Experimental Estimation of Quantum State Properties from Classical Shadows

G.I. Struchalin⦾,* Ya. A. Zagorovskii⦾, E.V. Kovlakov, S.S. Straupe⦾, and S.P. Kulik
Quantum Technology Centre and Faculty of Physics, M. V. Lomonosov Moscow State University, 119991 Moscow, Russia

Full quantum tomography of high-dimensional quantum systems is experimentally infeasible due to the exponential scaling of the number of required measurements on the number of qubits in the system. However, several ideas have been proposed recently for predicting the limited number of features for these states, or estimating the expectation values of operators, without the need for full state reconstruction. These ideas go under the general name of shadow tomography. Here, we provide an experimental demonstration of property estimation based on classical shadows proposed in Huang *et al.* [Nat. Phys. 16, 1050 (2020)] and study its performance in a quantum-optical experiment with high-dimensional spatial states of photons. We show by means of experimental data how this procedure outperforms conventional state reconstruction in fidelity estimation from a limited number of measurements.

I. INTRODUCTION

A full description of a quantum system state is provided by its density matrix ρ and conventional quantum tomography aims to provide an estimate $\hat{\rho}$ of a density matrix for an unknown quantum state given the measurement data [1]. The measurements have to be tomographically complete in the sense that they should allow unambiguous determination of all density matrix elements. Simple parameter counting shows that for a general mixed state of a system in a D-dimensional Hilbert space, the required number of measurements is at least D^2 [2]. This number may be reduced to $O(RD \log^2 D)$ if some prior information about the state rank R is known, by using the techniques of compressed sensing [3]. Otherwise, one has to stick to incomplete tomography methods [4] that include advanced estimators, e.g., the maximum-likelihood maximum-entropy estimator [5], or address adaptive measurements [6,7]. For pure states, protocols requiring as few as $5D$ measurements are known [8].

In any event, the recovery of a full density matrix of an n-qubit system requires an exponentially large number of measurements since $D = 2^n$, this being known as the curse of dimensionality. One of the ways around this problem is to assume some model for the quantum state, allowing for efficient representation, such as a matrix-product state model [9,10] or a neural-network-based model [11–13]. In general, however, there may be no *a priori* reason to assign such a model to an unknown state.

On the other hand, the exponential amount of information contained in a full density matrix may be redundant. Typically, a researcher is interested in a restricted number of state properties, such as fidelity to the given state that one intends to prepare or a mean value of some observable. This fact has led to a different approach called *shadow tomography*, which has been pioneered in the work of Aaronson [14]. It promises accurate estimation of exponentially many linear functions of ρ using only a polynomial number of state copies. However, the original method from Ref. [14] is very demanding for hardware implementation, as it involves measurements that act collectively on all copies. So despite significant experimental progress in approximate quantum learning [15], direct realization of the original shadow tomography is beyond the current technology.

Fortunately, the authors of Ref. [16] have proposed another procedure that requires only separable measurements on each copy and yet is powerful in estimating an exponentially large number of state properties. Here, we report an experimental realization of this procedure, demonstrating the estimation of mean values of operators and fidelity estimation from *classical shadows* of quantum states introduced in Ref. [16]. We experimentally access Hilbert spaces of dimensionality up to 32 and clearly demonstrate that the approach is applicable in the region of incomplete measurement sets, where general assumption-free methods for full tomography fail completely.

---

*struchalin.gleb@physics.msu.ru

II. METHOD

Shadow tomography is a tool for the effective prediction of quantum state properties. Let us note that in understanding the term in this broader sense, we also refer to the protocol of Ref. [16] as shadow tomography. While it is capable of recovering both linear and higher-order polynomial target functions in matrix elements of ρ, in the present work, we focus solely on linear ones. We explicitly describe the algorithm that we use in application to our experiment. For details on the general framework, nonlinear feature prediction, and proofs of performance guarantees, see the original paper [16].

The goal of the algorithm is to predict the expectation values $\{o_i\}$ for a set of M observables $\{O_i\}$:

$$o_i(\rho) = \operatorname{Tr} O_i \rho, \quad 1 \leq i \leq M, \tag{1}$$

where ρ is an n-qubit *true state*. Obviously, $o_i(\rho)$ are linear in matrix elements of ρ.

In the data-gathering stage, ρ is transformed by a unitary operator U, $\rho \rightarrow U\rho U^\dagger$ and then each qubit is measured in a computational basis. This procedure is repeated many times for different $U \in \mathcal{U}$, chosen at random from some matrix ensemble $\mathcal{U}$. The choice of $\mathcal{U}$ affects the tomography performance and determines a class of observables O_i that can be effectively estimated. The authors of Ref. [16] mainly consider two ensembles: stabilizer circuits, i.e., U belonging to the n-qubit Clifford group [17], and Pauli measurements, where each U is a tensor product of single-qubit operations. We have selected the first option as a more extensive alternative, yet our experimental setup can carry out any measurement.

The random unitary transformation, $\rho \rightarrow U\rho U^\dagger$, followed by a measurement in a computational basis $\{|b_i\rangle\}$, is equivalent to the projection onto a random vector $|\psi_i\rangle = U^\dagger |b_i\rangle \in \mathcal{S}$. Since U is a Clifford scheme, then by definition, $|\psi\rangle$ is a random *stabilizer* state and $\mathcal{S}$ is the set of all n-qubit stabilizer states. Later, such measurements are referred to as Clifford or stabilizer measurements. We resort to vectors, rather than Clifford gates, because our experiment lacks a natural decomposition of unitary transformations into a gate sequence. The algorithm for uniform sampling of random stabilizer states $|\psi\rangle \in \mathcal{S}$, is presented in the Appendix A.

When the measurement results are obtained, the *classical shadow* $\hat{\rho}$ of the n-qubit true state ρ is calculated:

$$\hat{\rho} = (2^n + 1) \sum_{i=1}^{P} f_i |\psi_i\rangle\langle\psi_i| - \mathbb{I}, \tag{2}$$

where P is the number of projections and f_i is the observed frequency for the outcome, corresponding to $|\psi_i\rangle$, $\sum_{i=1}^{P} f_i = 1$. The expression given in Eq. (2) is simply an explicit form of a linear inverse (least-squares) estimator for any spherical 2-design positive operator-valued

measure (POVM) [18]. Our choice, i.e., stabilizer states, forms a 3-design [19] and the expression is also applicable.

We emphasize that initially, in Ref. [16], each projection is assumed to be performed for a single copy of ρ. Therefore, the number of projections P coincides with the number N of measured copies, $P = N$, ($f_i = 1/N$). On the other hand, in our quantum-optical experiment, several photons can be detected for the same $|\psi_i\rangle$ during the acquisition time, so $P < N$. Moreover, we work in the regime of overexposure, for which $P \ll N$ (typically, $N/P \sim 10^4$–10^5, depending on the system dimensionality). This setting is common in compressive sensing experiments, where shot noise in the outcome probability estimation should be diminished [6,20,21]. Preliminary tomography simulations show that feature prediction accuracy is limited by the finite P even though $N = \infty$ (the observed frequency f_i is substituted for the exact outcome probability). In this sense, P is more important than N. When $P < N$, at least, P copies are measured with dissimilar projectors, so the theorems presented in Ref. [16] remain valid if N is replaced by P. However, the theorem statements can become pessimistic and the proofs may require further justification for the case $P < N$.

Once a classical shadow [see Eq. (2)] is obtained, an estimator $\hat{o}_i$ of o_i is simply

$$\hat{o}_i = \operatorname{Tr} O_i \hat{\rho}, \quad 1 \leq i \leq M. \tag{3}$$

Herein lies another discrepancy with the original algorithm: the authors of Ref. [16] propose the use of a *median-of-means* estimator. However, we omit the median evaluation and use a simple mean estimator throughout this work, because no valuable difference is found between the two approaches [22].

When $P = N$, the shadow-tomography protocol has the following sampling complexity [16]:

Theorem 1. *N stabilizer measurements suffice to predict M expectations $o_i = \operatorname{Tr} O_i \rho, 1 \leq i \leq M$, within an additive error ϵ, given that*

$$N \geq \mathcal{O}\left(\frac{\log M}{\epsilon^2} \max_i \operatorname{Tr} O_i^2\right). \tag{4}$$

Note that the number of copies N depends on the target operators O_i via $\operatorname{Tr} O_i^2$. This relation can lead to implicit dependence on the system dimension D for specific operators O_i; therefore, such operators cannot be efficiently estimated. For example, consider the estimation of the tensor products of the Pauli operators $O_i = \sigma_{i,1} \otimes \cdots \otimes \sigma_{i,n}$, where $\sigma_{i,j}$ is a single-qubit Pauli matrix. Since $\sigma_{i,j}^2 = \mathbb{I}$, then in this case $\operatorname{Tr} O_i^2 = D$. Nevertheless, in our experiments, we use rank-1 projectors, $\operatorname{Tr} O_i^2 = 1$, and this factor vanishes from Eq. (4).

III. EXPERIMENT

We use spatial degrees of freedom of photons to produce high-dimensional quantum states. The corresponding continuous Hilbert space is typically discretized using the basis of transverse modes. We choose Hermite-Gaussian (HG) modes $HG_{nm}(x,y)$, which are the solutions of the Helmholtz equation in Cartesian coordinates (x,y), and form a complete orthonormal basis. The *mode order k* is defined as a sum of mode indices: $k = n + m$. There exist $(k+1)(k+2)/2$ HG modes from zero to the kth order inclusive. We bound the beam order to prepare a D-dimensional system, i.e., the order k is limited by k_{max}, $k \leq k_{max}$, where k_{max} is the minimal integer fulfilling the inequality $(k_{max} + 1)(k_{max} + 2)/2 \geq D$. We test shadow tomography for dimensions $D = 2, 4, 8, 16$, and 32, which corresponds to 1–5 qubits.

In our setup (Fig. 1), attenuated light from an 808-nm diode laser is spatially filtered by a single-mode fiber (SMF-1) and collimated by an aspheric lens L2. The top half of a spatial light modulator (SLM, Holoeye Pluto) serves to prepare the desired state of the photon, the bottom half follows by focusing into a single-mode fiber (SMF-2), and single-photon detection implements projective measurements [23,24]. Lenses L3 and L4 have equal focal lengths $F = 100$ mm and are mounted 200 mm apart. Since holograms displayed on the SLM use a blazed grating for amplitude modulation [25], the pinhole in the focal plane is used for state selection in the first diffraction order. After a double pass through a telescope and a quarter-wave plate (QWP), the beam is reflected by a polarizing beam splitter (PBS) and directed back to the SLM without any additional alterations.

Note that the detected state differs from the prepared one due to the Gouy-phase incursion during beam propagation from one half of the SLM to another. The Gouy phase ϕ_G depends solely on the geometric parameters of the experimental setup, e.g., the beam Rayleigh range and traveling distance. It causes the following transformation of basis states: $|HG_{nm}\rangle \rightarrow e^{i(n+m+1)\phi_G}|HG_{nm}\rangle$. We use the Gouy phase as a fitting parameter to determine the "true" state.

IV. RESULTS

A. Correlation analysis

The expectations o_i can be estimated by shadow tomography via Eq. (3). On the other hand, the expression given in Eq. (1) has the form similar to Born's rule, so quantities o_i can be measured directly. It provides a method for independent experimental verification of shadow-tomography predictions. We denote the estimates given by shadow tomography as $\hat{o}_i^{est.}$ and the directly measured expectations as $\hat{o}_i^{meas.}$. These values are both subject to experimental imperfections and shot noise due to the finite statistics N. However, the latter factor is negligible, since in all experiments the total exposure corresponding to the value $o_i = 1$ is approximately 3×10^5 photons, with proportional scaling for other values of o_i.

First, we perform 10^4 stabilizer measurements to obtain the classical shadow $\hat{\rho}$. Then, 5000 projectors $O_i = |\phi_i\rangle\langle\phi_i|$ onto random Haar-distributed vectors $|\phi_i\rangle$ are measured, resulting in an array of $\hat{o}_i^{meas.}$. For the same operators O_i, we calculate the predictions $\hat{o}_i^{est.}$ using the classical shadow and plot them against $\hat{o}_i^{meas.}$. For each investigated dimension $D = 2^n, n = 1, \ldots, 5$, we probe five different Haar-distributed random pure true states to ensure that the procedure is a state-agnostic one. We observe a high Pearson correlation coefficient between the two quantities in all scenarios, manifesting the consistency of shadow tomography (see Table I).

A typical correlation plot is depicted in Fig. 2 for system dimension $D = 8$. The solid black line shows perfect matching—the dependence $\hat{o}_i^{est.} = \hat{o}_i^{meas.}$. As one can see, all points tend to concentrate near this line (the Pearson correlation coefficient is $r = 0.9758$). Note the existence of a small "nonphysical" region, where $\hat{o}_i^{est.} < 0$. It appears because the classical shadow $\hat{\rho}$ is not forced to be positive semidefinite as in conventional tomography, such as
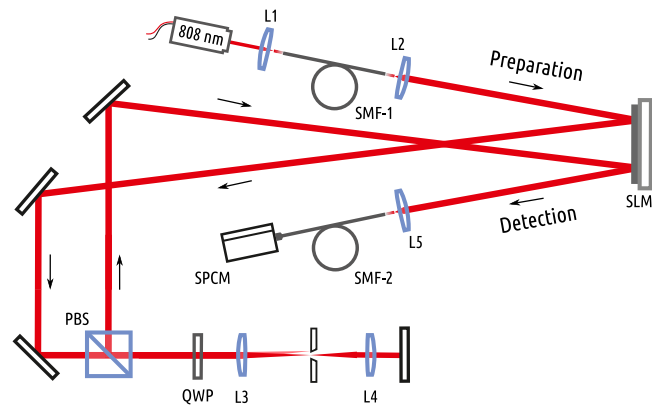

FIG. 1. The experimental setup. A spatial light modulator is used for preparation and projective measurements of arbitrary spatial states of photons in a basis of Hermite-Gaussian modes of dimensionality up to 32 (see text for details).

TABLE I. The Pearson correlation coefficient r and compensated preparation fidelity F, averaged over five random states, for different system dimensions D.

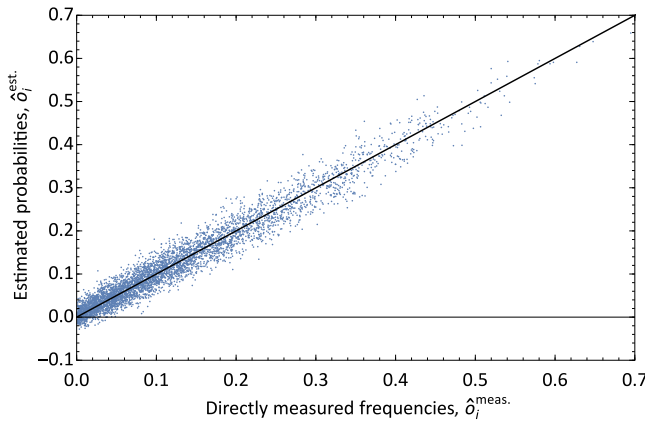| D	r	F
2	0.989 ± 0.002	0.981 ± 0.013
4	0.983 ± 0.001	0.974 ± 0.011
8	0.976 ± 0.002	0.899 ± 0.009
16	0.953 ± 0.003	0.920 ± 0.020
32	0.875 ± 0.006	0.807 ± 0.031

FIG. 2. A typical correlation plot (system dimension $D = 8$). The prediction of the operator mean values $\hat{o}_i^{\text{est.}}$ using shadow tomography versus directly measured quantities $\hat{o}_i^{\text{meas.}}$ is depicted. The solid black line corresponds to the equality $\hat{o}_i^{\text{est.}} = \hat{o}_i^{\text{meas.}}$.
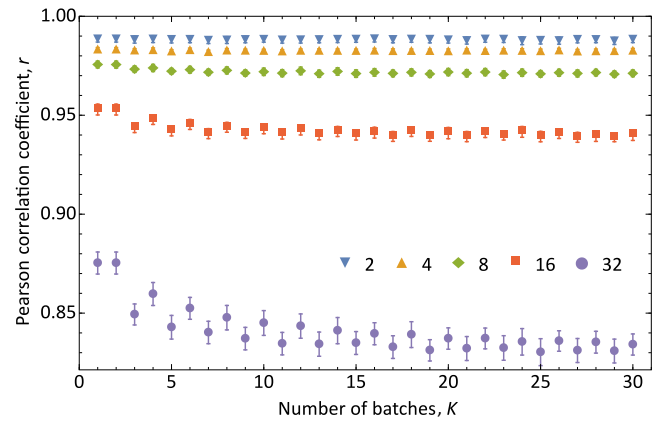

FIG. 3. The dependence of the Pearson correlation coefficient r between $\hat{o}_i^{\text{est.}}$ and $\hat{o}_i^{\text{meas.}}$ on the number of batches K in the median-of-means evaluation for different system dimensions D (see legends). Each data point is averaged over five true states. The error bars correspond to one standard deviation of the mean.

maximum-likelihood estimation (MLE). And, indeed, $\hat{\rho}$ contains negative eigenvalues due to experimental imperfections.

The visible shift of data points toward zero in Fig. 2 is a mere artifact of our choice for $O_i = |\phi_i\rangle\langle\phi_i|$. The quantity $o_i = \text{Tr}(O_i|\psi\rangle\langle\psi|) = |\langle\psi|\phi_i\rangle|^2$ can be treated as a random variable because $|\phi_i\rangle$ is a random Haar-distributed vector and $|\psi\rangle$ is the fixed true state. The probability density function $p_{o_i}(x)$ of such a squared dot product is known: $p_{o_i}(x) = (D-1)(1-x)^{D-2}$ [26]. As the dimensionality D increases, the mean value $\langle x \rangle = 1/D$ decreases. Both $\hat{o}_i^{\text{est.}}$ and $\hat{o}_i^{\text{meas.}}$ also concentrate near the axis origin, since they are estimates of o_i.

B. Effect of median-of-means estimator

It is said earlier that the authors of Ref. [16] suggest the use of the median-of-means estimator [27], which proceeds as follows:

(1) A sequence of P measurement results is divided into K batches of length $\lfloor P/K \rfloor$.
(2) An individual shadow $\hat{\rho}_k$ is calculated for each batch with index $k = 1, \ldots, K$, analogously to Eq. (2).
(3) A final assessment $\hat{o}_i$ is the median

$$\hat{o}_i = \text{median}(\text{Tr}\, O_i\hat{\rho}_1, \ldots, \text{Tr}\, O_i\hat{\rho}_K). \quad (5)$$

The median-of-means estimator is robust against outliers in the measured data. The number of batches K depends on the number of target operators M and the confidence probability $1 - \delta$: $K = 2\log(2M/\delta)$. For example, if the failure level is chosen to be $\delta = 0.01$ and $M = 5000$, the number of batches is $K \approx 28$.

In order to investigate how the number of batches K influences the overall tomography performance, we find

the median-of-means predictions $\hat{o}_i^{\text{est.}}$ for various K and calculate the Pearson correlation coefficient r between $\hat{o}_i^{\text{est.}}$ and $\hat{o}_i^{\text{meas.}}$. The obtained dependencies $r(K)$ are presented in Fig. 3 for different system dimensions D. Each curve is averaged over five tomography runs. The case $K = 1$ corresponds to the ordinary mean estimator, as used before. It can be seen that the dependencies $r(K)$ are almost flat and that the correlation even becomes slightly lower with the increase of K. This implies that in application to our experiment, the effect of the median-of-means estimator is negligible compared to the mean alone.

We connect the loose dependence of the correlation coefficient r on K with two facts. First, the statistics per measurement in our experiments are huge and the outliers hardly occur (for more detailed reasoning, see Appendix B). Second, systematic, deterministic errors in the measurement projectors dominate the statistical noise and medians cannot smooth away this source of imperfections.

C. Fidelity estimation

One of the important use cases for shadow tomography is the estimation of fidelity to some given pure state $|\psi\rangle$. In this case, the target operator O is simply a projector onto this state: $O = |\psi\rangle\langle\psi|$. In particular, one can find the fidelity of the state preparation. However, in our experiment, the prepared state $|\psi_{\text{prep.}}\rangle$ and the detected one $|\psi_{\text{det.}}\rangle$ differ significantly due to the Gouy-phase incursion during the beam propagation and we have to perform the corresponding correction (see Appendix C for details).

The obtained fidelities F of state preparation are listed in Table I. As one can see, F gradually decreases with the system dimension. This behavior is due to experimental imperfections and is not a limitation of shadow tomography. The degradation of fidelity is typical for experiments
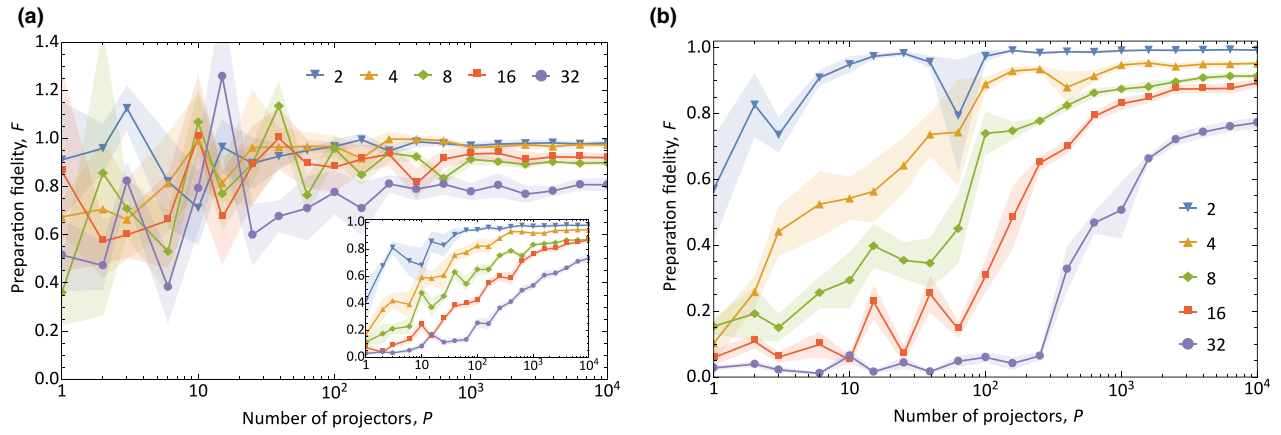
FIG. 4. The compensated preparation fidelity F on the number of stabilizer measurements P for different system dimensions D (see legends) obtained using (a) shadow tomography and (b) maximum-likelihood estimation (MLE). Each curve is averaged over five true states. The shaded area corresponds to one standard deviation of the mean. The inset of Fig. 4(a) shows the same dependencies but the classical shadow $\hat{\rho}$ is projected onto the set of physical density matrices.

with high-dimensional spatial states of light and has been observed before [23,25]. The state preparation and detection qualities may suffer for many reasons: optical-system aberrations, nonprecise alignment of digital holograms with respect to the incident beams, SLM display pixelation, and nonflatness. All of these factors become crucial with the increase of the beam order since the corresponding wave-front structure gets more complicated.

The results presented above are obtained using overcomplete measurement sets, since we use $P = 10^4$ projectors to construct the classical shadow $\hat{\rho}$. This number is far greater than the size of a minimal complete set, which has $D^2 - 1$ POVM elements, even for $D = 32$. However, the main distinguishing feature of shadow tomography is its ability to predict expectation values using much less than a tomographically complete set of measurements. Hence, we also study the performance of shadow tomography for the intermediate values of P, including the incomplete scenario, where $P < D^2 - 1$.

Figure 4(a) shows the averaged dependencies of the preparation fidelity F, estimated using shadow tomography, on the number of stabilizer measurements P for various system dimensions D. The fidelity is calculated with respect to the compensated prepared state, where the compensatory Gouy phase is found using the full data sequence (i.e., for $P = 10^4$). The averaging is done over five different states for each dimension.

In the beginning, for low P, the volatility of the curves is vast and the fidelity F can even lie outside the physical region $0 \leq F \leq 1$ due to the negative definiteness of a shadow matrix $\hat{\rho}$. As P increases, the fidelities start to stabilize near their final values. Nevertheless, the fidelity estimators are unbiased for any number of projectors P because shadow tomography is based on linear inversion, which is unbiased. Indeed, as one can see from Fig. 4(a),

the error bars cover the final values of the fidelity reasonably well for any P, which experimentally confirms the unbiasedness property.

It is interesting to see how the above fidelity estimates change if the shadow matrix $\hat{\rho}$ [see Eq. (2)] is forced to be positive semidefinite. To achieve this, we project the eigenvalues λ_i of $\hat{\rho}$ onto a canonical simplex $\Delta = \{(\lambda_1, \ldots, \lambda_D) \mid \lambda_i \geq 0 \wedge \sum_{i=1}^{D} \lambda_i = 1\}$, using the recipe from Ref. [28], while leaving the eigenvectors untouched. The obtained results are shown in the inset of Fig. 4(a). Now the estimators are biased: for incomplete measurement sets, $P \lesssim D^2$, the fidelity is underestimated and significantly shifted toward zero. When P becomes equal in order of magnitude to D^2, the assessments attain their final values. Note the apparent dependency on the system dimension D, which is not the case for ordinary shadow tomography.

The bias of the estimator leads to poor accuracy when a measurement set is incomplete. For example, consider the point with $D = 32$ and $P = 251$ in Fig. 4(a). Shadow tomography has already converged since the fidelity is $F = 0.81 \pm 0.04$, which coincides with the final value for $P = 10^4$ within the error bars, but after the projection of eigenvalues onto the positive simplex, the fidelity drops to $F = 0.36 \pm 0.03$. Unfortunately, the bias is unavoidable for any procedure that always yields positive density matrices [29]. An MLE is not an exclusion either. Figure 4(b) presents fidelity dependencies for the same measurement data, processed with an accelerated projective gradient MLE [30]. Qualitatively, the performance is the same as for the inset of Fig. 4(a).

D. Estimator biasedness

Preparation fidelity is not the only quantity estimated with heavy bias employing the MLE method when the
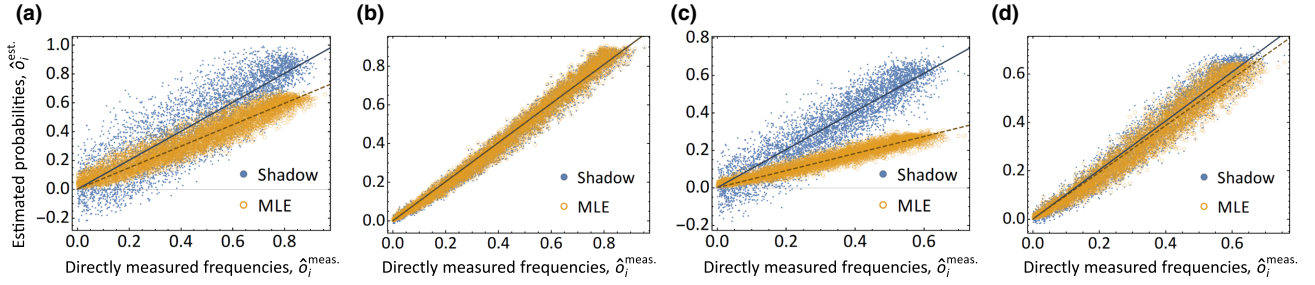
FIG. 5.   A comparison of the correlation plots obtained using shadow tomography and MLE for different system dimensions D and numbers of stabilizer measurements P. The prediction of the operator mean values $\hat{o}_i^{\text{est.}}$ using tomographic methods versus directly measured quantities $\hat{o}_i^{\text{meas.}}$ is depicted. The straight lines are best-fit dependencies of the form $\hat{o}_i^{\text{est.}} = \beta \hat{o}_i^{\text{meas.}}$ (solid lines, shadow tomography; dashed lines, MLE). For overcomplete numbers of measurements $P = 10^4$ [Figs. 5(b) and 5(d)], both methods result in the same unbiased predictions with a proportionality coefficient $\beta \approx 1$. For low values of P [Figs. 5(a) and 5(c)], the MLE predictions are highly biased and underestimate $\hat{o}_i^{\text{meas.}}$, while the classical shadow assessments are still unbiased. (a) $D = 8, P = 100$; (b) $D = 8, P = 10^4$; (c) $D = 32, P = 300$; (d) $D = 32, P = 10^4$.

number P of stabilizer measurements is low. All projectors O_i with the near-unity mean value $o_i \approx 1$ will be underestimated. To check this hypothesis, we carry out another correlationlike test, similar to those in Fig. 2. Both a classical shadow $\hat{\rho}_{\text{CS}}$ and an MLE $\hat{\rho}_{\text{MLE}}$ are calculated using the same stabilizer-measurement outcomes. Then, as usual, these estimators are substituted into Eq. (3) to give $\hat{o}_i^{\text{est.}}$ for a set of 5000 randomly chosen projectors O_i.

We note that the difference between the shadow and the MLE tomography is most visible in the region, where $o_i \approx 1$. At the same time, Haar-distributed projectors O_i tend to have low mean values o_i (on average, $\langle o_i \rangle = 1/D$), which are not well suited for this kind of test. Therefore, we select random projectors O_i with uniformly distributed expectations o_i. To do so, they should be adjusted to the true state. In particular, we use projectors $O_i = |\phi\rangle\langle\phi|$ onto a random vector $|\phi\rangle$:

$$|\phi\rangle = \sqrt{a}|\psi\rangle + \sqrt{1-a}\frac{|g\rangle - |\psi\rangle\langle\psi|g\rangle}{\||g\rangle - |\psi\rangle\langle\psi|g\rangle\|}, \qquad (6)$$

where $|g\rangle$ is a vector with the real and imaginary parts of its elements being independent Gaussian random variables with zero mean and unit variance and a is distributed uniformly on the interval $[0, 1]$. It is easy to verify that $|\langle\psi|\phi\rangle|^2 = a$, so if $|\psi\rangle$ is the true state, then, indeed, $o_i = a$

has a uniform distribution. We take a close approximation—a compensated prepared state—as the vector $|\psi\rangle$. The choice of distribution for $|g\rangle$ ensures that a "circle" determined by the equation $|\langle\psi|\phi\rangle|^2 = \text{const}$ is also populated uniformly [[31], Appendix C].

The obtained predictions $\hat{o}_i^{\text{est.}}$ against the directly measured mean values $\hat{o}_i^{\text{meas.}}$ are shown in Fig. 5 for system dimensions $D = 8$ and 32. We investigate two cases: estimates for a small number of measurements P (100 for $D = 8$ and 300 for $D = 32$) and large $P = 10^4$. As expected, shadow tomography gives unbiased estimates in all situations: $\hat{o}_i^{\text{est.}} \approx \hat{o}_i^{\text{meas.}}$. MLE performs differently, since it is only an *asymptotically* unbiased estimator. For small P, although the predictions are more condensed compared to classical shadows (there is less volatility), they are underestimated and they concentrate near a line $\hat{o}_i^{\text{est.}} = \beta \hat{o}_i^{\text{meas.}}$ with proportionality constant $\beta < 1$ (see Table II for the best-fit parameters). For large P, the behavior equalizes: MLE approaches the asymptotic and produces unbiased estimates that almost coincide with those calculated using classical shadows. We connect the observed flat-top cutoff under $o_i^{\text{est.}} = 1$ in Figs. 5(b) and 5(d) with the fact that our choice of $|\psi\rangle$ in Eq. (6) differs from the true state ρ.

V. CONCLUSION

We demonstrate experimentally that classical shadows, i.e., linear inversion estimators for quantum states, can be used to faithfully predict expectation values of observables from very few measurements. Specifically, we have shown that the estimator obtained from the classical shadow is unbiased and provides correct expectation values even when the number of measurements used for estimation is significantly less than that required for full state reconstruction. As a special case, we perform estimation of the fidelity with the "true" state and show that it is also possible with few measurements.

TABLE II.   The Pearson correlation coefficient r and proportionality coefficient β of the data in Fig. 5 obtained using classical shadows (CS) and MLE for different system dimensions D and numbers of stabilizer measurements P.

D	P	r_{CS}	r_{MLE}	β_{CS}	β_{MLE}
8	100	0.870	0.949	1.004 ± 0.004	0.745 ± 0.002
8	10^4	0.990	0.990	1.011 ± 0.001	1.010 ± 0.001
32	300	0.915	0.957	1.016 ± 0.003	0.455 ± 0.001
32	10^4	0.971	0.974	1.013 ± 0.002	0.967 ± 0.002

Our treatment reformulates the results of Ref. [16] in terms of a typical quantum-optical experiment and is then applied to experimental data for high-dimensional spatial states of photons. The versatility of the chosen experimental platform allows us to realize arbitrary projective measurements; however, we demonstrate that, in full accordance with the theoretical predictions, the procedure works well when the measurement set is restricted, for example, to projections on the stabilizer states. This is an important feature of the protocol, making it a scalable approach to quantum property estimation.

Shadow tomography is an inherently restricted method—it estimates only properties of the state, not the state itself. However, in the case of high dimensionality it may be the method of choice for an experimenter interested in estimating a limited set of observables and properties of the state. If, however, more information about the state is required, e.g., to understand the sources of decoherence in the system, other methods fit better if they are still feasible in terms of the required number of measurements.

The framework of shadow tomography has recently been extended with online learning protocols [32,33], which, from an operational point of view, are close in spirit to the one implemented in this work. Comparing the performance of these approaches on real experimental data is an interesting direction for further research.

ACKNOWLEDGMENTS

APPENDIX A: EXPLICIT PROCEDURE FOR THE GENERATION OF RANDOM STABILIZER STATES

In this section, we describe the procedure for the explicit generation of random uniformly distributed stabilizer states. By "explicit," we mean that the whole n-qubit state vector $|\psi\rangle$ of 2^n amplitudes is calculated. This requirement comes from the fact that in photonic experiments such as the one performed here, the preparation and measurement stage has no natural decomposition in terms of quantum gates and requires the explicit specification of the state vectors.

The set $\mathcal{S}$ of all stabilizer states is finite—its cardinality $C(n)$ is [34]:

$$C(n) = 2^n \prod_{k=1}^{n} (2^k + 1) \approx 2^{n^2/2}. \tag{A1}$$

Uniform sampling means that each state $|\psi_i\rangle \in \mathcal{S}$ is selected with equal probability. A naive approach would be to generate a random index $i = 1, \ldots, C(n)$, and pick the corresponding state $|\psi_i\rangle$ from a pregenerated set $\mathcal{S}$. However, the huge cardinality makes this infeasible.

When working with stabilizer states on a classical computer, one usually resorts to their stabilizer operators rather than vectors, since this implicit description allows a very efficient (polynomial in the number of qubits n) storage scheme and the simulation of Clifford gate actions. This fact is known as the Gottesman-Knill Theorem [34,35].

Therefore, an evident practical approach for constructing a random $|\psi\rangle$ is to generate a set of its stabilizers $\{g_i\}_{i=1}^{n} : g_i |\psi\rangle = |\psi\rangle$. This can be done efficiently by utilizing, e.g., a method from Ref. [36], which enumerates all possible stabilizer circuits $U: |\psi\rangle = U|0\rangle$. Then, given the stabilizers $\{g_i\}$, the state $|\psi\rangle$ is obtained using the relation

$$|\psi\rangle\langle\psi| = \prod_{i=1}^{n} \frac{1 + g_i}{2}. \tag{A2}$$

Thus, the conversion from a stabilizer formalism to an explicit form involves three exponentially hard routines:

(1) An explicit construction of stabilizer matrices g_i—$\mathcal{O}(n \cdot 2^{2n})$ operations
(2) Product evaluation—$\mathcal{O}(n \cdot 2^{3n})$
(3) Recovery of $|\psi\rangle$ from $|\psi\rangle\langle\psi|$—$\mathcal{O}(2^n)$

The overall complexity is dominated by the second stage (note the power $3n$).

Of course, the complexity of an explicit n-qubit state generation is always exponential and cannot be lower than $O(2^n)$—the number of operations required to address every element in the vector. But the power index dramatically affects the performance. For the above method, it is $3n$, while a decrease to n is possible. Below, we describe an algorithm that requires $\mathcal{O}[2^n \text{poly}(n)]$ operations.

Let us start with the universal form of any $|\psi\rangle \in \mathcal{S}$ [37,38]:

$$|\psi\rangle \propto \sum_{x \in \mathbb{F}_2^k} (-1)^{q(x)} i^{l(x)} |Rx + t\rangle, \tag{A3}$$

where $x \in \mathbb{F}_2^k$ and $t \in \mathbb{F}_2^n$ are, respectively, k- and n-dimensional binary vectors, $k \leq n$, $q(x)$ is a quadratic form on $\mathbb{F}_2^k$, $l(x)$ is a linear one, and $R \in \mathbb{F}_2^{n \times k}$ is an $n \times k$ binary matrix with rank k. Summation and multiplication in Eq. (A3) are carried modulo 2, since we work in a Galois field $\mathbb{F}_2$. Also, we identify a binary representation of a given integer number x with the corresponding binary vector and vice versa.

The representation given in Eq. (A3) reveals some properties of stabilizer states. Up to normalization, each element of the state can be ± 1, $\pm i$, or 0. The number of

nonzero elements is always 2^k, $0 \leq k \leq n$, which is simply the number of different vectors x in $\mathbb{F}_2^k$. By convention, we define $\mathbb{F}_2^0 = \{0\}$.

In our sampling algorithm, the set $\mathcal{S}_k$, which by definition contains all n-qubit stabilizer states with 2^k nonzero elements, plays an important role. In Theorem 2, we show how to sample $|\psi\rangle \in \mathcal{S}_k$ uniformly. Then, in Theorem 3, we calculate the cardinality $C(n,k)$ of $\mathcal{S}_k$. Finally, we combine these results in Theorem 4, where an algorithm for uniform sampling of the whole set $\mathcal{S} = \bigcup_{k=0}^n \mathcal{S}_k$ is presented.

Theorem 2. *Fix $k = 0, \ldots, n$. Let $|\psi\rangle$ be an n-qubit state of the form*

$$|\psi\rangle := |t\rangle, \quad if\ k = 0,$$

$$|\psi\rangle := \frac{1}{2^{k/2}} \sum_{x=0}^{2^k-1} (-1)^{x^T Q x} i^{c^T x} |Rx + t\rangle, \quad if\ k \neq 0, \quad (A4)$$

where $x \in \mathbb{F}_2^k$. Quantities $Q \in \mathbb{F}_2^{k \times k}$, $c \in \mathbb{F}_2^k$, $t \in \mathbb{F}_2^n$ are random with independent and identically distributed (i.i.d.) elements 0 or 1 appearing with probability $1/2$. $R \in \mathbb{F}_2^{n \times k}$ (rank $R = k$) is a random matrix sampled uniformly from the set of all rank-k matrices. Then $|\psi\rangle$ is uniformly sampled from $\mathcal{S}_k$.

Proof. Again, consider Eq. (A3). Every quadratic form $q(x)$ can be expressed as a sum: $q(x) = x^T Q x + b^T x + x_0$. The constant x_0 affects only the global phase of $|\psi\rangle$ and can be omitted. The linear term $b^T x$ is already enclosed in $x^T Q x$. Indeed, in the expansion of $x^T Q x$, there is a diagonal term $Q_{ii} x_i^2 = Q_{ii} x_i$, since $x_i^2 = x_i$ for any $x_i \in \mathbb{F}_2$. Therefore, without loss of generality, $q(x) = x^T Q x$, where Q is an arbitrary $k \times k$ binary matrix. Analogously, a constant term can be neglected in the linear form: $l(x) = c^T x + x_0 \sim c^T x$, where $c \in \mathbb{F}_2^k$ is an arbitrary binary vector. Starting from the form given in Eq. (A3), we already arrive at a more specific expression, as shown in Eq. (A4).

Let us prove that $|\psi\rangle$ is sampled uniformly. Quantities Q, c, R, and t are associated with their own structures—respectively, a quadratic form $\mathcal{Q}$, a linear form L, a k-dimensional vector subspace V_k, and an affine subspace A_k:

$$\mathcal{Q} = f_Q(Q) = \{(x, x^T Q x) \mid x \in \mathbb{F}_2^k\}, \quad (A5)$$

$$L = f_c(c) = \{(x, c^T x) \mid x \in \mathbb{F}_2^k\}, \quad (A6)$$

$$V_k = f_R(R) = \{Rx \mid x \in \mathbb{F}_2^k\}, \quad (A7)$$

$$A_k = f_t(t) = \{y + t \mid y \in V_k\}. \quad (A8)$$

The corresponding maps f_Q, f_c, f_R, and f_t are in general surjective, i.e., they may map many different quantities to a single structure. An affine subspace A_k determines the positions of the nonzero elements in $|\psi\rangle$, while forms $\mathcal{Q}$ and L define the order in which $\pm 1, \pm i$ appear. In this sense, $\mathcal{Q}, L$, and A_k act independently, so the state given in Eq. (A4) is uniformly distributed if each of these three structures is sampled uniformly.

Obviously, Q, c, and t are generated uniformly, because their elements are i.i.d. random variables with prob(0) = prob(1) = 1/2; R is sampled uniformly as the theorem condition states. However, for a general surjective map, $\omega = f(\xi), \xi \in \Xi, \omega \in \Omega$, a uniform sampling of the domain Ξ does not imply the same for its image Ω. Fortunately, for the maps given in Eqs. (A5)–(A8), the cardinality of a preimage for each element in a codomain is the same (see below):

$$|f^{-1}(\omega_1)| = |f^{-1}(\omega_2)|, \quad \forall \omega_{1,2} \in \Omega, \quad (A9)$$

where $|\cdot|$ denotes the cardinality evaluation. Any map f satisfying Eq. (A9) has the property that a uniform sampling of the domain Ξ results in a uniform sampling of its codomain Ω.

Let us start by proving that the map f_Q complies with Eq. (A9). In the expression $x^T Q x$, only the sum $(Q_{ij} + Q_{ji})x_i x_j, i < j$, matters . So adjoint nondiagonal elements Q_{ij} and Q_{ji} can be replaced by their equivalents: a pair $(Q_{ij}, Q_{ji}) = (0,1)$ is equivalent to $(1,0)$ and $(0,0) \sim (1,1)$. There are $k(k-1)/2$ such pairs in a $k \times k$ matrix Q and each pair has two equivalent values. Therefore, for every $\mathcal{Q}$, $|f_Q^{-1}(\mathcal{Q})| = 2^{k(k-1)/2}$.

There is a one-to-one correspondence between vectors c and linear forms L, so the map f_c is bijective and $|f_c^{-1}(L)| = 1$.

Any nondegenerate $n \times k$ matrix R defines some basis of a k-dimensional vector subspace V_k of a vector space V_n and vice versa. The number of different bases in a given subspace V_k depends solely on the dimension k, not on the contents of V_k. On the other hand, the number of bases is equal to $|f_R^{-1}(V_k)|$, so the condition given in Eq. (A9) holds for the map f_R.

Affine subspaces $f_t(t_1)$ and $f_t(t_2)$ [see Eq. (A8)], for different $t_1, t_2 \in \mathbb{F}_2^n, t_1 \neq t_2$, either coincide or do not intersect. Indeed, suppose that there is partial intersection and take $z \in f_t(t_1) \cap f_t(t_2)$; then $z = y_1 + t_1 = y_2 + t_2, y_1, y_2 \in V_k$. Consequently, $t_2 = t_1 + y_1 - y_2 = t_1 + \Delta y, \Delta y \in V_k$ and thus $f_t(t_2) \subset f_t(t_1)$. Analogously, one can prove that $f_t(t_1) \subset f_t(t_2)$. These two mutual inclusions mean that affine subspaces coincide, $f_t(t_1) = f_t(t_2)$, which contradicts the initial assumption.

It also follows from the above reasoning that two affine subspaces coincide, if and only if $t_2 = t_1 + y$, where $y \in V_k$. Therefore, for each $t_1 \in \mathbb{F}_2^n$, there exist 2^k vectors t_2 that result in the same affine subspace. So, cardinality

$|f_t^{-1}(A_k)| = 2^k$ is the same for all A_k and the condition given in Eq. (A9) is satisfied.

We have shown that affine subspaces A_k, viewed as a shift of the *fixed* vector subspace V_k by a random vector t, are uniformly distributed. Because, as we proved earlier, vector subspaces are also sampled uniformly, the set of *all* affine subspaces is sampled uniformly.

By pointing out that all necessary structures, namely, $\mathcal{Q}$, L, and A_k, are uniformly distributed, we complete the proof. ∎

Theorem 2 does not tell anything about how to sample matrices R. In our implementation, we use the simplest possible method. First, fill $n \times k$ matrix R with random bits, where $\text{prob}(0) = \text{prob}(1) = 1/2$, and compute the matrix rank over $\mathbb{F}_2$. If rank $R = k$, then stop; otherwise, repeat the procedure. We have taken a routine for the matrix rank calculation that requires $\mathcal{O}(n^2 k)$ operations.

Theorem 3. *The cardinality of $\mathcal{S}_k$ is*

$$C(n, k) = 2^{n + (k(k+1))/2} \binom{n}{k}_2, \qquad \text{(A10)}$$

where

$$\binom{n}{k}_2 = \prod_{j=0}^{k-1} \frac{2^n - 2^j}{2^k - 2^j}, \qquad \text{(A11)}$$

is a 2-binomial (Gaussian) coefficient.

Proof. As in the proof of Theorem 2, we can divide the evaluation of $C(n, k)$ by counting all distinct quadratic forms $\mathcal{Q}$ [see Eq. (A5)], linear forms L [see Eq. (A6)], and affine subspaces A_k [see Eq. (A8)] and multiplying the results.

The whole set of matrices $Q \in \mathbb{F}_2^{k \times k}$ has cardinality 2^{k^2}. But it is divided into groups of $2^{k(k-1)/2}$ matrices, where each group corresponds to the same quadratic form $\mathcal{Q}$. Therefore, the total number of distinct forms is given by the ratio of these quantities and is equal to $C_{\text{quad.}} = 2^{k(k+1)/2}$.

There are $C_{\text{lin.}} = 2^k$ different possible linear forms over $\mathbb{F}_2^k$.

The total number of k-dimensional vector subspaces V_k of an n-dimensional vector space over $\mathbb{F}_2$ is equal to a 2-binomial coefficient $\binom{n}{k}_2$ [39]. Each of these subspaces can be shifted in 2^k ways (by adding a vector $t \in V_k$; see proof of Theorem 2), resulting in the same affine subspace A_k. This gives $2^n/2^k$ different cosets $V_k + t$. Therefore, the total number of affine subspaces A_k is equal to the product $C_{\text{aff.}} = 2^{n-k} \binom{n}{k}_2$.

By multiplying the numbers $C_{\text{quad.}}$, $C_{\text{lin.}}$, and $C_{\text{aff.}}$, we obtain the expression given in Eq. (A10). ∎

Using the q-binomial Theorem [39], it is easy to check that, indeed, $\sum_{k=0}^{n} C(n, k) = C(n)$ [see Eq. (A1)].

Theorem 4. *Choose integer k at random with probability* $\text{prob}(k) = C(n, k)/C(n)$, $0 \leq k \leq n$, *and generate* $|\psi\rangle \in \mathcal{S}_k$ *according to Theorem 2. Then* $|\psi\rangle$ *is uniformly sampled from* $\mathcal{S}$.

Proof. The index k determines the set $\mathcal{S}_k$ to be sampled; hence, $\text{prob}(\psi \in \mathcal{S}_k) = \text{prob}(k)$. Conditional probability of sampling $|\psi\rangle$ from $\mathcal{S}_k$ is, $\text{prob}(\psi \mid \psi \in \mathcal{S}_k) = 1/C(n, k)$, because the algorithm from Theorem 2 produces uniformly distributed states. Then, the overall probability of obtaining the state $\psi \in \mathcal{S}$ is equal to

$$\text{prob}(\psi) = \text{prob}(\psi \mid \psi \in \mathcal{S}_k)\, \text{prob}(\psi \in \mathcal{S}_k) = \frac{1}{C(n)}. \qquad \text{(A12)}$$

Every state is produced with the same probability; therefore, the sampling is uniform. ∎

The overall complexity of the procedure from Theorem 4 is $\mathcal{O}[2^n \text{poly}(n)]$, because there are 2^n elements in $|\psi\rangle$ [see Eq. (A4)] and each element evaluation requires no more than $\text{poly}(n)$ operations. Actually, in our implementation $\text{poly}(n) = \mathcal{O}(n^3)$, since the complexity is dominated by the calculation of rank R.

We provide the PYTHON code for the explicit sampling of random stabilizer states, which is available at GitHub [40].

APPENDIX B: MEDIAN-OF-MEANS ESTIMATOR

Median-of-means estimation [27] is an enhancement over an empirical mean estimator that is robust against outlier corruption. Consider N samples $x_1, \ldots, x_N$ of a random variable x. The empirical mean $\hat{x}$ is defined as

$$\hat{x} = \frac{1}{N} \sum_i x_i. \qquad \text{(B1)}$$

According to the Chebyshev inequality, $\hat{x}$ deviates from the expectation $\mathbb{E}x$ by more than ϵ with probability at most δ:

$$\text{prob}(|\hat{x} - \mathbb{E}x| \geq \varepsilon) \leq \delta = \frac{\text{Var}\,x}{N\varepsilon^2}, \qquad \text{(B2)}$$

where $\text{Var}\,x$ denotes the variance of x. Therefore, the sampling complexity for a mean estimator is

$$N = \frac{\text{Var}\,x}{\varepsilon^2 \delta}. \qquad \text{(B3)}$$

To calculate the median-of-means estimator $\hat{x}_{\text{MM}}$ one first splits N samples into K batches, each containing $\lfloor N/K \rfloor$

representatives, and evaluates empirical means $\hat{x}_k$ over the group number k. Then, $\hat{x}_{\mathrm{MM}}$ is defined as follows:

$$\hat{x}_{\mathrm{MM}} = \mathrm{median}(\hat{x}_1, \ldots, \hat{x}_K). \tag{B4}$$

This estimator is substantially more robust, since for the choice $K = \log 1/\delta$, the following inequality holds:

$$\mathrm{prob}(|\hat{x}_{\mathrm{MM}} - \mathbb{E}x| \geq \varepsilon) \leq \delta = \exp\left(-\frac{N\varepsilon^2}{4\,\mathrm{Var}\,x}\right). \tag{B5}$$

The sample complexity for the median of means is

$$N = \frac{4\,\mathrm{Var}\,x}{\varepsilon^2} \log 1/\delta. \tag{B6}$$

Note the appearance of $\log 1/\delta$ instead of $1/\delta$ compared to Eq. (B3).

It turns out, however, that in the asymptotic limit $N \to \infty$, the central-limit theorem (CLT) holds and the empirical mean $\hat{x}$ is distributed normally: $\hat{x} \sim \mathcal{N}[\mathbb{E}x, \mathrm{Var}(x)/N]$. One can calculate the probability of deviation:

$$\mathrm{prob}(|\hat{x} - \mathbb{E}x| \geq \varepsilon) = \delta = 1 - \mathrm{erf}\left(\sqrt{\frac{N\varepsilon^2}{2\,\mathrm{Var}\,x}}\right),$$

$$\delta \leq \exp\left(-\frac{N\varepsilon^2}{2\,\mathrm{Var}\,x}\right), \tag{B7}$$

where $\mathrm{erf}(y)$ is the Gauss error function, which satisfies the inequality $1 - \mathrm{erf}(y) \leq \exp(-y^2)$ for $y \geq 0$. By expressing N, we obtain:

$$N \leq \frac{2\,\mathrm{Var}\,x}{\varepsilon^2} \log 1/\delta. \tag{B8}$$

In this case, the empirical mean is also robust because it contains the logarithmic dependence $\log 1/\delta$, which is similar to the one for the median-of-means estimator.

If the classical shadow $\hat{\rho}$ [see Eq. (2)] is substituted into Eq. (3), then the estimator $\hat{o}$ takes the form of a linear combination of random variables f_i. Each frequency f_i can be viewed as an empirical mean of roughly N/P single-shot measurements. Since in our experiment we work in the overexposure regime $N/P \to \infty$, the CLT conditions are satisfied with a high degree of accuracy and the above reasonings about the same performance of the mean and median-of-means estimators become valid.

APPENDIX C: COMPENSATION OF THE GOUY PHASE

The prepared state $|\psi_{\mathrm{prep.}}\rangle$ and the detected one $|\psi_{\mathrm{det.}}\rangle$ are tied by a unitary transformation $U(\phi_G)$ with one unknown parameter, namely, the Gouy phase ϕ_G: $|\psi_{\mathrm{det.}}\rangle =$
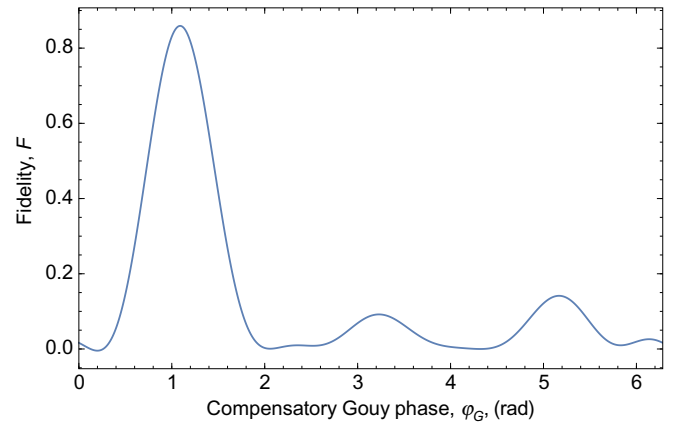

FIG. 6. A typical dependence of the preparation fidelity F on the compensatory Gouy phase for $D = 32$. The global maximum corresponds to the true phase value.

$U(\phi_G)|\psi_{\mathrm{prep.}}\rangle$, where $U(\phi_G)$ is a diagonal matrix and contains entities of unit magnitude only. The compensated fidelity of preparation F, or the *preparation fidelity* for short, is determined by maximizing the fidelity to $|\psi_{\mathrm{det.}}\rangle$ over ϕ_G:

$$F = \max_{\phi_G} \langle \psi_{\mathrm{prep.}} | U^\dagger(\phi_G) \hat{\rho} U(\phi_G) | \psi_{\mathrm{prep.}} \rangle. \tag{C1}$$

The state $U(\phi_G^{\mathrm{max}})|\psi_{\mathrm{prep.}}\rangle$, where ϕ_G^{max} maximizes Eq. (C1), is the *compensated prepared state*.

A typical dependence of the quantity under maximization in Eq. (C1) on ϕ_G is demonstrated in Fig. 6 for the system dimensionality $D = 32$. There is a "low-valued ripple" with local extrema (which are especially pronounced for higher D), but for all tested states and dimensions, a global maximum around $\phi_G \approx 1$ radians with a near-unity value is observed.

[1] *Quantum State Estimation*, edited by M. Paris and J. Řeháček, Lecture Notes in Physics, Vol. 649 (Springer-Verlag, Berlin, Heidelberg, 2004).

[2] K. Banaszek, G. M. D'Ariano, M. G. A. Paris, and M. F. Sacchi, Maximum-likelihood estimation of the density matrix, Phys. Rev. A **61**, 010304(R) (1999).

[3] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, Quantum state tomography via compressed sensing, Phys. Rev. Lett. **105**, 150401 (2010).

[4] Y. S. Teo, J. Řeháček, and Z. Hradil, Informationally incomplete quantum tomography, Quantum Meas. Quantum Metrol. **1**, 57 (2013).

[5] Y. S. Teo, H. Zhu, B.-G. Englert, J. Řeháček, and Z. C. V. Hradil, Quantum-state reconstruction by maximizing likelihood and entropy, Phys. Rev. Lett. **107**, 020404 (2011).

[6] Y. S. Teo, G. I. Struchalin, E. V. Kovlakov, D. Ahn, H. Jeong, S. S. Straupe, S. P. Kulik, G. Leuchs, and L.

L. Sánchez-Soto, Objective compressive quantum process tomography, Phys. Rev. A **101**, 022334 (2020).

[7] Y. Kim, Y. S. Teo, D. Ahn, D.-G. Im, Y.-W. Cho, G. Leuchs, L. L. Sánchez-Soto, H. Jeong, and Y.-H. Kim, Universal compressive characterization of quantum dynamics, Phys. Rev. Lett. **124**, 210401 (2020).

[8] D. Goyeneche, G. Cañas, S. Etcheverry, E. S. Gómez, G. B. Xavier, G. Lima, and A. Delgado, Five measurement bases determine pure quantum states on any dimension, Phys. Rev. Lett. **115**, 090401 (2015).

[9] M. Cramer, M. B. Plenio, S. T. Flammia, R. Somma, D. Gross, S. D. Bartlett, O. Landon-Cardinal, D. Poulin, and Y.-K. Liu, Efficient quantum state tomography, Nat. Commun. **1**, 1 (2010).

[10] B. Lanyon, C. Maier, M. Holzäpfel, T. Baumgratz, C. Hempel, P. Jurcevic, I. Dhand, A. Buyskikh, A. Daley, M. Cramer *et al.*, Efficient tomography of a quantum many-body system, Nat. Phys. **13**, 1158 (2017).

[11] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, Neural-network quantum state tomography, Efficient tomography of a quantum many-body system, Nat. Phys. **14**, 447 (2018).

[12] J. Carrasquilla, G. Torlai, R. G. Melko, and L. Aolita, Reconstructing quantum states with generative models, Nat. Mach. Intell. **1**, 155 (2019).

[13] E. S. Tiunov, V. Tiunova, A. E. Ulanov, A. Lvovsky, and A. Fedorov, Experimental quantum homodyne tomography via machine learning, Optica **7**, 448 (2020).

[14] S. Aaronson, in *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing* (Association for Computing Machinery, New York, NY, United States, 2018), p. 325.

[15] A. Rocchetto, S. Aaronson, S. Severini, G. Carvacho, D. Poderini, I. Agresti, M. Bentivegna, and F. Sciarrino, Experimental learning of quantum states, Sci. Adv. **5**, eaau1946 (2019).

[16] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, Nat. Phys. **16**, 1050 (2020).

[17] D. Gottesman, arXiv:quant-ph/9705052 [quant-ph] (1997).

[18] M. Guţă, J. Kahn, R. Kueng, and J. A. Tropp, Fast state tomography with optimal error bounds, J. Phys. A: Math. Theor. **53**, 204001 (2020).

[19] Z. Webb, The clifford group forms a unitary 3-design, Quantum Info. Comput. **16**, 1379 (2016).

[20] F. Tonolini, S. Chan, M. Agnew, A. Lindsay, and J. Leach, Reconstructing high-dimensional two-photon entangled states via compressive sensing, Sci. Rep. **4**, 6542 (2014).

[21] A. Steffens, C. A. Riofrío, W. McCutcheon, I. Roth, B. A. Bell, A. McMillan, M. S. Tame, J. G. Rarity, and J. Eisert, Experimentally exploring compressed sensing quantum tomography, Quantum Sci. Technol. **2**, 025005 (2014).

[22] The only exceptions are the results presented in Fig. 3.

[23] N. Bent, H. Qassim, A. A. Tahir, D. Sych, G. Leuchs, L. L. Sánchez-Soto, E. Karimi, and R. W. Boyd, Experimental realization of quantum tomography of photonic qudits via symmetric informationally complete positive operator-valued measures, Phys. Rev. X **5**, 041006 (2015).

[24] A. M. Palmieri, E. Kovlakov, F. Bianchi, D. Yudin, S. Straupe, J. D. Biamonte, and S. Kulik, Experimental neural network enhanced quantum tomography, npj Quantum Inf. **6**, 20 (2020).

[25] E. Bolduc, N. Bent, E. Santamato, E. Karimi, and R. W. Boyd, Exact solution to simultaneous intensity and phase encryption with a single phase-only hologram, Opt. Lett. **308**, 3546 (2013).

[26] K. Życzkowski and H.-J. Sommers, Average fidelity between random quantum states, Phys. Rev. A **71**, 032313 (2005).

[27] M. R. Jerrum, L. G. Valiant, and V. V. Vazirani, Random generation of combinatorial structures from a uniform distribution, Theor. Comput. Sci. **43**, 169 (1986).

[28] Y. Chen and X. Ye, arXiv:1101.6081 (2011).

[29] C. Schwemmer, L. Knips, D. Richart, H. Weinfurter, T. Moroder, M. Kleinmann, and O. Gühne, Systematic errors in current quantum state tomography tools, Phys. Rev. Lett. **114**, 080403 (2015).

[30] J. Shang, Z. Zhang, and H. K. Ng, Superfast maximum-likelihood reconstruction for quantum tomography, Phys. Rev. A **95**, 062336 (2017).

[31] G. I. Struchalin, I. A. Pogorelov, S. S. Straupe, K. S. Kravtsov, I. V. Radchenko, and S. P. Kulik, Experimental adaptive quantum tomography of two-qubit states, Phys. Rev. A **93**, 012103 (2016).

[32] S. Aaronson, X. Chen, E. Hazan, S. Kale, and A. Nayak, Online learning of quantum states, J. Stat. Mech.: Theory Exp. **2019**, 124019 (2019).

[33] Y. Chen and X. Wang, arXiv:2006.01013 (2020).

[34] S. Aaronson and D. Gottesman, Improved simulation of stabilizer circuits, Phys. Rev. A **70**, 052328 (2004).

[35] D. Gottesman, in *Proceedings of the XXII International Colloquium on Group Theoretical Methods in Physics*, edited by S. P. Corney, R. Delbourgo, and P. D. Jarvis (International Press, Cambridge, MA, 1999), p. 32.

[36] R. Koenig and J. A. Smolin, How to efficiently select an arbitrary clifford group element, J. Math. Phys. **55**, 122202 (2014).

[37] J. Dehaene and B. De Moor, Clifford group, stabilizer states, and linear and quadratic operations over gf(2), Phys. Rev. A **68**, 042318 (2003).

[38] M. V. den Nest, Classical simulation of quantum computation, the gottesman-knill theorem, and slightly beyond, Quant. Inf. Comput. **10**, 0258 (2010).

[39] J. Goldman and G.-C. Rota, On the foundations of combinatorial theory iv finite vector spaces and eulerian generating functions, Stud. Appl. Math. **49**, 239 (1970).

[40] https://github.com/qotlabs/randstab.