ \left(0 \right)_{0}^{\otimes (M+1)} \otimes \Pi_k^{(N)} \right] U_{\sigma}^{\dagger \otimes (M+N)} \left( U_{\sigma}^{\otimes (M+N)} \right) \] (74)
where we have introduced $U^{\dagger \otimes U}$ and its conjugate on the additional subsystem, then employed the invariance of $1_{M+1}^{\text{sym}}$ under $U_{\sigma}^{\otimes N}$. In order to proceed, we first compute the value of the following operator:
\[ A_1(M) = \text{Tr}_{1,M} \left[ 1_{M+1}^{\text{sym}} \otimes |0\rangle_{M}^{\otimes M} \right] = \frac{1}{(M+1)!} \sum_{i,j=0}^{d-1} \langle 0 | M \otimes (i) \otimes |0\rangle_{M}^{\otimes M} |j\rangle \langle j | \] (75)
where in the first equality we have taken the partial trace over $M$ subsystems, while in the second one we have written $1_{M+1}^{sym}$ as an average of permutations, like before, and written the explicit basis representation of the last subsystem. The third equality follows by evaluating the only non-zero elements in the sums: the first term contains all the permutations of $M$ subsystems times the identity on the remaining subsystem; the second term considers the additional permutations in the case $i = j = 0$, where the last subsystem can be exchanged with any of the other $M$ subsystems. By substituting this expression in Eq. (74) we obtain

$$\tilde{c}(k) = \frac{x_{\pi_v}(d)}{x_{\pi_{M+1}}(d)} \left( \frac{p(k)}{M+1} + \frac{M}{M+1} \text{Tr} \left[ \left( |0\rangle \langle 0| \otimes \Pi^{(N)}_k \right) \frac{1_{N+1}^{sym}}{x_{\pi_{M+1}}(d)} \right] \right),$$

(76)

where we have used Eq. (71). We then just need to compute the second term in the sum above, which is very similar to Eq. (72) with the change $|0\rangle \otimes k \rightarrow |0\rangle \otimes (k+1)$. We finally obtain

$$\tilde{c}(k) = \frac{(d-1)(d+N+k(M+1))N!(N-k+2)}{(d+M)(N-k)!d!}.\tag{77}$$

and the optimal Bayesian estimator for each $k$

$$c(k) = \frac{d + N + k(M + 1)}{(d+M)(d+N)}.$$  

(78)

The corresponding MSE is given by

$$v_{ep}^{bay} = \int dc p(c) c^2 - \sum_{k=0}^{N} p(k)c(k)^2 = \frac{(d-1)(d+M)^2 + (d+2M)N}{d(1+d)(d+M)^2(d+N)}.$$  

(79)

In the limit $M \rightarrow \infty$, $N$ constant we have

$$v_{ep}^{bay} \sim \frac{(d-1)}{d(d+1)(d+N)},$$

(80)

which again coincides with the optimal Bayesian strategy. In the limit $M + N \rightarrow \infty$, $M - N$ fixed we have instead

$$v_{ep}^{bay} \sim \frac{6(d-1)}{d(d+1)(M+N)},$$

(81)

which again is $3/2$ times larger than the optimal Bayesian strategy.

The EE strategy instead consists in estimating both states with a covariant measurement, hence it is described by overall POVM operators $E_{V,k}^{(ep)} = dVdW E_{V}^{(M)} \otimes E_{W}^{(N)}$, as mentioned in the main text. Its success probability can be written as

$$p(W|c) = dW \chi_{V}(d) \chi_{W}(d) \int _{U \in SU(d)} \text{d}U \text{Tr} \left[ \left( |0\rangle \langle 0| \otimes M \otimes (W|0\rangle \langle 0| W^{+})^N \right) \right]. \tag{82}$$

where again we could include one of the outcomes into the unitary average and, since by redefining $W \mapsto V^{+}W$ the dependence on $V$ is a constant, we averaged over $V$ without loss of generality.

In the Bayesian case for EE we proceed as before and compute first

$$p(W) = \int dc dcp(W|c) = dW \text{Tr} \left[ |0\rangle \langle 0| \otimes (M+N) \left( 1_{M}^{sym} \otimes 1_{N}^{sym} \right) \right] = dW, \tag{83}$$
then
\[ \tilde{c}(W) = \int dc p(c) p(W|c) = dW \frac{\chi_{\frac{d}{2}}(d)\chi_{\frac{d}{2}}(d)}{\chi_{\frac{d+2}{2}}(d)} \int_{U \in SU(d)} dUTr \left[ \left( |0\rangle\langle 0|^{(M+1)} \otimes (W|0\rangle \langle W^\dagger|)^N \right) \cdot \mathbf{1} \otimes U^{\otimes(M+N)} \left( \mathbf{1}^{\text{sym}}_{M+1} \otimes |0\rangle\langle 0|^N \right) \right]. \]

\[ = dW \frac{\chi_{\frac{d}{2}}(d)\chi_{\frac{d}{2}}(d)}{\chi_{\frac{d+2}{2}}(d)} \int_{U \in SU(d)} dUTr \left[ (U|0\rangle \langle 0| U^{\dagger} \otimes (W|0\rangle \langle W^\dagger|)^\otimes N) \right]. \]

\[ = dW \frac{\chi_{\frac{d}{2}}(d)\chi_{\frac{d}{2}}(d)}{\chi_{\frac{d+2}{2}}(d)(M+1)} \left( 1 + \frac{M}{M+1} \right) \int_{U \in SU(d)} dUTr \left[ \mathbf{1}^{\text{sym}}_{N+1}(W^\dagger|0\rangle \langle 0| W \otimes |0\rangle\langle 0|^N) \right]. \]

The second equality above comes from averaging over $U_c$, the third one from Eq. (75) and introducing $U^{\dagger}U$ and its conjugate on the additional subsystem, the fourth one from redefining $U \mapsto W^\dagger U$, switching the operators acting on the $N$ subsystems and averaging over $U$, while the fifth one from applying Eq. (75) again and defining $w = |(|0\rangle W|0\rangle)^2$. Then the optimal EE Bayesian estimator for each $W$ is simply $c(W) = \tilde{c}(W)/dW$ and the minimum average MSE attained by it is

\[ v^{\text{bay}}_{ee} = \int dc p(c)c^2 - \int_{W \in SU(d)} dW \tilde{c}_W^2 = \frac{(d-1)(d+M+N)(d^2 + 2MN + d(M+N))}{d(d+1)(d+M)^2(d+N)^2}, \]

where we have carried out the group averages in the usual way:

\[ \int_{W \in SU(d)} dW w^i = \int_{W \in SU(d)} dW Tr \left[ \frac{\chi_{\frac{d}{2}}}{\chi_{\frac{d}{2}} + |0\rangle\langle 0|} \right] = 1. \]

In the limit $M \to \infty$, $N$ constant we have

\[ v^{\text{bay}}_{ee} \sim \frac{(d-1)(d+2N)}{d(d+1)(d+N)^2}, \]

which is $(d + 2N)/(d + N)$ times larger than the optimal Bayesian strategy. In the limit $M+N \to \infty$, $M-N$ fixed we have instead

\[ v^{\text{bay}}_{ee} \sim \frac{8(d-1)}{d(d+1)(M+N)}, \]

which is 2 times larger than the optimal Bayesian strategy.

Finally, for the local EE estimation the estimator $c_W = w = |(|0\rangle W|0\rangle)^2$ is a natural guess. Its variance can be computed in terms of its first and second moments according to the $p(W|c)$ distribution:

\[ v^{\text{loc}}_{ee}(c) = \int_{W \in SU(d)} p(W|c)(w - c)^2 = c^2 - 2c \overline{w} + \overline{w}^2, \]

where

\[ \overline{w} = \int p(W|c)w^i = \frac{\chi_{\frac{d}{2}}(d)\chi_{\frac{d}{2}}(d)}{\chi_{\frac{d+2}{2}}(d)} \int_{U \in SU(d)} dW dUTr \left[ \left( |0\rangle\langle 0|^{(M+1)} \otimes (W|0\rangle \langle W^\dagger|)^{(N+i)} \right) \cdot U^{\otimes(M+N)} \otimes U_c^{\otimes(N+i)} \right]. \]

\[ = \frac{\chi_{\frac{d}{2}}(d)\chi_{\frac{d}{2}}(d)}{\chi_{\frac{d+2}{2}}(d)(N+1)} \int_{U \in SU(d)} dUTr \left[ \left( ((U_c|0\rangle \langle 0| U_c^\dagger)^{\otimes M} \otimes \mathbf{1}^{\text{sym}}_{N+1} \right) \cdot (U|0\rangle \langle 0| U^{\dagger} \otimes |0\rangle\langle 0|^N) \right]. \]

Then the first moment is straightforward to compute by inserting Eq. (75):

\[ \overline{w} = \frac{\chi_{\frac{d}{2}}(d)\chi_{\frac{d}{2}}(d)}{\chi_{\frac{d+2}{2}}(d)(N+1)} \left( \frac{1}{\chi_{\frac{d}{2}}(d)} + \frac{N}{\chi_{\frac{d}{2}}(d)} \right) \left[ \mathbf{1}^{\text{sym}}_{M+1}(0) \otimes U_c^\dagger |0\rangle\langle 0| U_c \right] \]

\[ = \frac{\chi_{\frac{d}{2}}(d)\chi_{\frac{d}{2}}(d)}{\chi_{\frac{d+2}{2}}(d)(N+1)} \left( \frac{1}{\chi_{\frac{d}{2}}(d)} + \frac{N}{\chi_{\frac{d}{2}}(d)} \right) \left[ 1 + Mc \right]. \]
for the second moment we first need to evaluate the following operator:

\[
A_2(N) = \text{Tr}(1, N) \left[ 1^\text{sym} \otimes \left| 0 \right\rangle \left\langle 0 \right| \otimes N \right] = \frac{1}{(N + 2)!} \sum_{\sigma} \sum_{i_1, i_2 \in (0, \ldots, d - 1) \times} \left| 0 \right\rangle \otimes N \left( \left| j \right\rangle \left\langle j \right| \right) \left| 0 \right\rangle \otimes N
\]

\[
= \frac{N!}{(N + 2)!} \left[ 21^\text{sym} + 2N \left( \left| 0 \right\rangle \otimes 1 + \left| 1 \otimes 0 \right\rangle \right) + N(N - 1) \left| 00 \right\rangle \left\langle 00 \right| \right],
\]

(92)

As before, the third equality follows by evaluating the only non-zero elements in the sums: the first term contains all the permutations of \( N \) subsystems times the identity and the swap on the remaining subsystems, which add up to the projector on the completely symmetric subspace of the two subsystems,

\[
1^\text{sym} = \frac{1}{2} \sum_{(i_1, i_2)} \left[ \left| i_1, i_2 \right\rangle \left\langle i_1, i_2 \right| + \left| i_1, i_2 \right\rangle \left\langle i_2, i_1 \right| \right];
\]

(93)

the second term considers the additional permutations in the case \( i_1 = j_1 = 0 \) and \( i_2 = j_2 = 0 \), where one of the remaining subsystems can be swapped or not with the other, then permuted with any of the other \( N \) subsystems; analogously, the third term considers the additional permutations in the case \( i = j = 0 \), where each remaining subsystem can be permuted respectively with \( N \) and \( N - 1 \) of the others. Hence the second moment of \( w \) can be written as

\[
\overline{w}^2 = \frac{\chi_2(d) \chi_2(d)}{\chi_{2 + 2}(d)} \int_{U \in SU(d)} dU Tr \left[ \left( U_c \left| 0 \right\rangle \left\langle 0 \right| U_c^\dagger \right) \otimes M \left( U^\dagger \left| 0 \right\rangle \left\langle 0 \right| U \right)^2 \right]
\]

\[
= \frac{\chi_2(d) \chi_2(d)}{\chi_{2 + 2}(d)} (N + 2)(N + 1) \left( \frac{2}{\chi_2(d)} + \frac{4N}{\chi_{2 + 1}(d)} \text{Tr} \left[ A_1(M) U_c^\dagger \left| 0 \right\rangle \left\langle 0 \right| U_c \right] \right)
\]

\[
+ \frac{N(N - 1)}{\chi_{2 + 2}(d)} \text{Tr} \left[ A_2(M) (U_c^\dagger \left| 0 \right\rangle \left\langle 0 \right| U_c)^2 \right]
\]

(94)

By plugging the expressions of Eqs. (91), (93) into Eq. (89) we finally get

\[
v^\text{loc}_{cc}(c) = \frac{d + M + N}{d + M + (d + 2)(d + 1) + 2c(1 - c)M \chi_{2 + 2}(d)(d + 1)(d + N)}.
\]

(95)

In the limit \( M \to \infty \), \( N \) finite we have

\[
v^\text{loc}_{cc}(c) \sim \frac{2 + c(d + 1)(d + 2) + 2cN - 2c^2N}{(d + N)(d + N + 1)} \cdot \frac{2c(1 - c)}{N} + O \left( \frac{1}{N^2} \right),
\]

(96)

which is twice as large as the optimal strategy in the leading order of \( N \). In the limit \( M \to N \to \infty \), \( M - N \) fixed we have instead

\[
v^\text{loc}_{cc}(c) \sim \frac{8c(1 - c)}{M + N},
\]

(97)

Bayesian estimation using the swap test

Here we derive the formulas employed for the plot of the swap-test performance in Fig. 4. The measurement statistics is given by a binomial distribution \( \text{Bin}(k, N, p(c)) \), of \( k \) events out of \( N \), with single-event probability \( p(c) = \frac{1 + c}{2} \). The corresponding optimal classical Bayesian estimator is given by Eq. (70), with \( p(k|c) \to \text{Bin}(k, N, p(c)) \). We have

\[
p(c) = \int dc p(c) \text{Bin}(k, N, p(c)) = \frac{(d - 1)(N)}{k} 2F_1(1, -k; d - 1k + N; -1),
\]

(98)

\[
c(k) = \int dc c p(c) \text{Bin}(k, N, p(c)) = (d - 1)2^{-N} \frac{(N)}{k} 2F_1(2, -k; d - 1k + N + 1; -1) \Gamma(d - k + N - 1),
\]

(99)

where \( 2F_1 \) and \( \tilde{2F}_1 \) are hypergeometric and regularized hypergeometric functions. Following the derivations of the previous section, the minimum average MSE attainable with the swap test can be written as

\[
v^\text{bay}_{\text{sw}} = \frac{2}{d(d + 1)} - \sum_{k=0}^N \frac{(d - 1)(N)(d - k + N - 1) \tilde{2F}_1(2, -k; d - k + N + 1; -1) \Gamma(d - k + N - 1)^2}{2N \tilde{2F}_1(1, -k; d - k + N; -1)}.
\]

(100)
Average post-measurement fidelity

Here we compute the average post-measurement fidelity for the optimal strategy and the swap test. For the former we have measurement operators $M_j = E_j = \mathbf{1}_j$, so that Eq. \ref{eq:overlap} reads out

$$F_{op} = \sum_{j=J_{\text{min}}}^{J_{\text{max}}} |\langle\Psi| \mathbf{1}_j |\Psi\rangle|^2 = \sum_j p(j|c)^2,$$

(101)

where we have used the SU($d$)-invariance of $\mathbf{1}_j$. For the swap test we restrict to $M = N$ as usual and we consider that the measurement is separable and identical on each couple of copies. Moreover, the measurement on a single couple of copies is a triplet/singlet projection, again SU($d$)-invariant, which succeeds/fails with probability $(1 \pm c)/2$. Hence we have

$$F_{sw} = \left[ \frac{1 + c}{2} \right]^2 + \left( \frac{1 - c}{2} \right)^2 = \left( \frac{1 + c^2}{2} \right)^N.$$

(102)

Estimating the overlap between two arbitrary mixed qubits

In this appendix we derive the optimal estimator and corresponding mean squared error for the case where we are given $N$ and $M$ copies of mixed states. We shall restrict our attention to qubit mixed states and for ease of notation we shall revert to the standard angular momentum notation for irrep labels.

The mixed states whose overlap we wish to estimate are

$$\rho_{\psi}^{\otimes N} (r_0) = \left( r_0 |\psi\rangle \langle\psi| + (1 - r_0) \frac{\mathbf{1}}{2} \right)^{\otimes N},$$

$$\rho_{\phi}^{\otimes M} (r_1) = \left( r_1 |\phi\rangle \langle\phi| + (1 - r_1) \frac{\mathbf{1}}{2} \right)^{\otimes M},$$

(103)

where $r_{0(1)}$ denotes the corresponding purity of the states. Following Ref. the states in Eq. \ref{eq:rho_pm} can be written in the total angular momentum basis, after tracing out the multiplicity space, as

$$\rho_{\psi}^{\otimes N} = \sum_{J_0=0}^N \sum_{J_{\alpha},k=J_0}^J \rho_{J_0} \tau_{J_0}^{(0)} (\vec{n}_0),$$

$$\rho_{\phi}^{\otimes M} = \sum_{J_1=0}^\infty \sum_{J_{\alpha},l=J_1}^J \rho_{J_1} \tau_{J_1}^{(1)} (\vec{n}_1),$$

(104)

where

$$\tau_{J_0}^{(0)} = \frac{1}{Z_{J_0}} \sum_{k=-J_0}^{J_0} R_{J_0}^{k} \langle J_0, k | J_0, k \rangle$$

$$\tau_{J_1}^{(1)} = \frac{1}{Z_{J_1}} \sum_{l=-J_1}^{J_1} R_{J_1}^{l} \sum_{\alpha,\beta=-J_1}^{J_1} d^{(J_1)}_{\alpha,l} (2 \cos^{-1} \sqrt{c}) d^{(J_1)}_{\beta,l} (2 \cos^{-1} \sqrt{c}) | J_1, \alpha \rangle \langle J_1, \beta |$$

(105)

with $R_{J} = \frac{1+e_j}{1-e_j}$, $Z_{J}^{(i)} = \frac{R_{J}^{i+1}-R_{-J}^{i-1}}{R_{0}^{i-1}}$, and just as for the case of pure states, we have chosen $\vec{n}_0 = \vec{z}$ without loss of generality. Moreover,

$$p_{J_0} = \left( \frac{1 - r^2}{4} \right)^{\frac{N}{2} - J_0} \frac{2}{N} \frac{2}{J_0 + 1} Z_{J_0}$$

(106)

and similarly for $p_{J_1}$. Using Eq. \ref{eq:overlap} we obtain

$$\rho(c) = G_{SU(2)} \left[ \rho_{\psi}^{\otimes N} \otimes \rho_{\phi}^{\otimes M} \right] = \sum_{J,J_0,J_1} p_{J_0,0,J_1} \sum_{k,l} \rho_{J_0}^{k} R_{J_0}^{k} \tau_{J_0}^{(0)} \sum_{\alpha,\gamma=-J_1}^{J_1} \left( C_{J_0,k,J_1,\alpha}^{J_1} d_{\alpha,l}^{(J_1)} (2 \cos^{-1} \sqrt{c}) \right)^2 \frac{\Pi_{J}^{2J+1}}{2J+1},$$

(107)
where $\Pi_J$ is the projector on the block with label $J$. To calculate the AvMSE estimator we need to compute the operators $\Gamma, \eta$ of Eq. (51). A similar calculation as in Eq. (54), (55) gives

$$
\Gamma = G_{SU(2)}^N \left[ \rho_\psi^N \right] \otimes G_{SU(2)}^M \left[ \rho_\psi^M \right] \\
= \sum_{J_0=0}^{\frac{N}{2}} \sum_{J_1=0}^{\frac{N}{2}} P_{J_0} \begin{pmatrix} \Pi_J \end{pmatrix}_{2J_0 + 1} \otimes \sum_{J_1=0}^{\frac{N}{2}} \sum_{J_1=0}^{\frac{N}{2}} P_{J_1} \begin{pmatrix} \Pi_{J_1} \end{pmatrix}_{2J_1 + 1} \\
= \sum_{J_0=0}^{\frac{N}{2}} \sum_{J_1=0}^{\frac{N}{2}} \sum_{J_0=0}^{\frac{N}{2}} \sum_{J_1=0}^{\frac{N}{2}} \frac{P_{J_0} P_{J_1}}{(2J_0 + 1)(2J_1 + 1)} \sum_{J=J_0+J_1}^{J_0+J_1} \begin{pmatrix} \Pi_J \end{pmatrix},
$$

(108)

whereas for $\eta$ one obtains

$$
\eta = \int_{g \in SU(2)} U(g)^{\otimes (N+M)} \left( \sum_{J_0=0}^{\frac{N}{2}} \sum_{J_1=0}^{\frac{N}{2}} \sum_{J_0=0}^{\frac{N}{2}} \sum_{J_1=0}^{\frac{N}{2}} R_{J_0}^k R_{J_1}^l \langle J_0, k \mid J_0, k \rangle \langle J_1, l \mid J_1, l \rangle D^{J_1}(h) |D(h)\rangle^+ |J, J, h\rangle \right)^\dagger U^\dagger(\eta)^{\otimes (N+M)}.
$$

(109)

We finally obtain

$$
\eta = \sum_{J_0=0}^{\frac{N}{2}} \sum_{J_1=0}^{\frac{N}{2}} \sum_{J_0=0}^{\frac{N}{2}} \sum_{J_1=0}^{\frac{N}{2}} P_{J_0} P_{J_1} \left( \frac{1}{(2J_0 + 1)(2J_1 + 1)} \sum_{J=J_0+J_1}^{J_0+J_1} \begin{pmatrix} \Pi_J \end{pmatrix} \right) \\
= \sum_{J=|J_0-J_1|=L=|J_1-J_0|=\frac{1}{2}}^{J_0+J_1} \{ \text{constant factors} \} \langle J, J, L, L, \rangle \langle J, J, L, L, \rangle
$$

(110)

For a given $J, J_0, J_1$ and overlap $c$ the estimator is given by

$$
c(J, J_0, J_1) = \frac{\text{Tr} \left[ \Pi_J \left( \Pi_{J_0} \otimes \Pi_{J_1} \eta \right) \right]}{\text{Tr} \left[ \Pi_J \left( \Pi_{J_0} \otimes \Pi_{J_1} \Gamma \right) \right]} \quad (111)
$$

and the average MSE reads

$$
v_{\text{op, mix}} = \int_0^1 p(c)c^2 - \sum_{J, J_0, J_1} p(J, J_0, J_1) c(J, J_0, J_1)^2.
$$

(112)

The sums in $L, L', m, k$ can be done exactly. The sums in $J$ at the leading order in $R_0, R_1$ can be done by keeping track of the non-exponentially decaying (in $J$) contributions. The final sum in $J_0$ and $J_1$ can be done using that $p_{J_0} R_{J_0}^{J_0} Z_{J_0}$ can be written as

$$
p_{J_0} Z_{J_0}^{J_0} = \left( 1 - r_0 \right) \frac{\Delta}{4} \left( \begin{pmatrix} N \end{pmatrix} _{2J_0} + 1 \right) \begin{pmatrix} N \end{pmatrix} _{2J_0 + 1} R_{J_0}^{J_0} = \left( \begin{pmatrix} N \end{pmatrix} _{2J_0} + 1 \right) \begin{pmatrix} N \end{pmatrix} _{2J_0 + 1} + \begin{pmatrix} N \end{pmatrix} _{2J_0 + 1} \langle J_0 \rangle \left( \frac{1 + r_0}{2} \right) \langle J_0 \rangle \left( 1 - r_0 \right) \langle J_0 \rangle
$$

(113)

and in the limit $M = \alpha Z, N = \beta Z, Z \rightarrow \infty$ one can approximate the average MSE expanding in moments around the mean of the binomial distribution. The final result reads

$$
v_{\text{op, mix}} \approx \frac{1}{6M r_0^2} + \frac{1}{6N r_1^2} + O(Z^{-1})
$$

(114)

in agreement with the pure state case for $d = 2, r_0 = r_1 = 1$. 

