Beyond the Swap Test: Optimal Estimation of Quantum State Overlap

M. Fanizza,[1,*] M. Rosati,[2,†] M. Skotiniotis,[2] J. Calsamiglia,[2] and V. Giovannetti[1]
[1]*NEST, Scuola Normale Superiore and Istituto Nanoscienze—CNR, I-56126 Pisa, Italy*
[2]*Física Teòrica: Informació i Fenòmens Quàntics, Departament de Física, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain*

(Received 27 June 2019; revised manuscript received 11 December 2019; accepted 12 December 2019; published 14 February 2020)

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 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.

DOI: 10.1103/PhysRevLett.124.060503

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], and communication without a shared reference frame [16–19] to quantum machine learning [20–29]. Recently, with the advent of quantum machine learning protocols [22–28], 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 [20,21]. In most applications, OVE is carried out through the swap test (SWT) [9,20,21]: given two systems in the state $|\psi\rangle|\phi\rangle$, the probability of projecting it onto 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 Letter, 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_σ being an unknown permutation of the qudits. Note that in this case OVE constitutes in itself an instance of an unsupervised quantum-classical learning problem, in a setting similar to that in Ref. [30].

The measurements optimizing the average information gain [16] and the average error [17,18] have been derived for the case of qubits, with only numerical solutions [19] 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) [31,32], whereas for Bayesian estimation [31,33] we provide an exact solution, generalizing the results of Ref. [19]. 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 local operations with classical communication (LOCC) strategies based on estimating either one or both of $|\psi\rangle$ and $|\phi\rangle$; see Fig. 1. Such strategies are useful in distributed scenarios where copies of $|\psi\rangle$ and $|\phi\rangle$ are produced in different and distant laboratories. We show that, in the limit of a large $M+N$ and $|M-N|$

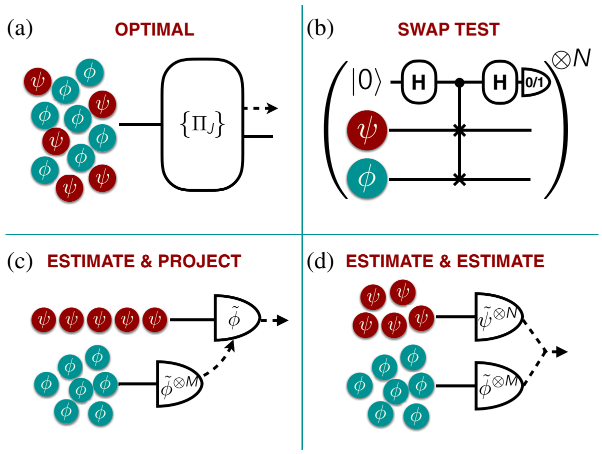

FIG. 1. Sketch of the OVE strategies studied in the Letter. (a) Optimal measurement, e.g., by Schur transform (see Ref. [34] for the circuit implementation). (b) Circuit for the SWT, to be repeated N times. (c) Estimate $|\phi\rangle$ and project $|\psi\rangle$ on the estimated direction. (d) Estimate both $|\psi\rangle$ and $|\phi\rangle$ and calculate the overlap.

constant, the optimal strategy displays a finite asymptotic gap with respect to all of 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:

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

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 $v = \sum_k \int dW [c(k) - c]^2 \mathrm{Tr}[E_k \rho(c)]$, where we have defined the effective state

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

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

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

where the distribution over overlaps is given by [30,35]

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

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 there exists a 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:

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

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) [36], the integral in Eq. (2) can be performed using SU(d) representation theory and SU(2) Clebsch-Gordan coefficients, obtaining the block-diagonal form

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

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

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

where $P_n^{(\alpha,\beta)}(x)$ is the nth-degree Jacobi polynomial. In the previous equations, for $d=2$, J is the familiar total-angular-momentum label, and $\mathbb{1}_J = \sum_{M=-J}^{J} |J,M\rangle\langle J,M|$ is the projector on the subspace of total angular momentum J, of dimension $\chi_J = 2J+1$. In general, for $d > 2$, $\mathbb{1}_J$ 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 a (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 with 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 of the information about the overlap is

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

Local est.	$v_{\mathrm{op}}(c)$	$v_{\mathrm{ep}}(c)$	$v_{\mathrm{ee}}(c)$
$M=N\to\infty$	$[4c(1-c)]/(M+N)$	$(3/2)v_{\mathrm{op}}(c)$	$2v_{\mathrm{op}}(c)$
$M\to\infty$	$[c(1-c)]/N$	$v_{\mathrm{op}}(c)$	$2v_{\mathrm{op}}(c)$

Bayesian est.	v_{op}	v_{ep}	v_{ee}
$M=N\to\infty$	$[4(d-1)]/[d(d+1)(M+N)]$	$(3/2)v_{\mathrm{op}}$	$2v_{\mathrm{op}}$
$M\to\infty$	$(d-1)/[d(d+1)(d+N)]$	v_{op}	$[(d+2N)/(2+N)]v_{\mathrm{op}}$

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 Π_J on the subspaces labeled with J, and it can be implemented via weak Schur sampling [34]. Indeed, for any positive operator valued measure (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 $p_{k,J} \coloneqq p(k|J)p(J|c)$. The same outcome probabilities $p_{k,J}$ can be generated by applying the POVM $\{\Pi_J\}_J$ directly followed by classical postprocessing. 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 for both local and global estimations, and one can replace $\mathrm{Tr}[E_J\rho(c)]$ with $p(J|c)$ in Eqs. (1) and (5), effectively reducing our problem to one of classical estimation, i.e., optimizing the function $c(J)$.

Local estimation.—The classical Cramer-Rao bound [37] 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 $M+N\to\infty$ and $M-N \ll (M+N)\sqrt{c}$, we can use an approximation of the Jacobi polynomial given in Ref. [38] to obtain the following asymptotically unbiased estimator and its associated MSE:

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

In the SM [36], we show that $v_{\mathrm{op}}(c)$ coincides with $H(c)^{-1}$ to leading order in $1/(M+N)$, and hence the Cramer-Rao bound is achievable in this limit. If instead $M\to\infty$ and N is finite, it is clear that $|\phi\rangle$ can be estimated perfectly, and hence the optimal strategy is to project the copies of $|\psi\rangle$ in this known direction, with resulting $v_{\mathrm{op}}(c) = [c(1-c)]/N$.

Bayesian estimation.—The optimal classical Bayesian (global) estimator is given [31] by $c_{\mathrm{op}}^{\mathrm{Bay}}(J) = [\int dc\, c\, p(c)p(J|c)]/[\int dc\, p(c)p(J|c)]$. Using graphical calculus techniques for the recoupling theory of Clebsch-Gordan coefficients [39] as explained in the SM [36], we obtain the following optimal global estimator and corresponding MSE:

$$c_{\mathrm{op}}^{\mathrm{Bay}}(J) = \frac{d+J+J^2+[(M+N)/2]-[(M+N)/2]^2+MN}{(d+M)(d+N)}, \quad (9)$$

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

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 uninformative with respect to the information that can be obtained by the actual measurement; indeed we can see that when $M+N\to\infty$, $M-N$ constant, $c_{\mathrm{op}}^{\mathrm{Bay}}(J) \approx c_{\mathrm{op}}^{\mathrm{loc}}(J)$, implying that the local optimal estimator is also a good Bayesian estimator, and vice versa; (ii) contrary 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_{op} decays as d^{-2} if 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 one-way LOCC (1-LOCC) on $|\psi\rangle^{\otimes N}$ and $|\phi\rangle^{\otimes M}$. The estimate-and-project (EP) strategy consists of 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 [37]. However, EP is not necessarily the optimal 1-LOCC strategy. The corresponding POVM elements can be written as $E_{V,k}^{(\mathrm{ep})} = dV\, E_V^{(M)} \otimes V^{\otimes N} \Pi_k^{(N)} V^{\dagger \otimes N}$, where $E_V^{(M)} = \chi_{M/2}(d)(V|0\rangle\langle 0|V^\dagger)^{\otimes M}$ is the optimal covariant measurement to estimate $|\phi\rangle$ [40,41], and $\Pi_k^{(N)}$ represents k successful projections of the copies of $|\psi\rangle$ on the estimate of $|\phi\rangle$. The estimator is $c_{\mathrm{ep}}^{\mathrm{loc}}(k) = k/N$. The estimate-and-estimate (EE) strategy instead consists of estimating both $|\psi\rangle$ and $|\phi\rangle$ separately, then computing the overlap between the estimated states. The corresponding POVM elements can be written as

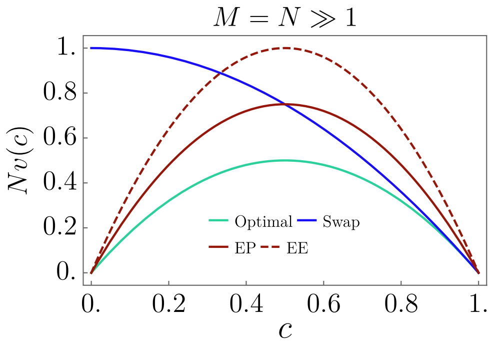

FIG. 2. Plot of the optimal local MSE scaling coefficient $Nv(c)$ vs the true value of the overlap c, at leading order in $M = N$, for the strategies analyzed in the Letter.

$E^{(\mathrm{ee})}_{V,W} = dV\, dW E^{(M)}_V \otimes E^{(N)}_W$, i.e., a product of two covariant measurements to estimate $|\phi\rangle$ and $|\psi\rangle$. We take as local estimator $c^{\mathrm{loc}}_{\mathrm{ee}}(V, W) = |\langle 0|V^\dagger W|0\rangle|^2$. In the SM [36], we provide the exact results for local and Bayesian estimations using EP and EE. Table I compares the performance of these strategies to the optimal one in two asymptotic limits. We find that, for both local and Bayesian estimations, the EE strategy is always worse than the optimal by a factor of 2, whereas the EP strategy attains a MSE equal to the optimal in the limit $M \to \infty$, N finite.

Performances comparison.—We now compare the strategies discussed so far with the traditional SWT [9]. Note that all of these strategies except the optimal one require labeling of the states. The latter consists of projecting the state $|\psi\rangle \otimes |\phi\rangle$ onto its triplet or 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 $p(c) = (1 + c)/2$ and the ensuing statistics of k successful projections out of N trials are given by the binomial distribution. The optimal local MSE attainable by this test is well known, $v_{\mathrm{sw}}(c) = (1 - c^2)/N$, while for the optimal global MSE v_{sw} one can derive an exact expression for each value of k, then compute the sum numerically, as detailed in the SM [36]. In the asymptotic limit of $M = N \gg d$, a good approximation is provided by averaging the optimal local MSE, $v_{\mathrm{sw}} \simeq \int dc\, p(c) v_{\mathrm{sw}}(c) = (d+2)(d-1)/[d(d+1)N]$, which is $\sim d$ times larger than v_{op}.

In the same limit, we can compare the local MSE of all of the strategies; see Fig. 2. First, we observe a gap between the optimal strategy that attains the QFI and all of 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 noncollective one. Second, we observe that the relative error $\sqrt{v(c)}/c$ for small c scales as $1/c\sqrt{N}$ for the SWT and as $1/\sqrt{cN}$ for the other strategies, implying a quadratic improvement in $1/\sqrt{c}$ in the number of copies needed to reach a fixed relative error, while the optimal measurement is still computationally efficient (see the 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 [42,43] for quantum variational circuits, and other types of strategies have been proposed to address this issue [44–46].

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 to 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, with the former being closer to the optimal one.

Gate complexity.—The advantage in the precision of the optimal estimation comes with the trade-off that the optimal measurement requires entangling operations over the whole system of $N + M$ qudits. The Schur transform [34,47,48] is a way to perform the optimal measurement, and it requires $O(\mathrm{poly}(N + M, \log d, \log 1/\epsilon))$ qudit gates for precision ϵ. The resulting algorithm is efficient but still unfeasible without error correction. The SWT instead requires N independent circuits of fixed depth, and it may still be convenient for large overlaps or very noisy gates.

A midterm solution is to divide input data into R groups of S copies of $|\phi\rangle$ and $|\psi\rangle$ such that S is the largest integer

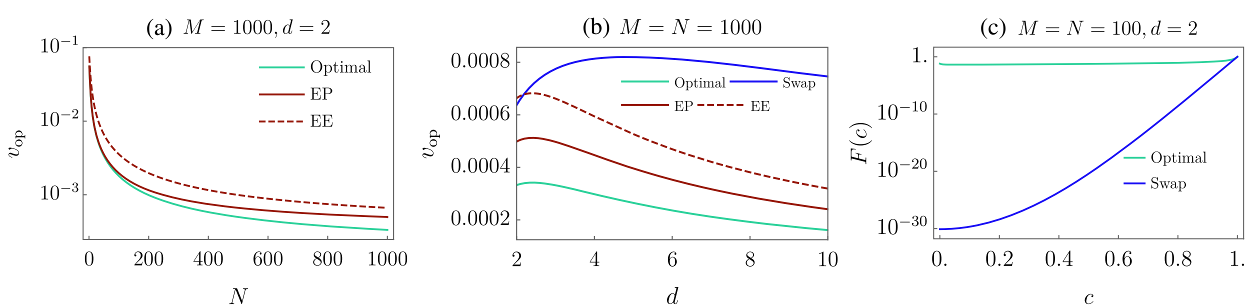

FIG. 3. (a) Plot of the optimal global MSE v_{op} vs the number of copies of one state N, for a fixed number of copies of the other $M = 1000$, in dimension $d = 2$, for the optimal, EP, and EE strategies. (b) Plot of the optimal global MSE v_{op} vs the dimension d, for a fixed number of copies $M =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

for which the given architecture can perform the optimal measurement with high fidelity, repeat the measurement R times, and do classical postprocessing. The performances of these intermediate protocols are between SWTs and the optimal measurement. See the SM [36] for a more detailed discussion of these issues.

Measurement invasiveness.—Another relevant figure of merit for applications is the fidelity between the postmeasurement state and the initial one, averaged over the measurement outcomes. Both the optimal measurement and the SWT are projective measurements. We assume that the postmeasurement states are given by the result of such projections, and hence the average postmeasurement fidelity can be written as

$$F(c) = \int dU \sum_k |\langle \Psi_U | E_k | \Psi_U \rangle|^2, \tag{11}$$

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

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

as shown in the SM [36]. 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 postprocessing 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 $\Delta_r = r\mathcal{I} + (1-r)(\mathbb{1}/d)\mathrm{Tr}$, 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 [36], we compute the optimal MSE in this case, restricting to $d = 2$ for simplicity. In the limit $M, N \to \infty$ with M/N finite, the global MSE at leading order is $v_{\mathrm{op,mix}} = (1/6Mr_0^2) + (1/6Nr_1^2)$, 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 of r_i^{-2} for each state.

Conclusions.—In this Letter, we computed the ultimate precision attainable when 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 of the copies and can be implemented efficiently using the Schur transform, although this remains experimentally challenging. A practical alternative is to do Schur sampling on subsets of the dataset, followed by classical postprocessing. 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 is 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, while also providing improvements in the run-time of quantum algorithms.

M. F. and V. G. acknowledge support by MIUR via PRIN 2017 (Progetto di Ricerca di Interesse Nazionale): project QUSHIP (2017SRNBRK). M. R., M. S., and J. C. acknowledge support from the Spanish MINECO, Project No. FIS2016-80681-P with the support of AEI/FEDER funds, and from the Generalitat de Catalunya, Project No. CIRIT 2017-SGR-1127. M. R. also acknowledges partial financial support from the Baidu-UAB collaborative project "Learning of quantum hidden Markov models." M. S. also acknowledges support from Spanish MINECO Project No. IJCI-2015-24643. J. C. also acknowledges the Catalan Government for the project QuantumCAT 001-P-001644 (RIS3CAT comunitats) co-financed by the European Regional Development Fund (FEDER). This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 845255. M. F. thanks Matthias Christandl for the helpful comments. M. S. and J. C. acknowledge the 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. Quantum. Inform. **04**, 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 J. Phys. **10**, 033023 (2008).
[5] I. Marvian and R. W. Spekkens, New J. Phys. **15**, 033001 (2013).
[6] I. Marvian and R. W. Spekkens, Phys. Rev. A **90**, 062110 (2014).
[7] I. Marvian and R. W. Spekkens, Commun. Math. Phys. **331**, 431 (2014).
[8] Y. Mo and G. Chiribella, 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 (2013).
[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. H. Lindner, P. F. Scudo, and D. Bruß, Int. J. Quantum. Inform. **04**, 131 (2006).
[19] N. Gisin and S. Iblisdir, Eur. Phys. J. D **39**, 321 (2006).
[20] L. Cincio, Y. Subaşı, A. T. Sornborger, and P. J. Coles, New J. Phys. **20**, 113022 (2018).
[21] U. Chabaud, E. Diamanti, D. Markham, E. Kashefi, and A. Joux, Phys. Rev. A **98**, 062318 (2018).
[22] S. Lloyd, M. Mohseni, and P. Rebentrost, arXiv:1307.0411.
[23] P. Rebentrost, M. Mohseni, and S. Lloyd, Phys. Rev. Lett. **113**, 130503 (2014).
[24] A. W. Harrow, A. Hassidim, and S. Lloyd, Phys. Rev. Lett. **103**, 150502 (2009).
[25] Z. Zhao, J. K. Fitzsimons, and J. F. Fitzsimons, Phys. Rev. A **99**, 052331 (2019).
[26] N. Wiebe, A. Kapoor, and K. Svore, Quantum Inf. Comput. **18**, 0318 (2015).
[27] V. Havlicek, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, Nature (London) **567**, 209 (2019).
[28] N. Liu and P. Rebentrost, Phys. Rev. A **97**, 042315 (2018).
[29] H.-Y. Huang, K. Bharti, and P. Rebentrost, arXiv:1909.07344.
[30] G. Sentís, A Monràs, R. Muñoz-Tapia, and J. Calsamiglia, Phys. Rev. X **9**, 041029 (2019).
[31] C. W. Helstrom, J. Stat. Phys. **1**, 231 (1969).
[32] M. G. A. Paris, Int. J. Quantum. Inform. **07**, 125 (2009).
[33] S. D. Personick, IEEE Trans. Inf. Theory **17**, 240 (1971).
[34] A. W. Harrow, Ph.D. thesis, Massachusetts Institute of Technology, 2005, arXiv:quant-ph/0512255.
[35] L. Alonso and T. Gorin, J. Phys. A **49**, 145004 (2016).
[36] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.124.060503 for detailed proofs and calculations.
[37] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982).
[38] G. Szego, *Orthogonal Polynomials*, 4th ed., Colloquium Publications Vol. 23 (American Mathematical Society, Providence, 1975).
[39] D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskii, *Quantum Theory of Angular Momentum* (World Scientific, Singapore, 1988).
[40] M. Hayashi, J. Phys. A **31**, 4633 (1998).
[41] M. Hayashi, *Group Representation for Quantum Theory* (Springer, Berlin, 2017).
[42] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Nat. Commun. **9**, 4812 (2018).
[43] A. G. R. Day, M. Bukov, P. Weinberg, P. Mehta, and D. Sels, Phys. Rev. Lett. **122**, 020601 (2019).
[44] X. Glorot and Y. Bengio, in *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics, Sardinia, Italy, 2010*, p. 249, http://proceedings.mlr.press/v9/glorot10a.html.
[45] M. Benedetti, D. Garcia-Pintos, O. Perdomo, V. Leyton-Ortega, Y. Nam, and A. Perdomo Ortiz, npj Quantum Inf. **5**, 45 (2019).
[46] S. Khatri, R. LaRose, A. Poremba, L. Cincio, A. T. Sornborger, and P. J. Coles, Quantum **3**, 140 (2019).
[47] D. Bacon, I. L. Chuang, and A. W. Harrow, in *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '07), New Orleans, 2007* (Society for Industrial and Applied Mathematics, Philadelphia, 2007), p. 1235.
[48] H. Krovi, Quantum **3**, 122 (2019).