Classical shadow tomography with mutually unbiased bases

Yu Wang[1,*] and Wei Cui[1,2,†]
[1]*Beijing Institute of Mathematical Sciences and Applications, Beijing 101408, China*
[2]*Yau Mathematical Sciences Center, Tsinghua University, Beijing 100084, China*

(Received 25 December 2023; accepted 15 May 2024; published 3 June 2024)

Classical shadow tomography, harnessing randomized informationally complete (IC) measurements, provides an effective avenue for predicting many properties of unknown quantum states. In n-qubit systems, projections onto 2^n+1 mutually unbiased bases (MUBs) are widely recognized as minimal and optimal IC measurements for full-state tomography. Mutually unbiased base circuits, structured as -CZ-S-H-, form the minimal subset of the whole $2^{n^2+2n}\prod_{j=1}^{n}(4^j-1)$ Clifford circuit ensemble. Each of them can be generated by n special MUB circuits. We study how to use MUB circuits as the ensemble in classical shadow tomography. For general observables, the variance to predict their expectation value is shown to be exponential to the number of qubits n. However, for a special class referred to as appropriate MUB-average (AMA) observables, the variance decreases to poly(n). Additionally, we find that through biased sampling of MUB circuits, the variance for non-AMA observables can again be reduced to poly(n) with the MUB-sparse condition. The performance and complexity of using the MUB and Clifford circuits as the ensemble in the classical shadow tomography are compared.

DOI: 10.1103/PhysRevA.109.062406

I. INTRODUCTION

In the realm of quantum information science, efficiently extracting information from unknown quantum states is pivotal. This is traditionally achieved through quantum state tomography [1–3]. Let $\{|k\rangle\}_{k=0}^{d-1}$ be the computational basis and $\{U_j\}_{j=0}^{L-1}$ be an informationally complete unitary ensemble. We usually perform projective measurements $\{U_j|k\rangle\langle k|U_j^\dagger\}$ to obtain experimental data $\{\mathrm{tr}(\rho U_j|k\rangle\langle k|U_j^\dagger)\}$ and then reconstruct density matrix ρ. It allows us to predict any observables on ρ, such as the state purity, entanglement entropy, and so on. [4–6]. These predictions are central to many-body physics and quantum information theory [7,8]. However, as quantum systems scale up, this approach becomes impractical and even infeasible due to the enormous memory requirements.

Sometimes, we are interested in some specific observables; it is not necessary to calculate all the components in the density matrix by exponentially many measurements. While shadow tomography was initially proposed with polynomial sampling [9,10], it required exponential-depth quantum circuits applied to copies of all quantum states, presenting challenges for quantum hardware. Subsequently, Huang *et al.* introduced classical shadow tomography [11], enabling random measurements on unknown quantum states and efficient prediction of various properties with a sampling complexity of $\log(M)\|\cdot\|^2_{\mathrm{shadow}}$, where M represents the number of observables and $\|\cdot\|^2_{\mathrm{shadow}}$ denotes the norm of the corresponding observables.

This norm also depends on the choice of the unitary ensemble $\{U_j\}$. The initial procedure applies random unitaries from a specific informationally complete (IC) ensemble to the system and then performs computational projective measurements, which is equivalent to performing randomly 3^n Pauli measurements or all Clifford measurements. Pauli measurements are ideal for predicting localized target functions, while Clifford measurements excel in estimating functions with constant Hilbert-Schmidt norms, both offering valuable tools for various quantum tasks. Subsequently, various other ensembles have been explored, including fermionic Gaussian unitaries [12], chaotic Hamiltonian evolutions [13], locally scrambled unitary ensembles [14–16], and Pauli-invariant unitary ensembles [17]. The random selection of multiple sets of IC projective measurements has been theoretically generalized to one IC positive-operator-valued measure (POVM) [18,19]. In alignment with this broadened perspective, dual measurement frames for IC POVMs are employed to analyze estimation errors in shadow tomography tasks [20,21]. Up to now, classical shadow tomography has found applications in diverse fields, including energy estimation [22], entanglement detection [23,24], quantum chaos [25], quantum gate engineering cycles [26], and quantum error mitigation [27], to name a few.

Without ancillas, the minimal-size IC ensemble contains 2^n+1 unitary operations [28]. Projective measurements onto the set of 2^n+1 mutually unbiased bases (MUBs) are recognized as the optimal approach for quantum tomography [28–30]. For a vector prepared within a specific MUB, a uniform distribution will be achieved when projecting it onto any other MUBs. These MUB measurements are regarded as having maximal incompatibility and complementarity [31], finding applications in various aspects of quantum information science, including quantum tomography [32,33], uncertainty relations [34–36], quantum key distribution [37,38], quantum error correction [39–41], and

*ming-jing-happy@163.com
†cwei@bimsa.cn

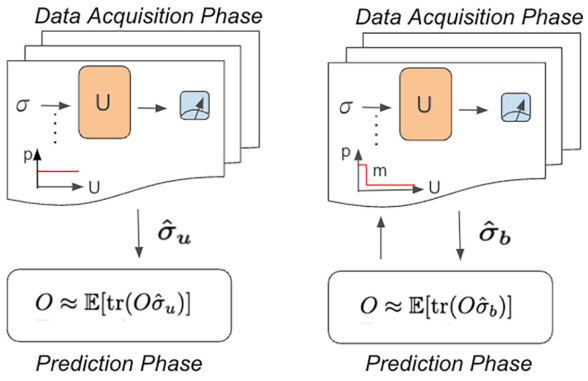

FIG. 1. Procedure of the classical shadow tomography based on uniform (left) and biased (right) sampling. Here σ is the unknown state, U is a unitary gate chosen from the MUB ensemble with probability p, and O is the observable that one wants to predict. For biased sampling, we choose one specific unitary gate with probability m and choose others uniformly. The classical shadows of the uniform and biased samplings are denoted by $\hat{\sigma}_u$ and $\hat{\sigma}_b$, respectively.

the identification of entanglement and other forms of quantum correlations [42–47].

In this work we use MUB circuits as the unitary ensemble for classical shadow tomography. The reconstruction channel and the variance for uniform and biased MUB samplings are computed. We find that the variance for some bounded norm observables can be exponential with the number of qubits, but it can become polynomial when the observables (or states) are approximate MUB-average (AMA) ones. For observables that are not AMA but obey the MUB-sparse condition, we show that by biased sampling one can reduce the variance to polynomial order. The procedure of the classical shadow tomography based on MUB circuits is summarized in Fig. 1. In the end, we compare the algorithmic and circuit complexity of the MUB and Clifford measurements.

We noticed that also Zhang *et al.* [48] studied how to perform classical shadow tomography with the MUB circuits. Different from us, their motivation was to identify the minimal subset of the Clifford ensemble that happens to be the MUB ensemble studied in our paper. Although some of the results like the reconstruction channel and variance analysis for uniform sampling are the same, as they should be, the methods to reach these results are different in many aspects. First, Ref. [48] employs properties of t-designs for calculating variances and reconstructing channels whereas we rely solely on properties inherent to MUBs. Second, the MUB circuits constructed in [48] are based on the Gottesman-Knill theorem while the circuits used in our paper are constructed directly using the representations of the calculations in the Galois field [49]. The circuit structure is the same, but the detailed gates are different. More information on MUB circuits is reviewed in Sec. II A. Third, regarding biased sampling, we directly compute the corresponding new reconstruction channel and its variance for the MUB-sparse case and thus the computational cost of estimation will be polynomial. The optimal sampling probabilities are given explicitly. While [48] employs postprocessing techniques for general observables, determination of the $2^n + 1$ optimal sampling probabilities can be exponentially hard but handled with some cases. Fourth, we find the sampling efficiency of approximately MUB-average and MUB-sparse observables (states), while [48] also found the sampling advantage of MUBs on estimating off-diagonal operators.

II. MUB CLASSICAL SHADOW TOMOGRAPHY

Let us consider a finite d-dimensional Hilbert space. When we perform a measurement on a quantum state ρ using an observable $O = \sum_{k=1}^{d} \lambda_k |\phi_k\rangle\langle\phi_k|$, the outcome λ_k is obtained with a probability of $\mathrm{tr}(\rho|\phi_k\rangle\langle\phi_k|)$. The projective eigenstates $\{|\phi_k\rangle\}_{k=1}^{d}$ are mutually orthonormal. The act of projective measurement onto $\{|\phi_k\rangle\}_{k=1}^{d}$ yields a projection-valued measure (PVM) $\{|\phi_k\rangle\langle\phi_k|\}_{k=1}^{d}$. These PVMs can be generalized into POVMs $\{E_k\}$, subject to the condition that E_k is positive and satisfies the normalization condition $\sum E_k = I$. Multiple PVMs can be amalgamated into a POVM by scaling each component with normalized probabilities. For instance, consider the PVMs corresponding to three Pauli observables: $\{|0\rangle\langle 0|, |1\rangle\langle 1|\}$, $\{|+\rangle\langle +|, |-\rangle\langle -|\}$, and $\{|\times\rangle\langle\times|, |\cdot\rangle\langle\cdot|\}$. These can be merged into a POVM $\{p_1|0\rangle\langle 0|, p_1|1\rangle\langle 1|, p_2|+\rangle\langle +|, p_2|-\rangle\langle -|, p_3|\times\rangle\langle\times|, p_3|\cdot\rangle\langle\cdot|\}$, where $p_1 + p_2 + p_3 = 1$.

A POVM $\{E_k\}$ is considered IC if it spans the entire Hermitian operators space $M_d(\mathbb{C})$ [50,51]. The probability distribution $\{\mathrm{tr}(\rho E_k)\}$ is then sufficient to uniquely determine any arbitrary unknown state ρ. Alternatively, given a probability distribution $\{\mathrm{tr}(\rho E_k)\}$, there exists no other state $\rho_2 \neq \rho$ such that $\mathrm{tr}(\rho E_k) = \mathrm{tr}(\rho_2 E_k)$ for some k.

An IC POVM should consist of at least d^2 elements. Symmetric informationally complete measurements (SIC POVMs) contain a minimum of d^2 rank-1 elements and have been employed in classical shadow tomography [52]. In the implementation of SIC POVMs, additional auxiliary systems are often needed [53–58]. If not, d^2 unitary operations are required [59]. Two intriguing yet tangential problems lie in the quest for the existence of SIC POVMs and $d + 1$ MUBs for arbitrary dimension d. These challenges have been identified as the first two open problems in quantum information theory by Horodecki *et al.* [60].

Without ancillary systems, a minimum of $d + 1$ PVMs is required to consolidate into an IC POVM. Here is a brief explanation. A normalized density operator ρ contains $d^2 - 1$ real parameters, reduced by 1 due to $\mathrm{tr}(\rho) = 1$. Each PVM generates $d - 1$ independent probabilities, as the final one can be expressed as 1 minus the sum of the others. For a set of eigenstates $\{|\phi_k^j\rangle\}_{k=1}^{d}$ of a PVM, there always exists a unitary operation U_j such that $U_j|k\rangle = |\phi_k^j\rangle$. Therefore, in the absence of ancillary systems, the ensemble $\{U_j\}$ must contain at least $d + 1$ unitary operations to ensure that $\{U_j|k\rangle\langle k|U_j^\dagger : j = 0, \ldots, d; k = 0, \ldots, d - 1\}$ is informationally complete. In the original classical shadow tomography scheme [11], when selecting the Pauli ensemble and Clifford ensemble, the number of unitary operations is 3^n and $2^{n^2+2n}\prod_{j=1}^{n}(4^j - 1)$, respectively, with dimension $d = 2^n$.

Informationally complete (or tomographically complete, as mentioned in [11]) is a crucial property of randomly selected measurements in classical shadow tomography. With

this property, the quantum channel for the random sampling process has a unique inverse.

Let A and B be two Hermitian operators with normalized eigenstates $\{|a_i\rangle\}_{i=1}^{d}$ and $\{|b_j\rangle\}_{j=1}^{d}$. These two bases are called mutually unbiased if they satisfy the property

$$|\langle a_i|b_j\rangle|^2 = \frac{1}{d} \tag{1}$$

for all i and j. For prime power d, there exists a maximum of $d+1$ MUBs.

In n-qubit systems, the Hilbert space has a dimension of $d = 2^n$. Denote the 2^n+1 MUBs by $\{\mathcal{B}_0, \ldots, \mathcal{B}_{2^n}\}$. Consider $\mathcal{B}_0$ as the canonical basis $\{|t\rangle\}_{t=0}^{2^n-1}$. Each additional basis is defined as $\mathcal{B}_j = \{|e_k^j\rangle\}_{k=0}^{2^n-1}$. Utilizing the Galois-Fourier method, all basis states $|e_k^j\rangle$ are explicitly constructed in Eq. (2.70) of [61].

One of the nice properties of the MUBs is that the projective measurements onto them are informationally complete [30]. By utilizing these projections, we can construct 4^n orthogonal operators according to the Hilbert-Schmidt inner product, $\mathrm{tr}(A^\dagger B) = (A, B)$. These operators can be constructed in the following manner.

(i) The initial 2^n operators stem from basis $\mathcal{B}_0$, represented as $\{|t\rangle\langle t|\}_{t=0}^{2^n-1}$.

(ii) The remaining $4^n - 2^n$ operators are generated from the bases $\{\mathcal{B}_j\}_{j=1}^{2^n}$. For each j, 2^n-1 operators are constructed as $\{|e_k^j\rangle\langle e_k^j| - I/2^n\}_{k=0}^{2^n-2}$.

The 4^n orthogonal operators are linearly independent and thus span the whole space $M_d(\mathbb{C})$.

A. Construction of MUB circuits

Recently, all nontrivial MUB circuits have been efficiently decomposed into $O(n^2)$ elementary gates within $O(n^3)$ time, structured as -H-S-CZ- [49]. We provide a brief review here.

The gates H and S represent the one-qubit Hadamard gate and the phase gate, respectively. The gate CZ denotes the two-qubit controlled-Z gate. The control and target qubits of the controlled-Z gate can vary and be any pair (j, k) where $0 \leqslant j < k \leqslant n-1$. Hence, there are $\binom{n}{2}$ different types of CZ gates.

Now let us delve into the decomposition of gates. The 2^n nontrivial unitary operations for MUBs are defined as $U(j) = \sum_{k=0}^{2^n-1} |f_k^j\rangle\langle k|$, where $j = 0, \ldots, 2^n-1$. For the mutually unbiased basis indexed by j, the 2^n orthonormal eigenstates $\{|f_k^j\rangle\}_{k=0}^{2^n-1}$ are defined as follows:

$$\left|f_k^j\right\rangle = \frac{1}{\sqrt{2^n}} \sum_{l=0}^{2^n-1} |l\rangle (-1)^{k\cdot l^T} \alpha_l^j. \tag{2}$$

The states $\{|f_k^j\rangle\}$ are the rearrangement of original MUB states $\{|e_k^j\rangle\}$. In addition, the coefficients α_l^j detailed in [61] are defined as

$$\alpha_l^j = \prod_{r=0}^{n-1} \overline{(\sqrt{-1})^{j\odot(l_r\times 2^r)\odot(l_r\times 2^r)}} \times \prod_{0\leqslant s<t\leqslant n-1} (-1)^{j\odot(l_s\times 2^s)\odot(l_t\times 2^t)}. \tag{3}$$

The key point is to change the representation of the coefficient α_l^j from the multiplication $\odot$ in the Galois field $\mathrm{GF}(2^n)$ to the traditional multiplication and addition of integers. After modification, the coefficient α_l^j is represented as follows:

$$\alpha_l^j = \prod_{r=0}^{n-1} (\sqrt{-1})^{a_r(j)l_r} \prod_{0\leqslant s<t\leqslant n-1} (-1)^{b_{s,t}(j)l_s l_t}. \tag{4}$$

Here $a_r(j) = 0, 1, 2, 3$ and $b_{s,t}(j) = 0, 1$.

With this representation, the structure of a three-stage MUB circuit is directly inferred. The operation $U(j)$ is expressed as

$$U(j) = \sum_{k=0}^{2^n-1} \left|f_k^j\right\rangle\langle k| = A(j)B(j)C, \tag{5}$$

which corresponds to the circuit structure -H-S-CZ-.

The operation $C = \frac{1}{\sqrt{2^n}} \sum_{k,l=0}^{2^n-1} (-1)^{k\cdot l^T} |l\rangle\langle k|$ represents the tensor product of n Hadamard gates, denoted by $H^{\otimes n}$. It is also utilized in the Deutsch-Jozsa algorithm, as shown in Eq. (2.55) of [62].

The diagonal operation

$$B(j) = \sum_{l=0}^{2^n-1} \prod_{r=0}^{n-1} (\sqrt{-1})^{a_r(j)l_r} |l\rangle\langle l| \tag{6}$$

is associated with the tensor product of S gates $S^{a_0(j)} \otimes \cdots \otimes S^{a_{n-1}(j)}$. The coefficient $a_l(j)$ means the number of times we should apply the S gate to qubit q_l, where $l = 0, \ldots, n-1$.

The diagonal operation

$$C(j) = \sum_{l=0}^{2^n-1} \prod_{0\leqslant s<t\leqslant n-1} (-1)^{b_{s,t}(j)l_s l_t} |l\rangle\langle l| = \prod_{0\leqslant s<t\leqslant n-1} \mathrm{CZ}(s,t)^{b_{s,t}(j)} \tag{7}$$

is associated with the tensor product of CZ gates, where $b_{s,t}(j) = 0, 1$. For coefficients $b_{s,t}(j)$ equal to 1, $\mathrm{CZ}(s,t)$ means that we should add a CZ gate at qubits q_s and q_t.

In addition to considering the circuit structure, further insights into the behavior of MUB circuits can be gained by examining the coefficients $a_l(j)$ and $b_{s,t}(j)$. For instance, the knowledge of only n special MUB circuits can determine all 2^n+1 MUB circuits, which can be observed as a peculiar and intriguing linear behavior. The entanglement aspect solely involves the CZ gates, which are only relevant to the behavior of coefficients $b_{s,t}(j)$. Specifically, if the $\mathrm{CZ}(2,5)$ gate appears in a certain MUB circuit, the $\mathrm{CZ}(3,4)$ gate should also appear in that circuit. Another behavior is related to the average gate number. For example, the total count of CZ gates is computed by summing all instances of $T_1 = \sum_{j=0}^{2^n-1} \sum_{0\leqslant s<t\leqslant n-1} b_{s,t}$.

It is noteworthy that the circuit design described in Ref. [48] utilizes the original MUB states $|e_k^j\rangle$ of Ref. [61]. Consequently, despite employing the same -H-S-CZ- structure where diagonal operations commute, the specific gate numbers and sequences could differ. It would be intriguing to discuss the performance among various styles of MUB circuits, particularly in more quantum information applications [63].

Regarding the application of classical shadow tomography, the numerical performance appears to be similar. Mutually unbiased base circuits are a subset of Clifford circuits, and the circuit design in [48] relies on the Gottesman-Knill theorem for the decomposition of Clifford circuits [64]. It is conceivable that the time complexity or circuit complexity may decrease with advancements in techniques for handling Clifford circuits. While Ref. [49] mentions that the time complexity is $O(n^3)$, considering computations over the Galois field, the maximum gate count is expressed as $(n^2+7n)/2$. Nevertheless, the direct decomposition method may offer fresh perspectives on various facets of MUB circuits from a global standpoint. These insights could include examining linear behavior, entanglement structure, average gate numbers, distributions of MUB state coefficients, and more.

B. Procedure

We can employ MUB circuits as a unitary ensemble for conducting classical shadow tomography [11]. Consider σ as the unknown quantum state and O as the observable for prediction. The general procedure encompasses two primary steps.

The initial step involves generating classical shadows of state σ utilizing MUB measurements. We randomly select a U_j from the 2^n+1 MUB circuits and rotate the unknown state, i.e., $\sigma \to U_j \sigma U_j^\dagger$. Here $U_j = U(j)^\dagger$. Subsequently, the qubits are measured on the computational basis. This measurement yields a 0/1 bit string of length n, denoted by $b_0 \cdots b_{n-1}$. Let $k = b_0 + 2b_1 + \cdots + b_{n-1}2^{n-1}$. We calculate the classical snapshot of σ, defined as

$$\hat{\sigma} = \mathcal{M}^{-1}(U_j^\dagger|k\rangle\langle k|U_j), \tag{8}$$

where $\mathcal{M}^{-1}$ represents the reconstruction channel depending on the chosen unitary ensemble. Interestingly, when uniformly sampling from MUB circuits, the reconstruction channel mirrors that of Clifford circuits and can be expressed for any operator X as

$$\mathcal{M}_u^{-1}(X) = (2^n+1)X - \mathrm{tr}(X)I_n. \tag{9}$$

The specific calculations are detailed in Appendix A. Repeat this rotation-measurement process N times. This yields a set of N classical snapshots, termed a classical shadow of σ, which will be stored in the classical memory.

The second step involves using the obtained classical shadows to predict observables $\{O_1, O_2, \ldots, O_M\}$ of the unknown quantum state σ. Their expectation values are given by

$$o_i = \mathrm{tr}(O_i\sigma), \quad 1 \leqslant i \leqslant M,$$

which can be approximated by the median of means of the expectation values

$$\hat{o}_i(N,K) = \mathrm{median}\{\hat{o}_i^{(1)}(L,1), \hat{o}_i^{(2)}(L,1), \ldots, \hat{o}_i^{(K)}(L,1)\},$$

where $L = \lfloor N/K \rfloor$ and

$$\hat{o}_i^{(k)}(L,1) = \frac{1}{L}\sum_{j=(k-1)L+1}^{kL} \mathrm{tr}(O_i\hat{\sigma}_j), \quad 1 \leqslant k \leqslant K.$$

Here, to reduce the variance, we split the shadow into K equally sized parts labeled j, with each part containing L snapshots. The quality of this approximation depends on the choices of the parameters L and K.

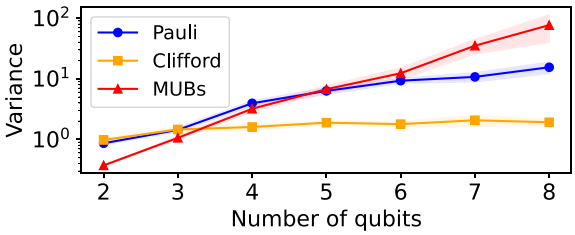

FIG. 2. Variance of fidelity estimation between two GHZ states obtained from the shadows with different unitary ensembles. For each case, we randomly generate ten different shadows. The number of measurements in each shadow is 1000. The shaded area is the statistical variance for ten independent experiments.

C. Performance

Suppose the unknown state is σ. To assess the performance of using MUB-based classical shadow tomography to predict an observable O under σ, one should examine the variance $\|O_0\|_\sigma^2$ in Eq. (10), where $O_0 = O - \frac{\mathrm{tr}(O)}{2^n}$ represents the traceless part of O. The sample complexity is linearly correlated with the variance,

$$\|O_0\|_\sigma^2 = \mathbb{E}_{U\sim\mathcal{U}} \sum_{k=0}^{2^n-1} \langle k|U\mathcal{M}^{-1}(O_0)U^\dagger|k\rangle^2 \langle k|U\sigma U^\dagger|k\rangle. \tag{10}$$

The variance for an arbitrary unknown state is defined by the shadow norm $\|O_0\|_{\mathrm{shadow}}^2 = \max_{\sigma\ \mathrm{state}} \|O_0\|_\sigma^2$. The variance of the shadow norm for Clifford and Pauli measurements has been studied in [11]. Now we consider the shadow variance of the MUB measurement. Using the result of the reconstruction channel in Eq. (9), it is straightforward to show that

$$\|O_0\|_{\sigma,u}^2 = (2^n+1)\sum_{j=0}^{2^n}\sum_{k=0}^{2^n-1} \mathrm{tr}^2(O_0P_{jk})\mathrm{tr}(\sigma P_{jk}), \tag{11}$$

where $P_{jk} = U_j^\dagger|k\rangle\langle k|U_j$. The subscript u means we uniformly sample MUB ensemble. This variance depends exponentially on n because the terms $\mathrm{tr}(\sigma P_{jk})$ and $\mathrm{tr}(O_0P_{jk})$ can take their maximal values to be approximately $O(1)$ at the same j,k. For example, when $O = |0\rangle\langle 0|$ and $\sigma = |0\rangle\langle 0|$, the variance is $\|O_0\|_{|0\rangle\langle 0|,u}^2 \geqslant (2^n+1)(1-1/2^n)^2$. Since each $0 \leqslant \mathrm{tr}(\sigma P_{jk}) \leqslant 1$ for all j,k, we can derive that the variance $\max_{\sigma\ \mathrm{state}} \|O_0\|_{\sigma,u}^2 \leqslant (2^n+1)\mathrm{tr}(O_0^2)$ for all σ and O. Thus, when the unitary ensemble in the shadow tomography is MUB circuits, to obtain an accurate enough prediction of $\langle O\rangle$, for the worst case, one needs to perform approximately 2^n samples.

A numerical experiment is performed. Consider the observable $O = |\mathrm{GHZ}\rangle\langle\mathrm{GHZ}|$ on the unknown state $|\mathrm{GHZ}\rangle\langle\mathrm{GHZ}|$. We predict the expectation value of $\langle O\rangle$, or equivalently the fidelity between two GHZ states, for n up to 8 using shadows generated by 1000 Pauli, Clifford, and MUB measurements. The experiments are independently performed ten times. We plot the variance $\|O\|_\sigma$ in Fig. 2, where the shaded area is the statistical variance between

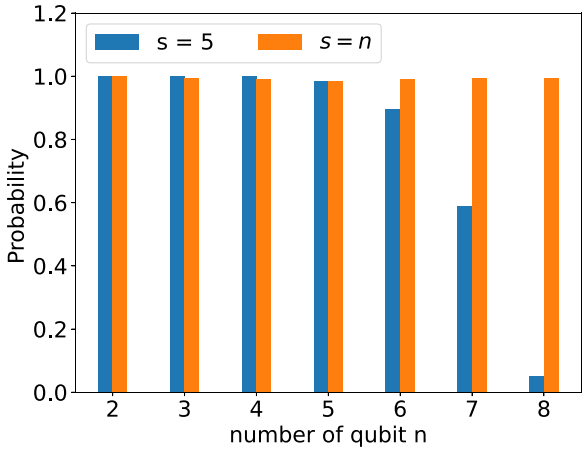

FIG. 3. Percentage of AMA states with $\epsilon = s/2^n$ in 1000 randomly chosen quantum states according to the Haar measure in $n = 2, 3, \ldots, 8$ qubit systems.

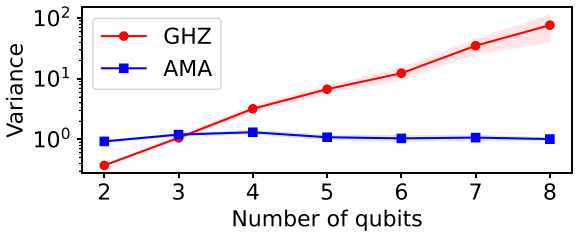

FIG. 4. Variance of fidelity estimation based on MUB measurements. The observable is $O_{\mathrm{GHZ}} = |\mathrm{GHZ}\rangle\langle\mathrm{GHZ}|$ (red) and $O_{\mathrm{AMA}} = |\phi\rangle\langle\phi|$ (blue), where $|\phi\rangle$ are randomly chosen AMA states with $\epsilon = 2/2^n$. The number of measurements in each experiment is 1000. The shaded area is the statistical variance for ten independent experiments.

ten different random experiments. The results show that the variance of the prediction with shadows of Clifford measurements is independent of the number of qubits, while for shadows obtained using Pauli and MUB measurements, $\|O\|_\sigma$ scales exponentially with n, which is consistent with the analysis made above.

Although when the unitary ensemble is a MUB circuit, the variance $\|O_0\|_{\sigma,u}$ of the prediction $\langle O\rangle$ using classical shadow tomography depends exponentially on n, it can be shown that this variance can decrease to polynomial with n when the interested observable O or the unknown state σ has the following property.

Definition 1 (approximately MUB average). A state σ (or observable O) is called approximately MUB average if it satisfies

$$|\mathrm{tr}(\sigma P_{jk}) - 1/2^n| \leqslant \epsilon \tag{12}$$

for all j, k and $\epsilon = O(\mathrm{poly}(n))/2^n \ll 1$.

In other words, its probability distribution under each basis of MUBs is approximately uniform. Alternatively, if we express σ with the 4^n orthogonal operators defined above, the matrix elements are all less than $\epsilon + 1/2^n$. If we sample $|\phi\rangle = U|0\rangle$ with the Haar measure, the average state will be [65]

$$\int_{U(2^n)} U|0\rangle\langle 0|U^\dagger \mathrm{d}\mu(U) = I/2^n. \tag{13}$$

Thus ϵ also reflects the deviation from the average state $I/2^n$ with $\epsilon = 0$.

We consider the special unknown state to be $\sigma = I/2^n$. Then $\mathrm{tr}(\frac{I}{2^n}P_{jk}) = 1/2^n$ for all j, k. Substituting this into Eq. (11), we can obtain

$$\|O_0\|^2_{I/2^n,u} = (1 + 1/2^n)\mathrm{tr}\big(O_0^2\big) \tag{14}$$

by Proposition 2.

We perform a numerical experiment using QISKIT to study the AMA states. The result is shown in Fig. 3. In particular, we randomly generate 1000 quantum states according to Haar measure in n-qubit system for $n = 2, 3, \ldots, 8$. If we set $s = 5$, we find that 89.7% of these random states are AMA for $n = 6$, 59% for $n = 7$, and 5.2% for $n = 8$, while for $n = 2, 3, 4, 5$, almost all these random states are AMA because $s = 5$ is comparable to 2^n. However, if we decrease it to $s = n$, almost all of these random states are AMA even for $n = 8$. Thus, it seems that as long as we choose an approximate s such that ϵ is large enough but still much less than 2^n, there is a large number of AMA states in the Hilbert space.

If we randomly select a density matrix σ, the values $\mathrm{tr}(\sigma P_{jk})$ are always less than 1 for all MUB states P_{jk}. If the sampling size is large enough, the average behavior will be like I/d and then $\mathrm{tr}(\sigma P_{jk}) \leqslant \frac{1}{2^n} + \epsilon$. It is an interesting question to rigorously discuss the percentage of states with different ϵ among all states.

Theorem 1. If the observable O is AMA, then for any unknown states σ, the upper bound of the variance is

$$\|O_0\|^2_{\sigma,\mathrm{u}} \leqslant \left(1 + \frac{1}{2^n}\right)^2 \mathrm{poly}(n). \tag{15}$$

On the other hand, if the unknown state σ is AMA, then for any observable O the variance is upper bounded by

$$\|O_0\|^2_{\sigma,u} \leqslant \big(1 + \tfrac{1}{2^n}\big)[1 + \mathrm{poly}(n)]\mathrm{tr}\big(O_0^2\big). \tag{16}$$

If $\mathrm{tr}(O_0^2)$ is a constant bounded norm, then the variance is bounded by a polynomial function of n.

The proof of the theorem is given in Appendix B. Thus, if the observable is AMA or the unknown state happens to be AMA, the MUB-based shadow tomography is an effective method to predict $\langle O\rangle$.

As an example, we study the expectation value $\langle O\rangle$ on the GHZ state for both AMA observable $O_{\mathrm{AMA}} = |\phi\rangle\langle\phi|$ and non-AMA observable $O_{\mathrm{GHZ}} = |\mathrm{GHZ}\rangle\langle\mathrm{GHZ}|$ for $n = 2, 3, \ldots, 8$. Here $|\phi\rangle$ are randomly chosen AMA states with $\epsilon = 2/2^n$ for each n. By 1000 MUB measurements, we plot the variance of the prediction $\langle O\rangle$ in Fig. 4. Compared with the O_{GHZ} case, the variance of O_{AMA} does not depend exponentially on n. Thus, it confirms our Theorem 1 and one can use the MUB-based classical shadow tomography to predict the AMA observables.

III. BIASED-MUB CLASSICAL SHADOW TOMOGRAPHY

When both the unknown state σ and observable O are not AMA, the variance $\|O_0\|^2_{\sigma,u}$ could be exponential with n. An accurate prediction of $\langle O\rangle$ requires approximately 2^n MUB

measurements in classical shadow tomography. In contrast, we find that if the observable $O = |\phi\rangle\langle\phi|$ or the unknown state σ satisfies the MUB-sparse condition, by the biased sampling, the variance can also be poly(n).

Definition 2 (MUB sparse). A state $|\phi\rangle$ is called MUB sparse if it has sparse expression under some basis $\mathcal{B}_j$. Precisely, $|\phi\rangle = \sum_{k=0}^{2^n-1} a_k U_j^\dagger |k\rangle$ for some j and the number of nonzero elements in the set $\{a_k : k = 0, \ldots, 2^n - 1\}$ is $t = O(\text{poly}(n))$.

If a state is MUB sparse under basis $\mathcal{B}_j$, it is not AMA, as $\epsilon = O(1)$ under the measurement of $\mathcal{B}_j$. However, for the other 2^n bases, we can prove $\epsilon = t/2^n$ as follows.

Proposition 1. Consider that $|\phi\rangle$ contains at most t nonzero amplitudes under the expression of $\mathcal{B}_j$. For the other 2^n MUBs, we can prove $|\langle\phi|U_{j'}|k\rangle|^2 \leqslant t/2^n$, where $j' \neq j$.

Proof. We use the expression $|\phi\rangle = \sum_{k=1}^{t} a_{l_k} U_j^\dagger |l_k\rangle$. With the definition of MUBs $|\langle k'|U_{j'} U_j^\dagger |l_k\rangle| = \frac{1}{\sqrt{2^n}}$ for $j \neq j'$,

$$|\langle\phi|U_{j'}|k\rangle|^2 = \frac{1}{2^n}\left|\sum_{k=1}^{t} a_{l_k} e^{i\theta_{jk}}\right|^2 \leqslant \frac{1}{2^n}\left(\sum_{k=1}^{t}|a_{l_k}|^2\right)\left(\sum_{k=1}^{t}|e^{i\theta_{jk}}|^2\right) = \frac{t}{2^n}.$$

■

As we will see in the following, for MUB-sparse observables, the variance of prediction can decrease significantly if one performs the MUB measurements in a biased way. Our motivation to consider the biased sampling is from the observations of fidelity estimation between two GHZ states. In this case, both the unknown states and observable are not AMA but they are MUB sparse with $t = 2$ under $\mathcal{B}_0$. If we sample the MUB circuits $\{U_i\}_{i=0}^d$ uniformly in shadow tomography, the variance is exponential with n from Fig. 2. Specifically, 2000 snapshots are enough to give 0.99 fidelity for $n \leqslant 7$, but when $n \geqslant 8$, ensuring a fidelity of 0.95 requires 10 000 samplings. Upon reviewing the numerical outcomes, we find that choosing more operations U_0 improves the estimated performance. Theoretically, $\text{tr}[O\mathcal{M}^{-1}(P_{0k})] = (2^n - 1)/2$ when U_0 is chosen. Otherwise, the expectation values could always be $1/2^{n+1} - 1/2$ or $1/2^n$. However, as n exceeds 8, the likelihood of uniformly selecting U_0 drastically decreases. Appropriately increasing the number of samples of U_0 will result in a faster approximation of the exact expected value 1.

Biased shadow tomography follows a process similar to the usual case, with the difference lying in the data acquisition phase, i.e., employing random MUB measurements based on a biased distribution. For instance, if the observable $O = |\phi\rangle\langle\phi|$ is MUB sparse under $\mathcal{B}_j$, adjusting probabilities can prioritize sampling from U_j while reducing others' likelihood, outlined as

$$p_{U_j} = \frac{1}{2^n+1} \to \frac{1+m}{2^n+1+m} \quad \text{for } U_j,$$
$$p_{U_k} = \frac{1}{2^n+1} \to \frac{1}{2^n+1+m} \quad \text{otherwise}, \tag{17}$$

where m is a real number. Note that when we use the biased shadow to predict $\langle O\rangle$, the variance for uniform sampling in Eq. (11) does not apply anymore. In fact, the correct variance depends on the parameter m. As we will see later, m is not a free parameter in our scheme but will be fixed to be the one that minimizes this variance.

When we randomly sample the MUB circuits according to Eq. (17), for any operator X, the reconstruction channel becomes

$$\mathcal{M}_b^{-1}(X) = (2^n + 1 + m)\left(X - \frac{m}{1+m}\sum_k \text{tr}(XP_{jk})P_{jk}\right) - \frac{\text{tr}(X)I}{1+m}. \tag{18}$$

When $m = 0$, it becomes the reconstruction channel for the uniform sampling case in Eq. (9) and $\mathcal{M}_b^{-1}(X) = \mathcal{M}_u^{-1}(X)$. The details of this derivation can be found in Appendix A.

Theorem 2. Given an observable $O = |\phi\rangle\langle\phi|$ and unknown quantum state σ, if $|\phi\rangle$ is MUB sparse, one can efficiently predict $\langle O\rangle$ using biased MUB sampling. The upper bound of variance is given by

$$\|O_0\|_{\sigma,b}^2 \leqslant t^2 \quad \text{when } m = \frac{2^n}{t-1} - 1 \tag{19}$$

and the optimal probabilities to choose U_j and the other unitary circuits in the MUB set are

$$p_{U_j} = \frac{1}{t}, \quad p_{U_k} = \frac{1}{2^n}\left(1 - \frac{1}{t}\right), \quad k \neq j. \tag{20}$$

On the other hand, given an observable O and unknown quantum state $\sigma = |\phi\rangle\langle\phi|$, if $|\phi\rangle$ is MUB sparse, one can predict $\langle O\rangle$ using biased MUB sampling. The subscript b indicates biased sampling from the MUB ensemble. The upper bound of variance is given by

$$\|O_0\|_{\sigma,b}^2 \leqslant (\sqrt{t}+1)^2\text{tr}(O_0^2) \quad \text{when } m = \frac{2^n}{\sqrt{t}} - 1 \tag{21}$$

and the optimal probabilities to choose U_j and the other unitary circuits in the MUB set are

$$p_{U_j} = \frac{1}{1+\sqrt{t}}, \quad p_{U_k} = \frac{1}{2^n}\left(\frac{1}{1+\sqrt{t}}\right), \quad k \neq j. \tag{22}$$

The proof is given in Appendix B. For both cases, the variance for predicting the MUB-sparse observable is upper bounded by a function independent of n. Thus, classical shadow tomography based on the MUB circuits works not only for AMA observables, but also for observables that are MUB sparse by biased sampling.

Let us consider the fidelity estimation between two GHZ states again. In Fig. 5 we compare the variance of the predictions of O_{GHZ} using the shadows obtained by Clifford, MUB, and biased MUB measurements. As discussed before, the observable is not AMA. Uniform sampling in shadow tomography leads to the exponential dependence of n in the variance. While since O_{GHZ} is MUB sparse with $t = 2$ under $U_0 = I$, according to Theorem 2, we perform the MUB measurements in a biased ways with the probability 0.5 on U_0 and $1/2^{n+1}$ on the others. By doing so, the variance decreases significantly, that is, even lower than in the Clifford case. In this simple example, we have shown that by biased sampling, the classical shadow tomography based on MUB circuits works for the MUB-sparse observables.

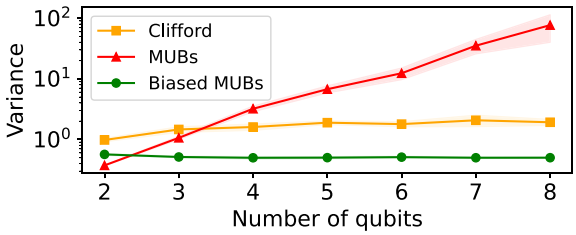

FIG. 5. Variance of fidelity estimation between two GHZ states using MUB shadows (red), Clifford shadows (orange), and biased MUB shadows (green). The shaded region is the statistical variance among ten randomly generated shadows. The number of measurements in each shadow is 1000.

Remark. Note that for observables that can be treated as the density matrix of mixed states, i.e., $O = \sum_{k=1}^{s} p_k |\phi_k\rangle\langle\phi_k|$ with $p_k \geqslant 0$ and $\sum_{k=1}^{s} p_k = 1$, if all $\{|\phi_k\rangle\}$ are MUB sparse under the same basis $\mathcal{B}_j$, then Theorem 2 still holds. It is easy to check that $\mathrm{tr}(OP_{j'k}) \leqslant \frac{t}{2^n}$ for each k when $j' \neq j$.

IV. COMPARISON WITH RANDOM CLIFFORD MEASUREMENTS

Based on the performance of the classical shadow tomography with a MUB ensemble, we classify the quantum states (observables) into three classes as depicted in Fig. 6. With uniform MUB circuit sampling, we observe that the worst-case variance is bounded by $(2^n + 1)\mathrm{tr}(O_0)^2$, while the average variance is $(1 + 1/2^n)\mathrm{tr}(O_0)^2$. Notably, for AMA states (observables), the variance remains polynomial. Additionally, with biased MUB circuit sampling, the variance for sparse states (observables) also stays polynomial. By contrast, when employing uniform sampling of Clifford circuits, the variance for all states (observables) is limited to $3\,\mathrm{tr}(O_0^2)$.

In addition to the performance, experiment implementation and postprocessing are two main procedures in classical shadow tomography. We compare them for Clifford ensembles and MUB ensembles.

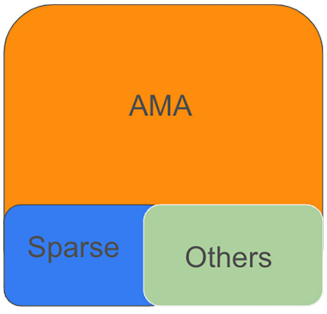

FIG. 6. States classification. We divide all n-qubit states into three parts: AMA states (observables), states (observables) with a sparse representation under some basis $\mathcal{B}_j$, and others. It is intriguing to explore if other states share similar properties with the sparse states and to assess the percentage of these failed states within the complete set.

A. Experiment implementation

In the data acquisition phase, one needs to randomly sample the unitary operations in the ensemble many times. The MUB ensemble has many fewer elements than the Clifford one, thus making the sampling process more direct. Also, the random MUB circuit structure is simpler to implement in the experiments than the random Clifford circuit. Surprisingly, we only need to consider n special MUB circuits to implement all.

1. Sampling process

As for the n-qubit Clifford ensemble, the number of elements is $2^{n^2+2n}\prod_{j=1}^{n}(4^j - 1)$. The quantity is too large to sample directly from the first to the last. Nevertheless, each Clifford circuit U is fully characterized by its action on 2^n Pauli operators [64]: $UX_jU^\dagger = (-1)^{r_j}\prod_{i=1}^{n} X_i^{\alpha_{ji}} Z_i^{\beta_{ji}}$ and $UZ_jU^\dagger = (-1)^{s_j}\prod_{i=1}^{n} X_i^{\gamma_{ji}} Z_i^{\delta_{ji}}$. The parameters that define U are $(\alpha, \beta, \gamma, \delta, r, s)$, where $\alpha, \beta, \gamma, \delta$ are $n \times n$ matrices of bits and r, s are n-bit vectors. Given these parameters, different methods can decompose U into elementary circuits [64,66–70]. The time complexity is $O(n^3)$ or $O(n^2)$, while the number of elementary gates is $O(n^2/\log n)$ or $O(n^2)$.

As for the n-qubit MUB ensemble, there are $2^n + 1$ circuits. This allows for a direct uniform (biased) sampling due to their significantly lower count compared to Clifford circuits. In Ref. [49], $2^n + 1$ circuits $\{I, U(0), U(1), \ldots, U(2^n - 1)\}$ produce $2^n + 1$ MUBs by acting on the computational basis. The sampled MUB circuits are $\{I, U^\dagger(0), \ldots, U^\dagger(2^n - 1)\}$. Each nontrivial circuit $U(j)$ is obtained within $O(n^3)$ time. Notably, every circuit comprises at most $(n^2 + 7n)/2$ elementary gates. On average, the counts of gates includes $3n/2$ S gates, $(n^2 - n)/4$ CZ gates, and $(n - u)/2$ CZ gates with a distance u.

2. Circuit structure

For the Clifford ensemble, the circuit is structured with 11-stage decomposition -H-CX-S-CX-S-CX-H-S-CX-S-CX- [66] or 7-stage decomposition -CX-CZ-S-H-S-CZ-CX- [68,70].

For the MUB ensemble, the circuit is structured as -H-S-CZ-. The entanglement CZ component consists of $2n - 3$ fixed modules. Furthermore, the structures among different MUB circuits exhibit a strong correlation. Linear relations exist in these $2^n + 1$ MUB circuits. Each $U(j)$ can be derived from the following n circuits: $\{U(2^0), U(2^1), \ldots, U(2^{n-1})\}$.

B. Postprocessing

In the prediction phase, we require two sets of data: the unitary operations used in the rotation, U_j, and the measurement outcomes $|k\rangle$. With them, one can estimate the expected value of observable O by

$$o \approx \mathrm{tr}[\mathcal{M}^{-1}(U_j^\dagger|k\rangle\langle k|U_j)O]. \tag{23}$$

In the worst case, as n grows, the evaluation of the above expression becomes exponentially slow. However, if the observable is MUB sparse in the MUB samplings or the observable has an efficient representation, such as an efficient stabilizer decomposition in Clifford sampling, the computation complexity decreases to a polynomial of n.

In the case of Clifford ensembles, the number of all stabilizer states $U_j^\dagger|k\rangle$ is $O(2^{n^2/2})$, while the number of MUB states is $2^n(2^n+1)$. Consequently, the number of possible classical snapshots in Eq. (8) decreases for MUB ensembles. The set of all MUB states is a subset of all stabilizer states. In general, the Clifford ensembles prove effective for a broader range of observables, considering the computational complexity involved in obtaining the estimation o.

The MUB circuit decomposition can be realized in all physical platforms. In a quantum optical experiment, one could perform the random Clifford measurements by uniformly projecting a stabilizer state $U_j^\dagger|k\rangle$. To compute all these 2^n coefficients, the time complexity is $O(n\times 2^{3n})$ when we use the expression of stabilizers generators $\{g_i\}_{i=1}^n$, where $U_j^\dagger|k\rangle\langle k|U_j^\dagger = \frac{1}{2^n}\prod_{i=1}^n(I+g_i)$. A new approach reduces the complexity to $O(2^n n^3)$ [71]. If we randomly project a MUB state $|f_k^j\rangle$ in Eq. (2), the time complexity is also $O(2^n n^3)$ and we can directly calculate the coefficients [49,61].

V. CONCLUSION

In this paper we explored classical shadow tomography by uniform and biased sampling MUB circuits. The 2^n+1 MUBs is the minimal and optimal set for full n-qubit state tomography, which also constitutes a subset of all Clifford circuits. A mutually unbiased base circuit is structured with three-stage decomposition -CZ-S-H- [49], a part of the Clifford circuit [70]. There are linear relations between these MUB circuits and the average number of different gates can be counted.

The reconstruction channel and variance were calculated for random uniform and biased sampling of MUB circuits. For the most general observable, the variance is bounded by $O(2^n)$, but when considering a special subset defined as AMA observables, we showed that the upper bound of the variance becomes poly(n), which is comparable to the Clifford case. Furthermore, we found that by biased sampling of MUB circuits, we could effectively decrease the variance to poly(n) when the observable (or the unknown state) is MUB sparse. All these results were demonstrated by numerical experiments.

To characterize whether a random state ρ (observable) is AMA or MUB sparse, we should calculate $d(d+1)$ expectation values $\{\mathrm{tr}(\rho P_{jk})\}$ by definition. However, this approach is not efficient when $d=2^n$ is large. One possible approach is checking the AMA- or MUB-sparse proprieties using random projection. If we randomly select the jth mutually unbiased bases and kth states, the value $\{\mathrm{tr}(\rho P_{jk})\}$ is probably less than $\frac{1}{2^n}+\epsilon$. Otherwise, the value is close to $O(1)$ and we can say that ρ is MUB sparse at that basis. An interesting topic to discuss is how many samples are needed to have a sufficiently high confidence level.

There are many possible research directions for the future regarding the classical shadow tomography with MUB circuits. First, we would like to find an efficient scheme to predict the general observables that are not AMA and MUB sparse. Second, it would be interesting to consider more types of observables such as two-point correlation functions and out-of-time-order correlations. Third, one can also study how to use the MUB circuits to predict nonlinear or polynomial observables such as entanglement entropy.

ACKNOWLEDGMENTS

We would like to thank Z. Chen, J. Sun, and D. Wu for helpful discussion. The work of Y.W. received financial support from the National Natural Science Foundation of China through Grants No. 62001260 and No. 42330707 and from the Beijing Natural Science Foundation under Grant No. Z220002. The work of W.C. was supported by the fellowship of China Postdoctoral Science Foundation through Grant No. 2022M720507 and in part by the Beijing Postdoctoral Research Foundation.

APPENDIX A: COMPUTATION DETAILS OF THE RECONSTRUCTION CHANNEL

In classical shadow tomography, after a MUB measurement $\{U_j^\dagger|k\rangle\langle k|U_j\}_{k=0}^{2^n-1}$, the unknown density matrix ρ can be viewed as collapsing to $P_{jk}=U_j^\dagger|k\rangle\langle k|U_j$, with $|k\rangle$ a state in the computational basis. To obtain the classical shadows $\hat{\rho}$, one needs to know the reconstruction channel $\mathcal{M}^{-1}$. We calculate the reconstruction channels of the MUB measurements for both uniform and biased samplings. The computation details are given below.

1. Uniform sampling

The channel of the MUB measurements is defined as

$$\mathcal{M}_u(\rho)=\frac{1}{2^n+1}\sum_{j=0}^{2^n}\sum_{k=0}^{2^n-1}\mathrm{tr}(\rho P_{jk})P_{jk}. \tag{A1}$$

As the MUB circuits are informationally complete, each ρ can be expressed in the form

$$\rho=\sum_{j=0}^{2^n}\sum_{k=0}^{2^n-1}x_{jk}P_{jk}.$$

Note here that the coefficients $\{x_{jk}\}$ may not be unique. It is straightforward to show that

$$\mathrm{tr}(\rho P_{jk})=\mathrm{tr}\left[\left(\sum_{a=0}^{2^n}\sum_{b=0}^{2^n-1}x_{ab}P_{ab}\right)P_{jk}\right] \tag{A2}$$

$$=\sum_{a=j,b=k}\mathrm{tr}(x_{ab}P_{ab}P_{jk}) + \sum_{a=j,b\neq k}\mathrm{tr}(x_{ab}P_{ab}P_{jk})+\sum_{a\neq j}\mathrm{tr}(x_{ab}P_{ab}P_{jk}) \tag{A3}$$

$$=x_{jk}+\underbrace{0+\cdots+0}_{2^n-1}+\frac{1}{2^n}\sum_{a\neq j}\sum_{b=0}^{2^n-1}x_{ab} \tag{A4}$$

$$=x_{jk}+\frac{1}{2^n}\sum_{a=0}^{2^n}\sum_{b=0}^{2^n-1}x_{ab}-\frac{1}{2^n}\sum_{b=0}^{2^n-1}x_{jb} \tag{A5}$$

$$=x_{jk}+\frac{1}{2^n}\mathrm{tr}(\rho)-\frac{1}{2^n}\sum_{b=0}^{2^n-1}x_{jb}. \tag{A6}$$

In Eq. (A4) we use the property in Eq. (1), i.e., the square of the inner product of two eigenstates in different MUBs is $1/2^n$:

$$\mathcal{M}_u(\rho) = \frac{1}{2^n+1}\sum_{j=0}^{2^n}\sum_{k=0}^{2^n-1}\left(x_{jk}+\frac{\mathrm{tr}(\rho)}{2^n}-\frac{1}{2^n}\sum_{b=0}^{2^n-1}x_{jb}\right)P_{jk} \tag{A7}$$

$$= \frac{1}{2^n+1}\left(\rho+\frac{\mathrm{tr}(\rho)(2^n+1)I}{2^n}-\sum_{j=0}^{2^n}\sum_{b=0}^{2^n-1}\frac{x_{jb}}{2^n}I\right) \tag{A8}$$

$$= \frac{1}{2^n+1}\left(\rho+\frac{\mathrm{tr}(\rho)(2^n+1)I}{2^n}-\frac{\mathrm{tr}(\rho)}{2^n}I\right) \tag{A9}$$

$$= \frac{1}{2^n+1}[\rho+\mathrm{tr}(\rho)I]. \tag{A10}$$

Then the inverse channel is given by

$$\mathcal{M}_u^{-1}(\rho) = (2^n+1)\rho-\mathrm{tr}(\rho)I. \tag{A11}$$

It is the same as the reconstruction channel for the Clifford measurements [11].

2. Biased sampling

Without loss of generality, we let the biased sampling basis be $\mathcal{B}_0$. The sampling probability is $\frac{1+m}{2^n+1+m}$ for $U_0 = I$ and the sampling probability for other U_j is $\frac{1}{2^n+1+m}$, where $j \neq 0$. The resulting reconstruction channel for this adjusted sampling process will be

$$\mathcal{M}_b(\rho) = \frac{1}{2^n+1+m}\left(\sum_{j=0}^{2^n}\sum_{k=0}^{2^n-1}\mathrm{tr}(\rho P_{jk})P_{jk} + m\sum_{k=0}^{2^n-1}\mathrm{tr}(\rho P_{0k})P_{0k}\right) \tag{A12}$$

$$= \frac{1}{2^n+1+m}\left(\rho+\mathrm{tr}(\rho)I+m\sum_{k=0}^{2^n-1}\mathrm{tr}(\rho P_{0k})P_{0k}\right). \tag{A13}$$

Then the updated inverse channel will be

$$\mathcal{M}_b^{-1}(\rho) = (2^n+1+m)\left(\rho-\frac{m}{1+m}\sum_k\mathrm{tr}(\rho P_{0k})P_{0k}\right) - \frac{\mathrm{tr}(\rho)I}{1+m}. \tag{A14}$$

APPENDIX B: COMPUTATIONAL DETAILS ON THE VARIANCE

The performance of classical shadow tomography for the unknown state is linearly dependent on the variance defined as

$$\left\|O-\frac{\mathrm{tr}(O)}{2^n}\right\|_{\mathrm{shadow}}^2 = \max_{\sigma\ \mathrm{state}}\mathbb{E}_{U\sim\mathcal{U}}\sum_{b\in\{0,1\}^n}\langle b|U\sigma U^\dagger|b\rangle \times\langle b|U\mathcal{M}^{-1}\left(O-\frac{\mathrm{tr}(O)}{2^n}\right)U^\dagger|b\rangle^2, \tag{B1}$$

where U is a randomly chosen circuit in the unitary ensemble $\mathcal{U}$. From the definition, the variance depends on the observable O and choice of $\mathcal{U}$, but it also depends implicitly on how one samples the circuits. In this Appendix we calculate this variance when $\mathcal{U}$ is MUB circuits for both uniform and biased sampling. When sampling uniformly MUBs for the unknown state σ, the variance is represented as $\|O_0\|_{\sigma,u}^2$. While sampling with bias, the variance is denoted by $\|O_0\|_{\sigma,b}^2$.

1. Uniform sampling

For the traceless part $O_0 = O-\frac{\mathrm{tr}(O)}{2^n}$, we know that $\mathrm{tr}(O_0)=0$. The reconstruction channel is

$$\mathcal{M}_u^{-1}(O_0) = (2^n+1)O_0-\mathrm{tr}(O_0)I = (2^n+1)O_0.$$

We obtain that

$$\|O_0\|_{\mathrm{shadow},u}^2 = \max_{\sigma\ \mathrm{state}}\|O_0\|_{\sigma,u}^2 \tag{B2}$$

$$= \max_{\sigma\ \mathrm{state}}\frac{1}{2^n+1}\sum_{j=0}^{2^n}\sum_{k=0}^{2^n-1}\mathrm{tr}(\sigma P_{jk})\mathrm{tr}^2\times[(2^n+1)O_0P_{jk}] \tag{B3}$$

$$= \max_{\sigma\ \mathrm{state}}(2^n+1)\sum_{j=0}^{2^n}\sum_{k=0}^{2^n-1}\mathrm{tr}^2(O_0P_{jk})\mathrm{tr}(\sigma P_{jk}). \tag{B4}$$

Proposition 2. $\sum_{j,k}\mathrm{tr}^2(O_0P_{jk}) = \mathrm{tr}(O_0^2)$.

Proof. By Eq. (A1) we have

$$\sum_{j=0}^{2^n}\sum_{k=0}^{2^n-1}\mathrm{tr}(O_0P_{jk})P_{jk} = (2^n+1)\mathcal{M}_u(O_0) = O_0-\mathrm{tr}(O_0)I = O_0. \tag{B5}$$

Thus $\sum_{j,k}\mathrm{tr}(O_0P_{jk}) = \mathrm{tr}(O_0) = 0$.

Define $O_{0,j} = \sum_{k=0}^{2^n-1}\mathrm{tr}(O_0P_{jk})P_{jk}$. Thus $O_0 = \sum_j O_{0,j}$. It is easy to prove that $\mathrm{tr}(O_{0,j}^2) = \sum_{k=0}^{2^n-1}\mathrm{tr}^2(O_0P_{jk})$. If $j \neq j'$ we have

$$\mathrm{tr}(O_{0,j}O_{0,j'}) = \frac{1}{2^n}\sum_{k,k'}\mathrm{tr}(O_0P_{jk})\mathrm{tr}(O_0P_{j'k'}) \tag{B6}$$

$$= \frac{1}{2^n}\mathrm{tr}\left(O_0\sum_k P_{jk}\right)\mathrm{tr}\left(O_0\sum_{k'}P_{j'k'}\right) \tag{B7}$$

$$= 0. \tag{B8}$$

Here we use $\mathrm{tr}(O_0I)=0$. Thus $\mathrm{tr}(O_0^2) = \sum_j\mathrm{tr}(O_{0,j}^2) = \sum_j\sum_k\mathrm{tr}^2(O_0P_{jk})$. ∎

Here, we give the proof of Eq. (15).

Proof. As the observable O is AMA, we have $|\mathrm{tr}(OP_{jk}) - 1/2^n| \leqslant \epsilon$ for all j,k. By $\mathrm{tr}(O_0P_{jk}) = \mathrm{tr}(OP_{jk}) - 1/2^n$, we have $\mathrm{tr}^2(O_0P_{jk}) \leqslant \epsilon^2$. For each j, $\sum_{k=0}^{2^n-1}\mathrm{tr}(\sigma P_{jk}) = 1$. Then $\|O_0\|^2_{\sigma,\mathrm{u}} \leqslant (2^n+1)\epsilon^2 \cdot \sum_{j,k}\mathrm{tr}(\sigma P_{jk}) = (2^n+1)^2\epsilon^2$. We know $\epsilon = O(\mathrm{poly}(n))/2^n$. Thus $\|O_0\|^2_{\sigma,\mathrm{u}} \leqslant (1+\frac{1}{2^n})^2\mathrm{poly}(n)$. ■

Now, we give the proof of (16).

Proof. As the unknown state σ is AMA, we have $\mathrm{tr}(\sigma P_{jk}) \leqslant \epsilon + \frac{1}{2^n}$ for all j,k. Then $\|O_0\|^2_{\sigma,\mathrm{u}} \leqslant (2^n+1)(\epsilon+\frac{1}{2^n})\sum_{j,k}\mathrm{tr}^2(O_0P_{jk}) = (2^n+1)(\epsilon+\frac{1}{2^n})\mathrm{tr}(O_0^2)$. Thus $\|O_0\|^2_{\sigma,\mathrm{u}} \leqslant (1+\frac{1}{2^n})(1+\mathrm{poly}(n))\mathrm{tr}^2(O_0)$.

When $\epsilon = O(1/2^n)$ and $\mathrm{tr}(O_0^2)$ is bounded, the variance for these states σ is at a constant level. When $\sigma = I/2^n$ we have $\epsilon = 0$ and $\|O_0\|^2_{\sigma,I/2^n} \leqslant \frac{2^n+1}{2^n}\mathrm{tr}(O_0^2)$. ■

2. Biased sampling

Here we give the proof of Theorem 2.

Proof. The inverse reconstruction channel for biased sampling is

$$\mathcal{M}_b^{-1}(O_0) = (2^n+1+m)\left(O_0 - \frac{m}{1+m}\sum_k \mathrm{tr}(O_0P_{0k})P_{0k}\right). \tag{B9}$$

Define $O_{0,m} = O_0 - \frac{m}{1+m}\sum_k \mathrm{tr}(O_0P_{0k})P_{0k}$. The variance for state σ changes to

$$\|O_0\|^2_{\sigma,\mathrm{b}} = (2^n+1+m)^2\sum_{j=1}^{2^n}\sum_{k=0}^{2^n-1}\mathrm{tr}^2(O_{0,m}P_{jk})\frac{\mathrm{tr}(\sigma P_{jk})}{2^n+1+m} + (2^n+1+m)^2\sum_{k=0}^{2^n-1}\mathrm{tr}^2(O_{0,m}P_{0k})\frac{(m+1)\mathrm{tr}(\sigma P_{0k})}{2^n+1+m}.$$

We have the following relations. When $j=0$, $\mathrm{tr}(O_{0,m}P_{0k}) = \frac{1}{1+m}\mathrm{tr}(O_0P_{0k})$. When $j \neq 0$,

$$\mathrm{tr}(O_{0,m}P_{jk}) = \mathrm{tr}(O_0P_{jk}) - \frac{1}{2^n}\frac{m}{1+m}\sum_{k'}\mathrm{tr}(O_0P_{0k'}) \tag{B10}$$

$$= \mathrm{tr}(O_0P_{jk}). \tag{B11}$$

With $\mathrm{tr}(O_0I) = 0$, we can rewrite the variance as

$$\|O_0\|^2_{\sigma,b} = (2^n+1+m)\sum_{j=1}^{2^n}\sum_{k=0}^{2^n-1}\mathrm{tr}^2(O_0P_{jk})\mathrm{tr}(\sigma P_{jk}) + \frac{1}{1+m}(2^n+1+m)\sum_{k=0}^{2^n-1}\mathrm{tr}^2(O_0P_{0k})\mathrm{tr}(\sigma P_{0k}). \tag{B12}$$

Now we give the proof of Eq. (19). Given the observable $O = |\phi\rangle\langle\phi|$ and any unknown state σ, we know that $\mathrm{tr}(O_0P_{0k}) \leqslant \max_{|k\rangle}[\mathrm{tr}(OP_{0k}) - \frac{1}{2^n}] \leqslant 1$ and $\mathrm{tr}(O_0P_{jk})^2 \leqslant \frac{(t-1)^2}{2^n2^n}$ for $j \neq 0$ and thus

$$\|O_0\|^2_{\sigma,b} \leqslant (2^n+1+m)\left[\frac{(t-1)^2}{2^n\times 2^n}\times 2^n + \frac{1}{1+m}\max_{|k\rangle}\left(\mathrm{tr}(OP_{0k}) - \frac{1}{2^n}\right)^2\right] \leqslant (2^n+1+m)\left(\frac{(t-1)^2}{2^n} + \frac{1}{1+m}\right). \tag{B13}$$

■

Now we give the proof of Eq. (21). Given the observable O and unknown state $\sigma = |\phi\rangle\langle\phi|$, we know that $\mathrm{tr}(\sigma P_{0k}) \leqslant 1$ and $\mathrm{tr}(\sigma P_{jk}) \leqslant \frac{t}{2^n}$ for $j \neq 0$ and thus

$$\|O_0\|^2_{\sigma,b} \leqslant (2^n+1+m)\left(\frac{t}{2^n}\sum_{j=1}^{2^n}\mathrm{tr}\big(O_{0,j}^2\big) + \frac{1}{1+m}\mathrm{tr}\big(O_{0,0}^2\big)\right) \leqslant (2^n+1+m)\left(\frac{t}{2^n} + \frac{1}{1+m}\right)\mathrm{tr}(O_0^2). \tag{B14}$$

[1] U. Leonhardt, Quantum-state tomography and discrete Wigner function, Phys. Rev. Lett. **74**, 4101 (1995).

[2] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, Measurement of qubits, Phys. Rev. A **64**, 052312 (2001).

[3] *Quantum State Estimation*, edited by M. G. A. Paris and J. Řeháček, Lecture Notes in Physics Vol. 649 (Springer Science + Business Media, New York, 2004).

[4] M. P. da Silva, O. Landon-Cardinal, and D. Poulin, Practical characterization of quantum devices without tomography, Phys. Rev. Lett. **107**, 210404 (2011).

[5] S. T. Flammia and Y.-K. Liu, Direct fidelity estimation from few Pauli measurements, Phys. Rev. Lett. **106**, 230501 (2011).

[6] T. Brydges, A. Elben, P. Jurcevic, B. Vermersch, C. Maier, B. P. Lanyon, P. Zoller, R. Blatt, and C. F. Roos, Probing Rényi entanglement entropy via randomized measurements, Science **364**, 260 (2019).

[7] J. Preskill, Quantum computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

[8] L. Amico, R. Fazio, A. Osterloh, and V. Vedral, Entanglement in many-body systems, Rev. Mod. Phys. **80**, 517 (2008).

[9] S. Aaronson, in *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, Los Angeles, 2018* (ACM, New York, 2018), pp. 325–338.

[10] S. Aaronson and G. N. Rothblum, in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, Phoenix, 2019* (ACM, New York, 2019), pp. 322–333.

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

[12] A. Zhao, N. C. Rubin, and A. Miyake, Fermionic partial tomography via classical shadows, Phys. Rev. Lett. **127**, 110504 (2021).

[13] H.-Y. Hu and Y.-Z. You, Hamiltonian-driven shadow tomography of quantum states, Phys. Rev. Res. **4**, 013054 (2022).

[14] H.-Y. Hu, S. Choi, and Y.-Z. You, Classical shadow tomography with locally scrambled quantum dynamics, Phys. Rev. Res. **5**, 023027 (2023).

[15] A. A. Akhtar, H.-Y. Hu, and Y.-Z. You, Scalable and flexible classical shadow tomography with tensor networks, Quantum **7**, 1026 (2023).

[16] M. Ippoliti, Y. Li, T. Rakovszky, and V. Khemani, Operator relaxation and the optimal depth of classical shadows, Phys. Rev. Lett. **130**, 230403 (2023).

[17] K. Bu, D. E. Koh, R. J. Garcia, and A. Jaffe, Classical shadows with Pauli-invariant unitary ensembles, npj Quantum Inf. **10**, 6 (2024).

[18] A. Acharya, S. Saha, and A. M. Sengupta, Shadow tomography based on informationally complete positive operator-valued measure, Phys. Rev. A **104**, 052418 (2021).

[19] H. C. Nguyen, J. L. Bönsel, J. Steinberg, and O. Gühne, Optimizing shadow tomography with generalized measurements, Phys. Rev. Lett. **129**, 220502 (2022).

[20] L. Innocenti, S. Lorenzo, I. Palmisano, F. Albarelli, A. Ferraro, M. Paternostro, and G. M. Palma, Shadow tomography on general measurement frames, PRX Quantum **4**, 040328 (2023).

[21] L. E. Fischer, T. Dao, I. Tavernelli, and F. Tacchino, Dual frame optimization for informationally complete quantum measurements, arXiv:2401.18071.

[22] C. Hadfield, Adaptive Pauli shadows for energy estimation, arXiv:2105.12207.

[23] A. Neven, J. Carrasco, V. Vitale, C. Kokail, A. Elben, M. Dalmonte, P. Calabrese, P. Zoller, B. Vermersch, R. Kueng, and B. Kraus, Symmetry-resolved entanglement detection using partial transpose moments, npj Quantum Inf. **7**, 152 (2021).

[24] A. Elben, R. Kueng, H.-Y. R. Huang, R. van Bijnen, C. Kokail, M. Dalmonte, P. Calabrese, B. Kraus, J. Preskill, P. Zoller, and B. Vermersch, Mixed-state entanglement from local randomized measurements, Phys. Rev. Lett. **125**, 200501 (2020).

[25] L. K. Joshi, A. Elben, A. Vikram, B. Vermersch, V. Galitski, and P. Zoller, Probing many-body quantum chaos with quantum simulators, Phys. Rev. X **12**, 011018 (2022).

[26] J. Helsen, M. Ioannou, J. Kitzinger, E. Onorati, A. H. Werner, J. Eisert, and I. Roth, Shadow estimation of gate-set properties from random sequences, Nat. Commun. **14**, 5039 (2023).

[27] A. Seif, Z.-P. Cian, S. Zhou, S. Chen, and L. Jiang, Shadow distillation: Quantum error mitigation with classical shadows for near-term quantum processors, PRX Quantum **4**, 010303 (2023).

[28] J. Schwinger, Unitary operator bases, Proc. Natl. Acad. Sci. USA **46**, 570 (1960).

[29] I. D. Ivonovic, Geometrical description of quantal state determination, J. Phys. A: Math. Gen. **14**, 3241 (1981).

[30] W. K. Wootters and B. D. Fields, Optimal state-determination by mutually unbiased measurements, Ann. Phys. (NY) **191**, 363 (1989).

[31] S. Designolle, P. Skrzypczyk, F. Fröwis, and N. Brunner, Quantifying measurement incompatibility of mutually unbiased bases, Phys. Rev. Lett. **122**, 050402 (2019).

[32] R. B. A. Adamson and A. M. Steinberg, Improving quantum state estimation with mutually unbiased bases, Phys. Rev. Lett. **105**, 030406 (2010).

[33] G. Lima, L. Neves, R. Guzmán, E. S. Gómez, W. A. T. Nogueira, A. Delgado, A. Vargas, and C. Saavedra, Experimental quantum tomography of photonic qudits via mutually unbiased basis, Opt. Express **19**, 3542 (2011).

[34] H. Maassen and J. B. M. Uffink, Generalized entropic uncertainty relations, Phys. Rev. Lett. **60**, 1103 (1988).

[35] M. A. Ballester and S. Wehner, Entropic uncertainty relations and locking: Tight bounds for mutually unbiased bases, Phys. Rev. A **75**, 022319 (2007).

[36] S. Massar and P. Spindel, Uncertainty relation for the discrete Fourier transform, Phys. Rev. Lett. **100**, 190401 (2008).

[37] N. J. Cerf, M. Bourennane, A. Karlsson, and N. Gisin, Security of quantum key distribution using d-level systems, Phys. Rev. Lett. **88**, 127902 (2002).

[38] M. Mafu, A. Dudley, S. Goyal, D. Giovannini, M. McLaren, M. J. Padgett, T. Konrad, F. Petruccione, N. Lütkenhaus, and A. Forbes, Higher-dimensional orbital-angular-momentum-based quantum key distribution with mutually unbiased bases, Phys. Rev. A **88**, 032305 (2013).

[39] A. R. Calderbank, E. M. Rains, P. W. Shor, and N. J. A. Sloane, Quantum error correction and orthogonal geometry, Phys. Rev. Lett. **78**, 405 (1997).

[40] A. R. Calderbank, E. M. Rains, P. M. Shor, and N. J. A. Sloane, Quantum error correction via codes over GF(4), IEEE Trans. Inf. Theory **44**, 1369 (1998).

[41] D. Gottesman, in *Quantum Computing and Quantum Communications*, edited by C. P. Williams, Lecture Notes in Computer Science (Springer, Berlin, 1998), Vol. 1509, pp. 302–313.

[42] C. Spengler, M. Huber, S. Brierley, T. Adaktylos, and B. C. Hiesmayr, Entanglement detection via mutually unbiased bases, Phys. Rev. A **86**, 022311 (2012).

[43] D. Giovannini, J. Romero, J. Leach, A. Dudley, A. Forbes, and M. J. Padgett, Characterization of high-dimensional entangled systems via mutually unbiased measurements, Phys. Rev. Lett. **110**, 143601 (2013).

[44] L. Maccone, D. Bruß, and C. Macchiavello, Complementarity and correlations, Phys. Rev. Lett. **114**, 130401 (2015).

[45] P. Erker, M. Krenn, and M. Huber, Quantifying high dimensional entanglement with two mutually unbiased bases, Quantum **1**, 22 (2017).

[46] J. Kaniewski, I. Šupić, J. Tura, F. Baccari, A. Salavrakos, and R. Augusiak, Maximal nonlocality from maximal entanglement and mutually unbiased bases, and self-testing of two-qutrit quantum systems, Quantum **3**, 198 (2019).

[47] A. Tavakoli, M. Farkas, D. Rosset, J.-D. Bancal, and J. Kaniewski, Mutually unbiased bases and symmetric informationally complete measurements in Bell experiments, Sci. Adv. **7**, eabc3847 (2021).

[48] Q. Zhang, Q. Liu, and Y. Zhou, Minimal Clifford shadow estimation by mutually unbiased bases, arXiv:2310.18749.

[49] W. Yu and W. Dongsheng, An efficient quantum circuit construction method for mutually unbiased bases in n-qubit systems, arXiv:2311.11698.

[50] G. M. d'Ariano, P. Perinotti, and M. F. Sacchi, Informationally complete measurements and group representation, J. Opt. B **6**, S487 (2004).

[51] S. T. Flammia, A. Silberfarb, and C. M. Caves, Minimal informationally complete measurements for pure states, Found. Phys. **35**, 1985 (2005).

[52] A. Acharya, S. Saha, and A. M. Sengupta, Informationally complete POVM-based shadow tomography, arXiv:2105.05992.

[53] J. Řeháček, B.-G. Englert, and D. Kaszlikowski, Minimal qubit tomography, Phys. Rev. A **70**, 052321 (2004).

[54] G. N. M. Tabia, Experimental scheme for qubit and qutrit symmetric informationally complete positive operator-valued measurements using multiport devices, Phys. Rev. A **86**, 062107 (2012).

[55] Z. Bian, J. Li, H. Qin, X. Zhan, R. Zhang, B. C. Sanders, and P. Xue, Realization of single-qubit positive-operator-valued measurement via a one-dimensional photonic quantum walk, Phys. Rev. Lett. **114**, 203602 (2015).

[56] Y.-Y. Zhao, N.-K. Yu, P. Kurzyński, G.-Y. Xiang, C.-F. Li, and G.-C. Guo, Experimental realization of generalized qubit measurements based on quantum walks, Phys. Rev. A **91**, 042101 (2015).

[57] X. Wang, X. Zhan, Y. Li, L. Xiao, G. Zhu, D. Qu, Q. Lin, Y. Yu, and P. Xue, Generalized quantum measurements on a higher-dimensional system via quantum walks, Phys. Rev. Lett. **131**, 150803 (2023).

[58] M. Cao, T. Deng, and Y. Wang, Dynamical quantum state tomography with time-dependent channels, J. Phys. A: Math. Theor. **57**, 215301 (2024).

[59] 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).

[60] P. Horodecki, Ł. Rudnicki, and K. Życzkowski, Five open problems in quantum information theory, PRX Quantum **3**, 010101 (2022).

[61] T. Durt, B.-G. Englert, I. Bengtsson, and K. Życzkowski, On mutually unbiased bases, Int. J. Quantum Inf. **08**, 535 (2010).

[62] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2010).

[63] W.-Z. Yan, Y. Li, Z. Hou, H. Zhu, G.-Y. Xiang, C.-F. Li, and G.-C. Guo, Experimental demonstration of inequivalent mutually unbiased bases, Phys. Rev. Lett. **132**, 080202 (2024).

[64] D. E. Gottesman, Stabilizer codes and quantum error correction, Ph.D. thesis, California Institute of Technology, 1997.

[65] L. Zhang, Matrix integrals over unitary groups: An application of Schur-Weyl duality, arXiv:1408.3782.

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

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

[68] D. Maslov and M. Roetteler, Shorter stabilizer circuits via Bruhat decomposition and quantum circuit transformations, IEEE Trans. Inf. Theory **64**, 4729 (2018).

[69] E. Van Den Berg, in *Proceedings of the 2021 IEEE International Conference on Quantum Computing and Engineering, Broomfield* (IEEE, Piscataway, 2021), pp. 54–59.

[70] S. Bravyi and D. Maslov, Hadamard-free circuits expose the structure of the Clifford group, IEEE Trans. Inf. Theory **67**, 4546 (2021).

[71] G. I. Struchalin, Y. A. Zagorovskii, E. V. Kovlakov, S. S. Straupe, and S. P. Kulik, Experimental estimation of quantum state properties from classical shadows, PRX Quantum **2**, 010307 (2021).