

# Beyond the swap test: optimal estimation of quantum state overlap

M. Fanizza,<sup>1,\*</sup> M. Rosati,<sup>2,†</sup> M. Skotiniotis,<sup>2</sup> J. Calsamiglia,<sup>2</sup> and V. Giovannetti<sup>1</sup>

<sup>1</sup>*NEST, Scuola Normale Superiore and Istituto Nanoscienze-CNR, I-56126 Pisa, Italy*

<sup>2</sup>*Física Teòrica: Informació i Fenòmens Quàntics, Departament de Física, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona) Spain*

We study the estimation of the overlap between two unknown pure quantum states of a finite dimensional system, given  $M$  and  $N$  copies of each type. This is a fundamental primitive in quantum information processing that is commonly accomplished from the outcomes of  $N$  swap-tests, a joint measurement on one copy of each type whose outcome probability is a linear function of the squared overlap. We show that a more precise estimate can be obtained by allowing for general collective measurements on all copies. We derive the statistics of the optimal measurement and compute the optimal mean square error in the asymptotic pointwise and finite Bayesian estimation settings. Besides, we consider two strategies relying on the estimation of one or both the states, and show that, although they are suboptimal, they outperform the 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 overlap between two unknown quantum states is an archetypical instance of quantum relative information [1–8] and the estimation of the overlap is a basic primitive in quantum information processing, with applications ranging from quantum fingerprinting [9–11], entanglement estimation [12–15], communication without a shared reference frame [1, 16–18] to quantum machine learning [19–28]. Recently, with the advent of quantum machine learning protocols [21–27], overlap estimation (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 [19, 20]. In most applications, OvE is carried out through the swap test (SWT) [9, 19, 20]: given two systems in the state  $|\psi\rangle|\phi\rangle$ , the probability of projecting it on its symmetric or antisymmetric part 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.

The task we consider is as follows: given  $N$  and  $M \geq N$  copies of two unknown pure states  $|\psi\rangle, |\phi\rangle$  of a  $d$ -dimensional quantum system we are requested to provide an estimate of their (squared) overlap  $|\langle\psi|\phi\rangle|^2$  which is fixed, but unknown to us. The task is carried out by a machine that performs a measurement on the state  $|\Psi\rangle = |\psi\rangle^{\otimes N} \otimes |\phi\rangle^{\otimes M}$  of  $M + N$  qudits and produces an estimate with maximum precision, as quantified by the mean square error (MSE). Furthermore we consider the case of unlabeled states, i.e., when the machine receives  $U_\sigma|\Psi\rangle$ , with  $U_\sigma$  an unknown permutation of the qudits. Note that in this case OvE constitutes in itself an instance of unsuper-

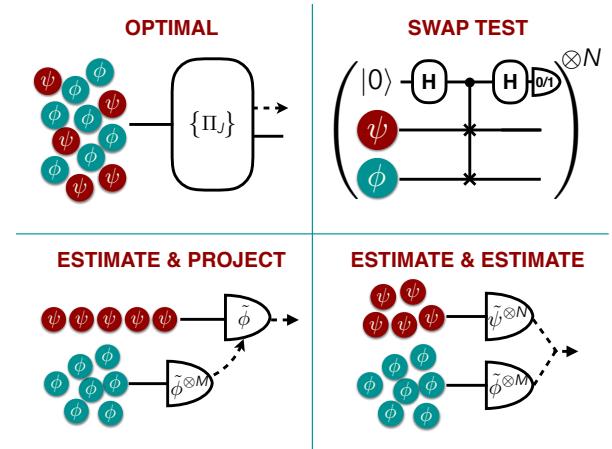


FIG. 1. Sketch of the OvE strategies studied in the article. a) Optimal measurement, e.g. by Schur transform (see Ref. [47] for the circuit implementation). b) Circuit for the SWT, to be repeated  $N$  times. c) Estimate  $|\phi\rangle$  and project  $|\psi\rangle$  on the estimated direction. d) Estimate both  $|\phi\rangle$  and  $|\psi\rangle$  and calculate the overlap.

vised quantum-classical learning problem, in a setting similar to [29].

The measurements optimizing the average information gain [16] and the average error [1, 17] have been derived for the case of qubits, with only numerical solutions [18] for higher dimensions. Here we tackle OvE in full generality, characterizing the optimal estimation within both local (pointwise) 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 [30, 33] we provide an exact solution, generalizing the results of [18]. We find that the optimal local strategy is also Bayesian-optimal asymptotically and that it performs identically in the labeled and unlabeled scenarios. We compare our results with the SWT and with two LOCC strategies based on estimating either one or both  $|\psi\rangle$  and  $|\phi\rangle$ , see Fig. 1. Such strategies are useful in distributed scenarios where copies of  $|\psi\rangle$  and  $|\phi\rangle$  are pro-

duced in different and distant laboratories. We show that in the limit of large  $M + N$  and  $|M - N|$  constant the optimal strategy displays a finite asymptotic gap with respect to all the others. Moreover, we show that the optimal measurement is less invasive than the SWT and robust against single-qubit noise.

*Assessing the machine's performance.*— The states  $|\psi\rangle = U|0\rangle$  and  $|\phi\rangle = V|0\rangle$  are drawn uniformly at random, i.e., with Haar-distributed  $U, V \in SU(d)$ . Upon performing a measurement  $\{E_k\}$  on  $|\Psi\rangle$  with outcome  $k$ , the machine outputs an estimate  $c(k)$  of the overlap  $c = |\langle\psi|\phi\rangle|^2$ , with squared error  $(c(k) - c)^2$ .

In the *global approach*, the machine's performance is quantified by averaging the squared error over all possible states and outcomes:

$$v = \sum_k \int dU dV (c(k) - c)^2 \text{Tr}[E_k |\Psi\rangle\langle\Psi|]. \quad (1)$$

We refer to  $v$  as *global* MSE. Writing  $V = UW$  and using the Haar-measure invariance  $dV = dW$  and  $c = |\langle 0|W|0\rangle|^2$  the average mean square error can be written as:  $v = \sum_k \int dW (c(k) - c)^2 \text{Tr}[E_k \rho(c)]$ , where we have defined the effective state

$$\rho(c) = \int dU U^{\otimes(N+M)} |\Psi_0\rangle\langle\Psi_0| U^{\dagger\otimes(N+M)}. \quad (2)$$

where  $|\Psi_0\rangle = \mathbf{1} \otimes W^{\otimes M} |0\rangle^{\otimes(N+M)}$ . In addition, we can write the above integral over  $W$  as an integral over the overlap such that

$$v = \sum_k \int dc p(c) (c(k) - c)^2 \text{Tr}[E_k \rho(c)] \quad (3)$$

where the distribution over overlaps is given by [29, 35]

$$p(c) = \int dU \delta(c - |\langle 0|U|0\rangle|^2) = (d-1)(1-c)^{d-2}. \quad (4)$$

From the above discussion we see that the average over both types of states, i.e. over  $U$  and  $V$ , is equivalent to an average over overlaps with weight  $p(c)$  and over different orientations,  $U^{\otimes(N+M)}$ . This is a direct consequence of the fact that if the states are completely unknown, then all pairs of states with equal overlap are equally probable and are related by a rigid unitary:  $|\langle\psi|\phi\rangle|^2 = |\langle\psi'|\phi'\rangle|^2$  if and only if it exists  $U$  such that  $|\psi'\rangle = U|\psi\rangle$  and  $|\phi'\rangle = U|\phi\rangle$ .

At variance with the global approach, where the overlap is a random variable, in the *local approach* the overlap is considered to be fixed. We then assess the performance of the machine by computing the average of the square error over all states with fixed overlap  $c$  and over all outcomes:

$$v(c) = \sum_k (c(k) - c)^2 \text{Tr}[E_k \rho(c)], \quad (5)$$

also in terms of the average state for a fixed overlap  $\rho(c)$ . We refer to  $v(c)$  as *local* MSE. As shown in the supplemental material (SM), the integral in Eq. (2)

can be performed using  $SU(d)$  representation theory and  $SU(2)$  Clebsch-Gordan coefficients, obtaining the block-diagonal form

$$p(c) = \sum_{J=J_{\min}}^{J_{\max}} p(J|c) \frac{\mathbf{1}_J}{\chi_J} \otimes |\sigma\rangle\langle\sigma|_J, \quad (6)$$

with  $J_{\min} = \frac{|M-N|}{2}$ ,  $J_{\max} = \frac{M+N}{2}$  and

$$p(J|c) = \frac{(2J+1)N!M!(1-c)^M P_{J+J_{\min}}^{(0,-2J_{\min})}\left(\frac{1+c}{1-c}\right)}{(J_{\max}-J)!(J_{\max}+1+J)!}, \quad (7)$$

where  $P_n^{(\alpha,\beta)}(x)$  is the  $n^{\text{th}}$ -degree Jacobi polynomial. In the previous equations, for  $d = 2$ ,  $J$  is the familiar total-angular-momentum label and  $\mathbf{1}_J = \sum_{M=-J}^J |J, M\rangle\langle J, M|$  is the projector on the subspace of total angular momentum  $J$ , of dimension  $\chi_J = 2J+1$ . In general, for  $d > 2$ ,  $\mathbf{1}_J$  are projectors over the subspaces of dimension  $\chi_J(d)$  hosting irreducible representations (irreps) of  $SU(d)$  arising from the tensor product of two completely symmetric representations of  $M$  and  $N$  qudits; these irreps are still indexed by an (half)-integer  $J \in [J_{\min}, J_{\max}]$ . Finally,  $|\sigma\rangle_J$  is a state representing the known labeling of the states and it belongs to the irrep-space of the permutation group, also labeled by  $J$ . Note that, in the unlabeled scenario, the average over qudit permutations acts only on  $|\sigma\rangle_J$  for each  $J$ , depolarizing it to a projector on the whole irrep space. Importantly, note that all the information about the overlap is contained in the  $J$ -statistics  $p(J|c)$ , which is independent of dimension and labeling. In particular, the optimal measurement is given by the projectors  $\Pi_J$  on the subspaces labeled by  $J$ , and it can be implemented via weak Schur sampling [47]. Indeed, for any POVM  $\{E_k\}_k$ , we can get the same outcome probability distribution if we use the POVM  $\{\Pi_J E_k \Pi_J\}_{k,J}$  and then ignore the  $J$  label. When this POVM is applied to  $\rho(c)$  the outcome probabilities are  $p_{k,J} := p(k|J)p(J|c)$ . The same outcome probabilities  $p_{k,J}$  can be generated by applying directly the POVM  $\{\Pi_J\}_J$  followed by classical post-processing. The latter can only increase the variance of the estimator, by convexity of the figure of merit:  $\sum_k p(k|J)(c_{k,J} - c)^2 \geq (\sum_k p(k|J)c_{k,J} - c)^2$ , which follows from the Cauchy-Schwartz inequality. Therefore  $\{\Pi_J\}_J$  is optimal both for local and global estimation and one can replace  $\text{Tr}[E_J \rho(c)]$  with  $p(J|c)$  in Eqs. (1,5), effectively reducing our problem to one of classical estimation, i.e., optimizing the function  $c(J)$ .

*Local estimation.*— The classical Cramer-Rao bound [31] places a lower bound on the MSE of all local unbiased estimators  $c(J)$  as  $v(c) \geq H(c)^{-1}$ , where  $H(c) = \sum_J (\partial_c p(J|c))^2 / p(J|c)$  is the Fisher information of the measurement statistics. In the limit  $M + N \rightarrow \infty$  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:

$$c_{\text{op}}^{\text{loc}}(J) = \left( \frac{2J}{M+N} \right)^2, \quad v_{\text{op}}(c) = \frac{4c(1-c)}{M+N}, \quad (8)$$

Local est.	$\mathbf{v}_{\text{op}}(\mathbf{c})$	$\mathbf{v}_{\text{ep}}(\mathbf{c})$	$\mathbf{v}_{\text{ee}}(\mathbf{c})$
$M = N \rightarrow \infty$	$\frac{4c(1-c)}{M+N}$	$\frac{3}{2}v_{\text{op}}(c)$	$2v_{\text{op}}(c)$
$M \rightarrow \infty$	$\frac{c(1-c)}{N}$	$v_{\text{op}}(c)$	$2v_{\text{op}}(c)$
Bayesian est.	$\mathbf{v}_{\text{op}}$	$\mathbf{v}_{\text{ep}}$	$\mathbf{v}_{\text{ee}}$
$M = N \rightarrow \infty$	$\frac{4(d-1)}{d(d+1)(M+N)}$	$\frac{3}{2}v_{\text{op}}$	$2v_{\text{op}}$
$M \rightarrow \infty$	$\frac{(d-1)}{d(d+1)(d+N)}$	$v_{\text{op}}$	$\frac{d+2N}{2+N}v_{\text{op}}$

TABLE I. Local MSE and global MSE attainable via the optimal, EP and EE strategies in two asymptotic limits. In all the cases the global MSEs coincide with the corresponding average local MSE values, apart from asymptotically vanishing corrections.

In the SM we show that  $v_{\text{op}}(c)$  coincides with  $H(c)^{-1}$  to leading order in  $\frac{1}{M+N}$  and hence the Cramer-Rao bound is achievable in this limit. If instead  $M \rightarrow \infty$  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  $v_{\text{op}}(c) = \frac{c(1-c)}{N}$ .

*Bayesian estimation.*— The optimal classical Bayesian (global) estimator is given [30] by  $c_{\text{op}}^{\text{bay}}(J) = \frac{\int dc c p(c)p(J|c)}{\int dc p(c)p(J|c)}$ . Using graphical calculus techniques for the recoupling theory of Clebsch-Gordan coefficients [39], as explained in the SM, we obtain the following optimal global estimator and corresponding MSE:

$$c_{\text{op}}^{\text{bay}}(J) = \frac{d + J + J^2 + \frac{M+N}{2} - (\frac{M+N}{2})^2 + MN}{(d+M)(d+N)}, \quad (9)$$

$$v_{\text{op}} = \frac{(d-1)(d+M+N)}{d(d+1)(d+M)(d+N)}. \quad (10)$$

We pause to highlight the following facts: i) when  $d$  is fixed and the number of copies is large, the prior distribution of the states is little informative with respect to the information that can be obtained by the actual measurement; indeed we can see that when  $M+N \rightarrow \infty$ ,  $M-N$  constant,  $c_{\text{op}}^{\text{bay}}(J) \approx c_{\text{op}}^{\text{loc}}(J)$  implying that the local optimal estimator is also a good Bayesian estimator and viceversa; ii) contrarily to the local estimation results, the global MSE of Eq. (10) is exact for all  $M, N$  and depends on  $d$  due to the prior, Eq. (4); iii) in particular,  $v_{\text{op}}$  decays as  $d^{-2}$  if one of either  $M$  or  $N$  is kept finite, whereas it decays only as  $d^{-1}$  when  $M, N \gg 1$ .

*1-LOCC strategies.*— We now consider a family of intermediate strategies that employ 1-LOCC on  $|\psi\rangle^{\otimes N}$  and  $|\phi\rangle^{\otimes M}$ . The estimate-and-project (EP) strategy consists in estimating  $|\phi\rangle$  from its  $M$  copies, then projecting each copy of  $|\psi\rangle$  on this estimate and counting the fraction of successful projections. When  $|\phi\rangle$  is known, projecting  $|\psi\rangle$  on  $|\phi\rangle$  is optimal [31]. However, EP is not necessarily the optimal 1-LOCC strategy. The corresponding POVM elements can be written as  $E_{V,k}^{(\text{ep})} = dVE_V^{(M)} \otimes V^{\otimes N}\Pi_k^{(N)}V^{\dagger\otimes N}$ , where  $E_V^{(M)} = \chi_{\frac{M}{2}}(d)(V|0\rangle\langle 0|V^{\dagger})^{\otimes M}$  is the optimal covariant measurement to estimate  $|\phi\rangle$  [36, 37] and  $\Pi_k^{(N)}$  represents  $k$  successful projections of the copies of  $|\psi\rangle$ .

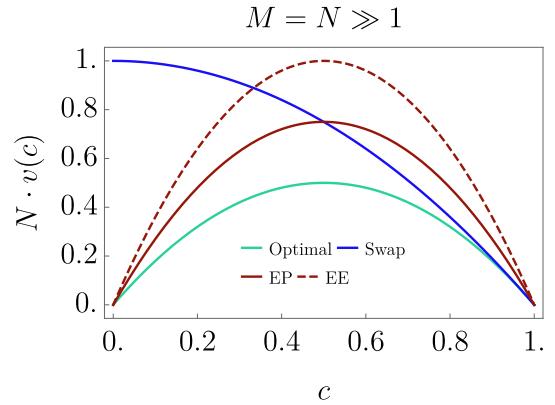


FIG. 2. Plot of the optimal local MSE scaling coefficient  $N \cdot v(c)$  vs. the true value of the overlap  $c$ , at leading order in  $M = N$ , for the strategies analyzed in the article.

on the estimate of  $|\phi\rangle$ . The estimator is  $c_{\text{ep}}^{\text{loc}}(k) = \frac{k}{N}$ . The estimate-and-estimate (EE) strategy instead consists in estimating both  $|\psi\rangle$  and  $|\phi\rangle$  separately, then computing the overlap between the estimated states. The corresponding POVM elements can be written as  $E_{V,W}^{(\text{ee})} = dVdWE_V^{(M)} \otimes E_W^{(N)}$ , i.e., a product of two covariant measurements to estimate  $|\phi\rangle$  and  $|\psi\rangle$ . We take as local estimator  $c_{\text{ee}}^{\text{loc}}(V,W) = |\langle 0|V^{\dagger}W|0\rangle|^2$ . 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  $M \rightarrow \infty$ ,  $N$  finite.

*Performances comparison.*— We now compare the strategies discussed so far with the traditional SWT [9]. Note that all these strategies except the optimal one require labeling of the states. The latter consists in projecting the state  $|\psi\rangle \otimes |\phi\rangle$  on its triplet/singlet components, hence it coincides with the optimal measurement for  $M = N = 1$ . As the SWT acts on couples of states, we restrict to the case  $M = N$ . The probability of a triplet projection  $p(c) = \frac{1}{2}(1+c)$  and the ensuing statistics of  $k$  successful projections out of  $N$  trials is given by the binomial distribution. The optimal local MSE attainable by this test is well-known,  $v_{\text{sw}}(c) = \frac{1-c^2}{N}$ , while for the optimal global MSE  $v_{\text{sw}}$  one can derive an exact expression for each value of  $k$ , then compute the sum numerically, as detailed in the SM. In the asymptotic limit of  $M = N \gg d$  a good approximation is provided by averaging the optimal local MSE:  $v_{\text{sw}} \simeq \int dc p(c)v_{\text{sw}}(c) = (d+2)(d-1)/(d(d+1)N)$ , which is  $\sim d$  times larger than  $v_{\text{op}}$ .

In the same limit, we can compare the local MSE of all the strategies, see Fig. 2. First, we observe a gap between the optimal strategy, that attains the QFI, and all the other strategies. This means that, even with a large number of copies, the collective measurement on  $|\psi\rangle^{\otimes N} \otimes |\phi\rangle^{\otimes M}$  has a clear advantage over a

non-collective one. Second, we observe that the relative error  $\frac{\sqrt{v(c)}}{c}$  for small  $c$  scales as  $\frac{1}{c\sqrt{N}}$  for the SWT and as  $\frac{1}{\sqrt{cN}}$  for the other strategies, implying a quadratic improvement in  $\frac{1}{\sqrt{c}}$  in the number copies needed to reach a fixed relative error, while the optimal measurement is still computationally efficient (see next section). This is particularly relevant since for large  $d$  small overlaps are exponentially more likely, see Eq. 4. This phenomenon is also at the source of the so-called “barren plateau” problem [40, 41] for quantum variational circuits, and other types of strategies have been proposed to address this issue [42–44].

We notice similar features for the global MSE, plotted in Fig. 3 as a function of  $N$  for  $M$  fixed and increasing  $d$  (inset). We observe that the SWT 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.

*Gate complexity.*— The advantage in the precision of the optimal estimation comes with the tradeoff that the optimal measurement requires entangling operations over the whole system of  $N + M$  qudits. The Schur transform [45–47] is a way to perform the optimal measurement, and requires  $O(\text{poly}(N + M, \log d, \log \frac{1}{\epsilon}))$  qudit gates for precision  $\epsilon$ . The resulting algorithm is efficient, but still unfeasible without error correction. The SWT instead requires  $N$  independent circuits of fixed depth, and may still be convenient for large overlaps or very noisy gates.

A mid-term solution is to divide input data in  $R$  groups of  $S$  copies of  $|\phi\rangle$  and  $|\psi\rangle$ , such that  $S$  is the largest integer for which the given architecture can perform the optimal measurement with high fidelity, repeat the measurement  $R$  times and do classical post-processing. The performances of these intermediate protocols are between SWTs and optimal measurement. See the SM for a more detailed discussion of these issues.

*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. Both the optimal measurement and the SWT 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

$$F(c) = \int_{U \in \text{SU}(d)} dU \sum_k |\langle \Psi_U | E_k | \Psi_U \rangle|^2, \quad (11)$$

with  $\{E_k \equiv \mathbf{1}_J\}$  for the optimal measurement and  $\{E_k = \mathcal{G}_{S_N}(\mathbf{1}_2^{S \otimes k} \otimes \mathbf{1}_2^{A \otimes N-k})\}$  for the SWT, where  $\mathbf{1}_2^{S/A}$  are the projectors on the singlet/triplet components of  $\mathcal{H}_2^{\otimes 2}$ . Then Eq. (11) is simply given by

$$F_{\text{op}}(c) = \sum_{J=J_{\min}}^{J_{\max}} p(J|c)^2, \quad F_{\text{sw}}(c) = \left(\frac{1+c^2}{2}\right)^N, \quad (12)$$

as shown in the SM. In Fig. 3 we plot these two quantities as a function of  $c$ , showing that the optimal measurement is less invasive than the SWT, especially for small overlap values.

*Noise-robustness.*— Finally, we consider how the optimal strategy changes when the states, which are expected to be pure, are affected by depolarizing noise acting independently on each qudit before reaching the measurement stage. Note that if the noisy channel is of a different kind, one can at least reach the optimal MSE for the depolarizing channel by performing a twirling operation, realizable by pre- and post-processing with random unitaries on each qudit plus classical forward communication. This operation is  $\int dU U^\dagger \mathcal{N}(U\rho U^\dagger) U = \Delta_r(\rho)$  for some  $r$ , where we have defined the depolarizing channel as  $\Delta_r = r\mathcal{I} + (1-r)\frac{1}{d}\text{Tr}$  and  $\mathcal{I}$  is the identity channel. After this operation the overall state of the system can now be written as  $\Delta_{r_0}(\psi)^{\otimes N} \otimes \Delta_{r_1}(\phi)^{\otimes M}$ .

In the SM we compute the optimal MSE in this case, restricting to  $d = 2$  for simplicity. In the limit  $M, N \rightarrow \infty$  with  $\frac{M}{N}$  finite, the global MSE at leading order is  $v_{\text{op,mix}} = \frac{1}{6Mr_0^2} + \frac{1}{6Nr_1^2}$ , which agrees with the previously found limit of Eq. (10) 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 the dimension of their Hilbert space and their number of copies. We showed that the commonly-used SWT 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, although it remains experimentally challenging. A practical alternative is to do Schur sampling on subsets of the dataset, followed by classical post-processing. In addition, we proposed two intuitive strategies that estimate separately one or both states and showed that they also outperform the SWT. Finally, we showed that the optimal measurement is less invasive than the SWT and robust to white noise. The strategies we introduced provide several clear advantages over the SWT, and they could become a standard tool for various quantum technologies, as well as providing improvements in the runtime of quantum algorithms.

*Note added after publication.*— We thank one of the referees of TQC 2020 for pointing out [48] to us, where the authors find a minimum variance unbiased estimator of the Hilbert-Schmidt distance of two unknown mixed states, and compute its variance. Our results, valid for more general estimators, complement their analysis in the case of pure states.

*Acknowledgments.*— M. F. and V. G. acknowledge support from PRIN 2017 “Taming complexity with quantum strategies”. M. R., M. S. and J. C. acknowledge support from the Spanish MINECO, project FIS2016-80681-P with the support of AEI/FEDER

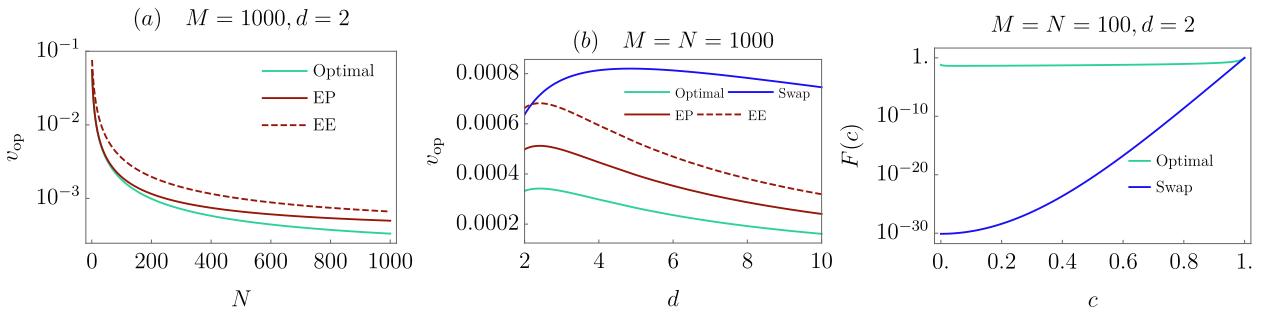


FIG. 3. (a) Plot of the optimal global MSE  $v_{opt}$  vs. the number of copies of one state  $N$ , for a fixed number of copies of the other  $M = 1000$ , in dimension  $d = 2$ , for the optimal, EP and EE strategies. (b) Plot of the optimal global MSE  $v_{opt}$  vs. the dimension  $d$ , for a fixed number of copies  $M = N = 1000$  for all the strategies studied. (c) 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 SWT.

funds; the Generalitat de Catalunya, project CIRIT 2017-SGR-1127. M. R. also acknowledges partial financial support by the Baidu-UAB collaborative project “Learning of Quantum Hidden Markov Models”. M. S. also acknowledges support from the Spanish MINECO project IJCI-2015-24643. M. F. thanks Matthias Christandl for helpful comments. M. S. and J. C. acknowledge useful discussion with Nana Liu.

\* marco.fanizza@sns.it

† Matteo.Rosati@uab.cat

- [1] N. H. Lindner, P. F. Scudo, and D. Bruß, Int. J. of Quantum Information **4**, 131 (2006).
- [2] A. Peres and P. F. Scudo, in *Quantum Theory: Reconsideration of Foundations*, edited by A. Khrennikov (Växjö University Press, Växjö Sweden, 2002), p.283.
- [3] S. D. Bartlett, T. Rudolph, and R. W. Spekkens, Rev. Mod. Phys. **79**, 555 (2007).
- [4] G. Gour, and R. W. Spekkens, New Journal of Physics **10** (3), 033023 (2008).
- [5] I. Marvian and R. W. Spekkens, New Journal of Physics **15** (3), 033001 (2013).
- [6] I. Marvian and R. W. Spekkens, Phys. Rev. A **90**, 062110 (2014).
- [7] I. Marvian and R. W. Spekkens, Communications in Mathematical Physics, 331(2), 431-475 (2014).
- [8] Y. Mo and G. Chiribella arxiv preprint, arXiv:1906.01300.
- [9] H. Buhrman, R. Cleve, J. Watrous, and R. de Wolf, Phys. Rev. Lett. **87**, 167902 (2001).
- [10] J. N. de Beaudrap, Phys. Rev. A **69**, 022307 (2004).
- [11] N. Kumar, E. Diamanti, and I. Kerenidis, Phys. Rev. A **95**, 032337 (2017).
- [12] A. K. Ekert, C. M. Alves, D. K. L. Oi, M. Horodecki, P. Horodecki, and L. C. Kwek, Phys. Rev. Lett. **88**, 217901 (2002).
- [13] F. Mintert, M. Kus, and A. Buchleitner, Phys. Rev. Lett., **95**, 260502 (2005).
- [14] A. W. Harrow and A. Montanaro, J. ACM **60**, 3 (2010).
- [15] S. P. Walborn, P. H. Souto Ribeiro, L. Davidovich, F. Mintert, and A. Buchleitner, Phys. Rev. A, **75**, 032338 (2007).
- [16] S. D. Bartlett, T. Rudolph, and R. W. Spekkens, Phys. Rev. A **70**, 032321 (2004).
- [17] E. Bagan, S. Iblisdir, and R. Muñoz-Tapia, Phys. Rev. A **73**, 022341 (2006).
- [18] N. Gisin and S. Iblisdir, European Physical Journal D **39**, 321-327 (2006).
- [19] L. Cincio, Y. Subaşı, A. T. Sornborger, and P. J. Coles, New J. Phys. **20**, 113022 (2018).
- [20] U. Chabaud, E. Diamanti, D. Markham, E. Kashefi, and A. Joux, Physical Review A **98** (6), 062318 (2018).
- [21] S. Lloyd, M. Mohseni, and P. Rebentrost, arXiv preprint quant-ph/1307.0411 .
- [22] P. Rebentrost, M. Mohseni, and S. Lloyd, Phys. Rev. Lett. **113**, 130503 (2014).
- [23] A. W. Harrow, A. Hassidim, and S. Lloyd, Phys. Rev. Lett., **103**, 150502 (2009).
- [24] Z. Zhao, J.K. Fitzsimons, and J.F. Fitzsimons, Phys. Rev. A **99**, 052331 (2019).
- [25] N. Wiebe, A. Kapoor, and K. Svore, Quantum Inf. Comput. **15**, 0318-0358 (2015).
- [26] V. Havlicek, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, Nature **567**, 209-212 (2019).
- [27] N. Liu and P. Rebentrost, Phys. Rev. A **97**, 042315 (2018).
- [28] H.-Y. Huang, K. Bharti, and Patrick Rebentrost, arXiv preprint arXiv:1909.07344.
- [29] G. Sentís, A. Monràs, R. Muñoz-Tapia, and J. Calsamiglia Phys. Rev. X **9**, 041029 (2019).
- [30] C. W. Helstrom, Quantum detection and estimation theory. Journal of Statistical Physics, 1(2), 231-252 (1969).
- [31] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982).
- [32] M. G. A. Paris, Int. J. Quant. Inf. **7**, 125-137 (2009).
- [33] S. D. Personick, IEEE Trans. Inf. Theory **17**, 240 (1971).
- [34] G. Szegő, *Orthogonal Polynomials*, 4th ed. Colloquium Publications Vol. 23 (American Mathematical Society, Providence, 1975).
- [35] L. Alonso and T. Gorin, Journal of Physics A: Mathematical and Theoretical **49**, 145004 (2016).
- [36] M. Hayashi, J. Phys. A. Math. Gen. **31**, 4633 (1998).
- [37] M. Hayashi, *Group Representation for Quantum Theory*. Berlin: (Springer, Berlin, 2017).
- [38] B. Gendra, E. Ronco-Bonvehi, J. Calsamiglia, R. Munoz-Tapia, and E. Bagan, New Journal of Physics **14**, 105015 (2012).
- [39] D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskii, *Quantum Theory of Angular Momen-*

- tum, (World Scientific, Singapore, 1988).
- [40] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, *Nature Communications* **9**, 4812 (2018).
- [41] A. G. R. Day, M. Bukov, P. Weinberg, P. Mehta, and D. Sels, *Physical Review Letters* **122**, 020601 (2019).
- [42] X. Glorot and Y. Bengio, in *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics, Sardinia, Italy, 2010*, p. 249.
- [43] M. Benedetti, D. Garcia-Pintos, O. Perdomo, V. Leyton-Ortega, Y. Nam, and A. Perdomo Ortiz, *npj Quantum Information* **5**, 45 (2019).
- [44] S. Khatri, R. LaRose, A. Poremba, L. Cincio, A. T. Sornborger, and P. J. Coles, *Quantum* **3**, 140 (2019).
- [45] D. Bacon, I. L. Chuang, and A. W. Harrow, *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, (SODA '07)*, (Society for Industrial and Applied Mathematics, Philadelphia, 2007).
- [46] H. Krovi, *Quantum* **3**, 122 (2019).
- [47] A. W. Harrow, Ph.D. thesis, (Massachusetts Institute of Technology, 2005). arXiv preprint: quant-ph/0512255.
- [48] C. Bădescu, R. O'Donnell, and J. Wright, in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pp. 503-514 (2019).
- [49] H. Weyl, *The classical groups: their invariants and representations*, (Princeton University Press, 1949).
- [50] S. Sternberg, *Group theory and physics*, (Cambridge University Press, 1995).
- [51] G. Darboux, *Journal de Mathmatiques Pures et Appliquées* (1878): 5-56.
- [52] I. N. Sanov, *Mat. Sbornik* **42**, 11-44 (1957) (in Russian). English translation: *Selected Translat. Math. Stat.* **1**, 213-244 (1961).

## SUPPLEMENTAL MATERIAL

### CONTENTS

References	5
Schur-Weyl duality and the irreducible representations of $SU(d)$ and $S_{N+M}$	6
Average state at fixed overlap	8
Fisher Information of $p(J c)$	10
Optimal global mean squared error	14
1-LOCC strategies	16
Estimate-and-project: local estimation	16
Estimate-and-project: Bayesian estimation	17
Estimate-and-estimate: Bayesian estimation	19
Estimate-and-estimate: local estimation	20
Bayesian estimation using the swap test	21
Comparison between Schur transform and swap test with imperfect implementations	21
Average post-measurement fidelity	22
Estimating the overlap between two arbitrary mixed qubits	22

### Schur-Weyl duality and the irreducible representations of $SU(d)$ and $S_{N+M}$

We begin by first reviewing some key concepts and techniques in group representation theory that we will frequently make use of throughout this Supplemental material. The most important ingredient is Schur-Weyl duality [49]. Consider the state space of  $N + M$ ,  $d$ -dimensional systems,  $\mathcal{H}_d^{\otimes(N+M)}$ . This space carries the action of two different groups; the special unitary group of  $d \times d$  complex matrices,  $SU(d)$ , and the permutation group of  $N + M$  objects,  $S_{N+M}$ . Specifically, the groups  $SU(d)$  and  $S_{N+M}$  act on a basis  $\{|i_1\rangle \otimes |i_2\rangle \otimes \dots \otimes |i_{N+M}\rangle\}_{i_1, i_2, \dots, i_{N+M}}$  of  $\mathcal{H}_d^{\otimes(N+M)}$  via unitary representations  $u_{N+M} : SU(d) \rightarrow U(\mathcal{H}_d^{\otimes(N+M)})$ , and  $s_{N+M} : S_{N+M} \rightarrow U(\mathcal{H}_d^{\otimes(N+M)})$  as follows

$$\begin{aligned} u_{N+M}(U) |i_1\rangle \otimes |i_2\rangle \otimes \dots \otimes |i_{N+M}\rangle &= U^{\otimes(N+M)} |i_1\rangle \otimes |i_2\rangle \otimes \dots \otimes |i_{N+M}\rangle \\ &= U |i_1\rangle \otimes U |i_2\rangle \otimes \dots \otimes U |i_{N+M}\rangle, \quad \forall U \in SU(d) \end{aligned} \quad (13)$$

$$s_{N+M}(\sigma) |i_1\rangle \otimes |i_2\rangle \otimes \dots \otimes |i_{N+M}\rangle = |\sigma^{-1}(i_1)\rangle \otimes |\sigma^{-1}(i_2)\rangle \otimes \dots \otimes |\sigma^{-1}(i_{N+M})\rangle, \quad \forall \sigma \in S_{N+M}. \quad (14)$$

Observe that  $[U^{\otimes(N+M)}, s(\sigma)] = 0$ ,  $\forall U \in \mathrm{SU}(d)$ , and  $\forall \sigma \in S_{N+M}$ . Schur-Weyl duality states that the total state space  $\mathcal{H}_d^{\otimes(N+M)}$  can be decomposed as

$$\mathcal{H}_d^{\otimes(N+M)} \cong \bigoplus_Y \mathcal{U}^{(Y)}(\mathrm{SU}(d)) \otimes \mathcal{U}^{(Y)}(S_{N+M}), \quad (15)$$

where  $\mathcal{U}^{(Y)}(\mathrm{SU}(d))$  is the space of dimension  $\chi_Y$  upon which the unitary irreducible representation (irrep)  $u^{(Y)}$  of  $\mathrm{SU}(d)$  acts and  $\mathcal{U}^{(Y)}(S_{N+M})$  is the space of dimension  $\omega_Y$  on which the irrep  $s^{(Y)}$  of  $S_{N+M}$  acts. The use of the congruence sign in Eq. (15) indicates that this block decomposition is accomplished by a unitary transformation; in the case considered here this unitary is the Schur transform [45, 46].

The label  $Y$  indexes the integer partitions of  $N+M$  in at most  $d$  parts written in descending order, pictorially represented by Young diagrams, where  $N+M$  boxes are arranged into at most  $d$  rows. For  $d=2$  the block decomposition of Eq. (15) is the familiar decomposition of  $N+M$  spin- $1/2$  systems into the total angular momentum label  $J$ . The latter is related to Young diagrams of at most two rows where the number of boxes in any of the two rows is given by  $(\frac{N+M}{2} + J, \frac{N+M}{2} - J)$ .

With the help of Eq. (15) we can express the action of  $u_{N+M}$  and  $s_{N+M}$  as

$$\begin{aligned} u_{N+M}(U) &= \bigoplus_Y u^{(Y)}(U) \otimes \mathbf{1}, \quad \forall U \in \mathrm{SU}(d), \\ s(\sigma) &= \bigoplus_Y \mathbf{1} \otimes s^{(Y)}(\sigma), \quad \forall \sigma \in S_{N+M}. \end{aligned} \quad (16)$$

where we have made implicit the fact that  $u^{(Y)}(U)$  acts on  $\mathcal{U}^{(Y)}(\mathrm{SU}(d))$  and  $s^{(Y)}(\sigma)$  acts on  $\mathcal{U}^{(Y)}(S_{N+M})$  respectively. Given a state  $\rho \in \mathcal{B}(\mathcal{H}_d^{\otimes(N+M)})$ , and applying  $u_{N+M}(U)$  with  $U$  extracted from the Haar measure  $dU$  of  $\mathrm{SU}(d)$  gives rise to the following completely positive, trace-preserving (CPTP) map

$$\mathcal{G}_{\mathrm{SU}(d)}[\rho] = \int dU U^{\otimes(N+M)} \rho U^{\dagger \otimes(N+M)}, \quad (17)$$

Similarly, applying  $s(\sigma)$  chosen uniformly at random gives rise to the following CPTP map

$$\mathcal{G}_{S_{N+M}}[\rho] = \frac{1}{(N+M)!} \sum_{\sigma} s_{N+M}(\sigma) \rho s_{N+M}^{\dagger}(\sigma). \quad (18)$$

By decomposing the representations  $u_{N+M}$ , and  $s_{N+M}$  into their irreps as in Eq. (16), and making use of Schur's lemmas [50] Eqs (17, 18) can be conveniently written as [3]

$$\begin{aligned} \mathcal{G}_{\mathrm{SU}(d)}[\rho] &= \bigoplus_Y (\mathcal{D}_{\mathcal{U}^{(Y)}(\mathrm{SU}(d))} \otimes \mathcal{I}_{\mathcal{U}^{(Y)}(S_{N+M})}) [\Pi_Y \rho] \\ \mathcal{G}_{S_{N+M}}[\rho] &= \bigoplus_Y (\mathcal{I}_{\mathcal{U}^{(Y)}(\mathrm{SU}(d))} \otimes \mathcal{D}_{\mathcal{U}^{(Y)}(S_{N+M})}) [\Pi_Y \rho], \end{aligned} \quad (19)$$

where  $\mathcal{I}$  is the identity map, i.e.,  $\mathcal{I}_{\mathcal{H}}[A] = A$ ,  $\mathcal{D}$  is the completely depolarizing map, i.e.,  $\mathcal{D}_{\mathcal{H}}[A] = \frac{\mathrm{Tr}[A]}{\dim(\mathcal{H})} \mathbf{1}_{\mathcal{H}}$ , and  $\Pi_Y$  is the projector onto the block  $\mathcal{U}^{(Y)}(\mathrm{SU}(d)) \otimes \mathcal{U}^{(Y)}(S_{N+M})$ .

In the following we will need some more observations. Define  $\mathrm{Sym}_N^{(d)}$  as the completely symmetric subspace of  $\mathcal{H}_d^{\otimes N}$ ; in particular, if taken  $Y_{max}$  to be the label for a Young diagram of just one row, one has  $\mathrm{Sym}_N^{(d)} = \mathcal{U}^{(Y_{max})}(\mathrm{SU}(d)) \otimes \mathcal{U}^{(Y_{max})}(S_N)$ . Moreover, the tensor product space

$$\mathrm{Sym}_{N,M}^{(d)} := \mathrm{Sym}_N^{(d)} \otimes \mathrm{Sym}_M^{(d)} \subset \mathcal{H}_d^{\otimes N+M}. \quad (20)$$

admits a decomposition

$$\mathrm{Sym}_{N,M}^{(d)} = \bigoplus_J \left( \mathcal{U}^{(J)}(\mathrm{SU}(d)) \otimes \mathcal{K}^J \right), \quad (21)$$

where  $J$  are the labels of Young diagram with two rows and  $\mathcal{K}^J := \mathrm{span}\{|J\rangle\}$ ,  $|J\rangle \in \mathcal{U}^{(J)}(S_{M+N})$ ,  $\mathcal{K}^J$  being a one dimensional space because the multiplicity of each irrep  $\mathcal{U}^{(J)}(\mathrm{SU}(d))$  is one. Notice that the label  $J$  can be indexed by an half-integer  $\frac{M-N}{2} \leq J \leq \frac{M+N}{2}$ .

The states of the form  $(|\psi\rangle\langle\psi|)^{\otimes N} \otimes (|\phi\rangle\langle\phi|)^{\otimes M}$  are supported on  $\mathrm{Sym}_{N,M}^{(d)}$ . Define also  $\mathcal{E} := \mathrm{span}\{|\psi\rangle, |\psi_{\perp}\rangle\} \subseteq \mathcal{H}_d$ . Since  $\mathcal{E}$  has dimension 2, it is isomorphic to  $\mathcal{H}_2$ . Using this isomorphism we can define  $\mathrm{Sym}_N^{\mathcal{E}} \cong \mathrm{Sym}_N^{(2)}$ .

Notice that  $(|\psi\rangle\langle\psi|)^{\otimes N} \otimes (|\phi\rangle\langle\phi|)^{\otimes M}$  is also supported on the product of the two completely symmetric subspaces of  $\mathcal{E}^{\otimes N}$  and  $\mathcal{E}^{\otimes M}$ ,  $\text{Sym}_{N,M}^{\mathcal{E}} := \text{Sym}_N^{\mathcal{E}} \otimes \text{Sym}_M^{\mathcal{E}}$ , which is isomorphic to

$$\text{Sym}_{N,M}^{(2)} := \bigoplus_J \left( \mathcal{U}^{(J)}(\text{SU}(2)) \otimes \mathcal{K}_J^{(2)} \right) \subseteq \text{Sym}_{N,M}^{(d)}. \quad (22)$$

where now  $K_J^{(2)} = \text{span}\{|J\rangle_{(2)}\}$ ,  $|J\rangle_{(2)} \in \mathcal{U}^{(J)}(S_{M+N})$ . Using this isomorphism, we can write the decomposition

$$\text{Sym}_{N,M}^{\mathcal{E}} := \bigoplus_J \left( \mathcal{U}_{\mathcal{E}}^{(J)}(\text{SU}(2)) \otimes \mathcal{K}_{\mathcal{E}}^J \right) \subseteq \text{Sym}_{N,M}^{(d)}. \quad (23)$$

and now  $K_{\mathcal{E}}^J = \text{span}\{|J\rangle_{\mathcal{E}}\}$ ,  $|J\rangle_{\mathcal{E}} \in \mathcal{U}^{(J)}(S_{M+N})$ . What is more,  $\mathcal{E}^{\otimes N} \otimes \mathcal{E}^{\otimes M}$  is an invariant subspace of  $\mathcal{H}_d^{\otimes M+N}$  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 irreps in the decomposition of the representation of  $S_{M+N}$  on  $\mathcal{H}_d^{\otimes M+N}$ , and it follows that  $\mathcal{U}_{\mathcal{E}}^{(J)}(\text{SU}(2)) \otimes \mathcal{K}_{\mathcal{E}}^J \subseteq \mathcal{U}^{(J)}(\text{SU}(d)) \otimes \mathcal{U}^{(J)}(\mathcal{S}_{N+M})$ . Finally, since  $\text{Sym}_{N,M}^{(2)} \subseteq \text{Sym}_{N,M}^{(d)}$  and each of the  $\mathcal{U}_{\mathcal{E}}^{(J)}(\text{SU}(2)) \otimes \mathcal{K}_{\mathcal{E}}^J$  and  $\mathcal{U}^{(J)}(\text{SU}(d)) \otimes \mathcal{K}^J$  are included in the same subspace  $\mathcal{U}^{(J)}(\text{SU}(d)) \otimes \mathcal{U}^{(J)}(\mathcal{S}_{N+M})$ , it follows that

$$\mathcal{U}_{\mathcal{E}}^{(J)}(\text{SU}(2)) \otimes \mathcal{K}_{\mathcal{E}}^J \subseteq \mathcal{U}^{(J)}(\text{SU}(d)) \otimes \mathcal{K}^J \quad (24)$$

and  $|J\rangle = |J\rangle_{\mathcal{E}}$  for each  $J$ .

### Average state at fixed overlap

In this section 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. Using group representation theory techniques we find a basis in which the average state is diagonal and compute the eigenvalues. We generalize the result to the average state in the unlabeled scenario.

Without loss of generality we may write the latter of the two states as

$$|\phi\rangle = V(\theta)|\psi\rangle \equiv \sqrt{c}|\psi\rangle + \sqrt{1-c}|\psi_{\perp}\rangle \quad (25)$$

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

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

$$\mathcal{G}_{\text{SU}(d)} \left[ (|\psi\rangle\langle\psi|)^{\otimes N} \otimes (|\phi\rangle\langle\phi|)^{\otimes M} \right] \equiv \int_{\text{SU}(d)} dU \left( U |\psi\rangle\langle\psi| U^\dagger \right)^{\otimes N} \otimes \left( U |\phi\rangle\langle\phi| U^\dagger \right)^{\otimes M} \quad (26)$$

where  $dU$  is the Haar measure of  $SU(d)$ . We can decompose the integral over  $SU(d)$  in Eq. (26) as follows. First we consider the two-dimensional subspace  $\mathcal{E} \equiv \text{span}\{|\psi\rangle, |\psi_{\perp}\rangle\} \subseteq \mathcal{H}_d$ . Using the invariance of the Haar measure, we first perform the group average over the  $SU(2)$  subgroup of  $SU(d)$  which acts non-trivially only on  $\mathcal{E}$ , which we denote as  $SU(\mathcal{E})$ . Afterwards, we can average over all  $SU(d)$ . This implies that Eq. (26) can be written as

$$\begin{aligned} \mathcal{G}_{\text{SU}(d)} \left[ (|\psi\rangle\langle\psi|)^{\otimes N} \otimes (|\phi\rangle\langle\phi|)^{\otimes M} \right] &= \int_{\text{SU}(d)} dU U^{\otimes(N+M)} \left[ (|\psi\rangle\langle\psi|)^{\otimes N} \otimes (|\phi\rangle\langle\phi|)^{\otimes M} \right] U^\dagger^{\otimes(N+M)} \\ &= \int_{\text{SU}(d)} dU U^{\otimes(N+M)} \left( \int_{\text{SU}(\mathcal{E})} dV V^{\otimes(N+M)} \left[ (|\psi\rangle\langle\psi|)^{\otimes N} \otimes (|\phi\rangle\langle\phi|)^{\otimes M} \right] V^\dagger^{\otimes(N+M)} \right) U^\dagger^{\otimes(N+M)} \\ &= \mathcal{G}_{\text{SU}(d)} \left[ \mathcal{G}_{\text{SU}(\mathcal{E})} \left[ (|\psi\rangle\langle\psi|)^{\otimes N} \otimes (|\phi\rangle\langle\phi|)^{\otimes M} \right] \right] \end{aligned} \quad (27)$$

First of all we compute the action of  $\mathcal{G}_{\text{SU}(2)}$  on  $(|\psi\rangle\langle\psi|)^{\otimes N} \otimes (|\phi\rangle\langle\phi|)^{\otimes M}$ , when  $d = 2$ . We use the addition rules for angular momentum on  $\mathcal{H}_2^{\otimes N+M}$  to write

$$\begin{aligned} |\psi\rangle^{\otimes N} \otimes |\phi\rangle^{\otimes M} &= \left| \frac{N}{2}, \frac{N}{2} \right\rangle \otimes \sum_{k=-\frac{M}{2}}^{\frac{M}{2}} d_{k\frac{M}{2}}^{\left(\frac{M}{2}\right)} \left| \frac{M}{2}, k \right\rangle \\ &= \sum_{J=J_{\min}}^{J_{\max}} \sum_{k=-\frac{M}{2}}^{\frac{M}{2}} C_{\frac{N}{2}, \frac{N}{2}; \frac{M}{2}, k}^{J, \frac{N}{2}+k} d_{k, \frac{M}{2}}^{\left(\frac{M}{2}\right)}(\theta) \left| J, \frac{N}{2} + k \right\rangle, \end{aligned} \quad (28)$$

where  $C_{\frac{N}{2}, \frac{N}{2}; \frac{M}{2}, k}^{J, \frac{N}{2} + k} = \langle J, \frac{N}{2} + k | \frac{N}{2}, \frac{N}{2}; \frac{M}{2}, k \rangle$  are the Clebsch-Gordan coefficients. Using Eq. (17) and the decomposition in Eq. (23) ( $(|\psi\rangle\langle\psi|)^{\otimes N} \otimes (|\phi\rangle\langle\phi|)^{\otimes M}$  is supported on  $\text{Sym}_{N,M}^{\mathcal{E}}$ ) the first Haar-measure group average of Eq. (27) reads

$$\mathcal{G}_{\text{SU}(2)} \left[ (|\psi\rangle\langle\psi|)^{\otimes N} \otimes (|\phi\rangle\langle\phi|)^{\otimes M} \right] = \sum_{J=J_{\min}}^{J_{\max}} p(J|c) \frac{\mathbf{1}_{\mathcal{U}^{(J)}(\text{SU}(2))}}{2J+1} \otimes |J\rangle\langle J|_{\mathcal{U}^{(J)}(S_{N+M})}, \quad (29)$$

where we have used the total angular momentum  $J$ , instead of the Young frame label  $Y$ ,  $\chi_J = 2J+1$ , and  $|J\rangle\langle J|_{\mathcal{U}^{(J)}(S_{N+M})}$  is a pure state. The coefficients of Eq. (29) are:

$$\begin{aligned} p(J|c) &= \sum_{k=-\frac{M}{2}}^{\frac{M}{2}} \left( C_{\frac{N}{2}, \frac{N}{2}; \frac{M}{2}, k}^{J, \frac{N}{2} + k} D_{k \frac{M}{2}}^{(\frac{M}{2})} (2 \arccos \sqrt{c}) \right)^2 \\ &= \frac{(2J+1)(J+J_{\min})!N!}{(J-J_{\min})!(J_{\max}-J)!(J_{\max}+1+J)!} \sum_{k=-\frac{M}{2}}^{J-\frac{N}{2}} \frac{(\frac{M}{2}-k)!(J+\frac{N}{2}+k)!}{(J-\frac{N}{2}-k)!(\frac{M}{2}+k)!} \left( D_{k \frac{M}{2}}^{(\frac{M}{2})} (2 \arccos \sqrt{c}) \right)^2 \\ &= \frac{(2J+1)(J+J_{\min})!N!M!}{(J-J_{\min})!(J_{\max}-J)!(J_{\max}+1+J)!} \sum_{k=-\frac{M}{2}}^{J-\frac{N}{2}} \frac{(J+\frac{N}{2}+k)!}{(J-\frac{N}{2}-k)!(\frac{M}{2}+k)!^2} (1-c)^{\frac{M}{2}-k} c^{\frac{M}{2}+k} \\ &= \frac{(2J+1)N!M!(1-c)^M}{(J_{\max}-J)!(J_{\max}+1+J)!} P_{J+J_{\min}}^{(0, -2J_{\min})} \left( \frac{1+c}{1-c} \right), \end{aligned} \quad (30)$$

and we have made use of the following expression for the Wigner  $D$  matrix in going from the second to the third line in Eq. (30)

$$D_{z', z}^{(J)}(\theta) = \sqrt{\frac{(J+z)!(J-z)!}{(J+z')!(J-z')!}} \sin^{(z-z')} \left( \frac{\theta}{2} \right) \cos^{(z+z')} \left( \frac{\theta}{2} \right) P_{(J-z)}^{(z-z', z+z')}(\cos \theta), \quad (31)$$

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

$$P_n^{(\alpha, \beta)}(x) = \frac{\Gamma(\alpha+n+1)}{n!\Gamma(\alpha+\beta+n+1)} \sum_{m=0}^n \binom{n}{m} \frac{\Gamma(\alpha+\beta+n+m+1)}{\Gamma(\alpha+m+1)} \left( \frac{x-1}{2} \right)^m. \quad (32)$$

Let's consider now  $d > 2$ . Notice that  $\mathcal{E}^{\otimes M+N} \cong \bigoplus_J \mathcal{U}_{\mathcal{E}}^{(J)}(\text{SU}(2)) \otimes \mathcal{U}_{\mathcal{E}}^{(J)}(S_{N+M})$ . Performing the average over  $\text{SU}(d)$  on the state given by Eq. (29) results in

$$\begin{aligned} \mathcal{G}_{\text{SU}(d)} \left[ \mathcal{G}_{\text{SU}(\mathcal{E})} \left[ (|\psi\rangle\langle\psi|)^{\otimes N} \otimes (|\phi\rangle\langle\phi|)^{\otimes M} \right] \right] &= \bigoplus_Y (\mathcal{D}_{\mathcal{U}^{(Y)}(\text{SU}(d))} \otimes \mathcal{I}_{\mathcal{U}^{(Y)}(S_{N+M})}) \left[ \Pi_Y \left( \sum_{J=J_{\min}}^{J_{\max}} p(J|c) \frac{\mathbf{1}_{\mathcal{U}_{\mathcal{E}}^{(J)}(\text{SU}(2))}}{2J+1} \right. \right. \\ &\quad \left. \left. \otimes |J\rangle\langle J|_{\mathcal{U}_{\mathcal{E}}^{(J)}(S_{N+M})} \right) \right] \\ &= \sum_{J=J_{\min}}^{J_{\max}} p(J|c) \frac{\mathbf{1}_{\mathcal{U}^{(J)}(\text{SU}(d))}}{\chi_J} \otimes |J\rangle\langle J|_{\mathcal{U}^{(J)}(S_{N+M})}, \end{aligned} \quad (33)$$

where in the last equality we used the fact that the support of  $\mathbf{1}_{\mathcal{U}_{\mathcal{E}}^{(J)}(\text{SU}(2))} \otimes |J\rangle\langle J|_{\mathcal{U}_{\mathcal{E}}^{(J)}(S_{N+M})}$  is a subspace of the support of  $\mathbf{1}_{\mathcal{U}^{(J)}(\text{SU}(d))} \otimes |J\rangle\langle J|_{\mathcal{U}^{(J)}(S_{N+M})}$ , as stated in Eq. 24.

Hitherto in the computation we tacitly assumed that the copies of states are labelled, i.e., that the first  $N$  states are all  $|\psi\rangle \in \mathcal{H}_d$  and the remaining  $M$  states are all  $|\phi\rangle \in \mathcal{H}_d$ . We now lift this assumption and derive the form of the average state in the case where the copies of the states are unlabelled. This is equivalent to averaging over all possible permutations of the  $N+M$  copies, i.e., by applying the map  $\mathcal{G}_{S_{N+M}}[\rho(c)]$  of Eq. (19). The final state can be easily shown to be

$$\begin{aligned} \rho(c)_{ul} := \mathcal{G}_{S_{N+M}}[\rho(c)] &= \mathcal{G}_{S_{N+M}} \circ \mathcal{G}_{\text{SU}(d)} \left[ (|\psi\rangle\langle\psi|)^{\otimes N} \otimes (|\phi\rangle\langle\phi|)^{\otimes M} \right] \\ &= \sum_{J=J_{\min}}^{J_{\max}} p(J|c) \frac{\mathbf{1}_{\mathcal{U}^{(J)}(\text{SU}(d))}}{\chi_J} \otimes \frac{\mathbf{1}_{\mathcal{U}^{(J)}(S_{M+N})}}{\omega_J}, \end{aligned} \quad (34)$$

by observing that the maps  $\mathcal{G}_{\text{SU}(d)}$ ,  $\mathcal{G}_{S_{N+M}}$  both project on the same irrep label  $J$ , but depolarize the states on different irrep subspaces  $\mathcal{U}^{(J)}(\text{SU}(d))$  and  $\mathcal{U}^{(J)}(S_{N+M})$  respectively. As all the information concerning the overlap between the two states is extracted by the projective measurement  $\{\Pi_J\}_J$ , and the action of any permutation—or global unitary rotation—does not alter the statistics  $p(J|c)$ . Consequently, all our results apply equally well to both labelled and unlabelled scenarios.

### Fisher Information of $p(J|c)$

In this section we derive the Fisher information of the probability distribution given by Eq. (30), and analyze its asymptotics. We can obtain the first non trivial order by using an asymptotic approximation for the Jacobi polynomials and large deviation evaluations. Using the same techniques, we identify an asymptotically unbiased estimator that saturates the Cramer-Rao bound.

The result we obtain is summarized by the following theorem:

**Theorem.** *Given  $N$  and  $M$  copies of two Haar-random states, respectively  $|\psi\rangle$  and  $|\phi\rangle$ , define  $|\Psi\rangle = |\psi\rangle^{\otimes N} \otimes |\phi\rangle^{\otimes M}$ . The mean square error on the estimation of the overlap  $c = |\langle\psi|\phi\rangle|^2$  attained by the weak Schur sampling projectors  $\{E_J\}_J$  with the estimator  $c_J = \left(\frac{J}{M+N}\right)^2$  is*

$$v(c) = \sum_J \text{Tr}[E_J |\Psi\rangle\langle\Psi|] (c_J - c)^2 = \frac{4c(1-c)}{N+M} + \mathcal{O}\left((N+M)^{-\frac{3}{2}}\right) \quad (35)$$

This measurement is optimal in the sense that saturates the quantum Cramer-Rao bound for the problem of estimating of  $c$  from the family of average states at fixed overlap

$$\rho(c) = \int dU U^{\otimes(N+M)} |\Psi\rangle\langle\Psi| U^{\dagger\otimes(N+M)}. \quad (36)$$

This implies that  $c$  can be estimated at a relative error  $\frac{\sqrt{v(c)}}{c} = \epsilon$  with  $N = M = \mathcal{O}(\frac{1}{\epsilon^2 c})$  samples.

We begin by recalling the definition of the Fisher information

$$H(c) = \sum_{J=J_{\min}}^{J_{\max}} p(J|c) \left( \frac{\frac{dp(J|c)}{dc}}{p(J|c)} \right)^2 = \sum_{J=J_{\min}}^{J_{\max}} \frac{\left( \frac{dp(J|c)}{dc} \right)^2}{p(J|c)}. \quad (37)$$

Using the identity

$$\frac{d^m P_n^{(\alpha,\beta)}(x)}{dx^m} = \frac{(\alpha + \beta + n + m)!}{2^m (\alpha + \beta + n)!} P_{n-m}^{(\alpha+m, \beta+m)}(x), \quad (38)$$

it follows that

$$\begin{aligned} \frac{dp(J|c)}{dc} &= \frac{(2J+1)M!N!(1-c)^{M-2}}{(J_{\max}-J)!(J+J_{\max}+1)!} \\ &\times \left( (J-J_{\min}+1)P_{J+J_{\min}-1}^{(1,1-2J_{\min})}\left(\frac{1+c}{1-c}\right) - (1-c)MP_{J+J_{\min}}^{(0,-2J_{\min})}\left(\frac{1+c}{1-c}\right) \right). \end{aligned} \quad (39)$$

For  $x \geq 1$  the following asymptotic expansion for the Jacobi polynomials holds [34, 51]. Defining the function

$$Q_n^{(\alpha,\beta)}(x) = \frac{(\sqrt{x+1} + \sqrt{x-1})^{\alpha+\beta}}{\sqrt{2\pi n}} \frac{(x + \sqrt{x^2-1})^{n+\frac{1}{2}}}{(\sqrt{x-1})^\alpha (\sqrt{x+1})^\beta \sqrt[4]{x^2-1}} \quad (40)$$

one has

$$P_n^{(\alpha,\beta)}(x) = Q_n^{(\alpha,\beta)}(x) \left( 1 + \mathcal{O}\left(\frac{1}{n}\right) \right), \quad (41)$$

where the convergence is uniform on any half-line  $x \in [1 + \delta, +\infty)$ ,  $\delta > 0$ . We want to use the leading order as an approximation to the Jacobi polynomial. Notice that the remainder in the expansion does not depend on  $J_{\max}$ . Take  $J_0$  such that for each  $J > J_0$  and a certain  $C > 0$

$$\left| \frac{P_{J+J_{\min}}^{(0,-2J_{\min})}(x)}{Q_{J+J_{\min}}^{(0,-2J_{\min})}(x)} - 1 \right| + \left| \frac{P_{J+J_{\min}-1}^{(1,1-2J_{\min})}(x)}{Q_{J+J_{\min}-1}^{(1,1-2J_{\min})}(x)} - 1 \right| \leq \frac{C}{J+J_{\min}}. \quad (42)$$

We can then split the sum in Eq. 37 in two parts. The first is

$$H(c)_1 := \sum_{J=J_{\min}}^{J_0} p(J|c) \left( \frac{\frac{dp(J|c)}{dc}}{p(J|c)} \right)^2. \quad (43)$$

Writting  $p(J|c)$  as

$$p(J|c) = \frac{2J+1}{2J_{\max}+1} \frac{\binom{2J_{\max}+1}{J_{\max}-J}}{\binom{2J_{\max}}{J_{\max}-J_{\min}}} (1-c)^{J_{\max}-J} \sum_{s=0}^{J-J_{\min}} \binom{J+J_{\min}}{s} \binom{J-J_{\min}}{s} c^s. \quad (44)$$

and noting that

$$\left( \frac{\frac{dp(J|c)}{dc}}{p(J|c)} \right) = -\frac{(J_{\max}-J)}{1-c} + \frac{(1-c)^{J_{\max}-J} \sum_{s=0}^{J-J_{\min}} s \binom{J+J_{\min}}{s} \binom{J-J_{\min}}{s} c^s}{cp(J|c)} \leq -\frac{(J_{\max}-J)}{1-c} + \frac{J-J_{\min}}{c}, \quad (45)$$

it follows that for  $J \leq J_0$

$$\left| \left( \frac{\frac{dp(J|c)}{dc}}{p(J|c)} \right) \right| \leq \frac{cJ_{\max} + J_0 + (1-c)J_{\min}}{c(1-c)}. \quad (46)$$

On the other hand, isolating in  $p(J|c)$  the terms depending on  $J_{\max}$ , and defining  $w(J, J_{\min}, c) = (2J+1) \sum_{s=0}^{J-J_{\min}} \binom{J+J_{\min}}{s} \binom{J-J_{\min}}{s} c^s$  which is bounded for  $J_{\min} \leq J \leq J_0$ , we have

$$p(J|c) = \frac{1}{2J_{\max}+1} \frac{\binom{2J_{\max}+1}{J_{\max}-J}}{\binom{2J_{\max}}{J_{\max}-J_{\min}}} (1-c)^{J_{\max}-J} w(J, J_{\min}, c) \quad (47)$$

By introducing the following binomial distribution

$$q(J) := \text{Bin}(2J_{\max}+1, J_{\max}-J, p) = \binom{2J_{\max}+1}{J_{\max}-J} p^{J_{\max}-J} (1-p)^{J_{\max}+J+1}, \quad (48)$$

with  $p = \frac{1-\sqrt{c}}{2}$ , whose mean and variance are given by

$$\begin{aligned} \mu &= (2J_{\max}+1) \left( \frac{1-\sqrt{c}}{2} \right) \\ \sigma^2 &= \frac{(2J_{\max}+1)}{4} (1-c). \end{aligned} \quad (49)$$

Eq. (47) can be written as

$$p(J|c) = \frac{1}{2J_{\max}+1} \frac{2^{2J_{\max}}}{\binom{2J_{\max}}{J_{\max}-J_{\min}}} q(J) w'(J, J_{\min}, c), \quad (50)$$

with  $w'(J, J_{\min}, c)$  still bounded.

Now by Sanov's theorem [52], the binomial distribution  $q(J)$  is exponentially suppressed as  $J_{\max} \rightarrow \infty$  and  $J < J_0$ . In particular, if  $J_{\max} > \frac{1}{\sqrt{c}} J_0$

$$q(J) \leq \exp^{-(2J_{\max}) D(\frac{J_{\max}-J_0}{2J_{\max}+1} || \frac{1-\sqrt{c}}{2})}, \quad (51)$$

where  $D(\frac{J_{\max}-J_0}{2J_{\max}+1} || \frac{1-\sqrt{c}}{2})$  is the relative entropy between the two Bernoulli distributions with probabilities  $p_1 = \frac{J_{\max}-J_0}{2J_{\max}+1}$  and  $p_2 = \frac{1-\sqrt{c}}{2}$ . For  $J_{\max} \rightarrow \infty$ ,  $p_1 \rightarrow \frac{1}{2}$ ,  $D(\frac{J_{\max}-J_0}{2J_{\max}+1} || \frac{1-\sqrt{c}}{2}) \rightarrow D(\frac{1}{2} || \frac{1-\sqrt{c}}{2}) > 0$  unless  $c = 0$ .

As the rest of the terms in both  $p(J|c)$  and  $\left| \left( \frac{\frac{dp(J|c)}{dc}}{p(J|c)} \right) \right|$  tend to a power law in  $J_{\max}$ ,  $H(c)_1$  is exponentially suppressed.

As an indication for which ratio  $\frac{J_{\max}}{J_{\min}}$  is expected to give exponential suppression when  $c$  is small, we can look at the first order of the Taylor expansion of  $D(p_1 || p_2)$  in  $\frac{1}{J_{\max}}$  and  $\sqrt{c}$  around zero:  $D(p_1 || p_2) = -\frac{1}{2} \log(1-c) - \frac{1}{4} \frac{2J_0+1}{J_{\max}} \log \frac{1+\sqrt{c}}{1-\sqrt{c}}$ , which for small  $c$  requires  $J_{\max}c >> 1$  from the zeroth order and  $J_{\max} >> \frac{1}{\sqrt{c}} J_0$  from the first order, implying also  $J_{\max} >> \frac{1}{\sqrt{c}} J_{\min}$ .

It remains to evaluate

$$H(c)_2 := \sum_{J=J_0+1}^{J_{\max}} p(J|c) \left( \frac{\frac{dp(J|c)}{dc}}{p(J|c)} \right)^2. \quad (52)$$

Using Eq. (41) one obtains, after some algebra

$$\begin{aligned} \frac{\left(\frac{dp(J|c)}{dc}\right)^2}{p(J|c)} &= \frac{(2J+1)M!N!(1+\sqrt{c})^{2J-1}(1-c)^{J_{\max}-J}}{2\sqrt{\pi}(1-\sqrt{c})^2c^{5/4}(J+J_{\min}-1)\sqrt{J+J_{\min}}(J_{\max}-J)!(J+J_{\max}+1)!} \\ &\times \left((J-J_{\min}+1)\sqrt{J+J_{\min}} - M\sqrt{c(J+J_{\min}-1)}\right)^2 \left(1 + \mathcal{O}\left(\frac{1}{J+J_{\min}}\right)\right), \end{aligned} \quad (53)$$

where the same binomial distribution,  $q(J)$  of Eq. (48), appears again. Writing

$$\frac{\left(\frac{dp(J|c)}{dc}\right)^2}{p(J|c)} = f(M, N, c)g(J, J_{\max}, J_{\min}, c)q(J) \left(1 + \mathcal{O}\left(\frac{1}{J+J_{\min}}\right)\right) \quad (54)$$

where

$$\begin{aligned} f(M, N, c) &= \frac{M!N!2^{2J_{\max}}}{\sqrt{\pi}(1-c)^2c^{5/4}(2J_{\max}+1)!} \\ g(J, J_{\max}, J_{\min}, c) &= \frac{(2J+1)\left((J-J_{\min}+1)\sqrt{J+J_{\min}} - M\sqrt{c(J+J_{\min}-1)}\right)^2}{(2J_{\max}+1)(J+J_{\min}-1)\sqrt{J+J_{\min}}} \\ &= \frac{(2J+1)(J-J_{\min}+1)}{(2J_{\max}+1)(J+J_{\min}+1)}\sqrt{J+J_{\min}} + c\frac{(2J+1)(J_{\max}+J_{\min})^2}{(2J_{\max}+1)\sqrt{J+J_{\min}}} \\ &- \sqrt{c}\frac{(2J+1)(J-J_{\min}+1)}{(2J_{\max}+1)\sqrt{J+J_{\min}-1}} \end{aligned} \quad (55)$$

we have

$$H(c)_2 = f(M, N, c) \sum_{J=J_0+1}^{J_{\max}} g(J, M, N, c) \text{Bin}(2J_{\max}+1, J_{\max}-J, p) \left(1 + \mathcal{O}\left(\frac{1}{J+J_{\min}}\right)\right). \quad (56)$$

In the limit  $J_{\max} \rightarrow \infty$

$$f(M, N, c) = \frac{1}{2(1-c)^2c^{5/4}} \frac{1}{\sqrt{J_{\max}}} + \mathcal{O}\left(J_{\max}^{-3/2}\right). \quad (57)$$

whilst  $g(J, J_{\max}, J_{\min}, c) = \mathcal{O}(J_{\max}^{3/2})$  by inspection.

Now consider the first  $\sqrt{d}J_{\max} - (J_0 + 1)$  terms in Eq. (52), where  $d < c$ ,

$$H(c)_{2a} := \sum_{J=J_0+1}^{\sqrt{d}J_{\max}-1} p(J|c) \left(\frac{\frac{dp(J|c)}{dc}}{p(J|c)}\right)^2. \quad (58)$$

Sanov's theorem tells us that these are exponentially suppressed,

$$H(c)_{2a} \leq \exp^{-(2J_{\max})D(\frac{1-\sqrt{d}}{2}||\frac{1-\sqrt{c}}{2})} \mathcal{O}(J_{\max}^2), \quad (59)$$

and as a result we are only left with having to evaluate

$$H(c)_{2b} := \sum_{J=\sqrt{d}J_{\max}}^{J_{\max}} q(J)g(J, J_{\max}, J_{\min}, c) \left(1 + \mathcal{O}\left(\frac{1}{J+J_{\min}}\right)\right). \quad (60)$$

As we are interested in computing the *QFI* in the limit  $J_{\max} \rightarrow \infty$  we will perform a Taylor series expansion of  $H(c)_{2b}$ , around the mean value  $\langle J \rangle = J_{\max} - \mu = \sqrt{c}(J_{\max} + \frac{1}{2}) - \frac{1}{2}$  of  $q(J)$ . For any analytical function  $h(J)$ , using the Lagrange remainder, an expansion about  $\langle J \rangle$  is given by

$$\sum_{J=\sqrt{d}J_{\max}}^{J_{\max}} q(J)h(J) = \sum_{J=\sqrt{d}J_{\max}}^{J_{\max}} q(J) \left( \sum_{s=0}^2 \frac{(J-\langle J \rangle)^s}{s!} \frac{d^s h(J)}{dJ^s} \Big|_{J=\langle J \rangle} + \frac{(J-\langle J \rangle)^3}{3!} \frac{d^3 h(J')}{dJ'^3} \Big|_{J'=\xi(J)} \right), \quad (61)$$

with  $J \leq \xi(J) \leq \langle J \rangle$  (so that in any case  $\sqrt{d}J_{\max} \leq \xi(J) \leq J_{\max}$ ). Using Eq. (61) and the fact that  $g(J, J_{\max}, J_{\min}, c)$  is analytic for  $J > J_{\min}$  by inspection, we have

$$\sum_{J=\sqrt{d}J_{\max}}^{J_{\max}} q(J)g(J, J_{\max}, J_{\min}, c) = \sum_{s=0}^2 \frac{1}{s!} \left. \frac{d^s g(J, J_{\max}, J_{\min}, c)}{dJ^s} \right|_{J=\langle J \rangle} \left( \sum_{J=\sqrt{d}J_{\max}}^{J_{\max}} q(J) (J - \langle J \rangle)^s \right) \\ + \frac{1}{3!} \left. \frac{d^3 g(J', J_{\max}, J_{\min}, c)}{dJ'^3} \right|_{J'=\xi(J)} \left( \sum_{J=\sqrt{d}J_{\max}}^{J_{\max}} q(J) (J - \langle J \rangle)^s \right), \quad (62)$$

where the change in the order of the summation is allowed since one of the sums is finite. Applying Sanov's theorem, Eq. (51), gives

$$\sum_{J=-J_{\max}-1}^{\sqrt{d}J_{\max}} q(J)(J - \langle J \rangle)^s \leq (2J_{\max} + 1)(J_{\max} + \langle J \rangle + 1)^s \exp^{-(2J_{\max})D(\frac{J_{\max}-\sqrt{d}J_{\max}}{2J_{\max}+1} || \frac{1-\sqrt{c}}{2})} \\ = (2J_{\max} + 1)(J_{\max} - \langle J \rangle)^s \left( \frac{J_{\max} + \langle J \rangle + 1}{J_{\max} - \langle J \rangle} \right)^s \exp^{-(2J_{\max})D(\frac{J_{\max}-\sqrt{d}J_{\max}}{2J_{\max}+1} || \frac{1-\sqrt{c}}{2})} \\ = (2J_{\max} + 1)(J_{\max} - \langle J \rangle)^s \left( \frac{1 + \sqrt{c}}{1 - \sqrt{c}} \right)^s \exp^{-(2J_{\max})D(\frac{J_{\max}-\sqrt{d}J_{\max}}{2J_{\max}+1} || \frac{1-\sqrt{c}}{2})}. \quad (63)$$

It follows that we can change the summation in  $J$  of Eq. (62) to start from  $-J_{\max} - 1$  just by adding a term exponentially suppressed in  $J_{\max}$ . For the Lagrange remainder, before extending the summation, we notice that for  $\sqrt{d}J_{\max} \leq \xi(J) \leq J_{\max}$ ,  $\left. \frac{d^s g(J', M, N, c)}{dJ'^s} \right|_{J'=\xi(J)} = \mathcal{O}(J_{\max}^{\frac{3}{2}-n})$  (by inspection).

Finally, using Eq. (49) and some algebra one obtains

$$\sum_J g(J, J_{\max}, J_{\min}, c)q(J) = g(\langle J \rangle, J_{\max}, J_{\min}, c) + \frac{1}{2}\sigma^2 \left. \frac{d^2 g(J, J_{\max}, J_{\min}, c)}{dJ^2} \right|_{J=\langle J \rangle} + \mathcal{O}(J_{\max}^{-\frac{1}{2}}) \\ = \frac{1}{2}(1-c)c^{\frac{1}{4}}J_{\max}^{\frac{3}{2}} + \mathcal{O}(J_{\max}^{\frac{1}{2}}). \quad (64)$$

Multiplying Eqs. ((57), (64)) gives the final result

$$H(c) = \frac{J_{\max}}{2c(1-c)} + \mathcal{O}(1) = \frac{M+N}{4c(1-c)} + \mathcal{O}(1). \quad (65)$$

The neglected term  $\mathcal{O}\left(\frac{1}{J+J_{\min}}\right)$  in Eq. (56) contributes to  $H(c)$  with a next to leading order term, by simple power counting, adapting the steps above.

Using similar techniques it can be shown that  $H(c)$  is achievable by the estimator of Eq. (7) of the main text. First note that the estimator is indeed asymptotically unbiased since

$$\langle c_{\text{op}}^{\text{loc}} \rangle := \sum_{J=J_{\min}}^{J_{\max}} \left( \frac{J}{J_{\max}} \right)^2 p(J|c) \\ = \frac{2^{2J_{\max}} M! N!}{c^{\frac{1}{4}} \sqrt{\pi} (2J_{\max} + 1)! J_{\max}^2} \sum_{J=\sqrt{d}J_{\max}}^{J_{\max}} \frac{J^2 (2J + 1)}{\sqrt{J - J_{\min}}} \text{Bin}(2J_{\max} + 1, J_{\max} - J, p), \quad (66)$$

apart from an exponentially suppressed term, and we have again used the asymptotic expansion of Eq. (41) and identified the binomial distribution of Eq. (48). In the limit  $J_{\max} \rightarrow \infty$  one can proceed as above, separating the  $J$ -dependence and applying Eq. (61), to obtain

$$\frac{2^{2J_{\max}} M! N!}{c^{\frac{1}{4}} \sqrt{\pi} (2J_{\max} + 1)! J_{\max}^2} = \frac{1}{2c^{\frac{1}{4}}} \frac{1}{J_{\max}^{\frac{5}{2}}} + \mathcal{O}(J_{\max}^{-\frac{7}{2}}), \\ \sum_{J=J_{\min}}^{J_{\max}} \frac{J^2 (2J + 1)}{\sqrt{J - J_{\min}}} \text{Bin}(2J_{\max} + 1, J_{\max} - J, p) = 2c^{\frac{5}{4}} J_{\max}^{\frac{5}{2}} + \mathcal{O}(J_{\max}^{\frac{3}{2}}). \quad (67)$$

It follows that

$$\langle c_{\text{op}}^{\text{loc}} \rangle = c + \mathcal{O}(J_{\max}^{-1}). \quad (68)$$

A similar calculation gives for the MSE of the estimator

$$\langle (c_{\text{op}}^{\text{loc}} - c)^2 \rangle := \frac{2c(1-c)}{J_{\max}} + \mathcal{O}\left(J_{\max}^{-\frac{3}{2}}\right). \quad (69)$$

### Optimal global mean squared error

In this section we derive the optimal estimator and corresponding global mean squared error (glMSE) for the case where the overlap  $c$  is a random variable with a distribution induced by the Haar-uniform measure of  $SU(d)$ .

The result we obtain is summarized by the following theorem:

**Theorem.** *Given  $N$  and  $M$  copies of two Haar-random states, respectively  $|\psi\rangle$  and the state  $|\phi\rangle$ , define  $|\Psi\rangle = |\phi\rangle^{\otimes N} \otimes |\psi\rangle^{\otimes M}$ . The global mean square error on the estimation of the overlap  $c = |\langle\psi|\phi\rangle|^2$  attained by a POVM  $\{E_k\}_k$  and an estimator  $c_k$  is*

$$v = \sum_k \int dU dV (c(k) - c)^2 \text{Tr}[E_k |\Psi\rangle\langle\Psi|]. \quad (70)$$

The minimum of  $v$  is attainable by a projective measurement and it reads

$$v_{\text{op}} = \frac{(d-1)(d+M+N)}{d(1+d)(d+M)(d+N)}. \quad (71)$$

We follow the standard treatment in [33] to compute the estimator that minimizes glMSE, and use representation theory to perform the integrations.

The probability distribution of the overlap, induced by the Haar measure, is [35, Eq. (13)]

$$p(c) = \int_{SU(d)} dU \delta(c - |\langle\psi| U |\psi\rangle|^2) = (d-1)(1-c)^{d-2} \quad (72)$$

Following [33] the optimal estimator,  $S$ , satisfies

$$\frac{S\Gamma + \Gamma S}{2} = \eta \quad (73)$$

where

$$\begin{aligned} \Gamma &\equiv \int p(c) \rho(c) dc \\ \eta &\equiv \int c p(c) \rho(c) dc. \end{aligned} \quad (74)$$

and is explicitly given by

$$S = \int_0^\infty e^{-\alpha\Gamma} \eta e^{-\alpha\Gamma} d\alpha \quad (75)$$

Plugging Eq. (29) into Eqs. (74,75) gives

$$\begin{aligned} S &= \sum_{J=J_{\min}}^{J_{\max}} \frac{\int cp(J,c)dc}{\int p(J,c)dc} \mathbf{1}_{\mathcal{U}^{(J)}(SU(d))} \otimes \mathbf{1}_{\mathcal{U}^{(J)}(S_{N+M})} = \sum_{J=J_{\min}}^{J_{\max}} \frac{\text{Tr}[\Pi_J \eta]}{\text{Tr}[\Pi_J \Gamma]} \mathbf{1}_{\mathcal{U}^{(J)}(SU(d))} \otimes \mathbf{1}_{\mathcal{U}^{(J)}(S_{N+M})} \\ &= \sum_{J=J_{\min}}^{J_{\max}} \int cp(c|J)dc \mathbf{1}_{\mathcal{U}^{(J)}(SU(d))} \otimes \mathbf{1}_{\mathcal{U}^{(J)}(S_{N+M})} \\ &= \sum_{J=J_{\min}}^{J_{\max}} \langle c \rangle_J \mathbf{1}_{\mathcal{U}^{(J)}(SU(d))} \otimes \mathbf{1}_{\mathcal{U}^{(J)}(S_{N+M})}, \end{aligned} \quad (76)$$

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. (76). Again, the optimal measurement corresponds to measuring the total irrep label  $J$ . Upon a given outcome the estimator that minimizes the glMSE is  $c_{\text{opt}}^{\text{bay}}(j) = \langle c \rangle_J$  where the expectation value is taken with respect to the conditional probability distribution  $p(c|J)$ .

The operators  $\Gamma$  and  $\eta$  are given by

$$\begin{aligned}
\Gamma &= \int_0^1 p(c) \left( \int_{\text{SU}(d)} dU (U |\psi\rangle\langle\psi| U^\dagger)^{\otimes N} \otimes (UT(c) |\psi\rangle\langle\psi| T^\dagger(c) U^\dagger)^{\otimes M} \right) dc \\
&= \int_{\text{SU}(d)} dU (U |\psi\rangle\langle\psi| U^\dagger)^{\otimes N} \otimes \int_0^1 dc \int_{\text{SU}(d)} dV \delta(c - |\langle\psi| V |\psi\rangle|^2) (UT(c) |\psi\rangle\langle\psi| T^\dagger(c) U^\dagger)^{\otimes M} \\
&= \int_{\text{SU}(d)} dU (U |\psi\rangle\langle\psi| U^\dagger)^{\otimes N} \otimes \int_{\text{SU}(d)} dV (UW_V V |\psi\rangle\langle\psi| U^\dagger(h) W_V^\dagger U^\dagger)^{\otimes M} \\
&= \int_{\text{SU}(d)} dU (U |\psi\rangle\langle\psi| U^\dagger)^{\otimes N} \otimes \int_{\text{SU}(d)} dV (V |\psi\rangle\langle\psi| U^\dagger(h))^{\otimes M} \\
&= \mathcal{G}_{\text{SU}(d)} [(|\psi\rangle\langle\psi|)^{\otimes N}] \otimes \mathcal{G}_{\text{SU}(d)} [(|\psi\rangle\langle\psi|)^{\otimes M}] \\
&= \frac{\mathbf{1}_{\mathcal{U}(\frac{N}{2})(\text{SU}(d))}}{\chi_{\frac{N}{2}}} \otimes \frac{\mathbf{1}_{\mathcal{U}(\frac{M}{2})(\text{SU}(d))}}{\chi_{\frac{M}{2}}} \\
&= \frac{1}{\chi_{\frac{N}{2}} \chi_{\frac{M}{2}}} \sum_{J=\lfloor \frac{|N-M|}{2} \rfloor}^{\frac{N+M}{2}} \mathbf{1}_{\mathcal{U}^{(J)}(\text{SU}(d))} \otimes \mathbf{1}_{\mathcal{U}^{(J)}(S_{N+M})}, \tag{77}
\end{aligned}$$

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

To compute  $\eta$  we make use of

$$\begin{aligned}
\int_{\text{SU}(2)} dg (d-1) \left( 1 - |D_{\frac{1}{2}, \frac{1}{2}}(\frac{1}{2})(g)|^2 \right)^{d-2} |D_{\frac{1}{2}, \frac{1}{2}}(\frac{1}{2})(g)|^2 (|0\rangle\langle 0|)^{\otimes N} \otimes \left( D(\frac{1}{2})(g)^\dagger |0\rangle\langle 0| D(\frac{1}{2})(g) \right)^{\otimes M} = \\
\int_{\text{SU}(2)} dg D_{-\frac{d-1}{2}, \frac{d-1}{2}}^{\frac{d-1}{2}}(g) D_{-\frac{d-3}{2}, \frac{d-1}{2}}^{\frac{d-1}{2}}(g)^* D_{k, \frac{M}{2}}^{\frac{M}{2}}(g) D_{k', \frac{M}{2}}^{\frac{M}{2}}(g)^* \left| \frac{N}{2}, \frac{N}{2} \right\rangle \left\langle \frac{N}{2}, \frac{N}{2} \right| \otimes \left| \frac{M}{2}, k \right\rangle \left\langle \frac{M}{2}, k' \right| = \\
\frac{1}{d+M} \sum_{k=-\frac{M}{2}}^{J-\frac{N}{2}} \left( C_{\frac{d-1}{2}, -\frac{d-3}{2}; \frac{M}{2}, k}^{\frac{d-1+M}{2}, -\frac{d-3}{2}+h} C_{\frac{N}{2}, \frac{N}{2}; \frac{M}{2}, k}^{J, \frac{N}{2}+k} \right)^2 \left| J, \frac{N}{2}+k \right\rangle \left\langle J, \frac{N}{2}+k \right|, \tag{78}
\end{aligned}$$

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

$$\begin{aligned}
\eta &= \int_0^1 p(c) c \left( \int_{\text{SU}(d)} dU (U |\psi\rangle\langle\psi| U^\dagger)^{\otimes N} \otimes (UT(c) |\psi\rangle\langle\psi| T^\dagger(c) U^\dagger)^{\otimes M} \right) dc \\
&= \int_{\text{SU}(d)} dU (U |\psi\rangle\langle\psi| U^\dagger)^{\otimes N} \otimes \int_0^1 dc p(c) c \int_{\text{SU}(\mathcal{E})} dV \delta(c - |\langle\psi| V |\psi\rangle|^2) (UT(c) |\psi\rangle\langle\psi| T^\dagger(c) U^\dagger)^{\otimes M} \\
&= \mathcal{G}_{\text{SU}(d)} \left[ \int_{\text{SU}(\mathcal{E})} dV (d-1)(1 - |\langle\psi| V |\psi\rangle|^2)^{d-2} |\langle\psi| V |\psi\rangle|^2 |\psi\rangle\langle\psi|^{\otimes N} \otimes (V |\psi\rangle\langle\psi| V^\dagger)^{\otimes M} \right] \\
&= \frac{1}{d+M} \sum_{J=J_{\min}}^{J_{\max}} \sum_{k=-J-\frac{N}{2}}^{J-\frac{N}{2}} \left( C_{\frac{d-1}{2}, -\frac{d-3}{2}; \frac{M}{2}, k}^{\frac{d-1+M}{2}, -\frac{d-3}{2}+h} C_{\frac{N}{2}, \frac{N}{2}; \frac{M}{2}, k}^{J, \frac{N}{2}+k} \right)^2 \frac{\mathbf{1}_{\mathcal{U}^{(J)}(\text{SU}(d))}}{\chi_J} \otimes |J\rangle\langle J|_{\mathcal{U}^{(J)}(S_{N+M})}.
\end{aligned}$$

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

$$c(J) = \frac{\text{Tr}[(\Pi_J \eta)]}{\text{Tr}[(\Pi_J \Gamma)]} = \frac{\frac{\chi_J}{\chi_{\frac{N}{2}} \chi_{\frac{M}{2}}}}{\frac{1}{d+M} \sum_{k=-J-\frac{N}{2}}^{J-\frac{N}{2}} \left( C_{\frac{d-1}{2}, -\frac{d-3}{2}; \frac{M}{2}, k}^{\frac{d-1+M}{2}, -\frac{d-3}{2}+h} C_{\frac{N}{2}, \frac{N}{2}; \frac{M}{2}, k}^{J, \frac{N}{2}+k} \right)^2}, \tag{79}$$

$\chi_J$  can be calculated with the Hook formula [37]. For Young diagrams with two rows one has

$$\chi_J = (2J+1) \frac{(d+J+\frac{N+M}{2}-1)! (d-J+\frac{N+M}{2}-2)!}{(d-1)!(d-2)! (\frac{N+M}{2}+J+1)! (\frac{N+M}{2}-J)!}.$$

To simplify the denominator we employ the graphical calculus techniques in [39]:

$$\begin{aligned} & \frac{1}{d+M} \sum_{k=-J-\frac{N}{2}}^{J-\frac{N}{2}} \left( C_{\frac{d-1}{2}, -\frac{d-3}{2}; \frac{M}{2}, k}^{d-1+M, -d-3+k} C_{\frac{N}{2}, \frac{N}{2}; \frac{M}{2}, k}^{J, \frac{N}{2}+k} \right)^2 = (2J+1) \sum_{L=\frac{d-3+N}{2}}^{\frac{d-1+N}{2}} \left( C_{\frac{d-1}{2}, \frac{d-3}{2}; \frac{N}{2}, \frac{N}{2}}^{L, \frac{d-3+N}{2}} \right)^2 \left\{ \frac{M}{2} \quad \frac{d-1}{J} \quad \frac{d+n-1}{\frac{N}{2}} \right\}^2 \\ & = (d-1)(2J+1)(4d+4J+4J^2+2N-N^2+2M+2NM-M^2) \\ & \times \frac{N!M!(d-1+J+\frac{N+M}{2})!(d-2-J+\frac{N+M}{2})!}{4(d+N)!(d+M)!(-J+\frac{N+M}{2})!(1+J+\frac{N+M}{2})!} \end{aligned} \quad (80)$$

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

$$c_{opt}^{bay}(J) = \frac{d+J+J^2+\frac{M+N}{2}-\left(\frac{M+N}{2}\right)^2+MN}{(d+M)(d+N)} \quad (81)$$

with its corresponding glMSE

$$v_{op} = \langle (c_{op}^{bay} - c)^2 \rangle = \int_0^1 p(c)c^2 - \sum_J p(J)c(J)^2 = \frac{(d-1)(d+M+N)}{d(1+d)(d+M)(d+N)}. \quad (82)$$

### 1-LOCC strategies

Here we derive the pointwise MSE attainable by the estimate-and-project (EP) and estimate-and-estimate (EE) strategies, and study the optimal Bayesian estimator for both.

#### *Estimate-and-project: local estimation*

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],  $\{M_{dV}\}$ , that produces an estimate  $|\phi_V\rangle = V|0\rangle$ ,  $V \in \text{SU}(d)$ , with probability density

$$d\mu(V) = \chi_{\frac{M}{2}} |\langle \phi_V | \phi \rangle|^{2M} dV. \quad (83)$$

Finally, we perform the projective measurement  $\{|\phi_V\rangle \langle \phi_V|, \mathbb{1} - |\phi_V\rangle \langle \phi_V|\}$  on each copy of  $|\psi\rangle$ . This succeeds with probability

$$p_V(c) = |\langle \phi_V | \psi \rangle|^2. \quad (84)$$

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

$$p(k|c) = \chi_{\frac{M}{2}} \int_{\text{SU}(d)} dV \text{Tr} \left[ \left( |0\rangle \langle 0|^{\otimes M} \otimes \Pi_k^{(N)} \right) V^{\otimes(M+N)} \left( \left( U_c |0\rangle \langle 0| U_c^\dagger \right)^{\otimes M} \otimes |0\rangle \langle 0|^{\otimes N} \right) V^{\dagger \otimes(M+N)} \right], \quad (85)$$

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. We now consider both local and Bayesian estimation with this strategy.

For local estimation with EP, we employ as an estimator  $c_{ep}^{\text{loc}}(k) = \frac{k}{N}$ , 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 but provides a natural guess for the overlap given our strategy. Its MSE is

$$\begin{aligned} v_{ep}(c) &= \int_{\text{SU}(d)} d\mu(V) \sum_{k=0}^N \text{Bin}(k, N, p_V(c)) \left( \frac{k}{N} - c \right)^2 \\ &= \chi_{\frac{M}{2}} \int_{\text{SU}(d)} d\mu(V) \left( c^2 - 2cp_V(c) + \frac{1}{N} p_V(c)(1-p_V(c)) + p_V(c)^2 \right) \\ &= c^2 - \left( 2c - \frac{1}{N} \right) I_1(M) + \left( 1 - \frac{1}{N} \right) I_2(M), \end{aligned} \quad (86)$$

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

$$I_i(M) = \chi_{\frac{M}{2}} \int_{\text{SU}(d)} dV |\langle \phi_V | \phi \rangle|^{2M} |\langle \phi_V | \psi \rangle|^{2i}. \quad (87)$$

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. (26), with the substitution  $N \mapsto i$ :

$$I_i(M) = \chi_{\frac{M}{2}} \langle 0 |^{\otimes M+i} \mathcal{G}_{\text{SU}(d)} \left[ |\phi\rangle \langle \phi|^M \otimes |\psi\rangle \langle \psi|^i \right] |0\rangle^{\otimes M+i}, \quad (88)$$

where we have eliminated the integral over  $V$  by including this rotation into the group average. The result is then given by Eqs. (30,33) with the same substitution. Plugging in the expression of  $\chi_J = \binom{2J+d-1}{d-1}$ , Ref. [36], we have

$$I_i(M) = \binom{M+i+d-1}{i}^{-1} (1-c)^i P_i^{(0,M-i)} \left( \frac{1+c}{1-c} \right), \quad (89)$$

which can be computed explicitly for  $i = 1, 2$ . Inserting these expressions in Eq. (86) we finally obtain

$$v_{ep}(c) = \frac{c^2(dM + d^2M + N - 2MN - N^2) + c(-2M - 2dM - 3N + dN + 2MN + N^2) + d - 1 + 2M + N}{M(d+N)(1+d+N)}. \quad (90)$$

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

$$v_{ep}(c) \sim \frac{c(1-c)}{N}, \quad (91)$$

which coincides with the optimal strategy, corresponding to a projection on the known direction of  $|\phi\rangle$ . In the limit  $M + N \rightarrow \infty$ ,  $M - N$  fixed we have instead

$$v_{ep}(c) \sim \frac{6c(1-c)}{(M+N)}, \quad (92)$$

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

#### *Estimate-and-project: Bayesian estimation*

For Bayesian estimation with EP, as in the previous section, the optimal classical estimator is given by

$$c(k) = \int dc c p(c|k) = \frac{\tilde{c}(k)}{p(k)}, \text{ with } \tilde{c}(k) := \int dc c p(c)p(k|c). \quad (93)$$

We start by computing the probability distribution of the outcomes, using Eq. (85):

$$\begin{aligned} p(k) &= \int dc p(c)p(k|c) \\ &= \chi_{\frac{M}{2}} \int_0^1 dc p(c) \int_{\text{SU}(d)} dU \text{Tr} \left[ \left( |0\rangle \langle 0|^{\otimes M} \otimes \Pi_k^{(N)} \right) U^{\otimes(M+N)} \left( \left( U(c) |0\rangle \langle 0| U^\dagger(c) \right)^{\otimes M} \otimes |0\rangle \langle 0|^{\otimes N} \right) U^{\dagger \otimes(M+N)} \right] \\ &= \int_{\text{SU}(d)} dU \text{Tr} \left[ \left( |0\rangle \langle 0|^{\otimes M} \otimes \Pi_k^{(N)} \right) \left( \mathbf{1}_M^{\text{sym}} \otimes \left( U |0\rangle \langle 0| U^\dagger \right)^{\otimes N} \right) \right] \\ &= \frac{1}{\chi_{\frac{N}{2}}} \text{Tr} \left[ \Pi_k^{(N)} \mathbf{1}_N^{\text{sym}} \right], \end{aligned} \quad (94)$$

where we performed the average over  $c \in \text{SU}(d)$  in the third equality, defining  $\mathbf{1}_M^{\text{sym}} = \mathbf{1}_{U(\frac{M}{2})(\text{SU}(d))}$  as the projector on the completely symmetric subspace of  $M$  qudits, and employed its invariance under  $U^{\otimes M}$ . We are then left to compute the overlap of  $\Pi_k^{(N)}$  with  $\mathbf{1}_N^{\text{sym}}$ . In order to do so, we recall that the latter can be written

as the average of all permutation operators  $V(\sigma)$ ,  $\sigma \in S_N$  of  $N$   $d$ -level systems. Then the symmetrization of  $\Pi_k^{(N)}$  gives a trivial binomial factor and we can write

$$\begin{aligned} p(k) &= \frac{1}{\chi_{\frac{N}{2}} N!} \binom{N}{k} \sum_{\underline{i} \in \{1, \dots, d-1\}^{N-k}} \sum_{\sigma} \langle 0 |^{\otimes k} \langle \underline{i} | V(\sigma) | 0 \rangle^{\otimes k} | \underline{i} \rangle \\ &= \frac{1}{\chi_{\frac{N}{2}} N!} \binom{N}{k} \sum_{\underline{i} \in \{1, \dots, d-1\}^{N-k}} k! \prod_{j=1}^{d-1} \beta_j(i)!, \end{aligned} \quad (95)$$

where  $\beta_j(i)$  is the number of times the integer  $j$  appears in the sequence  $\underline{i}$ . The sum over  $\underline{i}$  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  $\underline{\beta}$  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  $j$ 's:

$$p(k) = \frac{1}{\chi_{\frac{N}{2}} N!} \binom{N}{k} \sum_{\underline{\beta}: \underline{1}=N-k} \frac{(N-k)!}{\prod_{j=1}^{d-1} \beta_j!} k! \prod_{j=1}^{d-1} \beta_j! = \frac{1}{\chi_{\frac{N}{2}}} \binom{N-k+d-2}{d-2}, \quad (96)$$

where in the last equality we summed over  $\underline{\beta}$ . For the first moment of the distribution we get, similarly,

$$\begin{aligned} \tilde{c}(k) &= \chi_{\frac{M}{2}} \int_0^1 dc p(c) \int_{SU(d)} dU \text{Tr} \left[ \left( |0\rangle \langle 0|^{\otimes(M+1)} \otimes \Pi_k^{(N)} \right) \right. \\ &\quad \times \mathbf{1} \otimes U^{\otimes(M+N)} \left( (U(c)|0\rangle \langle 0|U(c))^{\otimes(M+1)} \otimes |0\rangle \langle 0|^{\otimes N} \right) \mathbf{1} \otimes U^{\dagger \otimes(M+N)} \Big] \\ &= \frac{\chi_{\frac{M}{2}}}{\chi_{\frac{M+1}{2}}} \int_{SU(d)} dU \text{Tr} \left[ \left( |0\rangle \langle 0|^{\otimes(M+1)} \otimes \Pi_k^{(N)} \right) U^\dagger \otimes \mathbf{1}^{\otimes M} \otimes U^{\otimes N} \left( \mathbf{1}_{M+1}^{\text{sym}} \otimes |0\rangle \langle 0|^{\otimes N} \right) U \otimes \mathbf{1}^{\otimes M} \otimes U^{\dagger \otimes N} \right], \end{aligned} \quad (97)$$

where we have introduced  $U^\dagger U$  and its conjugate on the additional subsystem, then employed the invariance of  $\mathbf{1}_{M+1}^{\text{sym}}$  under  $U^{\otimes(M+1)}$ . In order to proceed, we first compute the value of the following operator:

$$\begin{aligned} A_1(M) &= \text{Tr}_{(1,M)} \left[ \mathbf{1}_{M+1}^{\text{sym}} \cdot |0\rangle \langle 0|^{\otimes M} \right] = \frac{1}{(M+1)!} \sum_{\sigma} \sum_{i,j=0}^{d-1} \langle 0 |^{\otimes M} \langle i | V(\sigma) | 0 \rangle^{\otimes M} | j \rangle | i \rangle \langle j | \\ &= \frac{M!}{(M+1)!} (\mathbf{1} + M |0\rangle \langle 0|), \end{aligned} \quad (98)$$

where in the first equality we have taken the partial trace over  $M$  subsystems, while in the second one we have written  $\mathbf{1}_{M+1}^{\text{sym}}$  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. (97) we obtain

$$\tilde{c}(k) = \frac{\chi_{\frac{M}{2}}}{\chi_{\frac{M+1}{2}}} \left( \frac{p(k)}{M+1} + \frac{M}{M+1} \text{Tr} \left[ \left( |0\rangle \langle 0| \otimes \Pi_k^{(N)} \right) \frac{\mathbf{1}_{M+1}^{\text{sym}}}{\chi_{\frac{N+1}{2}}} \right] \right), \quad (99)$$

where we have used Eq. (94). We then just need to compute the second term in the sum above, which is very similar to Eq. (95) with the change  $|0\rangle^{\otimes k} \mapsto |0\rangle^{\otimes(k+1)}$ . We finally obtain

$$\tilde{c}(k) = \frac{(d-1)(d+N+k(M+1))N!(N-k+d-2)!}{(d+M)(N-k)!(d+N)!} \quad (100)$$

and the optimal Bayesian estimator for each  $k$  is

$$c(k) = \frac{d+N+k(M+1)}{(d+M)(d+N)}. \quad (101)$$

The corresponding glMSE is given by

$$v_{ep} = \int dc p(c) c^2 - \sum_{k=0}^N p(k) c(k)^2 = \frac{(d-1)((d+M)^2 + (d+2M)N)}{d(1+d)(d+M)^2(d+N)}. \quad (102)$$

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

$$v_{ep} \sim \frac{(d-1)}{d(d+1)(d+N)}, \quad (103)$$

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

$$v_{ep} \sim \frac{6(d-1)}{d(d+1)(M+N)}, \quad (104)$$

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

#### Estimate-and-estimate: Bayesian estimation

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

$$\begin{aligned} p(W|c) &= dW \chi_{\frac{M}{2}} \chi_{\frac{N}{2}} \int_{\mathrm{SU}(d)} dU \mathrm{Tr} \left[ \left( |0\rangle \langle 0|^{\otimes M} \otimes (W|0\rangle \langle 0| W^\dagger)^N \right) \right. \\ &\quad \times U^{\otimes(M+N)} \left( (U(c)|0\rangle \langle 0| U^\dagger(c))^{\otimes M} \otimes |0\rangle \langle 0|^{\otimes N} \right) U^{\dagger \otimes(M+N)} \left. \right], \end{aligned} \quad (105)$$

where again we could include one of the outcomes into the unitary average and, since by redefining  $W \mapsto V^\dagger W$  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

$$p(W) = \int dc p(c) p(W|c) = dW \mathrm{Tr} \left[ |0\rangle \langle 0|^{\otimes(M+N)} (\mathbf{1}_M^{sym} \otimes \mathbf{1}_N^{sym}) \right] = dW, \quad (106)$$

then

$$\begin{aligned} \tilde{c}(W) &= \int dc p(c) p(W|c) c = dW \frac{\chi_{\frac{M}{2}} \chi_{\frac{N}{2}}}{\chi_{\frac{M+1}{2}}} \int_{\mathrm{SU}(d)} dU \mathrm{Tr} \left[ \left( |0\rangle \langle 0|^{\otimes(M+1)} \otimes (W|0\rangle \langle 0| W^\dagger)^N \right) \cdot \right. \\ &\quad \cdot \mathbf{1} \otimes U^{\otimes(M+N)} \left( \mathbf{1}_{M+1}^{sym} \otimes |0\rangle \langle 0|^{\otimes N} \right) \mathbf{1} \otimes U^{\dagger \otimes(M+N)} \left. \right] \\ &= dW \frac{\chi_{\frac{M}{2}} \chi_{\frac{N}{2}}}{\chi_{\frac{M+1}{2}}} \int_{\mathrm{SU}(d)} dU \mathrm{Tr} \left[ (U|0\rangle \langle 0| U^\dagger \otimes (W|0\rangle \langle 0| W^\dagger)^{\otimes N}) \cdot \right. \\ &\quad \cdot \left( \frac{\mathbf{1}}{M+1} + \frac{M}{M+1} |0\rangle \langle 0| \right) \otimes (U|0\rangle \langle 0| U^\dagger)^{\otimes N} \left. \right] \\ &= dW \frac{\chi_{\frac{M}{2}} \chi_{\frac{N}{2}}}{\chi_{\frac{M+1}{2}} (M+1)} \left( \frac{1}{\chi_{\frac{N}{2}}} + \frac{M}{\chi_{\frac{N+1}{2}}} \mathrm{Tr} \left[ \mathbf{1}_{N+1}^{sym} (W^\dagger|0\rangle \langle 0| W \otimes |0\rangle \langle 0|^{\otimes N}) \right] \right) \\ &= dW \frac{\chi_{\frac{M}{2}} \chi_{\frac{N}{2}}}{\chi_{\frac{M+1}{2}} (M+1)} \left( \frac{1}{\chi_{\frac{N}{2}}} + \frac{M}{\chi_{\frac{N+1}{2}} (N+1)} (1+Nw) \right) = \frac{d+M+N+MNw}{(d+M)(d+N)}. \end{aligned} \quad (107)$$

The second equality above comes from averaging over  $U_c$ , the third one from Eq. (98) 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. (98) again and defining  $w = |\langle 0| W |0\rangle|^2$ .

Then the optimal EE Bayesian estimator for each  $W$  is simply  $c(W) = \tilde{c}(W)/dW$  and the minimum gMSE attained by it is

$$v_{ee} = \int dc p(c) c^2 - \int_{\mathrm{SU}(d)} dW \tilde{c}(W)^2 = \frac{(d-1)(d+M+N)(d^2+2MN+d(M+N))}{d(d+1)(d+M)^2(d+N)^2}, \quad (108)$$

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

$$\int_{\mathrm{SU}(d)} dW w^i = \int_{\mathrm{SU}(d)} dW \mathrm{Tr} \left[ \frac{\mathbf{1}_i^{sym}}{\chi_{\frac{i}{2}}} |0\rangle \langle 0|^{\otimes i} \right] = \frac{1}{\chi_{\frac{i}{2}}}. \quad (109)$$

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

$$v_{ee} \sim \frac{(d-1)(d+2N)}{d(d+1)(d+N)^2}, \quad (110)$$

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

$$v_{ee} \sim \frac{8(d-1)}{d(d+1)(M+N)}, \quad (111)$$

which is 2 times larger than the optimal Bayesian strategy.

#### Estimate-and-estimate: local estimation

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

$$v_{ee}(c) = \int_{\mathrm{SU}(d)} p(W|c) (w - c)^2 = c^2 - 2c\bar{w} + \overline{w^2}, \quad (112)$$

where

$$\begin{aligned} \overline{w^i} &= \int p(W|c) w^i = \chi_{\frac{M}{2}} \chi_{\frac{N}{2}} \int_{W,U \in \mathrm{SU}(d)} dW dU \mathrm{Tr} \left[ \left( |0\rangle \langle 0|^{\otimes M} \otimes (W |0\rangle \langle 0| W^\dagger)^{\otimes(N+i)} \right) \right. \\ &\quad \times U^{\otimes(M+N)} \otimes \mathbf{1}^{\otimes i} \left( (U_c |0\rangle \langle 0| U_c^\dagger)^{\otimes M} \otimes |0\rangle \langle 0|^{\otimes(N+i)} \right) U^{\dagger \otimes(M+N)} \otimes \mathbf{1}^{\otimes i} \left. \right] \\ &= \frac{\chi_{\frac{M}{2}} \chi_{\frac{N}{2}}}{\chi_{\frac{N+i}{2}}} \int_{\mathrm{SU}(d)} dU \mathrm{Tr} \left[ \left( (U_c |0\rangle \langle 0| U_c^\dagger)^{\otimes M} \otimes \mathbf{1}_{N+i}^{sym} \right) \cdot \right. \\ &\quad \cdot \left. \left( (U^\dagger |0\rangle \langle 0| U)^{\otimes(M+i)} \otimes |0\rangle \langle 0|^{\otimes N} \right) \right]. \end{aligned} \quad (113)$$

Then the first moment is straightforward to compute by inserting Eq. (98):

$$\begin{aligned} \overline{w} &= \frac{\chi_{\frac{M}{2}} \chi_{\frac{N}{2}}}{\chi_{\frac{N+1}{2}} (N+1)} \left( \frac{1}{\chi_{\frac{M}{2}}} + \frac{N}{\chi_{\frac{M+1}{2}}} \mathrm{Tr} \left[ \mathbf{1}_{M+1}^{sym} \left( |0\rangle \langle 0|^{\otimes M} \otimes U_c^\dagger |0\rangle \langle 0| U_c \right) \right] \right) \\ &= \frac{\chi_{\frac{M}{2}} \chi_{\frac{N}{2}}}{\chi_{\frac{N+1}{2}} (N+1)} \left( \frac{1}{\chi_{\frac{M}{2}}} + \frac{N}{\chi_{\frac{M+1}{2}}} \frac{1+Mc}{1+M} \right). \end{aligned} \quad (114)$$

For the second moment we first need to evaluate the following operator:

$$\begin{aligned} A_2(N) &= \mathrm{Tr}_{(1,N)} \left[ \mathbf{1}_{N+2}^{sym} \cdot |0\rangle \langle 0|^{\otimes N} \right] = \frac{1}{(N+2)!} \sum_{\sigma} \sum_{\underline{i}, \underline{j} \in \{0, \dots, d-1\}^2} \langle 0|^{\otimes N} \langle \underline{i} | V(\sigma) |0\rangle^{\otimes N} | \underline{j} \rangle | \underline{i} \rangle \langle \underline{j} | \\ &= \frac{N!}{(N+2)!} [2\mathbf{1}_2^{sym} + 2N (|0\rangle \langle 0| \otimes \mathbf{1} + \mathbf{1} \otimes |0\rangle \langle 0|) + N(N-1) |00\rangle \langle 00|]. \end{aligned} \quad (115)$$

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,

$$\mathbf{1}_2^{sym} = \frac{1}{2} \sum_{(i_1, i_2)} (|i_1, i_2\rangle \langle i_1, i_2| + |i_1, i_2\rangle \langle i_2, i_1|); \quad (116)$$

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  $\underline{i} = \underline{j} = \underline{0}$ , 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

$$\begin{aligned}
\overline{w^2} &= \frac{\chi_{\frac{M}{2}} \chi_{\frac{N}{2}}}{\chi_{\frac{N+2}{2}}} \int_{\text{SU}(d)} dU \text{Tr} \left[ (U_c |0\rangle \langle 0| U_c^\dagger)^{\otimes M} (U^\dagger |0\rangle \langle 0| U)^{\otimes M} \otimes A_2(N) (U^\dagger |0\rangle \langle 0| U)^2 \right] \\
&= \frac{\chi_{\frac{M}{2}} \chi_{\frac{N}{2}}}{\chi_{\frac{N+2}{2}}(N+2)(N+1)} \left( \frac{2}{\chi_{\frac{M}{2}}} + \frac{4N}{\chi_{\frac{M+1}{2}}} \text{Tr} [A_1(M) U_c^\dagger |0\rangle \langle 0| U_c] \right. \\
&\quad \left. + \frac{N(N-1)}{\chi_{\frac{M+2}{2}}} \text{Tr} [A_2(M) (U_c^\dagger |0\rangle \langle 0| U_c)^{\otimes 2}] \right) \\
&= \frac{\chi_{\frac{M}{2}} \chi_{\frac{N}{2}}}{\chi_{\frac{N+2}{2}}(N+2)(N+1)} \left( \frac{2}{\chi_{\frac{M}{2}}} + \frac{4N(1+Mc)}{\chi_{\frac{M+1}{2}}(M+1)} + \frac{N(N-1)(2+4Mc+M(M-1)c^2)}{\chi_{\frac{M+2}{2}}(M+2)(M+1)} \right).
\end{aligned} \tag{117}$$

By plugging the expressions of Eqs. (114,117) into Eq. (112) we finally get

$$v_{ee}(c) = \frac{(d+M+N)((2+c(cd-2)(d+1))(d+M+N+1)+2c(1-c)MN)}{(d+M)(d+M+1)(d+N)(d+N+1)}. \tag{118}$$

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

$$v_{ee}(c) \sim \frac{2+c(d+1)(cd-2)+2cN-2c^2N}{(d+N)(d+N+1)} \underset{N \gg 1}{\approx} \frac{2c(1-c)}{N} + O\left(\frac{1}{N^2}\right), \tag{119}$$

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

$$v_{ee}(c) \sim \frac{8c(1-c)}{(M+N)}, \tag{120}$$

which is 2 times larger than the optimal strategy.

### Bayesian estimation using the swap test

Here we derive the formulas employed for the plot of the swap-test performance in Fig. 3 of the main text. 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. (93), with  $p(k|c) \rightarrow \text{Bin}(k, N, p(c))$ . We have

$$p(c) = \int dc p(c) \text{Bin}(k, N, p(c)) = \frac{(d-1)\binom{N}{k} {}_2F_1(1, -k; d-1k+N; -1)}{2^N(d-k+N-1)}, \tag{121}$$

$$\tilde{c}(k) = \int dc c p(c) \text{Bin}(k, N, p(c)) = (d-1)2^{-N}\binom{N}{k} {}_2\tilde{F}_1(2, -k; d-1k+N+1; -1)\Gamma(d-k+N-1), \tag{122}$$

where  ${}_2F_1$  and  ${}_2\tilde{F}_1$  are hypergeometric and regularized hypergeometric functions respectively. Following the derivations of the previous section, the minimum glMSE attainable with the swap test can be written as

$$v_{sw}^{\text{bay}} = \frac{2}{d(d+1)} - \sum_{k=0}^N \frac{(d-1)\binom{N}{k}(d-k+N-1) {}_2\tilde{F}_1(2, -k; d-k+N+1; -1)^2 \Gamma(d-k+N-1)^2}{2^N {}_2F_1(1, -k; d-k+N; -1)}. \tag{123}$$

### Comparison between Schur transform and swap test with imperfect implementations

In this section we sketch an evaluation of the effect of imperfect gates on the accuracy of the estimate of the overlap. First of all we model the error of each iteration of the swap test as white noise for each iteration:  $\mathcal{N}_{sw}(p(c)) = (1 - \epsilon_{sw})p(c) + \epsilon_{sw}\frac{1}{2}$ , the Fisher information becomes

$$H(\mathcal{N}_{sw}(p(c))) = \frac{(1 - \epsilon_{sw})^2}{1 - c^2(1 - \epsilon_{sw})^2}, \tag{124}$$

For  $N$  repetitions, one gets a resulting MSE

$$v_{sw,noisy}(c) = \frac{1 - c^2(1 - \epsilon_{sw})^2}{(1 - \epsilon_{sw})^2 N}. \tag{125}$$

We model the noise on the Schur transform measurement outcomes also as mixing with a probability distribution  $q(c)$ :  $\mathcal{N}_{sw}(p(J|c)) = (1 - \epsilon_{Sch})p(J|c) + \epsilon_{Sch}q(c)$ , with a probability of mixing that scales exponentially in the number of gates,  $1 - \epsilon_{Sch} \approx (1 - \epsilon)^g$ , where  $\epsilon$  is the error per gate, and  $g$  is the total number of gates. We recall the joint convexity property of the Fisher Information, coming from its monotonicity:

$$F(\lambda p(c) + (1 - \lambda)q(c)) \leq \lambda F(p(c)) + (1 - \lambda)F(q(c)), \quad (126)$$

If we assume  $q(c)$  to be overlap independent, we obtain the bound

$$F(\mathcal{N}_{Sch}(p(J|c))) \leq (1 - \epsilon_{Sch})F(p(J|c)), \quad (127)$$

so that

$$v_{Sch,noisy}(c) \geq \frac{2c(1 - c)}{(1 - \epsilon_{Sch})N}. \quad (128)$$

This is a very conservative estimate, as we are assuming we acquiring useful information with exponentially small probability. Hence the Swap test outperforms our optimal strategy, based on the Schur transform, when the respective implementation errors satisfy the following relation

$$\frac{(1 - \epsilon_{Sw})^2}{(1 - c^2)(1 - \epsilon_{Sw}^2)} \geq \frac{1 - \epsilon_{Sch}}{2c(1 - c)}. \quad (129)$$

One can express  $\epsilon_{Sch}$  and  $\epsilon_{Sw}$  in terms of the error per gate,  $\epsilon$ , raised to gate complexity of their respective circuits. An intermediate strategy could be to divide the  $N$  copies of both  $|\phi\rangle$  and  $|\psi\rangle$ , into  $R$  groups of  $S$  copies, and perform the optimal measurement on each group, followed by classical post-processing. If  $N = M = RS$  and  $F(J|c, S)$  is the optimal Fisher information for the case with  $M = N = S$  copies, the Cramer-Rao bound reads

$$v(c) \geq \frac{1}{RF(J|c, S)}. \quad (130)$$

The best option would be to choose  $S$  as the highest number of copies such that the architecture can perform the optimal measurement in a sufficiently precise way. On the other hand, if one requires to be in the asymptotic regime of the approximation for  $c > c_0$ , one can just find the minimum  $S$  for which the approximation works, and perform the optimal measurement with  $S$  copies  $R$  times. The classical post processing will have the optimal asymptotic performance for  $c > c_0$ . In any case the bound (130) is asymptotically achieved by a maximum likelihood estimator when  $R \rightarrow \infty$ .

### 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  $\{\Pi_J\}_J$ , so that Eq. (8) of the main text reads

$$F_{op} = \sum_{J=J_{\min}}^{J_{\max}} |\langle \Psi | \Pi_J | \Psi \rangle|^2 = \sum_J p(J|c)^2, \quad (131)$$

where we have used the  $SU(d)$ -invariance of  $\Pi_J$ . 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 pair of copies is a triplet/singlet projection, which is again  $SU(d)$ -invariant, and succeeds/fails with probability  $(1 \pm c)/2$ . Hence we have

$$F_{sw} = \left[ \left( \frac{1+c}{2} \right)^2 + \left( \frac{1-c}{2} \right)^2 \right]^N = \left( \frac{1+c^2}{2} \right)^N. \quad (132)$$

### 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

$$\begin{aligned}\rho_{\psi}^{\otimes N}(r_0) &= \left( r_0 |\psi\rangle\langle\psi| + (1 - r_0) \frac{\mathbf{1}}{2} \right)^{\otimes N}, \\ \rho_{\phi}^{\otimes M}(r_1) &= \left( r_1 |\phi\rangle\langle\phi| + (1 - r_1) \frac{\mathbf{1}}{2} \right)^{\otimes M},\end{aligned}\quad (133)$$

where  $r_{0(1)}$  denotes the corresponding purity of the states. Following [38] the states in Eq. (133) can be written in the total angular momentum basis, after tracing out multiplicities, as

$$\begin{aligned}\tilde{\rho}_{\psi}^{\otimes N} &= \sum_{J_0=0}^{\frac{M}{2}} p_{J_0} \tau_{J_0}^{(0)}(\vec{n}_0) \\ \tilde{\rho}_{\phi}^{\otimes M} &= \sum_{J_1=0}^{\frac{N}{2}} p_{J_1} \tau_{J_1}^{(1)}(\vec{n}_1),\end{aligned}\quad (134)$$

where

$$\begin{aligned}\tau_{J_0}^{(0)} &= \frac{1}{Z_{J_0}^{(0)}} \sum_{k=-J_0}^{J_0} R_0^k |J_0, k\rangle\langle J_0, k| \\ \tau_{J_1}^{(1)} &= \frac{1}{Z_{J_1}^{(1)}} \sum_{l=-J_1}^{J_1} R_1^l \sum_{\alpha, \beta=-J_1}^{J_1} d_{\alpha, l}^{(J_1)}(2 \cos^{-1} \sqrt{c}) d_{l, \beta}^{(J_1)}(2 \cos^{-1} \sqrt{c}) |J_1, \alpha\rangle\langle J_1, \beta|\end{aligned}\quad (135)$$

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

$$p_{J_0} = \left( \frac{1-r^2}{4} \right)^{\frac{N}{2}} \binom{N}{\frac{N}{2} - J_0} \frac{2J_0 + 1}{\frac{N}{2} + J_0 + 1} Z_{J_0}\quad (136)$$

and similarly for  $p_{J_1}$ . Using Eq. (17) we obtain

$$\rho(c) = \mathcal{G}_{\text{SU}(2)} \left[ \rho_{\psi}^{\otimes N} \otimes \rho_{\phi}^{\otimes M} \right] = \sum_J \sum_{J_0, J_1} p_{J_0} p_{J_1} \sum_{k, l} \frac{R_0^k R_1^l}{Z_{J_0}^{(0)} Z_{J_1}^{(1)}} \sum_{\alpha=-J_1}^{J_1} \left( C_{J_0, k; J_1, \alpha}^{J, k+\alpha} d_{\alpha, l}^{(J_1)}(2 \cos^{-1} \sqrt{c}) \right)^2 \frac{\mathbf{1}_{\mathcal{U}^{(J)}(\text{SU}(2))} \otimes \sigma^{(J_0, J_1)}}{2J+1}, \quad (137)$$

where  $\sigma^{(J_0, J_1)} \in \mathcal{B}(\mathcal{U}^{(J)}(S_{N+M}))$  and they are orthogonal for different couples  $(J_0, J_1)$ . To calculate the gMSE we need to compute the operators  $\Gamma, \eta$  of Eq. (74). A similar calculation as in Eq. (77), (78) gives

$$\begin{aligned}\Gamma &= \mathcal{G}_{\text{SU}(2)} \left[ \rho_{\psi}^{\otimes N} \right] \otimes \mathcal{G}_{\text{SU}(2)} \left[ \rho_{\phi}^{\otimes M} \right] \\ &= \left( \sum_{J_0=0}^{\frac{N}{2}} p_{J_0} \frac{\mathbf{1}_{\mathcal{U}^{(J_0)}(\text{SU}(2))}}{2J_0 + 1} \otimes \frac{\mathbf{1}_{\mathcal{U}^{(J)}(S_{M+N})}}{\omega_J} \right) \otimes \left( \sum_{J_1=0}^{\frac{M}{2}} p_{J_1} \frac{\mathbf{1}_{\mathcal{U}^{(J_1)}(\text{SU}(2))}}{2J_1 + 1} \otimes \frac{\mathbf{1}_{\mathcal{U}^{(J)}(S_{M+N})}}{\omega_J} \right) \\ &= \sum_{J_0=0}^{\frac{N}{2}} \sum_{J_1=0}^{\frac{M}{2}} \frac{p_{J_0} p_{J_1}}{(2J_0 + 1)(2J_1 + 1)} \sum_{J=|J_0 - J_1|}^{J_0 + J_1} \mathbf{1}_{\mathcal{U}^{(J)}(\text{SU}(d))} \otimes \sigma^{(J_0, J_1)},\end{aligned}\quad (138)$$

For  $\eta$  one obtains

$$\begin{aligned}\eta &= \int_{\text{SU}(2)} U^{\otimes(N+M)} \left( \sum_{J_0=0}^{\frac{N}{2}} \sum_{J_1=0}^{\frac{M}{2}} \sum_{k=-J_0}^{J_0} \sum_{l=-J_1}^{J_1} \frac{R_0^k R_1^l}{Z_{J_0}^{(0)} Z_{J_1}^{(1)}} |J_0, k\rangle\langle J_0, k| \right. \\ &\quad \left. \otimes \int_{\text{SU}(2)} dh |D_{\frac{1}{2} \frac{1}{2}}^{(\frac{1}{2})}(h)|^2 D^{(J_1)}(h) |J_1, l\rangle\langle J_1, l| D^{(J_1)}(h)^\dagger \right) U^{\dagger \otimes(N+M)}.\end{aligned}\quad (139)$$

We finally obtain

$$\begin{aligned}
\eta &= \sum_{J_0=0}^{\frac{N}{2}} \sum_{J_1=0}^{\frac{M}{2}} \frac{p_{J_0} p_{J_1}}{(2J_0+1)(2J_1+1)} \sum_{J=|J_0-J_1|}^{J_0+J_1} \left( 1 - \frac{(2J_0+1)(2J_1+1)}{p_{J_0} p_{J_1}} \sum_{k=-J_0}^{J_0} \sum_{l=-J_1}^{J_1} \frac{R_0^k R_1^l}{Z_{J_0}^{(0)} Z_{J_1}^{(1)}} \right. \\
&\quad \left. \sum_{L=|J_1-\frac{1}{2}|}^{J_1+\frac{1}{2}} \sum_{h=-J_1}^{J_1} \frac{\left( C_{\frac{1}{2}, -\frac{1}{2}; J_1, h}^{L, -\frac{1}{2}+h} C_{\frac{1}{2}, \frac{1}{2}; J_1, l}^{L, \frac{1}{2}+l} C_{J_0, k; J_1, h}^{J, k+h} \right)^2}{(2L+1)} \right) \frac{\mathbf{1}_{\mathcal{U}^{(J)}(\mathrm{SU}(2))}}{(2J+1)} \otimes \sigma^{(J_0, J_1)} \\
&= \sum_{J_0=0}^{\frac{N}{2}} \sum_{J_1=0}^{\frac{M}{2}} \frac{p_{J_0} p_{J_1}}{(2J_0+1)(2J_1+1)} \sum_{J=|J_0-J_1|}^{J_0+J_1} \left( 1 - \frac{(2J_0+1)(2J_1+1)}{p_{J_0} p_{J_1}} \sum_{L=|J_1-\frac{1}{2}|}^{J_1+\frac{1}{2}} \sum_{L'=|J_0-\frac{1}{2}|}^{J_0+\frac{1}{2}} (2J+1) \right. \\
&\quad \times \left. \sum_{k=-J_0}^{J_0} \sum_{l=-J_1}^{J_1} \frac{R_0^k \left( C_{\frac{1}{2}, \frac{1}{2}, J_0, k}^{L', \frac{1}{2}+k} \right)^2 R_1^l \left( C_{\frac{1}{2}, \frac{1}{2}, J_1, l}^{L, \frac{1}{2}+l} \right)^2}{Z_{J_0}^{(0)} Z_{J_1}^{(1)}} \left\{ \begin{matrix} J_1 & \frac{1}{2} & L \\ L' & J & J_0 \end{matrix} \right\}^2 \right) \frac{\mathbf{1}_{\mathcal{U}^{(J)}(\mathrm{SU}(2))}}{(2J+1)} \otimes \sigma^{(J_0, J_1)}. \tag{140}
\end{aligned}$$

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

$$c(J, J_0, J_1) = \frac{\mathrm{Tr} [\Pi_J (\Pi_{J_0} \otimes \Pi_{J_1} \eta)]}{\mathrm{Tr} [\Pi_J (\Pi_{J_0} \otimes \Pi_{J_1} \Gamma)]} \tag{141}$$

and the glMSE reads

$$v_{op, mix} = \int_0^1 p(c) c^2 - \sum_{J, J_0, J_1} p(J, J_0, J_1) c(J, J_0, J_1)^2. \tag{142}$$

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 the fact that  $p_{J_0} \frac{R_0^{J_0}}{Z_{J_0}}$  can be written as

$$\begin{aligned}
p_{J_0} \frac{R_0^{J_0}}{Z_{J_0}} &= \left( \frac{1-r^2}{4} \right)^{\frac{N}{2}} \binom{N}{\frac{N}{2}-J_0} \frac{2J_0+1}{\frac{N}{2}+J_0+1} R_0^{J_0} = \binom{N}{\frac{N}{2}-J_0} \frac{2J_0+1}{\frac{N}{2}+J_0+1} \left( \frac{1+r_0}{2} \right)^{\frac{N}{2}+J_0} \left( \frac{1-r_0}{2} \right)^{\frac{N}{2}-J_0} \\
&= \frac{2J_0+1}{\frac{N}{2}+J_0+1} \mathrm{Bin}(N, \frac{N}{2}-J_0, \frac{1+r}{2}),
\end{aligned} \tag{143}$$

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

$$v_{op, mix} = \frac{1}{6Mr_0^2} + \frac{1}{6Nr_1^2} + o(Z^{-1}) \tag{144}$$

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