Minimal-Clifford shadow estimation by mutually unbiased bases

Qingyue Zhang,[1] Qing Liu,[1] and You Zhou[1,2,*]

[1]*Key Laboratory for Information Science of Electromagnetic Waves (Ministry of Education),*
Fudan University, Shanghai 200433, China
[2]*Hefei National Laboratory, Hefei 230088, China*

Predicting properties of large-scale quantum systems is crucial for the development of quantum science and technology. Shadow estimation is an efficient method for this task based on randomized measurements, where many-qubit random Clifford circuits are used for estimating global properties like quantum fidelity. Here we introduce the minimal Clifford measurement (MCM) to reduce the number of possible random circuits to the minimum, while keeping the effective postprocessing channel in shadow estimation. In particular, we show that MCM requires $2^n + 1$ distinct Clifford circuits, and it can be realized by mutually unbiased bases, with n as the total qubit number. By applying the Z-tableau formalism, this ensemble of circuits can be synthesized to the $-\mathrm{S} - \mathrm{CZ} - \mathrm{H}-$ structure, which can be decomposed to $2n - 1$ *fixed* circuit modules, and the total circuit depth is at most $n + 1$. Compared to the original Clifford measurements, our MCM reduces the circuit complexity and the compilation costs. In addition, we find the sampling advantage of MCM on estimating off-diagonal operators, and extend this observation to the biased-MCM scheme to enhance the sampling improvement further.

I. INTRODUCTION

Learning quantum systems is of both fundamental and practical interests for quantum physics and quantum information processing [1,2]. With the increase on the qubit number in various platforms, from quantum networks to quantum simulators and computers [3,4], it is crucial to develop efficient tools to benchmark them [5,6], from characterizing quantum noises [7] to measuring interesting properties, such as entanglement entropy [8] and out-of-time-ordered correlator [9]. Shadow tomography [10] is a recently proposed framework to predict many properties of quantum objects, like quantum states and channels with randomized measurements [11]. Compared to traditional quantum tomography aiming to reconstruct quantum objects [12,13], the key point of shadow estimation is to build a few classical snapshots of the original quantum object [14], which are utilized to estimate many observables in a multiplex way.

The performance of shadow estimation depends on the applied random unitary evolution on the unknown state and the observables to be predicted [14–16]. There are two primary random unitary ensembles [14]. Pauli measurements enabled by independent single-qubit random unitary rotations are efficient in predicting local observables with

a constant nontrivial support. On the other hand, Clifford measurements using random Clifford evolution on all n qubits, are efficient for global low-rank observables, such as the fidelity to some entangled states. The Pauli measurement and its variants are more feasible to realize [17–22], with a few experimental demonstrations [23–25].

In contrast, the development and realization of the Clifford measurement is more challenging, which is mainly due to the complex structure of the Clifford group with an astronomical number of about $O(4^{n^2})$ of elements [26,27]. In addition, it generally needs $O(n)$-depth quantum circuit with considerable sampling [28,29] and compiling efforts [30], which further hinders the applications on the near-term quantum platform. There are positive progresses, for example, applying random local quantum gates sequentially to approximate full Clifford measurements [15,31] or using emergent design with the assistance of a large number of auxiliary qubits [32,33]. However, the performance is generally guaranteed in the average case [31], and the classical postprocessing is in an approximate and empirical manner [15].

To accelerate the application of Clifford measurement, in this work, we propose the minimal Clifford measurement (MCM) framework for shadow estimation, to avoid the exhaustion of the full Clifford group. More specifically, we reduce the number of sampled Clifford circuits to its minimum, while the postprocessing maintains the same simple form as the original one. In particular, we prove

*Corresponding author: you_zhou@fudan.edu.cn, zyqphy@gmail.com

064001-1

that such a minimal set should contain $2^n + 1$ elements, which can be realized by mutually unbiased bases (MUBs) (Sec. III). We further give a general routine to synthesize these Clifford circuits with the help of tableau formalism, and the final circuit structure is in the $-S - CZ - H-$ form with the depth at most $n + 1$ (actually can be further reduced to $n/2$ [30]) over the all-to-all architecture. Even though the circuit depth still scales with n, S and CZ gates are diagonal in the computational basis, which may be realized simultaneously [34], potentially with the help of an Ising-type Hamiltonian [35]. Most significantly, the $-S - CZ-$ part can be further decomposed to $2n - 1$ fixed modules with an explicit subcircuit structure. Such module decomposition enables the sampling and synthesis of random Clifford circuits on the module level in a unified manner, which is more feasible to benchmark and improve experimentally [36] (Sec. IV). We give a thorough performance analysis of MCM shadow analytically and numerically. In particular, we find our approach shows some advantage for estimating off-diagonal observables, and relate the variance of the estimation to the *coherence* of the observable (Sec. V). Furthermore, we develop the biased-MCM protocol as a mimic of biased-Pauli measurement [17,18] in the Clifford scenario and demonstrate its advantages. In particular, we show that the optimal variance of biased-MCM is directly quantified by the stabilizerness norm [37] of the observable, which is a useful measure of magic [38,39] that may lead to potential applications in fidelity estimation (Sec. VI). To visually illustrate the process of our work, we refer to Fig. 1 as the roadmap of MCM shadow estimation.

II. PRELIMINARIES FOR SHADOW ESTIMATION AND CLIFFORD MEASUREMENTS

Here we first give a brief review [14] on the paradigm of shadow estimation and Clifford group [40]. For the quantum experiment, an unknown n-qubit quantum state $\rho \in \mathcal{H}_d$ with $d = 2^n$ is evolved under a unitary U, which is randomly selected from some ensemble $\mathcal{E}$ to $\rho \mapsto U\rho U^\dagger$. After that, it is measured in the computational basis to get the result $\mathbf{b} \in \{0, 1\}^n$. According to Born's rule, the corresponding probability is $\Pr(\mathbf{b} \mid U) = \langle \mathbf{b}|U\rho U^\dagger|\mathbf{b}\rangle$. Note that both U and $\mathbf{b}$ are random variables and the whole process is indeed random.

For the classical postprocessing, one "prepares" $U^\dagger |\mathbf{b}\rangle \langle \mathbf{b}| U$ on the classical computer, and the effective process can be written as a quantum channel. For a fixed U, that is, by taking the expectation first on $\mathbf{b}$, we denote the effective channel as

$$\mathcal{M}(\rho \mid U) := \mathbb{E}_{\mathbf{b}} \, U^\dagger |\mathbf{b}\rangle \langle \mathbf{b}| U$$
$$= \sum_{\mathbf{b}} \mathrm{tr}[\rho U^\dagger |\mathbf{b}\rangle \langle \mathbf{b}| U] U^\dagger |\mathbf{b}\rangle \langle \mathbf{b}| U, \quad (1)$$
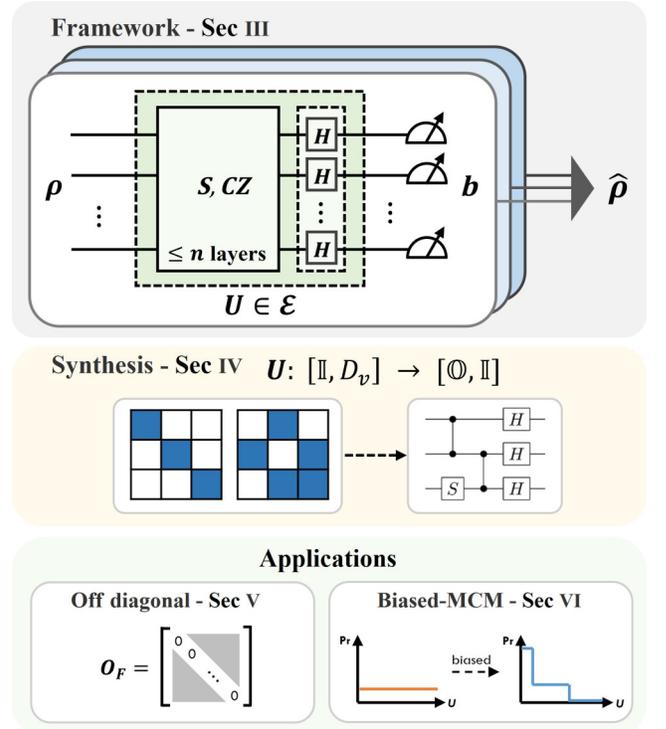
FIG. 1.   The outline of MCM shadow estimation.

where the conditional probability $\Pr(\mathbf{b} \mid U)$ is in the trace form for later convenience. Thus the whole channel is obtained by further taking the expectation on U,

$$\mathcal{M}_{\mathcal{E}}(\rho) := \mathbb{E}_{U \in \mathcal{E}} \, \mathcal{M}(\rho \mid U). \quad (2)$$

If the measurement is tomographically (over)complete, i.e., one takes sufficient distinct U for evolution, the whole information of ρ should be preserved. In other words, the channel $\mathcal{M}_{\mathcal{E}}$ can be reversed mathematically, and one can construct the classical snapshot as

$$\hat{\rho} = \mathcal{M}_{\mathcal{E}}^{-1} \, (U^\dagger |\mathbf{b}\rangle \langle \mathbf{b}| U). \quad (3)$$

It is direct to check that $\mathbb{E}\hat{\rho} = \rho$ by Eqs. (1) and (2), and $\hat{O} := \mathrm{tr}[\hat{\rho}O]$ is an unbiased estimator of $\bar{O} := \mathrm{tr}[\rho O]$. For applications, one can construct the shadow set containing a few of these independent snapshots $\{\hat{\rho}_i\}$ to predict many properties $\{O_j\}$.

There are two prominent unitary ensembles, that is, n-qubit random Clifford ensemble $\mathcal{E}_{\mathrm{Cl}}$ and the tensor product of random single-qubit Clifford gate ensemble $\mathcal{E}_{\mathrm{Pauli}}$. The Pauli measurement by the ensemble $\mathcal{E}_{\mathrm{Pauli}}$ can be realized efficiently in experiments, but it works poorly on estimating global observables, such as the fidelity to some many-qubit entangled states. Clifford measurement $\mathcal{E}_{\mathrm{Cl}}$ is good at this task, with the effective channel and its inverse

for Clifford measurement being

$$\mathcal{M}_{Cl}(A) = (2^n + 1)^{-1}[A + \mathrm{tr}(A)\mathbb{I}],$$
$$\mathcal{M}_{Cl}^{-1}(A) = (2^n + 1)A - \mathrm{tr}(A)\mathbb{I}. \tag{4}$$

Every coin has two sides. Random Clifford unitary is challenging to realize on current quantum platforms, as it requires quite a few two-qubit gates. Current methods evenly sample and compile a Clifford to $O(n)$-depth quantum circuit with time complexity $O(n^2)$ [41,42]. Note that the synthesis of Clifford circuits is intricately tied to the qubit connectivity of platforms. Over linear-nearest-neighbor architecture, the upper bound of two-qubit circuit depth is $7n - 4$, while for all-to-all architectures, it suggests $1.5n + O(\log^2(n))$ [43]. In the following results, Secs. III and IV, we aim to simplify the circuit construction of Clifford measurement to its minimal form with the help of MUB for the all-to-all architecture.

At the end of this section, we recast the essentials of Clifford unitary and Pauli group for later use. The Pauli group for n-qubit quantum system is $\mathbb{P}^n = \{\pm\sqrt{\pm 1} \otimes \{\mathbb{I}_2, X, Y, Z\}\}^{\otimes n}$, with $\mathbb{I}_2$, X, Y, Z being the identity and Pauli operators for single qubit. For convenience, we denote the quotient group without the phase as $\mathbf{P}^n = \mathbb{P}^n / \pm \sqrt{\pm 1}$, and all nonidentity Pauli operator as $\mathbf{P}_*^n = \mathbf{P}^n \setminus \mathbb{I}$. Clifford unitary is the normalizer of $\mathbb{P}^n$, i.e., $U^\dagger P U \in \mathbb{P}^n$, $\forall P \in \mathbb{P}^n$, and a specific Clifford unitary is determined by its action on $U^\dagger X_i U$ and $U^\dagger Z_i U$ for $i \in [n]$, as X_i and Z_i can generate all Pauli operators. Hereafter we use $Z_i(X_i)$ for short to denote an n-qubit operator $Z_i \otimes \mathbb{I}_{[n]/\{i\}}$ ($X_i \otimes \mathbb{I}_{[n]/\{i\}}$) with the ith qubit being nonidentity.

III. MINIMAL CLIFFORD MEASUREMENT FROM MUTUALLY UNBIASED BASES

To simplify the full Clifford measurement in the original shadow estimation, we raise the following task, which aims to minimize the size of the Clifford group while preserving the effective quantum channel $\mathcal{M}_{Cl}$.

Task 1. Suppose $\mathcal{E}$ is a subset of the full Clifford group $\mathcal{E}_{Cl}$, the task reads

$$\textbf{min:} \quad |\mathcal{E}|$$
$$\textbf{such that:} \quad \forall \rho, \frac{1}{|\mathcal{E}|}\sum_{U \in \mathcal{E}} \mathcal{M}(\rho \mid U) = \mathcal{M}_{Cl}(\rho), \tag{5}$$

where $\mathcal{M}_{Cl}(\rho)$ is the effective quantum channel shown in Eq. (4).

First of all, we show a lower bound of $|\mathcal{E}|$ by introducing a lemma that accounts for the effect of individual Clifford unitary. After the Clifford U rotation, the final measurement is conducted in the computational basis, that is, the Z

basis. As a result, we need only to care about the action of U on the Z basis. Define the generators in the Heisenberg picture as

$$g_i = U^\dagger Z_i U, \quad i \in [n]. \tag{6}$$

In other words, different Clifford U with the same (ignoring the phase) $\langle g_i \rangle_{i=0}^{n-1}$ lead to the same measurement setting in the shadow estimation, as shown explicitly in the following Lemma 1.

Lemma 1. Denote the operators generated by $\langle g_i \rangle_{i=0}^{n-1}$ given in Eq. (6) as $S_{\mathbf{m}} = \prod_{i=0}^{n-1} g_i^{m_i}$ with $\mathbf{m}$ an n-bit vector, the quantum channel $\mathcal{M}(\rho \mid U)$ defined in Eq. (1) shows

$$\mathcal{M}(\rho \mid U) = 2^{-n} \sum_{\mathbf{m} \in \{0,1\}^n} \mathrm{tr}(\rho S_{\mathbf{m}}) S_{\mathbf{m}}. \tag{7}$$

The proof is left to Appendix A 1, which transforms the summation of computational basis $\mathbf{b}$ to the operators $S_{\mathbf{m}}$. In fact, $S_{\mathbf{m}}$ are the stabilizers for the state $|\Psi_0\rangle = U^\dagger |0\rangle$ with $S_{\mathbf{m}} |\Psi_0\rangle = |\Psi_0\rangle$, $\forall \mathbf{m}$. g_i and $S_{\mathbf{m}}$ are Hermitian operators by definition, thus the possible phase does not affect the measurement setting in Eq. (7). Hereafter, we ignore the global phase of them, that is $g_i, S_{\mathbf{m}} \in \mathbf{P}^n$.

Proposition 1. The cardinality of the unitary ensemble $\mathcal{E}$ in Task 1 is lower bounded by $|\mathcal{E}| \geq 2^n + 1$.

The proof given in Appendix A 2 is based on the fact that one individual Clifford unitary acquires $2^n - 1$ nonidentity Pauli information as shown in Lemma 1, and there are totally $4^n - 1$ ones. In this case, the Clifford elements in $\mathcal{E}$ do not share nonidentity $S_{\mathbf{m}}$, one has $|\mathcal{E}| = (4^n - 1)/(2^n - 1) = 2^n + 1$ to cover all the Paulis for tomographic completeness.

Definition 1. Suppose a subset of the Clifford group denoted by $\mathcal{E}_{\min}$ reaches the lower bound in Proposition 1, we call the corresponding measurement the minimal Clifford measurement (MCM).

Note that MCM simplifies only the Clifford measurement, but keeps the postprocessing. As a result, the realization of MCM shadow estimation follows the same routine as the original one, except the random Clifford circuit ensemble applied before the final projective measurement. Next, by introducing the MUBs [44–46], we find a typical set $\mathcal{E}_{\min} = \mathcal{E}_{\text{MUB}}$ that saturates the lower bound. Here the Clifford unitaries in $\mathcal{E}_{\text{MUB}}$ are written in the form of the stabilizer generators $\langle g_i \rangle$.

Proposition 2. The MUB in Ref. [44] is an MCM, and the generators for all $2^n + 1$ different Clifford unitaries are written by the stabilizer generators as follows. The 0th

element of $\mathcal{E}_{\text{MUB}}$ is $\langle Z_i \rangle_{i=0}^{n-1}$, i.e., the Z-basis measurement; and the other 2^n elements labeled by $v = 0, 1, \ldots, 2^n - 1$ are

$$\left\langle g_i = \sqrt{-1}^{\alpha_{v,i,i}} X_i \bigotimes_{j=0}^{n-1} Z_j^{\alpha_{v,i,j}} \right\rangle, \qquad (8)$$

with $\alpha_{v,i,j} = [(v \odot 2^i) M_n^{(0)}]_j$ being a binary value. Here $\odot$ denotes the multiplication in Galois field $GF(2^n)$, $(\cdot)$ transforms the number to an n-bit string, and $M_n^{(0)}$ is an $n \times n$ binary matrix.

The specific definition of $M_n^{(0)}$ is in Appendix B 1, and the detailed proof of $\mathcal{E}_{\text{MUB}}$ being an MCM is left to Appendix B 2. In fact, Eq. (8) offers another way to represent the MUB previously introduced in Ref. [46].

Some remarks about the quantum design are demonstrated as follows. Note that the measurement channel in Eqs. (1), (2), and (7) are second-order functions of the quantum state, so any (projective) 2-design ensemble returns the same channel as the full Clifford measurement in Eq. (4). It was shown that any covariant-Clifford 2-design of an n-qubit system reaches its minimum as an MUB [47]. As a result, MUB leads to the same channel as the full Clifford. In some sense, our approach here manifests this result from the perspective of the action of individual Clifford elements via Lemma 1.

IV. EFFICIENT CIRCUIT SYNTHESIS OF MCM

In this section, we focus on the sampling and synthesis of Clifford circuit of MCM shadow estimation, especially for the unitary ensemble $\mathcal{E}_{\text{MUB}}$ given in Proposition 2. The synthesis applies tableau formula under the framework of Gottesman-Knill theorem [40,48]. Interestingly, as $|\mathcal{E}_{\text{MUB}}| \ll |\mathcal{E}_{\text{Cl}}|$, there is possible room for further simplification on the quantum circuits. Finally, we give a unified method to sample and synthesize on the module level, which makes the realization of MCM shadow very efficient.

Here we first briefly introduce the simplified Z tableau and leave more details in Appendix B 3. As shown in Sec. III, the essential information is the stabilizer generators $\langle g_i \rangle$ in Eq. (6) without considering the phase factor. In this way, we write

$$\tilde{g}_i = \bigotimes_{j=0}^{n-1} X_j^{\gamma_{ij}} Z_j^{\delta_{ij}}, \quad i \in [n], \qquad (9)$$

where γ_{ij} and δ_{ij} are elements of two $n \times n$ binary matrices C and D, and $T = [C, D]$ is called the Z tableau. The action of Clifford gates $V\tilde{g}_i V^\dagger$ on $\tilde{g}_i$ is thus recorded as the transformation of the matrix T. As a result, our task is to find V such that it takes all $\tilde{g}_i$ back to the original Z_i,

i.e., take T to $T_0 = [\mathbb{O}, \mathbb{I}]$, with $C = \mathbb{O}$ the null matrix, and $D = \mathbb{I}$ the identity matrix. We choose the basic generating Clifford gates as the single-qubit Hadamard gate H, phase gate S, and two-qubit controlled-Z gate CZ, with the corresponding update rules of tableau T listed as follows. For all i,

- $H(a)$: exchange $\gamma_{i,a}$ and $\delta_{i,a}$;
- $S(a)$: $\delta_{i,a} := \gamma_{i,a} + \delta_{i,a}$;
- $CZ(a, b)$: $\delta_{i,a} := \delta_{i,a} + \gamma_{i,b}$, $\delta_{i,b} := \delta_{i,b} + \gamma_{i,a}$,

on qubit a, and qubit pair (a, b), respectively [29,40].

In MCM, one needs to evenly sample a unitary from the ensemble $\mathcal{E}_{\text{MUB}}$ given in Proposition 2. For elements in $\mathcal{E}_{\text{MUB}}$, the tableau of the 0 th element is already $[\mathbb{O}, \mathbb{I}]$, thus we do not need to synthesize; tableaus of the remaining 2^n ones in Eq. (8) are all in the form of $[\mathbb{I}, D_v]$ with v denoting the element label, and $\alpha_{v,i,j} = [D_v]_{i,j}$. In particular, the i th row of D_v is the vector $(v \odot 2^i) M_n^{(0)}$ [46], as shown in Proposition 2.

The circuit sampling and synthesis for all 2^n tableaus in the form of $[\mathbb{I}, D_v]$ is shown as follows. First we introduce a special property for D_v to simplify the procedure. A matrix D is said to be a Hankel matrix if $D_{i,j} = D_{i',j'}$ for any $i + j = i' + j'$. We prove the following result in Appendix B 4.

Proposition 3. The matrix D_v from the tableau $[\mathbb{I}, D_v]$ of the element in $\mathcal{E}_{\text{MUB}}$ given in Eq. (8) is a Hankel matrix, for all $v = 0, 1, \ldots, 2^n - 1$.

Since D_v is a Hankel matrix, it can be decomposed as

$$D_v = \sum_{k=0}^{2n-2} \beta_k^v \mathbb{H}_k, \qquad (10)$$

where $[\mathbb{H}_k]_{i,j} = \delta_{i+j,k}$, i.e., it only has 1's on the k th antidiagonal. In this way, any D_v can be written as a linear combination of at most $(2n - 1)$ Hankel matrices $\{\mathbb{H}_k\}_{k=0}^{2n-2}$, with $(2n - 1)$-bit vector $\vec{\beta}^v$ characterizing the information of elements from Eq. (8).

Circuit sampling. Suppose $v = \sum_{i=0}^{n-1} v_i 2^i$, and the binary representation shows $v = \{v_0, \ldots, v_{n-1}\}$. We randomly sample an n-bit string v, and actually, we need *not* calculate the coefficient β_k^v in Eq. (10) for each D_v one by one, which will determine the final quantum circuit. Denote the corresponding coefficient of each basis vector $\{0, \ldots, 1_i, \ldots, 0\}$ with $i \in [n]$ as $\beta_k^{(i)}$, then by linearity the coefficient of D_v reads

$$\beta_k^v = \sum_i \beta_k^{(i)} v_i, \qquad (11)$$

which is a matrix multiplication with the binary matrix $\{\beta_k^{(i)}\}$ being $(2n - 1) \times n$. Such sampling consumes only n random bits, compared with $O(n^2)$ for full Clifford circuits [29]. We also remark that it needs $O(n \log n)$ complexity
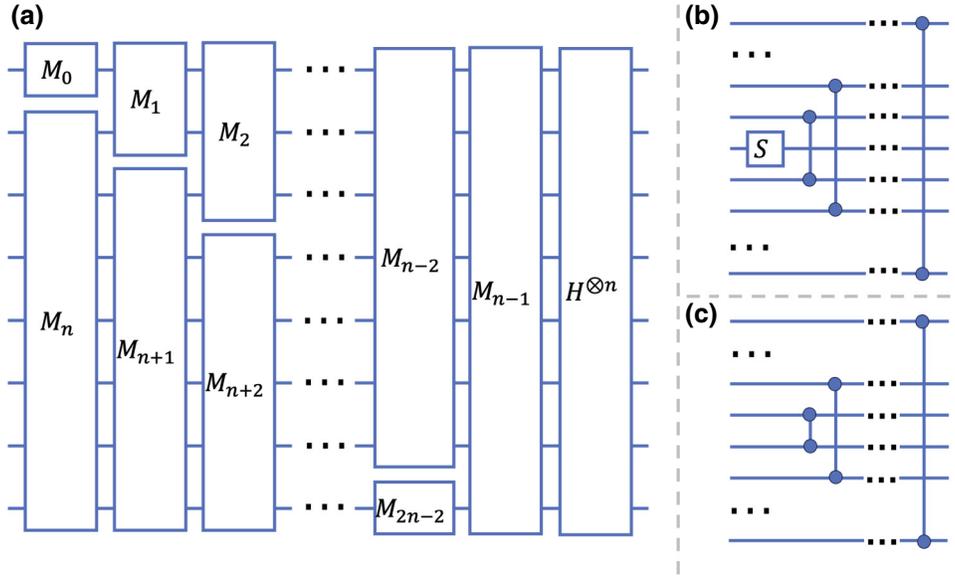
FIG. 2.   An illustration of the quantum circuit structure after the synthesis for tableau $[\mathbb{I}, D_v]$ with $D_v = \sum_{k=0}^{2n-2} \beta_k^v \mathbb{H}_k$. If $\beta_k^v = 1$, apply module M_k as shown in (a), and the module M_k and module M_{2n-2-k} share the same quantum gates. If k is even, the construction of the corresponding modules are shown in (b), with both S and CZ gates; if k is odd, the construction of the corresponding modules is shown in (c), which contains only CZ gates.

to calculate Eq. (11) using fast Fourier transforming since $\{\beta_k^{(i)}\}$ is a Hankel matrix [49].

Circuit synthesis. For each $\mathbb{H}_k$ with $\beta_k = 1$ in the summation in Eq. (10), one can use a circuit module M_k in Fig. 2 satisfying $[\mathbb{I}, \mathbb{H}] \rightarrow [\mathbb{I}, \mathbb{H} - \mathbb{H}_k]$ to transform the tableau from the initial $[\mathbb{I}, D_v]$ gradually to $[\mathbb{I}, \mathbb{O}]$. After that, by applying the Hadamard gates $\mathrm{H}^{\otimes n}$, one can transform the tableau $[\mathbb{I}, \mathbb{O}] \rightarrow [\mathbb{O}, \mathbb{I}]$ to complete the synthesis. Hence we directly obtain a three-stage form $-\mathrm{S} - \mathrm{CZ} - \mathrm{H}-$, and the whole synthesis process is listed in Algorithm 1.

Figure 2 shows the explicit circuit construction using Algorithm 1. The module M_k may contain S and CZ gates on the first k qubit with $k < n$. M_k and M_{2n-2-k} are identical to each other but act on qubits in reverse order. There are in total $2n - 1$ modules, and the circuit depth for all modules can be parallelized to n by combining M_k and M_{n+k} to the same layer.

We provide Fig. 3 as a visual explanation of how to eliminate the Z tableau and synthesize the corresponding quantum circuit. Moreover, we give an example of MUB for $n = 3$, starting from Eq. (8) to the synthesis of quantum circuits, which is left to Appendix B 5.

The whole circuit can be summarized in three stages in the form of $-\mathrm{S} - \mathrm{CZ} - \mathrm{H}-$, by observing that S gates and CZ gates commute. The total circuit depth can be further reduced, by using the result that the upper bound of the circuit depth is $(n/2) + O(\log^2(n))$ for the CZ layer [30]. In addition, the full Clifford group followed by projective measurements can also be effectively synthesized

to a one CZ-layer form, by the canonical form of Ref. [41] with proper mapping on the computational basis vector. However, the modular design as shown Fig. 2 streamlines the sampling and synthesis, and may lead to advantages in the calibration and realization. We also remark that the layer of intermediate CZ gates could be implemented

---

ALGORITHM 1.   Circuit synthesis for Z tableau $[\mathbb{I}, D_v]$.

Input: The qubit number n, and the parameter vector of the Hankel matrix $\{\beta_k\}_{k=0}^{2n-2}$.
Output: The Clifford Circuit $\mathbf{C}$, initialized as $\mathbf{C} = \emptyset$.
1: **for** $k = 0$ **to** $2n - 2$ **do**
2:      **if** $\beta_k = 1$ **then**
3:          **if** k is even **then**
4:              $\mathbf{C} \leftarrow \mathbf{C} \cup \mathbf{S}(\frac{k}{2})$
5:          **end if**
6:          Initialize $p, q = 0$.
7:          **if** $k < n - 1$ **then**
8:              $p = 0, q = k$.
9:          **else**
10:             $p = k - n + 1, q = n - 1$.
11:         **end if**
12:         **while** $p < q$ **do**
13:             $\mathbf{C} \leftarrow \mathbf{C} \cup \mathbf{CZ}(\mathbf{p}, \mathbf{q})$
14:             $p = p + 1, q = q - 1$
15:         **end while**
16:     **end if**
17: **end for**
18: **for** $i = 0$ **to** $n - 1$ **do**
19:     $\mathbf{C} \leftarrow \mathbf{C} \cup \mathbf{H}(\mathbf{i})$
20: **end for**
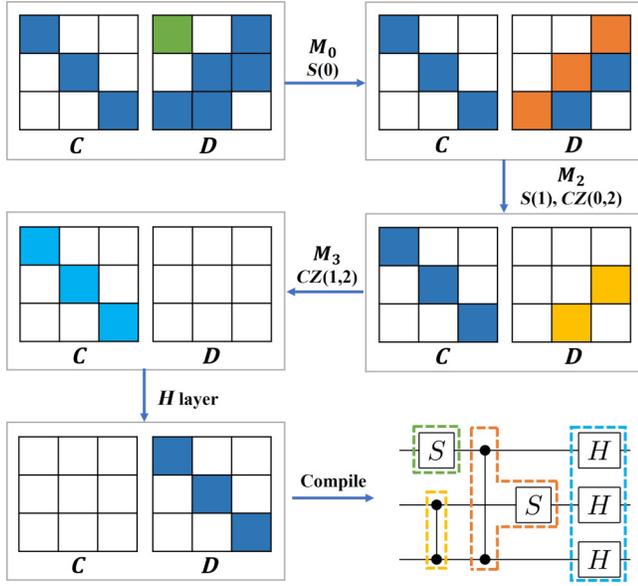
FIG. 3. Illustration of the circuit synthesis in Algorithm 1 given input Z tableau $[C, D]$ when $n = 3$. The D matrix of the input Z tableau is a Hankel matrix. To eliminate the k th antidiagonal of the D matrix, we make use of the k th module called M_k. Finally a fully H layer is utilized to transform the Z tableau $[\mathbb{I}, \mathbb{O}] \to [\mathbb{O}, \mathbb{I}]$.

using one global Mølmer–Sørensen gate [50] systems like on trapped-ion system simultaneously [51]. Finally, we should point out that the frequent use of long-range CZ gates in modules would hinder its application on quantum systems with linear-nearest-neighbor architecture like the superconducting platform.

V. PERFORMANCE ANALYSIS AND OFF-DIAGONAL ADVANTAGE

With the synthesized quantum circuits for MUB at hand, one can proceed the MCM shadow estimation via Eqs. (1)–(3), using the same postprocessing scheme as the full Clifford measurements in Eq. (4). In this section, we analyze the performance of the introduced MCM shadow in depth by investigating the variance for estimating some observable O. In shadow estimation, such variance shows

$$\mathrm{Var}_{\mathcal{E}}(\hat{O}) = \mathrm{Var}_{\mathcal{E}}(\hat{O}_0) \le \max_{\rho} \mathbb{E}\mathrm{tr}(O_0\hat{\rho})^2 = \|O_0\|_{\mathrm{s},\mathcal{E}}^2. \quad (12)$$

Here, $O_0 = O - \mathrm{tr}(O)\mathbb{I}/d$ is the traceless part of O. Note that the variance is further bounded by the square of the shadow norm $\|O_0\|_{\mathrm{s},\mathcal{E}}$ [14]. Shadow norm is a function of the observable O and the unitary ensemble $\mathcal{E}$, by taking the maximization over all possible input states ρ. Another variance quantification is called the locally scrambled shadow norm $\|O\|_{\mathrm{ls},\mathcal{E}} \le \|O\|_{\mathrm{s},\mathcal{E}}$ [15,31,52], by considering

the *average* on ρ from a 1-design ensemble, or equivalently, the input state $\rho = \mathbb{I}/2^n$ being the maximally mixed state.

Theorem 1. For MCM shadow estimation, if one takes the unitary ensemble $\mathcal{E}_{\mathrm{MUB}}$ determined by some MUB, the corresponding shadow norm and locally scrambled shadow norm can be upper bounded by

$$\|O_0\|_{\mathrm{s},\mathcal{E}_{\mathrm{MUB}}}^2 \le (2^n + 1)\mathrm{tr}(O_0^2),$$
$$\|O_0\|_{\mathrm{ls},\mathcal{E}_{\mathrm{MUB}}}^2 = \frac{2^n + 1}{2^n}\mathrm{tr}(O_0^2). \quad (13)$$

Proof. The proof is mainly based on the 2-design property [47] of MUB. Here, we denote the MUB state as $\left|\Phi_{U,\mathbf{b}}\right\rangle = U^\dagger \left|\mathbf{b}\right\rangle$, and the density matrix as $\Phi = \left|\Phi\right\rangle \left\langle\Phi\right|$. By utilizing the inverse channel in Eq. (4), and recalling Eq. (12), the variance can be upper bounded by

$$\mathbb{E}\mathrm{tr}(O_0\hat{\rho})^2 = \mathbb{E}\mathrm{tr}[O_0 \, \mathcal{M}^{-1}(\Phi_{U,\mathbf{b}})]^2$$
$$= (2^n + 1)^2 \, \mathbb{E}\mathrm{tr}[O_0\Phi_{U,\mathbf{b}}]^2$$
$$= (2^n + 1) \sum_{U\in\mathcal{E}_{\mathrm{MUB}},\mathbf{b}} \mathrm{tr}[O_0\Phi_{U,\mathbf{b}}]^2\mathrm{tr}(\rho\Phi_{U,\mathbf{b}})$$
$$\le (2^n + 1)\mathrm{tr}[O_0^{\otimes 2} \sum_{U\in\mathcal{E}_{\mathrm{MUB}},\mathbf{b}} \Phi_{U,\mathbf{b}}^{\otimes 2}]$$
$$= (2^n + 1)\mathrm{tr}[O_0^{\otimes 2} \, (\mathbb{S} + \mathbb{I}^{\otimes 2})] = (2^n + 1)\mathrm{tr}(O_0^2). \quad (14)$$

Here, $\mathbb{S}$ and $\mathbb{I}^{\otimes 2}$ are SWAP and identity operators on the 2-copy space. The inequality is due to the fidelity $\mathrm{tr}(\rho\Phi_{U,\mathbf{b}}) \le 1$. And the final equality is because $\sum \Phi_{U,\mathbf{b}}^{\otimes 2}$ is a projective 2-design, thus the summation result is proportional to the projection to the symmetric subspace $\Pi_{\mathrm{sym}} = 2^{-n}(2^n + 1)^{-1}(\mathbb{S} + \mathbb{I}^{\otimes 2})$ on $\mathcal{H}_d^{\otimes 2}$.

For the locally scrambled shadow norm, one just takes $\rho = \mathbb{I}/2^n$ in Eq. (14), which contributes a 2^{-n} to the final result. ∎

Some remarks on the comparison to the original full Clifford shadow are illustrated as follows. For the worst case, i.e., the original shadow norm is $\|O_0\|_{\mathrm{s},\mathcal{E}_{\mathrm{Cl}}}^2 \sim O(1)\mathrm{tr}(O_0^2)$ and our MCM shadow is exponentially worse than it, the average behavior of both protocols are the same by the 2-design property $\|O_0\|_{\mathrm{ls},\mathcal{E}_{\mathrm{MUB}}}^2 = \|O_0\|_{\mathrm{ls},\mathcal{E}_{\mathrm{Cl}}}^2$.

On the other hand, the bound on $\|O_0\|_{\mathrm{s},\mathcal{E}_{\mathrm{MUB}}}^2$ is tight in some sense, that is, the exponential scaling with the qubit number n is not by mathematical derivation, but the essence of MCM. Let us denote $\langle O_0\rangle_{U,\mathbf{b}} := \mathrm{tr}[O_0\Phi_{U,\mathbf{b}}], \langle\rho\rangle_{U,\mathbf{b}} := \mathrm{tr}(\rho\Phi_{U,\mathbf{b}})$ for short, the third line of
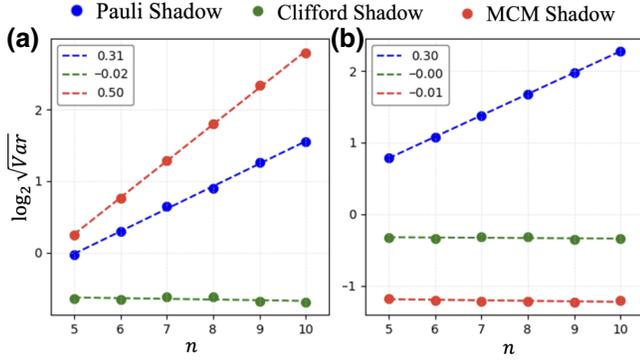
FIG. 4. The statistical variance of shadow estimation with a total number of snapshots $N = 10\,000$ using Pauli measurement (blue), Clifford measurement (green) and minimal Clifford measurement (red), respectively. In both (a) and (b) the state is the standard Greenberger-Horne-Zeilinger (GHZ) state $|GHZ\rangle = \frac{1}{\sqrt{2}}(|0\rangle^{\otimes n} + |1\rangle^{\otimes n})$. In (a) $O = |GHZ\rangle\langle GHZ|$, and in (b) $O_F = [(|0\rangle\langle 1|)^{\otimes n} + |1\rangle\langle 0|)^{\otimes n}]/2$, which is the off-diagonal term of the GHZ state. The dotted lines show the fitting curves with the numerical values, and the numbers on the left top represent the corresponding slopes of these curves.

Eq. (14) can be written as

$$\text{Var}_{\mathcal{E}_{\text{MUB}}}(\hat{O}) \leq (2^n + 1) \sum_{U,\mathbf{b}} \langle O_0 \rangle_{U,\mathbf{b}}^2 \langle \rho \rangle_{U,\mathbf{b}}. \qquad (15)$$

Let us consider the fidelity estimation task, which is the main application of the original shadow estimation. In this task, $O = \Psi$ is some pure (stabilizer) state, and thus $\text{tr}(O_0^2)$ is a constant, then we have the following result.

Observation 1. Suppose there is some $\Phi_{U,\mathbf{b}}$, such that both $\langle O_0 \rangle_{U,\mathbf{b}} = \Theta(1)$ and $\langle \rho \rangle_{U,\mathbf{b}} = \Theta(1)$, then the variance of the MUB-based shadow would scale as $\Theta(2^n)$.

The observation is direct to see, as in this case one of the terms in the summation in Eq. (15) $\langle O_0 \rangle_{U,\mathbf{b}}^2 \langle \rho \rangle_{U,\mathbf{b}} = \Theta(1)$. That is, if O_0 and ρ both share constant overlap with some basis state $\Phi_{U,\mathbf{b}}$, the final result should be exponential with n. Indeed, it is not hard to find examples to support this observation.

Here, in the numerical simulation we take the observable and the input state $O = \rho = |GHZ\rangle\langle GHZ|$ as an example. In this case, there exists $\Phi_{\mathbb{I},\mathbf{0}}$, such that $\langle O_0 \rangle_{\mathbb{I},\mathbf{0}}, \langle \rho \rangle_{\mathbb{I},\mathbf{0}} = \Theta(1)$. That is, the state $\Phi_{\mathbb{I},\mathbf{0}}$ is a diagonal state corresponding to the 0 th element of MUB—the Z basis with $U = \mathbb{I}$. As a result, the variance should be $\Theta(2^n)$, which is also manifested by the numerics in Fig. 4(a).

Observation 1 motivates us to relieve such worst-case exponential-scaling variance of MCM by considering the diagonal and off-diagonal part [53] of the observable O separately, under a chosen basis determined by $U \in \mathcal{E}_{\text{MUB}}$ from all possible $(2^n + 1)$ elements. We demonstrate

the variance for estimating only the off-diagonal part as follows.

Theorem 2. For a given basis from one element of MUB, i.e., $\{|\Phi_{U,\mathbf{b}}\rangle\}$ with a fixed $U \in \mathcal{E}_{\text{MUB}}$, consider the off-diagonal part of observable O, $O_F = \sum_{\mathbf{b} \neq \mathbf{b}'} O_{\mathbf{b},\mathbf{b}'} |\Phi_{U,\mathbf{b}}\rangle\langle\Phi_{U,\mathbf{b}'}|$. The variance of estimating O_F with MCM shadow is upper bounded by

$$\text{Var}_{\mathcal{E}_{\text{MUB}}}(\hat{O}_F) \leq \frac{2^n + 1}{2^n} C_{l_1}(O)^2, \qquad (16)$$

where $C_{l_1}(O) := \sum_{\mathbf{b} \neq \mathbf{b}'} |O_{\mathbf{b},\mathbf{b}'}|$ denotes the l_1 norm of quantum coherence [54].

The proof is left to Appendix C 2. Theorem 2 links the scaling variance to an intrinsic quantum resource, quantum coherence [55]. In terms of applications, it guarantees the performance of MCM via diagonal and off-diagonal strategy when estimating the fidelity of states with a modest degree of superposition, like GHZ states and W states. As such, the variance is at most polynomial if $C_{l_1}(O) = O(\text{poly}(n))$. In practice, one can estimate the diagonal part of O by directly selecting the measurement basis determined by some preferred U, which is easy to conduct. For the challenging part, i.e., the off-diagonal part [16,56], one can apply the MCM-based shadow to estimate it. Finally, by combining the diagonal and off-diagonal results, the whole estimation procedure is finalized.

In the example of estimating the fidelity of a GHZ state, the chosen basis is the Z basis, i.e., $U = \mathbb{I}$. $O_F = \frac{1}{2}(|1\rangle\langle 0|^{\otimes n} + |0\rangle\langle 1|^{\otimes n})$ with $C_{l_1}(O) = 1$. One can measure O_F with MCM shadow and the diagonal part with the Z-basis measurement, respectively. The final variance would not scale with n, which is further manifested by Fig. 4(b).

The off-diagonal advantage shown in Theorem 2 and the showcase of GHZ state lies on the essence that we choose a proper basis from all MUBs as the "proper" diagonal basis, then separately measure the diagonal and off diagonal in standard and shadow measurements. The "proper" basis is chosen such that the constant overlap $\langle O_0 \rangle_{U,\mathbf{b}} = \Theta(1)$ in Observation 1 can be avoided in the shadow estimation. In general, one would like to measure a complex observable, spanning quite a few different MUBs. This situation makes the selection of the "proper" basis frustrating, and this motivates us to consider the biased MCM in the next section.

VI. BIASED-MCM SHADOW ESTIMATION

In this section, we further propose the biased-MCM scheme as an extension and enhancement of the off-diagonal advantage discussed in the previous section.

In the off-diagonal advantage, a reference basis is chosen deterministically in some sense, compared to the other bases that are selected randomly. Here, by leveraging prior

knowledge about the observable to be predicted, we extend this kind of unbiased treatment to another framework that samples MUB elements with varying probabilities. Intuitively, one can increase the probability of choosing some basis if it can reveal more information about the underlying observable. Note that there are biased schemes for Pauli measurements [17,18], but none for Clifford measurements, on account of the large cardinality of the Clifford group. Here our MCM approach makes the biased scheme possible for Clifford measurements.

Compared to the (uniform) MCM shadow, two modifications are made as follows. First, instead of being sampled uniformly, now a unitary $U \in \mathcal{E}_{\mathrm{MUB}}$ is sampled with a given probability P_U. Second, the formula of postprocessing is changed to

$$\widehat{O_0} = \mathrm{tr}(O_0 \frac{U^\dagger |\mathbf{b}\rangle\langle\mathbf{b}| U}{P_U}), \tag{17}$$

with the estimator of O as $\hat{O} = \widehat{O_0} + 2^{-n}\mathrm{tr}(O)$. We demonstrate the unbiasedness of $\hat{O}$ in Appendix D 1.

In general, an arbitrary observable can be decomposed into MUBs as [53,57]

$$O = -\mathrm{tr}(O)\mathbb{I} + \sum_{U \in \mathcal{E}_{\mathrm{MUB}}} \sum_{\mathbf{b} \in \{0,1\}^n} \alpha_{U,\mathbf{b}} \Phi_{U,\mathbf{b}}, \tag{18}$$

where $\alpha_{U,\mathbf{b}} = \mathrm{tr}(O\Phi_{U,\mathbf{b}})$ with $\Phi_{U,\mathbf{b}} = U^\dagger |\mathbf{b}\rangle\langle\mathbf{b}| U$. Based on this canonical decomposition, we analytically find the "optimal" probability distribution to conduct biased MCM. Here optimal means that we can find a solution to optimize the upper bound of the variance in the estimation. In particular, we choose

$$P_U = \frac{B_U}{\sum_{U' \in \mathcal{E}_{\mathrm{MUB}}} B_{U'}}, \tag{19}$$

with

$$B_U := \max_{\mathbf{b} \in \{0,1\}^n} |\mathrm{tr}(\Phi_{U,\mathbf{b}} O_0)|$$
$$= \max_{\mathbf{b} \in \{0,1\}^n} |\alpha_{U,\mathbf{b}} - 2^{-n}\mathrm{tr}(O)|. \tag{20}$$

Based on the biased-MCM approach according to the optimal probability for sampling unitaries in Eqs. (19) and (20), we show the following upper bound of estimation variance.

Theorem 3. For an observable O, if one applies the biased-MCM shadow estimation using the sampling probability P_U given in Eq. (19), the variance of estimation is

upper bounded by

$$\mathrm{Var}_{\mathrm{biased}-\mathcal{E}_{\mathrm{MUB}}}(\hat{O}) \leq \left(\sum_{U \in \mathcal{E}_{\mathrm{MUB}}} B_U \right)^2 \leq \mathcal{D}(O_0)^2, \tag{21}$$

where $\mathcal{D}(A) := 2^{-n} \sum_{P \in \mathbf{P}_n} |\mathrm{tr}(PA)|$ is defined as the stabilizer norm [37].

The proof is in Appendix D 2. Here we utilize a norm of nonstabilizerness (magic) [37] to evaluate the estimation variance of shadow estimation, along a similar line with Theorem 2 about quantum coherence, is also a measure of quantum resource [38,39,58]. In particular, for $O = \rho$ being some quantum state, $\mathcal{D}(O_0) = D(\rho) - 1/d$, where $\mathcal{D}(\rho)$ serves as a lower bound of the robustness of magic, a key measure of nonstabilizerness that quantifies the complexity of classical simulation cost [38]. Therefore, any observable with a constant or polynomial magic can be estimated with a feasible sample complexity if utilizing biased MCM.

For biased MCM, an additional expense for classical computation complexity is to sample unitary U and to compute the probability P_U, which unfortunately tends to be exponentially hard. However, inspired by Ref. [59], a solution to stabilizer states is given, where the unitaries U can be efficiently sampled and P_U can be efficiently calculated, with the estimation variance being reduced to constant. More details on efficient sampling and computing P_U are left to Appendix D 5.

We further provide the variance results for the case where the observable O is a stabilizer state in the following Proposition 4, by applying the interesting properties as shown in Lemma 2.

Proposition 4. Suppose the observable is $O = V^\dagger |0\rangle\langle 0| V$ for any Clifford unitary V, one has

$$\mathrm{Var}_{\mathrm{biased}-\mathcal{E}_{\mathrm{MUB}}}(\hat{O}) \leq 1, \tag{22}$$

by applying the biased-MCM with the optimized probability given in Eq. (19). Particularly, when the quantum state $\rho = O$.

$$\mathrm{Var}_{\mathrm{biased}-\mathcal{E}_{\mathrm{MUB}}}(\hat{O}) = 0. \tag{23}$$

The proof is left to Appendix D 4.

Lemma 2. Suppose the observable $O = V^\dagger |\mathbf{0}\rangle\langle\mathbf{0}| V$, where V is an arbitrary Clifford, then the element of the canonical MUB decomposition $\alpha_{U,\mathbf{b}} = |\langle\mathbf{b}|UV^\dagger|\mathbf{0}\rangle|^2$ satisfies

$$\sum_{U \in \mathcal{E}_{\mathrm{MUB}}} \max_{\mathbf{b}} \alpha_{U,\mathbf{b}} = 2. \tag{24}$$

Furthermore, suppose the Z tableau of $UV^\dagger$ is $T_{UV^\dagger} = [C, D]$, the distribution of $\alpha_{U,\mathbf{b}}$ for different $\mathbf{b} \in \{0,1\}^n$ are
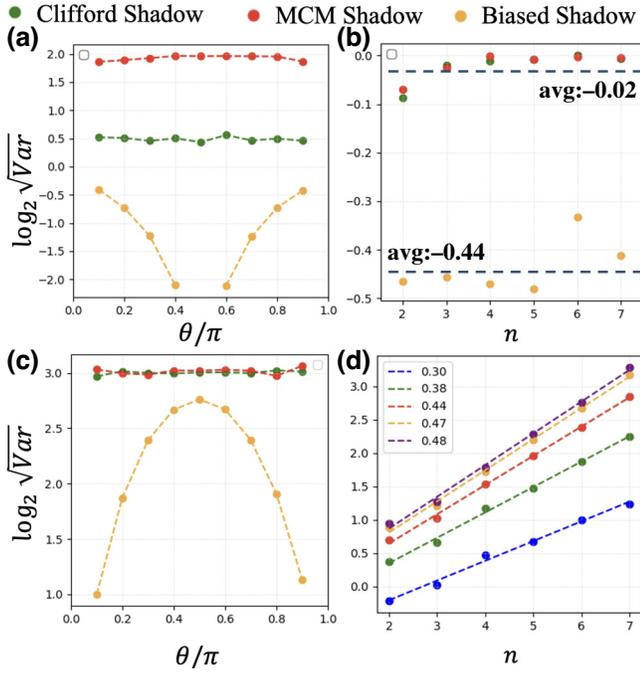
FIG. 5. The statistical variance of shadow estimation with the total number of snapshots $N = 10\,000$ using Clifford measurement (green), MCM (red), and biased MCM (orange) in three different scenarios. In (a), the quantum state ρ is parametered as the phased-GHZ states $|\text{GHZ}(\theta)\rangle = 1/\sqrt{2}[\cos(\theta/2)|0\rangle^{\otimes 6} + \sin(\theta/2)|1\rangle^{\otimes 6}]$, with $\theta \in [0.1\pi, 0.9\pi]$. And the observable O is set to be a six-qubit GHZ state. In (b), we estimate the fidelity of a set of Haar random states $\rho_i = \Phi_i$ for $i \in [1, 100]$ with some Clifford stabilizer $O = \Phi_C$. The variance is the average of these 100 cases. In (c) and (d), the observable $O = [\cos(\theta/2)X + \sin(\theta/2)Z]^{\otimes n}$ and the state $\rho = |\mathbf{0}\rangle\langle\mathbf{0}|$, with $n = 6$ and $\theta \in [0.1\pi, 0.9\pi]$ in (c), and in (d) the lines representing $\theta = 0.1$ to 0.5 from bottom to top. The dotted lines in (d) show the fitting curves with the numerical values, and the numbers on the left top represent the corresponding slopes of these curves.

as follows: totally 2^{r_U} of them taking 2^{-r_U} and the remaining $2^n - 2^{r_U}$ being 0. Here $r_U = \text{rank}_{\mathbb{F}_2}(C)$ on the binary field.

This lemma characterizes the projection of the stabilizer states to the MUBs, which is of independent interest. And the proof is left to Appendix D 3.

In addition, we numerically demonstrate the variance of estimating observables through different estimation protocols as shown in Fig. 5. Figure 5(a) showcases the significant sampling advantage of the biased MCM by considering the fidelity estimation of GHZ-like states. In particular, when $\rho = O = |\text{GHZ}\rangle\langle\text{GHZ}|$ at $\theta = 0.5\pi$, the variance disappears. Specifically, in Fig. 5(b), we randomly generate 100 quantum states $\mathcal{E} = \{|\Phi_i\rangle\}_{i=1}^{100}$ according to the Haar measure, along with one Clifford stabilizer state $|\Phi_C\rangle$. Then we estimate the fidelity with $O = \Phi_C$ as $\text{tr}(\Phi_i \Phi_C) = |\langle\Phi_i|\Phi_C\rangle|^2$ via different shadow protocols.

The experimental result as an average from these 100 experiments highlights a constant advantage of the biased MCM over other protocols. Moreover, the estimation variance of the observable $O = [\cos(\theta/2)X + \sin(\theta/2)Z]^{\otimes n}$ is discussed in (c) and (d). The stabilizer norm of observables shows $\mathcal{D}(O_0)^2 = (1 + \sin\theta)^n$, and it grows exponentially with a base of 2 when $\theta = 0.5\pi$, which is supported by the numerics in Fig. 5(d).

Note that the observable O in Eq. (18) is a linear combination of several rank-1 projections $\Phi_{U,\mathbf{b}}$. In this framework, the role of projections $\Phi_{U,\mathbf{b}}$ in biased MCM can be considered as a mimic of Pauli terms in local-biased shadow estimation [18].

VII. CONCLUSION AND OUTLOOK

In this work, we propose the MCM shadow framework to simplify the original Clifford measurement to the greatest extent. By applying the canonical MUB and the tableau formalism, we give an explicit and efficient method to synthesize the random unitary ensemble, where the budgets of both the circuit synthesizing and the experimental realization are reduced. For example, the quantum circuit can be realized in the platforms with the all-to-all architectures, such as the Rydberg atoms with optical tweezers [60] and the trapped-ion system [50]. The performance analysis shows the limits and advantages of MCM shadow, especially in estimating off-diagonal observables. The updated biased-MCM shadow protocol adjusts the sampling probability of the random Clifford circuit based on the target observable, which can further enhance the performance of the framework.

From the framework presented here, there are a few intriguing directions to explore in the future. First, it is very interesting to apply the MCM shadow to detect multipartite entanglement [61,62], as genuine entanglement [63,64] and more detailed structures [65–67] can be revealed by the fidelities to some target entangled states. Second, the quantum circuit structure is mainly determined by the Z tableau according to the selection of MUBs. It is thus very worth studying how different circuit architectures [68] and ensemble definitions [69–71] would enhance the performance of the whole shadow protocol. In addition, it is worth exploring the performance of MCM on near-term noisy quantum platforms using quantum error mitigation techniques [72], such as the robust shadow method [73,74], considering the specific structure of our quantum circuits. Third, in the biased-MCM, we need the projection results of the target observables to all MUBs. Consequently, it is pivotal to develop additional adaptive or real-time methods to ease this [75,76], and find applications where the decomposition on MUB is of polynomial terms, like the Pauli operator summation for general Hamiltonian. Fourth, it is possible to extend our MCM to higher-dimensional systems, where the existence of MUB is not fully known

[46]. Finally, the extension to boson and fermion systems [77–79] with a similar spirit of MCM is also intriguing.

APPENDIX

In this Appendix, we provide the proofs, additional discussions, and generalizations of the results presented in the main text. In Appendix. A, we prove Lemma 1 and Proposition 1 from the main text. Moving on to Appendix. B, we offer further details and examples on mutually unbiased bases and the circuit synthesis. Appendix C discusses the variance of estimation and we also show two applications of MCM-shadow estimation. Finally, in Appendix D, we provide the details and proofs of the main results of biased MCM.

APPENDIX A: MINIMAL UNBIASED CLIFFORD SUBSET

1. Proof of Lemma 1

Proof. First, we define the matrix $Z_{\mathbf{m}} = \bigotimes_{i=0}^{n-1} Z_i^{m_i}$ with $\mathbf{m}$ an n-bit binary vector. Hence

$$Z_{\mathbf{m}} = \bigotimes_{i=0}^{n-1} Z_i^{m_i} = \bigotimes_{i=1}^{n} |0\rangle\langle 0| + (-1)^{m_i}|1\rangle\langle 1|$$
$$= \sum_{\mathbf{b}\in\{0,1\}^n} (-1)^{\mathbf{m}\cdot\mathbf{b}}|\mathbf{b}\rangle\langle\mathbf{b}|, \qquad (A1)$$

where $\mathbf{b}$ is also an n-bit binary vector. Reversely, we have

$$\frac{1}{2^n}\sum_{\mathbf{m}\in\{0,1\}^n}(-1)^{\mathbf{b}\cdot\mathbf{m}}Z_{\mathbf{m}} = \frac{1}{2^n}\sum_{\mathbf{m}\in\{0,1\}^n}(-1)^{\mathbf{b}\cdot\mathbf{m}}\sum_{\mathbf{m}'\in\{0,1\}^n}(-1)^{\mathbf{m}'\cdot\mathbf{m}}|\mathbf{m}'\rangle\langle\mathbf{m}'|$$
$$= \frac{1}{2^n}\sum_{\mathbf{m},\mathbf{m}'\in\{0,1\}^n}(-1)^{\mathbf{m}\cdot(\mathbf{b}+\mathbf{m}')}|\mathbf{m}'\rangle\langle\mathbf{m}'|$$
$$= \frac{1}{2^n}\sum_{\mathbf{m}'=\mathbf{b},\mathbf{m}}|\mathbf{m}'\rangle\langle\mathbf{m}'| + \frac{1}{2^n}\sum_{\mathbf{m}'\neq\mathbf{b},\mathbf{m}}(-1)^{\mathbf{m}\cdot(\mathbf{b}+\mathbf{m}')}|\mathbf{m}'\rangle\langle\mathbf{m}'|$$
$$= |\mathbf{b}\rangle\langle\mathbf{b}| + \frac{1}{2^n}\sum_{\mathbf{m}'\neq\mathbf{b}}|\mathbf{m}'\rangle\langle\mathbf{m}'|\sum_{\mathbf{m}}(-1)^{\mathbf{m}\cdot(\mathbf{b}+\mathbf{m}')}$$
$$= |\mathbf{b}\rangle\langle\mathbf{b}|. \qquad (A2)$$

After that, we define $\Phi_{U,\mathbf{b}} = U^\dagger|\mathbf{b}\rangle\langle\mathbf{b}|U$, thus we have

$$\Phi_{U,\mathbf{b}} = \frac{1}{2^n}\sum_{\mathbf{m}\in\{0,1\}^n}(-1)^{\mathbf{b}\cdot\mathbf{m}}S_{\mathbf{m}}. \qquad (A3)$$

Now we can compute that

$$\mathcal{M}(\rho\,|\,U) = \sum_{\mathbf{b}\in\{0,1\}^n}\mathrm{tr}\left[\rho U^\dagger|\mathbf{b}\rangle\langle\mathbf{b}|U\right]U^\dagger|\mathbf{b}\rangle\langle\mathbf{b}|U$$
$$= \frac{1}{2^{2n}}\sum_{\mathbf{b}\in\{0,1\}^n}\left[\sum_{\mathbf{m}\in\{0,1\}^n}(-1)^{\mathbf{b}\cdot\mathbf{m}}S_{\mathbf{m}}\right]\left[\sum_{\mathbf{m}'\in\{0,1\}^n}(-1)^{\mathbf{b}\cdot\mathbf{m}'}\mathrm{tr}(\rho S_{\mathbf{m}'})\right]$$
$$= \frac{1}{2^{2n}}\sum_{\mathbf{b}\in\{0,1\}^n}\left[\sum_{\mathbf{m},\mathbf{m}'\in\{0,1\}^n}S_{\mathbf{m}}\mathrm{tr}(\rho S_{\mathbf{m}'})(-1)^{(\mathbf{m}+\mathbf{m}')\cdot\mathbf{b}}\right]$$

$$= \frac{1}{2^{2n}} \sum_{\mathbf{m},\mathbf{m}'\in\{0,1\}^n} S_{\mathbf{m}} \mathrm{tr}(\rho S_{\mathbf{m}'}) \sum_{\mathbf{b}\in\{0,1\}^n} (-1)^{(\mathbf{m}+\mathbf{m}')\cdot\mathbf{b}}$$

$$= \frac{1}{2^{2n}} \sum_{\mathbf{m},\mathbf{m}'\in\{0,1\}^n} S_{\mathbf{m}} \mathrm{tr}(\rho S_{\mathbf{m}'}) \prod_{i=0}^{n-1} \left[1 + (-1)^{m_i+m_i'}\right]$$

$$= \frac{1}{2^{2n}} \sum_{\mathbf{m}=\mathbf{m}'} S_{\mathbf{m}} \mathrm{tr}(\rho S_{\mathbf{m}'}) \prod_{i=0}^{n-1} \left[1 + (-1)^{m_i+m_i'}\right] = 2^{-n} \sum_{\mathbf{m}\in\{0,1\}^n} S_{\mathbf{m}} \mathrm{tr}(\rho S_{\mathbf{m}}). \tag{A4}$$

$\square$

2. Proof of Proposition 1

Proof. Firstly, the representation of $\mathcal{M}_{Cl}$ is converted to

$$\rho = \frac{2^n+1}{|\mathcal{E}|} \sum_{U\in\mathcal{E}} \mathcal{M}(\rho\,|\,U) - \mathbb{I} = \frac{1}{2^n} \sum_{\sigma\in\mathbf{P}^n} \mathrm{tr}(\rho\sigma)\,\sigma. \tag{A5}$$

By introducing Lemma 1, we have $\mathcal{M}(\rho\,|\,U) = (1/2^n)$ $\sum_{\mathbf{m}\neq 0} S_{\mathbf{m}} \mathrm{tr}(\rho S_{\mathbf{m}}) + (1/2^n)\mathbb{I}$. Then we obtain that

$$\frac{2^n+1}{|\mathcal{E}|} \sum_{U\in\mathcal{E}} \sum_{\mathbf{m}\neq 0} \mathrm{tr}(\rho S_{\mathbf{m}}) S_{\mathbf{m}} = \sum_{\sigma\in\mathbf{P}_*^n} \mathrm{tr}(\rho\sigma)\,\sigma. \tag{A6}$$

Equation (A6) must be satisfied for all quantum state ρ. If $|\mathcal{E}| < 2^n + 1$, there exists $\sigma \in \mathbf{P}_*^n$ that for all $\mathbf{m} \neq 0$ and $U \in \mathcal{E}$, $\sigma \neq S_{\mathbf{m}}$. Then let $\rho = (\sigma + \mathbb{I})/2^n$, we have $0 = \sigma$, which is obviously a paradox! Therefore, $|\mathcal{E}| \geq 2^n + 1$. $\blacksquare$

APPENDIX B: MUTUALLY UNBIASED BASES AND CIRCUIT SYNTHESIS

1. Introduction to Galois field

The example of minimal unbiased Clifford subset employs the construction of mutually unbiased bases. Here, we first introduce the concept of Galois field GF(2^n) as a mathematical tool to explain Eq. (8).

A Galois field GF(2^n) has 2^n elements, which can be represented in the row vector form or polynomial form.

$$a = (a_0, a_1, \ldots, a_{n-1}) \quad \text{or} \quad a = \sum_{i=0}^{n-1} a_i 2^i, \tag{B1}$$

where a_i is a binary number. So a GF(2^n) element is also associated with a number from 0 to $2^n - 1$ in the polynomial form.

There are two operations in a Galois field GF(2^n), the addition and the multiplication. Both of them can only be performed on two elements that belong to the same Galois field. All binary additions below are actually the XOR operations.

Galois-field addition is relatively simple. Let a and b be two elements in the Galois field. The addition is defined as

$$a \oplus b = (a_0 \oplus b_0, a_1 \oplus b_1, \ldots, a_{n-1} \oplus b_{n-1})$$

$$= \sum_{i=0}^{n-1} (a_i \oplus b_i) 2^i. \tag{B2}$$

To perform Galois-field multiplication correctly and ensure unique results, it is necessary to use irreducible polynomials. There are several efficient algorithms available [80,81] for generating these polynomials. An irreducible polynomial in binary format with $(n+1)$ bits is represented by $P_n = 2^n + \sum_{i=1}^{n-1} c_i 2^i + 1$, and the multiplication of the GF (2^n) elements a and b is defined as

$$a \odot b = \sum_{k=0}^{n-1} a_i 2^i \times \sum_{i=0}^{n-1} b_i 2^i \bmod P_n. \tag{B3}$$

For example, an irreducible polynomial for GF (2^2) is $P_2 = 2^2 \oplus 2 \oplus 1$, then $2 \odot 2 = 2^2 \bmod P_2 = 2 \oplus 1 = 3$. It is easy to prove that the addition and multiplication in a Galois field satisfy the property of commutativity and associativity, respectively, and they satisfy the distributive law with each other [46].

After introducing the fundamental operation of Galois field, we introduce the definition of the matrix $M_n^{(j)}$ in Eq. (8) in the main text. $M_n^{(j)}$ is a symmetric matrix originally introduced for easy computation of the multiplication in the Galois field. It is originally defined to calculate the multiplication of GF(2^n) elements in the matrix form. Generally, for GF(2^n) elements a, b, we have

$$[a \odot b]_j = a M_n^{(j)} b^T. \tag{B4}$$

The calculation method for matrix $M_n^{(j)}$ is as follows. First, we define $2^k \bmod P_n = (\Gamma_{k,0}, \Gamma_{k,1}, \ldots, \Gamma_{k,n-1})$, $k = 0, 1, \ldots, 2n-2$ and construct a $(2n-1) \times n$ binary

matrix Γ_n:

$$\Gamma_n = \begin{pmatrix} \Gamma_{0,0} & \Gamma_{0,1} & \cdots & & \Gamma_{0,n-1} \\ \Gamma_{1,0} & \Gamma_{1,1} & \ddots & & \vdots \\ \cdots & \ddots & \ddots & & \Gamma_{n-1,n-1} \\ \Gamma_{2n-2,0} & \cdots & \Gamma_{2n-2,n-2} & & \Gamma_{2n-2,n-1} \end{pmatrix}. \quad (B5)$$

According to the definition of Γ_n, the first n rows of the elements in Γ_n is easily computed, and the nth row of Γ_n contains the coefficients of the irreducible polynomial P_n. there exists a certain recursive relationship between the elements in later rows, which can be listed as follows:

$$\Gamma_{i,j} = \begin{cases} \delta_{i,j} & i = 0, 1, \ldots, n-1 \\ \Gamma_{n,j} & i = n \\ (1 - \delta_{j,0})\Gamma_{i-1,j-1} + \Gamma_{n,j}\Gamma_{i-1,n-1} & i = n+1, \ldots, 2n-2, \end{cases} \quad (B6)$$

where $\Gamma_{n,j}$ denotes the jth coefficient of P_n. We can extract the elements in a certain column of Γ_n to form a symmetric matrix $M_n^{(j)}$, where j is the column number, and the elements in the matrix satisfy

$$[M_n^{(j)}]_{p,q} = \Gamma_{(p+q),j}. \quad (B7)$$

2. Further details on mutually unbiased bases

In fact, Eq. (8) in the main text is basically a rewrite of the construction of the MUB in Ref. [46]. Here we give a brief introduction to the property of this $\mathcal{E}_{\mathrm{MUB}}$. Suppose the first element of $\mathcal{E}_{\mathrm{MUB}}$ is $\mathbb{I}$. For the next 2^n element, we define the vth element of $\mathcal{E}_{\mathrm{MUB}}$ is U_v, and U_v is generated by 2^n Paulis. Hence we define a Pauli $S_{\mathbf{m},v} = U_v^\dagger Z_{\mathbf{m}} U_v \in \mathbf{P}^n$, where $\mathbf{m}$ is an n-bit binary vector, and m_i represents the ith bit of $\mathbf{m}$.

Notice that Pauli matrices $\{S_{\mathbf{m},v}\}_{\mathbf{m} \in \{0,1\}^n}$ form a maximally commuting set (MCS) [82] and therefore can generate Clifford elements. Thus, we show an interesting fact as follows.

Observation 2. $\mathbf{P}_*^n = \{S \mid S = U^\dagger Z_{\mathbf{m}} U, U \in \mathcal{E}_{\mathrm{MUB}}, \mathbf{m} \neq 0\}$.

Proof. Firstly, we define $X(a), Z(a)$, respectively, as

$$X(a) = \sum_{k=0}^{2^n-1} |k \oplus a\rangle\langle k|,$$
$$Z(a) = \sum_{k=0}^{2^n-1} (-1)^{k \odot a} |k\rangle\langle k|. \quad (B8)$$

If v and $m = \sum_{k=0}^{n-1} m_k 2^k$ are viewed as two GF(2^n) elements, the Pauli matrices can be specified as

$$S_{\mathbf{m},v} = \omega_m^v X(m) Z(m \odot v). \quad (B9)$$

Equation (B9) is actually a rewrite of Eq. (2.54) in [46], where the parameters $\omega_m^v \in \{\pm 1, \pm i\}$, and $S_{\mathbf{m},v} \in \mathbf{P}^n$. If $\mathbf{m}_1 = \mathbf{m}_2 = 0$, then $S_{\mathbf{m}_1,v_1} = S_{\mathbf{m}_2,v_2}$, else $S_{\mathbf{m}_1,v_1} = S_{\mathbf{m}_2,v_2} = \mathbb{I}$ if and only if $\mathbf{m}_1 = \mathbf{m}_2 \neq 0$, $v_1 = v_2$. Consequently, the set $\{S_{\mathbf{m},v} \mid \mathbf{m} \neq 0, v = 0, 1, \ldots, 2^n - 1\}$ have $(2^n - 1)2^n$ distinct Paulis. Compared with $\mathbf{P}_*^n$, the set is still short of $2^n - 1$ elements, which happen to be the nonidentity generators of $\mathbb{I}$: $\{Z_{\mathbf{m}}\}_{\mathbf{m} \neq 0}$. This is because $S_{\mathbf{m},v} = U_v^\dagger Z_{\mathbf{m}} U_v, U_v \neq \mathbb{I}$. Hence the union of nonidentity generators of MUB elements forms the set $\mathbf{P}_*^n$. ∎

Observation 3.

$$\frac{1}{|\mathcal{E}_{\mathrm{MUB}}|} \sum_{U \in \mathcal{E}_{\mathrm{MUB}}} \mathcal{M}(\rho \mid U) = \mathcal{M}_{Cl}(\rho), \quad \forall \rho.$$

Proof. We can use Lemma 1 to decompose these Cliffords.

$$\frac{1}{|\mathcal{E}_{\mathrm{MUB}}|} \sum_{U \in \mathcal{E}_{\mathrm{MUB}}} \mathcal{M}(\rho \mid U)$$
$$= \frac{1}{2^n + 1} \sum_{U \in \mathcal{E}_{\mathrm{MUB}}} \mathcal{M}(\rho \mid U)$$
$$= \frac{1}{2^n(2^n + 1)} \sum_{U \in \mathcal{E}_{\mathrm{MUB}}} \sum_{\mathbf{m} \neq 0} [S_{\mathbf{m},v} \mathrm{tr}(\rho S_{\mathbf{m},v}) + \mathbb{I}]$$
$$= \frac{1}{2^n + 1} \left[\frac{1}{2^n} \sum_{\sigma \in \mathbf{P}^n} \sigma \mathrm{tr}(\rho \sigma) \right] + \frac{\mathbb{I}}{2^n + 1}$$
$$= \frac{\rho + \mathbb{I}}{2^n + 1} = \mathcal{M}_{Cl}(\rho). \quad (B10)$$

∎

3. Introduction to the *Z* tableau

Here, we introduce the language of tableau [40]. An n-qubit Clifford element can be specified with four $n \times n$

binary matrices $(\alpha, \beta, \gamma, \delta)$ and two n-dimensional binary vectors $(\mathbf{r}, \mathbf{s})$, such that

$$U^\dagger X_i U = (-1)^{r_i} \prod_{j=0}^{n-1} X_j^{\alpha_{ij}} Z_j^{\beta_{ij}} \quad \& \quad U^\dagger Z_i U$$

$$= (-1)^{s_i} \prod_{j=0}^{n-1} X_j^{\gamma_{ij}} Z_j^{\delta_{ij}}. \tag{B11}$$

The matrix X_i, Z_i are, respectively, defined as a Pauli matrix with X applied to the i th qubit, and as a Pauli matrix with Z applied to the i th qubit. The parameters form a $2n \times (2n + 1)$ binary matrix, which is called the tableau of a Clifford element

$$\begin{pmatrix} [\alpha_{i,j}]_{n\times n} & [\beta_{i,j}]_{n\times n} & [r_{i,j}]_{n\times 1} \\ [\gamma_{i,j}]_{n\times n} & [\delta_{i,j}]_{n\times n} & [s_{i,j}]_{n\times 1} \end{pmatrix}. \tag{B12}$$

According to Lemma 1, if we substitute Clifford elements U with the parameters (γ, δ) shown above, we still have the same $\mathcal{M}(\rho \,|\, U)$. Following this thread, we develop an $n \times 2n$ binary matrix called the Z-tableau matrix

$$\begin{pmatrix} \gamma_{0,0} & \gamma_{0,1} & \cdots & \gamma_{0,n-1} & \delta_{0,0} & \delta_{0,1} & \cdots & \delta_{0,n-1} \\ \gamma_{1,0} & \gamma_{1,1} & \ddots & \vdots & \delta_{1,0} & \delta_{1,1} & \ddots & \vdots \\ \cdots & \ddots & \ddots & \gamma_{n-2,n-1} & \vdots & \ddots & \ddots & \delta_{n-2,n-1} \\ \gamma_{n-1,0} & \cdots & \gamma_{n-1,n-2} & \gamma_{n-1,n-1} & \delta_{n-1,0} & \cdots & \delta_{n-1,n-2} & \delta_{n-1,n-1} \end{pmatrix}. \tag{B13}$$

According to the Gottesman-Knill theorem [48], Clifford circuits consist of three fundamental quantum gates: the Hadamard gate, the phase gate, and the CNOT gate, and can be efficiently simulated by classical computers in polynomial time. This means that the action of the three gates in U from $U^\dagger Z_i U$ can be expressed in terms of changes to the elements in the Z tableau.

Now Clifford circuit $\mathbb{I}$ corresponds to the Z tableau $[\mathbb{O}, \mathbb{I}]$, since $\mathbb{I}^\dagger Z_i \mathbb{I} = \prod_{j=0}^{n-1} X_j^{\gamma_{ij}=0} Z_j^{\delta_{ij}=\delta_{i,j}}$. Moreover, it can be inferred that if a Clifford circuit U can transform the Z tableau from $[C, D]$ to $[\mathbb{O}, \mathbb{I}]$, then it is a Clifford circuit that satisfies the condition $U^\dagger Z_i U = \prod_{j=0}^{n-1} X_j^{\gamma_{ij}} Z_j^{\delta_{ij}}$.

4. Proof of Proposition 3

Proof. We use the following three equations to demonstrate the proposition, where the first equation is determined by the property of the irreducible polynomial, and the next two come from the recurrence relation in Eq. (B6).

- $\Gamma_{n,0} = 1$,
- $\Gamma_{i+1,0} = \Gamma_{i,n-1}$,
- $\Gamma_{i+1,j} = \Gamma_{i,j-1} + \Gamma_{n,j}\Gamma_{i,n-1} (j \geq 1)$.

By the definition of Hankel matrix, the linear combination of Hankel matrices is also a Hankel matrix. Considering that a $\mathrm{GF}(2^n)$ element v can be represented as $v = \sum_{i=0}^{n-1} v_i 2^i$, if a matrix $\mathbb{D}_i$ is a Hankel matrix with its j th row vector being $(2^i \odot 2^j)M_n^{(0)}$, then D_v is also a Hankel matrix. This is evident from the fact that $D_v = \sum_{i=0}^{n-1} v_i \mathbb{D}_i$. The matrix $\mathbb{D}_i$ is defined such that its j th row vector corresponds to the $(i+j)$ th row vector of the matrix $\Gamma_n M_n^{(0)}$. Consequently, if $\Gamma_n M_n^{(0)}$ is a Hankel matrix, all $\mathbb{D}_i$ s are also Hankel matrices.

Let $M_n = \Gamma_n M_n^{(0)}$. The first n rows of M_n are equivalent to $M_n^{(0)}$, which is a Hankel matrix. We need to prove that $M_{i,j} = M_{i-1,j+1}, j = 0, 1, \ldots, n-2$ for the remaining elements. When $i = n$, we have

$$\begin{aligned} M_{n,j} &= \sum_{k=0}^{n-1} \Gamma_{n,k} M_{k,j}^{(0)} = \sum_{k=0}^{n-1} \Gamma_{n,k} \Gamma_{k+j,0} = \sum_{k=n}^{n+j-1} \Gamma_{n,k-j} \Gamma_{k,0} \\ &= \Gamma_{n,n-j} + \sum_{k=n+1}^{n+j-1} \Gamma_{n,k-j} \Gamma_{k,0} \\ &= \Gamma_{n,n-j} + \sum_{k=n+1}^{n+j-1} \Gamma_{k,k-j} - \Gamma_{k-1,k-j-1} \\ &= \Gamma_{n+j-1,n-1} = \Gamma_{n+j,0} = M_{n-1,j+1}, \end{aligned} \tag{B14}$$

when $i \geq n + 1$, we have

$$
\begin{aligned}
M_{i,j} &= \sum_{k=0}^{n-1} \Gamma_{i,k} M_{k,j}^{(0)} = \sum_{k=0}^{n-1} \Gamma_{i,k} \Gamma_{k+j,0} = \sum_{k=n}^{n+j-1} \Gamma_{i,k-j} \Gamma_{k,0} \\
&= \sum_{k=n}^{n+j-1} \Gamma_{i-1,k-j-1} \Gamma_{k,0} + \sum_{k=n}^{n+j-1} \Gamma_{n,k-j} \Gamma_{i-1,n-1} \Gamma_{k,0} \\
&= M_{i-1,j+1} + \Gamma_{n+j,0} \Gamma_{i-1,n-1} + \Gamma_{n,n-j} \Gamma_{i-1,n-1} + \sum_{k=n+1}^{n+j-1} \Gamma_{n,k-j} \Gamma_{k-1,n-1} \Gamma_{i-1,n-1} \\
&= M_{i-1,j+1} + \Gamma_{n+j,0} \Gamma_{i-1,n-1} + \Gamma_{n,n-j} \Gamma_{i-1,n-1} + \Gamma_{i-1,n-1}(\Gamma_{n+j-1,n-1} - \Gamma_{n,n-j}) \\
&= M_{i-1,j+1}.
\end{aligned}
\tag{B15}
$$

Therefore, we demonstrate that M_n is a Hankel matrix, then by definition, the matrix D_v is also a Hankel. ∎

5. The $n = 3$ example of MUB

In this part, we give an example of $\mathcal{E}_{\mathrm{MUB}}$ in the $n = 3$ case, where the total number of elements is 9. The 0th element of $\mathcal{E}_{\mathrm{MUB}}$ is undoubtedly $\mathbb{I}$, and the rest of the eight elements are constructed with the help of the Z-tableau language. Suppose the Z tableau of the v th element is $[\mathbb{I}, D_v]$, where $v = 0, 1, \ldots, 7$. Equation (8) in the main text shows that the j th row of matrix D_v is $(v \odot 2^j) M_n^{(0)}$. Since $v = \sum_{i=0}^{2} v_i 2^i$, D_v can be rewritten as $D_v = \sum_{i=0}^{2} v_i \mathbb{D}_i$, where the j th row of matrix $\mathbb{D}_i$ is $(2^{j+i}) M_3^{(0)}$. Moreover, considering $(j + i) = 0, 1, \ldots, 4$. The binary matrix Γ_n in Eq. (B5) can be used to compute $\mathbb{D}_i$ easily.

Here is an example for the case of $n = 3$, where $P_3 = 2^3 \oplus 2 \oplus 1$. First, we compute Γ_3 according to Eq. (B6)

$$
\Gamma_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}.
\tag{B16}
$$
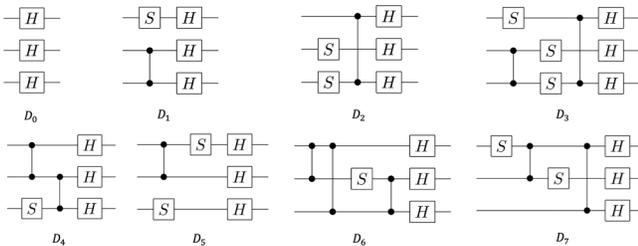
FIG. 6. Illustrations of all nonidentity MCM circuits for $n = 3$, with $P_3 = 2^3 \oplus 2 \oplus 1$.

And $M_3^{(0)}$ can be constructed from the first row of Γ_3

$$
M_3^{(0)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.
\tag{B17}
$$

Therefore, we have the matrix

$$
M_3 = \Gamma_3 M_3^{(0)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix},
\tag{B18}
$$

and $\mathbb{D}_i$ consists of the elements from the i th to $(i + 2)$ th row of matrix M_3. Finally, we can conduct that

$$
\mathbb{D}_0 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad \mathbb{D}_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}
$$

$$
\mathbb{D}_2 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix},
\tag{B19}
$$

and every Z-tableau matrix D_v can be decomposed as $D_v = \sum_{i=0}^{2} v_i \mathbb{D}_i$ for $v = \sum_{i=0}^{2} v_i 2^i$. Notice that the choice of Z tableau can also be variant considering the choice of irreducible polynomials, and the transformation from Z tableau to the Clifford circuit can be easily implemented with a three-stage computation of $-\mathrm{S} - \mathrm{CZ} - \mathrm{H}-$. The corresponding Clifford circuits are illustrated in Fig. 6.

APPENDIX C: PERFORMANCE ANALYSIS OF MCM SHADOW ESTIMATION

1. Variance of the estimation

First, we give a brief introduction to the variance of shadow estimation. In this paper, when calculating the

variance, we consider only the linear physical properties and do not take into account the median-of-means method in quantum shadow tomography. Let O be a fixed observable, and ρ be the quantum state. After quantum shadow tomography, the classical snapshot is denoted as $\hat{\rho}$. And the desired property $\hat{O} = \text{tr}(O\hat{\rho})$ should satisfy $\mathbb{E}(\hat{O}) = \text{tr}(O\rho)$. Let $O_0 = O - (\text{tr}(O)/2^n)\mathbb{I}$, one can calculate its variance as

$$\text{Var}(\hat{O}) = \mathbb{E}\left[(\hat{O} - \mathbb{E}(\hat{O}))^2\right] = \mathbb{E}\left[(\text{tr}(O\hat{\rho}))^2\right] - (\text{tr}(O\rho))^2$$

$$= \mathbb{E}\left[(\text{tr}(O_0\hat{\rho}))^2\right] - (\text{tr}(O_0\rho))^2$$

$$= \mathbb{E}\text{tr}[O_0\mathcal{M}^{-1}(\Phi_{U,\mathbf{b}})]^2 - (\text{tr}(O_0\rho))^2. \quad (C1)$$

Note that $(\text{tr}(O_0\rho))^2$ is a fixed value, thus does not require detailed discussion. On the other hand, $\mathbb{E}\text{tr}[O_0\mathcal{M}^{-1}(\Phi_{U,\mathbf{b}})]^2$ is related to the specific measurement method, including the random measurement set $\mathcal{E}$ and the channel function $\mathcal{M}$,

$$\mathbb{E}\text{tr}[O_0\mathcal{M}^{-1}(\Phi_{U,\mathbf{b}})]^2$$
$$= \mathbb{E}_{U\in\mathcal{U}} \sum_{\mathbf{b}\in\{0,1\}^n} \langle\mathbf{b}|U\rho U^\dagger|\mathbf{b}\rangle\langle\mathbf{b}|U\mathcal{M}^{-1}(O_0)U^\dagger|\mathbf{b}\rangle^2. \quad (C2)$$

Here, we briefly introduce the theoretical upper limits of the variance under two conventional methods, i.e., the Clifford measurement and the Pauli measurement. The specific derivation can be found in Ref. [14]. For Clifford measurement,

$$\mathbb{E}\text{tr}[O_0\mathcal{M}^{-1}(\Phi_{U,\mathbf{b}})]^2 \leq 3tr(O^2). \quad (C3)$$

And for Pauli measurement,

$$\mathbb{E}\text{tr}[O_0\mathcal{M}^{-1}(\Phi_{U,\mathbf{b}})]^2 \leq 4^{locality(O)}||O||_\infty^2, \quad (C4)$$

where $||\cdot||_\infty$ is the spectral norm (or operator norm) of a matrix, and $locality(O)$ represents the number of nonunit observables of each qubit in the observable O. Meanwhile, it is noteworthy that for Clifford measurement, the average upper bound is

$$\mathbb{E}_{U\sim\mathcal{E}_{\text{Cl}},\rho\sim\mathcal{E}}\text{tr}[O_0\mathcal{M}^{-1}(\Phi_{U,\mathbf{b}})]^2$$
$$= \frac{2^n+1}{2^n+2}\mathbb{E}_{\rho\sim\mathcal{E}}(\text{tr}(\rho)\text{tr}(O_0^2) + 2\text{tr}(\rho O_0^2))$$
$$= \frac{2^n+1}{2^n+2}(\text{tr}(O_0^2) + \text{tr}(O_0^2)/2^{n-1}) = \frac{2^n+1}{2^n}\text{tr}(O_0^2). \quad (C5)$$

And for Pauli measurement, define $O = \widetilde{O} \otimes \mathbb{I}^{\otimes n-k}$ to emphasize the local observables, and $\widetilde{O} = \sum_{\mathbf{p}} \alpha_{\mathbf{p}} P_{\mathbf{p}}$ where $P_{\mathbf{p}} \in \{\mathbb{I}, X, Y, Z\}^{\otimes k}$, $|\mathbf{p}|$ denote the number of nonidentity

Paulis in $P_{\mathbf{p}}$. The average upper bound is

$$\mathbb{E}_{U\sim\mathcal{E},\rho\sim\mathcal{E}_{\text{Pauli}}}\text{tr}[O_0\mathcal{M}^{-1}(\Phi_{U,\mathbf{b}})]^2 = \sum_{\mathbf{p}} \alpha_{\mathbf{p}}^2 3^{|\mathbf{p}|}. \quad (C6)$$

2. Off-diagonal fidelity

In this section, we discuss four specific cases for the performance of MCM and give proof for Theorem (2) in the main text. After that, we use the example of fidelity estimation to illustrate the connection between the off-diagonal observables and the original observables. Here we provide more details on the observation of the off-diagonal fidelity. The four specific cases are listed as follows.

For $U \in \mathcal{E}_{\text{MUB}}$ and $\mathbf{b} \in \{0,1\}^n$, $\langle\rho\rangle_{U,\mathbf{b}} = \text{tr}[\rho\Phi_{U,\mathbf{b}}] \geq 0$, $\langle O_0\rangle_{U,\mathbf{b}} = \text{tr}[O_0\Phi_{U,\mathbf{b}}] \neq 0$. In the following discussion, if there exists no extra clues, we assume $\langle\rho\rangle_{U,\mathbf{b}}, \langle O_0\rangle_{U,\mathbf{b}} = O(2^{-n})$.

- **Case I:** There exists $U \in \mathcal{E}_{\text{MUB}}$ and $\mathbf{b} \in \{0,1\}^n$ such that both $\langle\rho\rangle_{U,\mathbf{b}}, \langle O_0\rangle_{U,\mathbf{b}} = \Theta(1)$. Then the upper bound is $\mathbb{E}\text{tr}[O_0\mathcal{M}^{-1}(\Phi_{U,\mathbf{b}})]^2 = \Theta(2^n)$.
- **Case II:** There exists $U \in \mathcal{E}_{\text{MUB}}$ and $\mathbf{b} \in \{0,1\}^n$, such that $\langle\rho\rangle_{U,\mathbf{b}} = \Theta(1)$, but $\langle O_0\rangle_{U,\mathbf{b}} = O(2^{-n})$. Then the upper bound is $\mathbb{E}\text{tr}[O_0\mathcal{M}^{-1}(\Phi_{U,\mathbf{b}})]^2 = \Theta(1)$.
- **Case III:** There exists $U \in \mathcal{E}_{\text{MUB}}$ and $\mathbf{b} \in \{0,1\}^n$, such that $\langle O_0\rangle_{U,\mathbf{b}} = \Theta(1)$, but $\langle\rho\rangle_{U,\mathbf{b}} = O(2^{-n})$. Then the upper bound is $\mathbb{E}\text{tr}[O_0\mathcal{M}^{-1}(\Phi_{U,\mathbf{b}})]^2 = \Theta(1)$.
- **Case IV:** For all $U \in \mathcal{E}_{\text{MUB}}$ and $\mathbf{b} \in \{0,1\}^n$, $\langle O_0\rangle_{U,\mathbf{b}} = O(2^{-n})$ as well as $\langle\rho\rangle_{U,\mathbf{b}} = O(2^{-n})$. Then the upper bound is $\mathbb{E}\text{tr}[O_0\mathcal{M}^{-1}(\Phi_{U,\mathbf{b}})]^2 = \Theta(1)$.

These four cases are based on the hypothesis that $\text{tr}(O_0^2)$ is constant, which is satisfied when estimating fidelities. The proofs of the four cases are demonstrated as follows.

Case I:

Proof. We set $a_1 \leq \langle\rho\rangle_{U,\mathbf{b}} \leq b_1, a_2 \leq \langle O_0\rangle_{U,\mathbf{b}} \leq b_2$.

$$\mathbb{E}\text{tr}[O_0\mathcal{M}^{-1}(\Phi_{U,\mathbf{b}})]^2$$
$$= (2^n+1) \sum_{U\in\mathcal{E}_{\text{MUB}}} \sum_{\mathbf{b}\in\{0,1\}^n} \text{tr}[O_0\Phi_{U,\mathbf{b}}]^2\text{tr}[\rho\Phi_{U,\mathbf{b}}]$$
$$\geq (2^n+1)a_2^2 a_1. \quad (C7)$$

Since the upper bound is generally $O(2^n)$, then it is necessarily $\Theta(2^n)$. ∎

Case II:

Proof. We set $\langle O_0 \rangle_{U,\mathbf{b}} \leq (a/2^n)$ for all $U, \mathbf{b}$,

$$\mathbb{E}\mathrm{tr}[O_0 \mathcal{M}^{-1}(\Phi_{U,\mathbf{b}})]^2$$

$$= (2^n + 1) \sum_{U \in \mathcal{E}_{\mathrm{MUB}}} \sum_{b \in \{0,1\}^n} \mathrm{tr}[O_0 \Phi_{U,\mathbf{b}}]^2 \mathrm{tr}[\rho \Phi_{U,\mathbf{b}}]$$

$$\leq (2^n + 1) \sum_{U \in \mathcal{E}_{\mathrm{MUB}}} \frac{a^2}{4^n} \sum_{b \in \{0,1\}^n} \mathrm{tr}[\rho \Phi_{U,\mathbf{b}}]$$

$$= \frac{(2^n + 1)^2 a^2}{4^n} \sim O(1). \tag{C8}$$

∎

Case III:

Proof. We set $\langle \rho \rangle_{U,\mathbf{b}} \leq (b/2^n)$ for all $U, \mathbf{b}$,

$$\mathbb{E}\mathrm{tr}[O_0 \mathcal{M}^{-1}(\Phi_{U,\mathbf{b}})]^2$$

$$= (2^n + 1) \sum_{U \in \mathcal{E}_{\mathrm{MUB}}} \sum_{b \in \{0,1\}^n} \mathrm{tr}[O_0 \Phi_{U,\mathbf{b}}]^2 \mathrm{tr}[\rho \Phi_{U,\mathbf{b}}]$$

$$\leq \frac{b(2^n + 1)}{2^n} \sum_{U \in \mathcal{E}_{\mathrm{MUB}}} \sum_{b \in \{0,1\}^n} \mathrm{tr}[O_0 \Phi_{U,\mathbf{b}}]^2. \tag{C9}$$

Now the equation is at the same scale with the average upper bound of minimal Clifford measurement. ∎

Case IV:

Proof. The conclusion is easy to proof since the scale of the upper bound in this case is certainly smaller than that of both case II and case III. ∎

Here is the proof of Theorem 2.

Proof. For observable $O_0 = O_F = \sum_{\mathbf{b_1} \neq \mathbf{b_2}} O_{\mathbf{b_1},\mathbf{b_2}} |\Phi_{U,\mathbf{b_1}}\rangle\langle\Phi_{U,\mathbf{b_2}}|$, Eq. (C2) can be written as

$$\mathbb{E}\mathrm{tr}[O_0 \mathcal{M}^{-1}(\Phi_{U',\mathbf{b}})]^2 = (2^n + 1) \sum_{U' \in \mathcal{E}_{\mathrm{MUB}}, \mathbf{b}} \mathrm{tr}[O_0 \Phi_{U',\mathbf{b}}]^2 \mathrm{tr}(\rho \Phi_{U',\mathbf{b}})$$

$$= (2^n + 1) \sum_{U' \in \mathcal{E}_{\mathrm{MUB}}, \mathbf{b}} \mathrm{tr}\left[\sum_{\mathbf{b_1} \neq \mathbf{b_2}} O_{\mathbf{b_1},\mathbf{b_2}} |\Phi_{U,\mathbf{b_1}}\rangle\langle\Phi_{U,\mathbf{b_2}}| \Phi_{U',\mathbf{b}} \right]^2 \mathrm{tr}(\rho \Phi_{U',\mathbf{b}})$$

$$= (2^n + 1) \sum_{U' \in \mathcal{E}_{\mathrm{MUB}}, \mathbf{b}} \sum_{\mathbf{b_1} > \mathbf{b_2}} \left[O_{\mathbf{b_1},\mathbf{b_2}} \langle\Phi_{U',\mathbf{b}}|\Phi_{U,\mathbf{b_1}}\rangle\langle\Phi_{U,\mathbf{b_2}}|\Phi_{U',\mathbf{b}}\rangle \right.$$

$$\left. + O_{\mathbf{b_2},\mathbf{b_1}} \langle\Phi_{U',\mathbf{b}}|\Phi_{U,\mathbf{b_2}}\rangle\langle\Phi_{U,\mathbf{b_1}}|\Phi_{U',\mathbf{b}}\rangle \right]^2 \mathrm{tr}(\rho \Phi_{U',\mathbf{b}}). \tag{C10}$$

After reviewing the definition of MUB [46], we have

$$\langle\Phi_{U',\mathbf{b}}|\Phi_{U,\mathbf{b_1}}\rangle = \begin{cases} \delta_{\mathbf{b},\mathbf{b_1}}, & U' = \mathbb{I} \\ \frac{1}{\sqrt{2^n}} e^{i\theta_{\mathbf{b},\mathbf{b_1}}}, & U' \neq \mathbb{I} \end{cases}, \quad \langle\Phi_{U',\mathbf{b}}|\Phi_{U,\mathbf{b_2}}\rangle = \begin{cases} \delta_{\mathbf{b},\mathbf{b_2}}, & U' = \mathbb{I} \\ \frac{1}{\sqrt{2^n}} e^{i\theta_{\mathbf{b},\mathbf{b_2}}}, & U' \neq \mathbb{I} \end{cases}, \tag{C11}$$

where $\theta_{\mathbf{b},\mathbf{b_1}}, \theta_{\mathbf{b},\mathbf{b_2}}$ are phases for certain U, U'. Then,

$$O_{\mathbf{b_1},\mathbf{b_2}} \langle\Phi_{U',\mathbf{b}}|\Phi_{U,\mathbf{b_1}}\rangle\langle\Phi_{U,\mathbf{b_2}}|\Phi_{U',\mathbf{b}}\rangle + O_{\mathbf{b_2},\mathbf{b_1}} \langle\Phi_{U',\mathbf{b}}|\Phi_{U,\mathbf{b_2}}\rangle\langle\Phi_{U,\mathbf{b_1}}|\Phi_{U',\mathbf{b}}\rangle$$

$$= \frac{1}{2^n} \left[O_{\mathbf{b_1},\mathbf{b_2}} e^{i(\theta_{\mathbf{b},\mathbf{b_1}} - \theta_{\mathbf{b},\mathbf{b_2}})} + O_{\mathbf{b_2},\mathbf{b_1}} e^{i(\theta_{\mathbf{b},\mathbf{b_2}} - \theta_{\mathbf{b},\mathbf{b_1}})} \right]$$

$$\leq \frac{1}{2^n} \left(|O_{\mathbf{b_1},\mathbf{b_2}}| + |O_{\mathbf{b_2},\mathbf{b_1}}| \right). \tag{C12}$$

Now since $\sum_{\mathbf{b} \in \{0,1\}^n} \mathrm{tr}(\rho \Phi_{U',\mathbf{b}}) = 1$, we can conclude that

$$\mathbb{E}\text{tr}[O_0\mathcal{M}^{-1}(\Phi_{U',\mathbf{b}})]^2$$

$$\leq \frac{2^n+1}{(2^{2n})}\sum_{U'\neq\mathbb{I}}\sum_{\mathbf{b}\in\{0,1\}^n}\left(\sum_{\mathbf{b}_1\neq\mathbf{b}_2}|O_{\mathbf{b}_1,\mathbf{b}_2}|\right)^2\text{tr}[\rho\Phi_{U',\mathbf{b}}]$$

$$= \frac{2^n+1}{2^n}(C_{l_1}(O))^2, \tag{C13}$$

where $C_{l_1}(O) = \sum_{\mathbf{b}_1\neq\mathbf{b}_2}|O_{\mathbf{b}_1,\mathbf{b}_2}|$ denotes the l_1 norm of quantum coherence [54]. Therefore, the scale of the theoretical upper bound is no larger than that of the norm if the norm remains a constant level. ∎

Finally, we introduce the off-diagonal fidelity as the observable in our numerical simulation. We define the observable $O(a) = a(|0\rangle\langle1|^{\otimes n} + |1\rangle\langle0|^{\otimes n}) + (1-a)(|0\rangle\langle0|^{\otimes n} + |1\rangle\langle1|^{\otimes n})$. It shows that when $a = 0$, $O(a)$ turns to the off-diagonal fidelity, and when $a = 0.5$, $O(a)$ turns to the fidelity of the GHZ state. Therefore, we build a connection between the fidelity and off-diagonal fidelity.

As is shown in Fig. 7, the fitting curve in $a \geq 0.2$ is approximately straight, with a slope of 0.5. However in $a = 0$, the variance can be considered as a constant. The sudden transformation of the variance with the parameter a demonstrates the tremendous impact of the diagonal terms on the variance itself, especially when the qubit number increases. This is because the diagonal terms of the observable make an exponential contribution to the variance with the qubit number. Hence the significance of dividing the off-diagonal part is demonstrated.

3. Example 1: Estimating local observables with MCM

In this example, we focus on the observable being a k-local operator $O_k = O(\theta)^{\otimes k} \otimes \mathbb{I}^{\otimes(n-k)}$, where $O(\theta) = \cos(\theta)Z + \sin(\theta)X$, and the processed state $\rho = (|0\rangle\langle0|)^{\otimes n}$ with the qubit number $n = 8$. We compare numerically the performance of estimating such an observable using our MCM approach with the Clifford and Pauli measurement as shown in Fig. 8. Note that each experiment repeats $N = 10\,000$ trials.

We find that the performance of the MCM protocol behaves almost the same as the Clifford measurement. Specifically, the estimation variance of Pauli measurement grows exponentially with the locality number k, while the other two remain constant. Figure 8(c) shows that the intersection point of these three approaches lies in around $k = 5$. Moreover, the variance of single Pauli observables in our approach can be theoretically computed. Generally, we have $O = O_0 = P$, where P is a nonidentity Pauli matrix. For any MUB, there exists a Clifford element V and an n-bit binary vector $\mathbf{b}'$ such that $V^\dagger Z_{\mathbf{b}'}V = P$. Thus, we
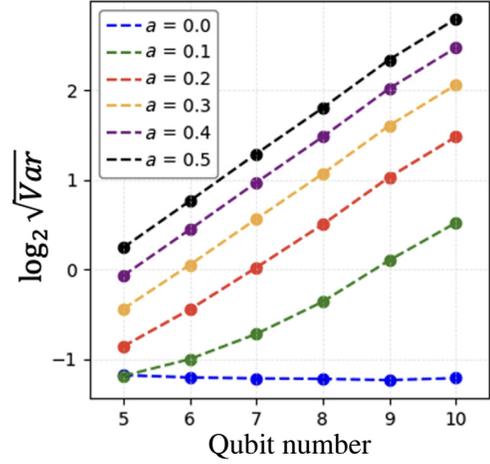
FIG. 7. The statistical variance of shadow estimation with $N = 10\,000$ using MCM with different parameters a.

have

$$\mathbb{E}\text{tr}[O_0\mathcal{M}^{-1}(\Phi_{U,\mathbf{b}})]^2 = (2^n+1)\left(\sum_{U\neq V}+\sum_{U=V}\right)$$

$$\times\sum_{\mathbf{b}\in\{0,1\}^n}(\langle\mathbf{b}|UPU^\dagger|\mathbf{b}\rangle)^2\langle\mathbf{b}|U\rho U^\dagger|\mathbf{b}\rangle. \tag{C14}$$

After reviewing Eq. (A1) and knowing that $\mathbf{b}' \neq 0$, we have

$$\langle\mathbf{b}|UPU^\dagger|\mathbf{b}\rangle = \langle\mathbf{b}|UV^\dagger Z_{\mathbf{b}'}VU^\dagger|\mathbf{b}\rangle$$

$$= \frac{1}{2^n}\sum_{l\in\{0,1\}^n}(-1)^{l\cdot\mathbf{b}'}\langle\mathbf{b}|UV^\dagger|l\rangle\langle l|VU^\dagger|\mathbf{b}\rangle$$

$$= \begin{cases} 0, & U \neq V \\ (-1)^{\mathbf{b}\cdot\mathbf{b}'}, & U = V. \end{cases} \tag{C15}$$

Therefore, Eq. (C15) can be converted to $\mathbb{E}\text{tr}[O_0\mathcal{M}^{-1}(\Phi_{U,\mathbf{b}})]^2 = 2^n + 1$, which grows exponentially with the qubit number. On the other hand, the theoretical bound of the Clifford variance is at a scale of $\text{tr}(O^2)$, which is also $O(2^n)$. The theoretical result shows that the variance of both approaches grow with the same scale of the qubit number, and the numerical result shows that at a wider range, the performance of the Clifford measurement and the MCM are almost the same.

4. Example 2: Haar random fidelities

In this example, we utilize the random fidelities to access the average performance of different random measurements. The procedure for the simulation experiments is as follows. First we randomly generate 100 quantum states, denoted as $\mathcal{E} = \{|\Phi_i\rangle\}_{i=1}^{100}$ on the Haar measure, then we select the first state as the observable to estimate the fidelity $|\langle\Phi_1|\Phi_i\rangle|^2$ through shadow tomography.
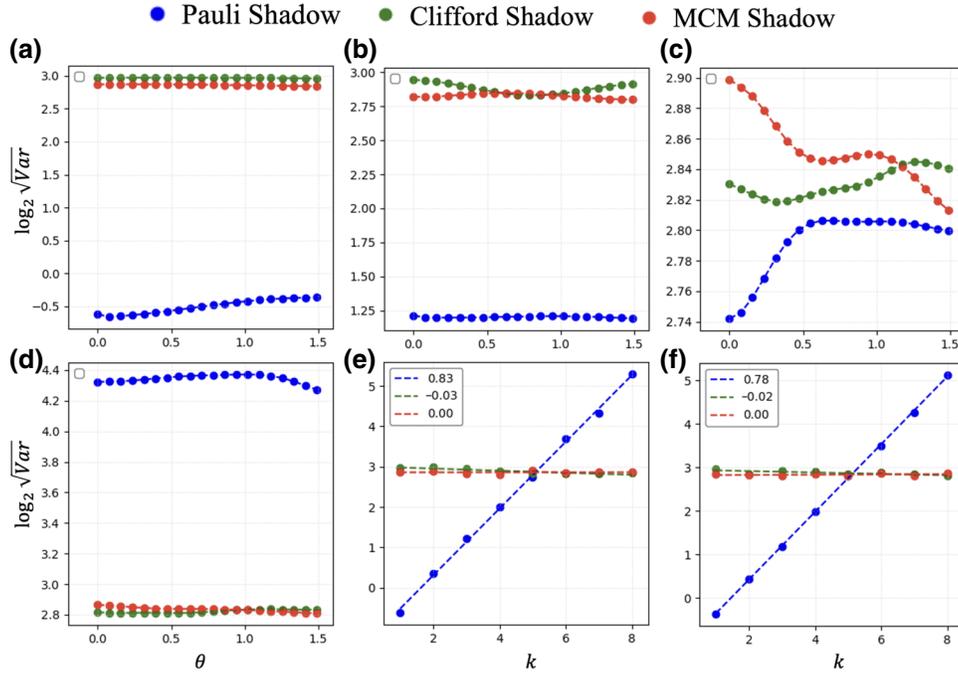
FIG. 8. The statistical variance of estimating local observable through Pauli measurement (blue), Clifford measurement (green), and our MCM approach (red), respectively, with $N = 10\,000$. In (a)–(d), we investigate the dependence of $\log_2 \sqrt{Var}$ with the phase factor θ in the observable O_k, with locality $k = 1, 3, 5, 7$, respectively. In (e),(f), we show the dependence of the estimation variance with the locality k, with $\theta = 0, \pi/2$, respectively. The dotted lines in (e),(f) show the fitting curves and the corresponding slopes of these cases.

This process is repeated $10\,000$ times, and the logarithmic variance is calculated to evaluate the performance of the random measurements. Due to the property of Haar measure, the expectation $\mathbb{E}_{\rho \sim \mathcal{E}} = \mathbb{I}/2^n$, which indicates that the variance of our approach in estimating random quantum states is on the same scale as the Clifford measurement. This is illustrated in Fig. 9.

FIG. 9. The statistical variance of shadow estimation with $N = 10\,000$ using Pauli measurement (blue), Clifford measurement (green), and MCM protocol (red), respectively. The dotted lines show the fitting curves and the corresponding slopes of the logarithmic variance of these cases.

APPENDIX D: BIASED-MCM SHADOW ESTIMATION

1. The unbiasedness of the estimator

In this section, we prove the unbiasedness of the estimator shown in the main text. Here we assume that all the probability $P_U \neq 0$ for all possible $U \in \mathcal{E}_{\text{MUB}}$. By definition,

$$
\begin{aligned}
\mathbb{E}_{U,\mathbf{b}} \widehat{O_0} &= \sum_{U \in \mathcal{E}_{\text{MUB}}, \mathbf{b}} \frac{\langle O_0 \rangle_{U,\mathbf{b}}}{P_U} \cdot P_U \cdot \langle \rho \rangle_{U,\mathbf{b}} \\
&= \sum_{U \in \mathcal{E}_{\text{MUB}}, \mathbf{b}} \langle O_0 \rangle_{U,\mathbf{b}} \langle \rho \rangle_{U,\mathbf{b}} \\
&= \text{tr}(O_0 \rho) + \text{tr}(O_0)\text{tr}(\rho) = \text{tr}(O_0 \rho),
\end{aligned} \tag{D1}
$$

where in the third line we apply the 2-design of MUB as that for Eq. (14) in the main text. It is clear that $\hat{O} = \widehat{O_0} + 2^{-n}\text{tr}(O)$ is also unbiased by some constant shift.

2. Proof of Theorem 3

Here we calculate the variance of the estimator in this section. By definition,

$$
\text{Var}_{\text{biased}-\mathcal{E}_{\text{MUB}}}(\hat{O}) = \text{Var}_{\text{biased}-\mathcal{E}_{\text{MUB}}}(\widehat{O_0})
$$
$$
= \mathbb{E}\widehat{O_0}^2 - \bar{O_0}^2. \qquad (D2)
$$

By maximizing on possible unknown state ρ, the first term is bounded by

$$
\mathbb{E}_{U,\mathbf{b}}\widehat{O_0}^2 \leq \max_{\rho} \mathbb{E}_{U,\mathbf{b}}\widehat{O_0}^2
$$
$$
= \max_{\rho} \mathbb{E}_{U,\mathbf{b}} \frac{\langle O_0 \rangle_{U,\mathbf{b}}^2}{P_U^2}
$$
$$
= \max_{\rho} \sum_U \sum_{\mathbf{b}} \frac{\langle O_0 \rangle_{U,\mathbf{b}}^2}{P_U^2} \times P_U \times \langle \rho \rangle_{U,\mathbf{b}}
$$
$$
= \max_{\rho} \sum_U \sum_{\mathbf{b}} \frac{\langle O_0 \rangle_{U,\mathbf{b}}^2 \langle \rho \rangle_{U,\mathbf{b}}}{P_U}. \qquad (D3)
$$

Considering the constraint that $\sum_{\mathbf{b}} \langle \rho \rangle_{U,\mathbf{b}} = 1$, for a fixed U, one can further bound it using the maximal term of $\langle O_0 \rangle_{U,\mathbf{b}}^2$, and finally has

$$
\text{Var}_{\text{biased}-\mathcal{E}_{\text{MUB}}}(\hat{O}) \leq \sum_{U \in \mathcal{E}_{\text{MUB}}} \frac{\max_{\mathbf{b}\in\{0,1\}^n} \langle O_0 \rangle_{U,\mathbf{b}}^2}{P_U}. \qquad (D4)
$$

And our optimization problem on this upper bound is as follows:

$$
\min P_U: \quad \sum_{U \in \mathcal{E}_{\text{MUB}}} \frac{\max_{\mathbf{b}\in\{0,1\}^n} \langle O_0 \rangle_{U,\mathbf{b}}^2}{P_U},
$$
$$
\text{such that:} \quad \sum_{U \in \mathcal{E}_{\text{MUB}}} P_U = 1. \qquad (D5)
$$

Using the Lagrange dual method, the solution to this problem is $(\sum_{U\in\mathcal{E}_{\text{MUB}}} B_U)^2$, and the corresponding $P_U = B_U / \sum_{U'\in\mathcal{E}} B_{U'}$, with B_U being short for $\max_{\mathbf{b}\in\{0,1\}^n} |\langle O_0 \rangle_{U,\mathbf{b}}|$. For the observable that is decomposed in the canonical form in Eq. (18) in the main text, one has $\langle O_0 \rangle_{U,\mathbf{b}} = \alpha_{U,\mathbf{b}} - 2^{-n}\text{tr}(O)$.

The optimal solution can be further simplified to the following form:

$$
\sum_{U \in \mathcal{E}_{\text{MUB}}} B_U = \sum_{U \in \mathcal{E}_{\text{MUB}}} \max_{\mathbf{b}\in\{0,1\}^n} |\text{tr}(O_0 \Phi_{U,\mathbf{b}})|
$$
$$
= \sum_{U \in \mathcal{E}_{\text{MUB}}} \max_{\mathbf{b}\in\{0,1\}^n} \left| 2^{-n} \sum_{\mathbf{m}\in\{0,1\}^n} (-1)^{\mathbf{b}\cdot\mathbf{m}} \text{tr}(S_{\mathbf{m}} O_0) \right|
$$
$$
\leq \sum_{U \in \mathcal{E}_{\text{MUB}}} 2^{-n} \sum_{\mathbf{m}\in\{0,1\}^n} |\text{tr}(S_{\mathbf{m}} O_0)|
$$
$$
= 2^{-n} \sum_{P \in \mathbf{P}_n^*} |\text{tr}(P O_0)| =: \mathcal{D}(O_0). \qquad (D6)
$$

Here in the second line, we express the state $\Phi_{U,\mathbf{b}}$ by its stabilizers using Eq. (A3), and the third line is by the fact that MUBs can cover all the Paulis, and also $\text{tr}(\mathbb{I}O_0) = 0$. The final line we use the definition of stabilizer norm [37]. As a result, the optimal solution is upper bounded by $\mathcal{D}(O_0)^2$. We also remark that $\mathcal{D}(O_0) = \mathcal{D}(O) - |\text{tr}(O)|/d$.

Some remarks on the possibility of $P_U = 0$. Suppose for some U, $P_U = 0$ and we have $\max_{\mathbf{b}\in\{0,1\}^n} |\langle O_0 \rangle_{U,\mathbf{b}}| = 0$. That is, $\langle O_0 \rangle_{U,\mathbf{b}} = 0$, $\forall \mathbf{b}$ under this U. Let us denote the set with $P_U \neq 0$ as $\mathcal{E}' \subset \mathcal{E}_{\text{MUB}}$. We show as follows that the possibility of $P_U = 0$ does *not* affect the unbiasedness proved in Appendix D 1. In particular, for Eq. (D1), we can rewrite its second line by only summing the unitary in $\mathcal{E}'$,

$$
\mathbb{E}_{U,\mathbf{b}}\widehat{O_0} = \sum_{U \in \mathcal{E}',\mathbf{b}} \langle O_0 \rangle_{U,\mathbf{b}} \langle \rho \rangle_{U,\mathbf{b}}
$$
$$
= \sum_{U \in \mathcal{E}_{\text{MUB}},\mathbf{b}} \langle O_0 \rangle_{U,\mathbf{b}} \langle \rho \rangle_{U,\mathbf{b}} = \text{tr}(\rho O_0), \qquad (D7)
$$

where in the second line we extend the range of summation on U by the fact that when $U \in \mathcal{E}_{\text{MUB}}/\mathcal{E}'$, $\langle O_0 \rangle_{U,\mathbf{b}} = 0$, $\forall \mathbf{b}$.

3. Proof of Lemma 2

First, we prove the first property as follows.

Proof.

$$
\alpha_{U,\mathbf{b}} = \langle 0|VU^\dagger|\mathbf{b}\rangle\langle\mathbf{b}|UV^\dagger|0\rangle
$$
$$
= \frac{1}{2^n} \sum_{\mathbf{m}\in\{0,1\}^n} (-1)^{\mathbf{m}\cdot\mathbf{b}} \langle 0|VU^\dagger Z_{\mathbf{m}} UV^\dagger|0\rangle. \qquad (D8)
$$

According to Observation 2, we have $\bigcup_{U\in\mathcal{E}_{\text{MUB}}} \{U^\dagger Z_{\mathbf{m}} U\}_{\mathbf{m}\neq 0} = \mathbf{P}_*^n$. And by the definition of Clifford, rotate a Clifford and we have $\bigcup_{U\in\mathcal{E}_{\text{MUB}}} \{|VU^\dagger Z_{\mathbf{m}} UV^\dagger|\}_{\mathbf{m}\neq 0} = \mathbf{P}_*^n$. Also, the intersection of any two sets $\{|VU^\dagger Z_{\mathbf{m}} UV^\dagger|\}_{\mathbf{m}\neq 0}$ is $\varnothing$.

Suppose $VU^\dagger Z_{\mathbf{m}} UV^\dagger = (-1)^{p_{U,\mathbf{m}}} S_{U,\mathbf{m}}$, where $p_{U,\mathbf{m}}$ is a binary indicator of phases, and $S_{U,\mathbf{m}}$ is a Pauli string. Then

$$\langle 0|VU^\dagger Z_{\mathbf{m}} UV^\dagger|0\rangle = (-1)^{p_{U,\mathbf{m}}} \times \mathbf{1}\{S_{U,\mathbf{m}} \rhd Z^{\otimes n}\}. \quad \text{(D9)}$$

$A \rhd Z^{\otimes n}$ [17] means that the Pauli string A consists of solely $\mathbb{I}$ and Z. For example, $IZ, ZI, ZZ \rhd ZZ$ and $XI \not\rhd ZZ$. Generally $\alpha_{U,\mathbf{b}}$ can be rewritten as

$$\alpha_{U,\mathbf{b}} = \frac{1}{2^n} \sum_{\mathbf{m}\in\{0,1\}^n} (-1)^{\mathbf{m}\cdot\mathbf{b}+p_{U,\mathbf{m}}} \times \mathbf{1}\{S_{U,\mathbf{m}} \rhd Z^{\otimes n}\}$$

$$\leq \frac{1}{2^n} \sum_{\mathbf{m}\in\{0,1\}^n} \mathbf{1}\{S_{U,\mathbf{m}} \rhd Z^{\otimes n}\}. \quad \text{(D10)}$$

The inequality holds if and only if $\mathbf{m} \cdot \mathbf{b} + p_{U,\mathbf{m}} = 0$ for all $\mathbf{m} : \{S_{U,\mathbf{m}} \rhd Z^{\otimes n}\}$. Since $\mathbf{m}$ that satisfies these solutions can form a group, where the solution for $\mathbf{b}$ must exist. Therefore

$$\max_{\mathbf{b}\in\{0,1\}^n} \alpha_{U,\mathbf{b}} = \frac{1}{2^n} \sum_{\mathbf{m}\in\{0,1\}^n} \mathbf{1}\{S_{U,\mathbf{m}} \rhd Z^{\otimes n}\}$$

$$= \frac{1}{2^n} + \frac{1}{2^n} \sum_{\mathbf{m}\neq\mathbf{0}} \mathbf{1}\{S_{U,\mathbf{m}} \rhd Z^{\otimes n}\}. \quad \text{(D11)}$$

And

$$\sum_{U\in\mathcal{E}_{\text{MUB}}} \max_{\mathbf{b}\in\{0,1\}^n} \alpha_{U,\mathbf{b}} = \frac{2^n+1}{2^n} + \frac{1}{2^n} \sum_{S\in\mathbf{P}_*^n} \mathbf{1}\{S \rhd Z^{\otimes n}\} = 2. \quad \text{(D12)}$$

■

Next, we prove the second property as follows.

Proof. In the beginning of the proof, we consider the Z tableau of $UV^\dagger$: $T_{UV^\dagger} = [C, D]$. We call several rowsum operations in Ref. [40] to perform Gaussian elimination on the matrix C. The result of the Gaussian elimination is the Clifford W, whose matrix C is turned to the row simplest form $\begin{pmatrix} C'_{r_U\times n} \\ \mathbb{0} \end{pmatrix}$.

A plain conclusion is that call rowsum of Z tableau will change only the order of summation in Eq. (D8), not the value of $\alpha_{U,\mathbf{b}}$. Therefore,

$$\alpha_{U,\mathbf{b}} = \frac{1}{2^n} \sum_{\mathbf{m}\in\{0,1\}^n} (-1)^{\mathbf{m}\cdot\mathbf{b}} \langle 0|W^\dagger Z_{\mathbf{m}} W|0\rangle$$

$$= \frac{1}{2^n} \sum_{\mathbf{m}'\in\{0,1\}^{n-r_U}} (-1)^{\mathbf{m}'\cdot\mathbf{b}} \langle 0|W^\dagger Z_{\mathbf{m}'} W|0\rangle. \quad \text{(D13)}$$

The vector $\mathbf{m}' = [0,\dots,0,m_{r_U},\dots,m_{n-1}]$. The equation holds true because $\langle 0|W^\dagger Z_{\mathbf{m}} W|0\rangle = 0$ for all $W^\dagger Z_{\mathbf{m}} W$ that

is composed of more than $\{\mathbb{I}, Z\}$, which is $\{W^\dagger Z_{\mathbf{m}'} W \not\rhd Z^{\otimes n}\}$. In this scenario, the first r_U bits of $\mathbf{m}$ cannot be all zero. Since $W^\dagger Z_{\mathbf{m}'} W$ now consists of solely Z Paulis, we set $W^\dagger Z_{\mathbf{m}'} W = \prod_{j=r_U}^{n-1} (-1)^{p_j m_j} Z_j^{m_j}$. Hence

$$\langle 0|W^\dagger Z_{\mathbf{m}'} W|0\rangle = \langle 0| \prod_{j=r_U}^{n-1} (-1)^{p_j m_j} Z_j^{m_j} |0\rangle. \quad \text{(D14)}$$

A special property for arbitrary Z-Pauli matrices Z_1, Z_2 is that $\langle 0|Z_1 Z_2|0\rangle = \langle 0|Z_1|0\rangle \langle 0|Z_2|0\rangle$. Hence the upper formula can be further simplified to

$$\langle 0|W^\dagger Z_{\mathbf{m}'} W|0\rangle = \langle 0| \prod_{j=r_U}^{n-1} (-1)^{p_j m_j} Z_j^{m_j} |0\rangle$$

$$= \prod_{j=r_U}^{n-1} (-1)^{m_j \cdot (p_j + \sum_k^{n-1} \delta_{jk})}, \quad \text{(D15)}$$

where δ_{jk} is the element of matrix D in the Z tableau of W (after Gaussian elimination). Now as a summarization,

$$\alpha_{U,\mathbf{b}} = \frac{1}{2^n} \sum_{\mathbf{m}'\in\{0,1\}^{n-r_U}} (-1)^{\mathbf{m}'\cdot\mathbf{b}} \langle 0|W^\dagger Z_{\mathbf{m}'} W|0\rangle$$

$$= \frac{1}{2^n} \sum_{\mathbf{m}'\in\{0,1\}^{n-r_U}} (-1)^{\mathbf{m}'\cdot\mathbf{b}} \prod_{j=r_U}^{n-1} (-1)^{m_j \cdot (p_j + \sum_k^{n-1} \delta_{jk})}$$

$$= \frac{1}{2^n} \sum_{\mathbf{m}'\in\{0,1\}^{n-r_U}} \prod_{j=r_U}^{n-1} (-1)^{m_j \cdot (p_j + \sum_k^{n-1} \delta_{jk} + b_j)}$$

$$= \frac{1}{2^n} \prod_{j=r_U}^{n-1} 1 + (-1)^{b_j + p_j + \sum_k^{n-1} \delta_{jk}}$$

$$= \begin{cases} 2^{-r_U}, b_j + p_j + \sum_k^{n-1} \delta_{jk} = 0 & \text{for all } j, \\ 0, & \text{otherwise.} \end{cases} \quad \text{(D16)}$$

Therefore, for any $\mathbf{b} \in \{0,1\}^n$, $\alpha_{U,\mathbf{b}}$ has 2^{r_U} in 2^{-r_U} and the remaining $2^n - 2^{r_U}$ in 0. ■

4. Proof of Proposition 4

Proof. Suppose for a single shot, the unitary rotation is $U \in \mathcal{E}_{\text{MUB}}$ and the measurement outcome is $|\mathbf{b}\rangle$. Before calculating the variance, we discuss the property of the probability P_U. According to the definition, $P_U = B_U/\sum_{U'\in\mathcal{E}_{\text{MUB}}} B_{U'}$. By introducing Eq. (D12), we conclude

that

$$\sum_{U' \in \mathcal{E}_{\text{MUB}}} B_{U'} = \sum_{U' \in \mathcal{E}_{\text{MUB}}} (\max_{\mathbf{b}} \alpha_{U,\mathbf{b}} - 2^{-n}) = 1 - 2^{-n}.$$

(D17)

By applying Eq. (21) in the main text, we calculate the variance

$$\text{Var}_{\text{biased}-\mathcal{E}_{\text{MUB}}}(\hat{O}) \le \left(\sum_{U' \in \mathcal{E}_{\text{MUB}}} B_{U'} \right)^2 \le 1.$$

(D18)

A specific example is that for $\rho = O$ are Clifford stabilizers, the variance is zero. Here we prove it.

Also according to Eq. (D16), $\max_{\mathbf{b}} \alpha_{U,\mathbf{b}} = 2^{-r_U}$. Thus,

$$\sum_{U \in \mathcal{E}_{\text{MUB}}} 2^{-r_U} = 2.$$

(D19)

In this task, the result of each snapshot for U and $\mathbf{b}$ is defined as

$$\hat{O} = \frac{\langle O_0 \rangle_{U,\mathbf{b}}}{P_U} + \frac{\text{tr}(O)}{2^n}.$$

(D20)

Hence the estimation is of no variance. Since $O = \rho$ is a stabilizer state, the result of $\alpha_{U,\mathbf{b}}$ has two options: 2^{-r_U} and 0, which also equals the probability of the result $\mathbf{b}$ being measured. This means that the only possible result of $\alpha_{U,\mathbf{b}}$ is 2^{-r_U}. As a result,

$$\text{Var}[\hat{O}] = \mathbb{E}_{\mathbf{U},\mathbf{b}}(\hat{O} - 1)^2$$

$$= \sum_{U \in \mathcal{E}_{\text{MUB}}} P_U \sum_{\mathbf{b} \in \{0,1\}^n} \langle \rho \rangle_{U,\mathbf{b}} \left[\frac{\langle O_0 \rangle_{U,\mathbf{b}}}{P_U} + 2^{-n} - 1 \right]^2$$

$$= \sum_{U \in \mathcal{E}_{\text{MUB}}} P_U \sum_{\mathbf{b} \in \{0,1\}^n} \alpha_{U,\mathbf{b}} \left[\frac{\alpha_{U,\mathbf{b}} - 2^{-n}}{B_U / \sum_{U'} B_{U'}} + 2^{-n} - 1 \right]^2$$

$$= \sum_{U \in \mathcal{E}_{\text{MUB}}} P_U \sum_{\mathbf{b} \in \{0,1\}^n} \alpha_{U,\mathbf{b}} \left[\frac{\alpha_{U,\mathbf{b}} - 2^{-n}}{|\max_{\mathbf{b}} \alpha_{U,\mathbf{b}} - 2^{-n}|} (1 - 2^{-n}) + 2^{-n} - 1 \right]^2$$

$$= \sum_{U \in \mathcal{E}_{\text{MUB}}} P_U \sum_{\mathbf{b} \in \{0,1\}^n} \alpha_{U,\mathbf{b}} \left[\frac{\alpha_{U,\mathbf{b}} - 2^{-n}}{|\max_{\mathbf{b}} \alpha_{U,\mathbf{b}} - 2^{-n}|} (1 - 2^{-n}) + 2^{-n} - 1 \right]^2$$

$$= \sum_{U \in \mathcal{E}_{\text{MUB}}} P_U 2^{r_U} 2^{-r_U} \left[\frac{2^{-r_U} - 2^{-n}}{2^{-r_U} - 2^{-n}} (1 - 2^{-n}) + 2^{-n} - 1 \right]^2$$

$$= 0.$$

(D21)

∎

ALGORITHM 2.　　Find $U \in \mathcal{E}_{\text{MUB}}$ for Pauli operator P with $P \blacktriangleright U$.

Input: The qubit number n, the non-identity Pauli $P \in \mathbf{P}_n^*$, and the ensemble $\mathcal{E}_{\text{MUB}}$.
Output: MUB element $U \in \mathcal{E}_{MUB}$.
1: **if** $P \triangleright Z^{\otimes n}$ **then** $U = \mathbb{I}$
2: **else** Decompose P as $P = \bigotimes_{i=0}^{n} i^{a_j b_j} X^{a_j} Z^{b_j}$.
3:　　Solve the linear system

$$\sum_i (\sum_{j'} \beta_{j+j'}^{(i)} a_{j'}) v_i = b_j \text{ for all } j,$$

(D22)

　　to obtain the parameter v so that the Z-Tableau of MUB element is $[\mathbb{I}, D_v]$.
4: **end if**

5. Efficient sampling of P_U for stabilizer states

The main content of this subsection is to introduce a method to efficiently sample unitary $U \in \mathcal{E}_{\mathrm{MUB}}$ with respect to P_U defined in Eqs. (19) and (20) in the main text for stabilizer states.

Initially, we give an algorithm to identify the unique $U \in \mathcal{E}_{\mathrm{MUB}}$ for a given Pauli operator $P \in \mathbf{P}_n^*$, such that there exists $\mathbf{m} \in \{0, 1\}^n$ satisfying $U^\dagger Z_{\mathbf{m}} U = P$. And we denote this relation as $P \blacktriangleright U$.

Equation (D22) is indeed an explicit form of $D_v \mathbf{a} = \mathbf{b}$ via Eq. (11), where $\mathbf{a}$ and $\mathbf{b}$ are vectors of parameters $\{a_i\}$ and $\{b_i\}$. The algorithm requires $O(n^\omega)$ (with $\omega < 2.375512$) [83] complexity for solving a system of linear equations [84].

After introducing Algorithm 2, we illustrate how to efficiently sample Clifford stabilizers. For a stabilizer state $O = V^\dagger |\mathbf{0}\rangle\langle\mathbf{0}| V$, the sampling probability is

$$
\begin{aligned}
P_U &= \frac{B_U}{\sum_{U' \in \mathcal{E}_{\mathrm{MUB}}} B_{U'}} \\
&= \frac{\max_{\mathbf{b} \in \{0,1\}^n} \alpha_{U,\mathbf{b}} - 2^{-n}}{1 - 2^{-n}} \\
&= \frac{1}{2^n - 1} \sum_{\mathbf{m} \neq 0} \mathbf{1}\{V U^\dagger Z_{\mathbf{m}} U V^\dagger \triangleright Z^{\otimes n}\} \\
&= \frac{1}{2^n - 1} \sum_{\mathbf{m} \neq 0} \mathbf{1}\{U^\dagger Z_{\mathbf{m}} U \blacktriangleright V\} \quad \text{(D23)}
\end{aligned}
$$

on account of the definition of P_U and Eq. (D11). Therefore, the sampling protocol is as follows. First, sample a nonidentity stabilizer P_V of O with equal probability. Second, find the sole MUB element U so that $P_V \blacktriangleright U$ using Algorithm 2.

Third, compute the probability

$$
P_U = \frac{2^{-r_U} - 2^{-n}}{1 - 2^{-n}} \quad \text{(D24)}
$$

by Eq. (D16). Here $r_U = \mathrm{rank}_{\mathbb{F}_2}(C)$ on the binary field, and the Z tableau of $U V^\dagger$ is $T_{UV^\dagger} = [C, D]$.

In conclusion, the classical computation complexity for each sampling process is $O(n^3)$.

---

[1] V. Gebhart, R. Santagati, A. A. Gentile, E. M. Gauger, D. Craig, N. Ares, L. Banchi, F. Marquardt, L. Pezzè, and C. Bonato, Learning quantum systems, Nat. Rev. Phys. **5**, 141 (2023).

[2] A. Anshu and S. Arunachalam, A survey on the complexity of learning quantum states, Nat. Rev. Phys. **6**, 59 (2024).

[3] E. Altman, *et al.*, Quantum simulators: Architectures and opportunities, PRX Quantum **2**, 017003 (2021).

[4] Y. Alexeev, *et al.*, Quantum computer systems for scientific discovery, PRX Quantum **2**, 017001 (2021).

[5] J. Eisert, D. Hangleiter, N. Walk, I. Roth, D. Markham, R. Parekh, U. Chabaud, and E. Kashefi, Quantum certification and benchmarking, Nat. Rev. Phys. **2**, 382 (2020).

[6] M. Kliesch and I. Roth, Theory of quantum system certification, PRX Quantum **2**, 010201 (2021).

[7] J. Emerson, R. Alicki, and K. Życzkowski, Scalable noise estimation with random unitary operators, J. Opt. B: Quantum Semiclassical Opt. **7**, S347 (2005).

[8] 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).

[9] X. Mi, P. Roushan, C. Quintana, S. Mandra, J. Marshall, C. Neill, F. Arute, K. Arya, J. Atalaya, R. Babbush, *et al.*, Information scrambling in quantum circuits, Science **374**, 1479 (2021).

[10] H.-Y. Huang, Learning quantum states from their classical shadows, Nat. Rev. Phys. **4**, 81 (2022).

[11] A. Elben, S. T. Flammia, H.-Y. Huang, R. Kueng, J. Preskill, B. Vermersch, and P. Zoller, The randomized measurement toolbox, Nat. Rev. Phys. **5**, 9 (2023).

[12] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, Sample-optimal tomography of quantum states, IEEE Trans. Inf. Theory **63**, 5628 (2017).

[13] S. T. Flammia, D. Gross, Y.-K. Liu, and J. Eisert, Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators, New J. Phys. **14**, 095022 (2012).

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

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

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

[17] H.-Y. Huang, R. Kueng, and J. Preskill, Efficient estimation of Pauli observables by derandomization, Phys. Rev. Lett. **127**, 030503 (2021).

[18] C. Hadfield, S. Bravyi, R. Raymond, and A. Mezzacapo, Measurements of quantum hamiltonians with locally-biased classical shadows, Commun. Math. Phys. **391**, 951 (2022).

[19] B. Wu, J. Sun, Q. Huang, and X. Yuan, Overlapped grouping measurement: A unified framework for measuring quantum states, Quantum **7**, 896 (2023).

[20] K. Van Kirk, J. Cotler, H.-Y. Huang, and M. D. Lukin, Hardware-efficient learning of quantum many-body states, arXiv:2212.06084.

[21] Y. Zhou and Q. Liu, Performance analysis of multi-shot shadow estimation, Quantum **7**, 1044 (2023).

[22] Y. Zhou and Z. Liu, A hybrid framework for estimating nonlinear functions of quantum states, arXiv:2208.08416.

[23] T. Zhang, J. Sun, X.-X. Fang, X.-M. Zhang, X. Yuan, and H. Lu, Experimental quantum state measurement with classical shadows, Phys. Rev. Lett. **127**, 200501 (2021).

[24] R. Stricker, M. Meth, L. Postler, C. Edmunds, C. Ferrie, R. Blatt, P. Schindler, T. Monz, R. Kueng, and M. Ringbauer,

Experimental single-setting quantum state tomography, PRX Quantum **3**, 040310 (2022).

[25] K. An, Z. Liu, T. Zhang, S. Li, Y. Zhou, X. Yuan, L. Wang, W. Zhang, G. Wang, and H. Lu, Efficient characterizations of multiphoton states with ultra-thin integrated photonics, arXiv:2308.07067.

[26] M. Ozols, Clifford group, Essays at University of Waterloo, Spring (2008), http://home.lu.lv/~sd20008/papers/essays/Clifford%20group%20[paper].pdf.

[27] S. Bravyi, J. A. Latone, and D. Maslov, 6-qubit optimal Clifford circuits, npj Quantum Inf. **8**, 79 (2022).

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

[29] E. Van Den Berg, in *2021 IEEE International Conference on Quantum Computing and Engineering (QCE)* (IEEE, 2021), p. 54, https://ieeexplore.ieee.org/document/9519499.

[30] D. Maslov and B. Zindorf, Depth optimization of CZ, CNOT, and Clifford circuits, IEEE Trans. Quantum Eng. **3**, 1 (2022).

[31] C. Bertoni, J. Haferkamp, M. Hinsche, M. Ioannou, J. Eisert, and H. Pashayan, Shallow shadows: Expectation estimation using low-depth random clifford circuits, arXiv:2209.12924.

[32] M. C. Tran, D. K. Mark, W. W. Ho, and S. Choi, Measuring arbitrary physical properties in analog quantum simulation, Phys. Rev. X **13**, 011049 (2023).

[33] M. McGinley and M. Fava, Shadow tomography from emergent state designs in analog quantum simulators, Phys. Rev. Lett. **131**, 160601 (2023).

[34] M. J. Bremner, A. Montanaro, and D. J. Shepherd, Average-case complexity versus approximate simulation of commuting quantum computations, Phys. Rev. Lett. **117**, 080501 (2016).

[35] Y. Nakata, C. Hirche, M. Koashi, and A. Winter, Efficient quantum pseudorandomness with nearly time-independent hamiltonian dynamics, Phys. Rev. X **7**, 021006 (2017).

[36] G. Liu, Z. Xie, Z. Xu, and X. Ma, Group twirling and noise tailoring for multi-qubit controlled phase gates, arXiv:2309.15651.

[37] E. T. Campbell, Catalysis and activation of magic states in fault-tolerant architectures, Phys. Rev. A **83**, 032317 (2011).

[38] M. Howard and E. Campbell, Application of a resource theory for magic states to fault-tolerant quantum computing, Phys. Rev. Lett. **118**, 090501 (2017).

[39] L. Leone, S. F. Oliviero, and A. Hamma, Stabilizer Rényi entropy, Phys. Rev. Lett. **128**, 050402 (2022).

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

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

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

[43] D. Maslov and W. Yang, CNOT circuits need little help to implement arbitrary Hadamard-free Clifford transformations they generate, npj Quantum Inf. **9**, 96 (2023).

[44] S. Bandyopadhyay, P. O. Boykin, V. Roychowdhury, and F. Vatan, A new proof for the existence of mutually unbiased bases, Algorithmica **34**, 512 (2002).

[45] T. Durt, About mutually unbiased bases in even and odd prime power dimensions, J. Phys. A: Math. Gen. **38**, 5267 (2005).

[46] T. Durt, B.-G. Englert, I. Bengtsson, and K. Życzkowski, On mutually unbiased bases, Int. J. Quantum Inf. **8**, 535 (2010).

[47] H. Zhu, Mutually unbiased bases as minimal Clifford covariant 2-designs, Phys. Rev. A **91**, 060301 (2015).

[48] D. Gottesman, in *Proceedings of Symposia in Applied Mathematics*, Vol. 58 (2002), p. 221, https://arxiv.org/pdf/quant-ph/0004072.pdf.

[49] G. Beliakov, On fast matrix-vector multiplication with a Hankel matrix in multiprecision arithmetics, arXiv:1402.5287.

[50] C. Figgatt, A. Ostrander, N. M. Linke, K. A. Landsman, D. Zhu, D. Maslov, and C. Monroe, Parallel entangling operations on a universal ion-trap quantum computer, Nature **572**, 368 (2019).

[51] S. Bravyi, D. Maslov, and Y. Nam, Constant-cost implementations of Clifford operations and multiply-controlled gates using global interactions, Phys. Rev. Lett. **129**, 230501 (2022).

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

[53] J. Morris and B. Dakić, Selective quantum state tomography, arXiv:1909.05880.

[54] T. Baumgratz, M. Cramer, and M. B. Plenio, Quantifying coherence, Phys. Rev. Lett. **113**, 140401 (2014).

[55] A. Streltsov, G. Adesso, and M. B. Plenio, Colloquium: Quantum coherence as a resource, Rev. Mod. Phys. **89**, 041003 (2017).

[56] O. Gühne, C.-Y. Lu, W.-B. Gao, and J.-W. Pan, Toolbox for entanglement detection and fidelity estimation, Phys. Rev. A **76**, 030305 (2007).

[57] M. Wieśniak, T. Paterek, and A. Zeilinger, Entanglement in mutually unbiased bases, New J. Phys. **13**, 053047 (2011).

[58] L. Leone and L. Bittel, Stabilizer entropies are monotones for magic-state resource theory, arXiv:2404.11652.

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

[60] S. J. Evered, D. Bluvstein, M. Kalinowski, S. Ebadi, T. Manovitz, H. Zhou, S. H. Li, A. A. Geim, T. T. Wang, N. Maskara, *et al.*, High-fidelity parallel entangling gates on a neutral atom quantum computer, Nature **622**, 268 (2023).

[61] O. Guhne and G. Toth, Entanglement detection, Phys. Rep. **474**, 1 (2009).

[62] N. Friis, G. Vitagliano, M. Malik, and M. Huber, Entanglement certification from theory to experiment, Nat. Rev. Phys. **1**, 72 (2019).

[63] G. Tóth and O. Gühne, Detecting genuine multipartite entanglement with two local measurements, Phys. Rev. Lett. **94**, 060501 (2005).

[64] Y. Zhou, B. Xiao, M.-D. Li, Q. Zhao, Z.-S. Yuan, X. Ma, and J.-W. Pan, A scheme to create and verify scalable

entanglement in optical lattice, npj Quantum Inf. **8,** 99 (2022).

[65] H. Lu, Q. Zhao, Z.-D. Li, X.-F. Yin, X. Yuan, J.-C. Hung, L.-K. Chen, L. Li, N.-L. Liu, C.-Z. Peng, *et al.*, Entanglement structure: Entanglement partitioning in multipartite systems and its experimental detection using optimizable witnesses, Phys. Rev. X **8,** 021072 (2018).

[66] Y. Zhou, Q. Zhao, X. Yuan, and X. Ma, Detecting multipartite entanglement structure with minimal resources, npj Quantum Inf. **5,** 1 (2019).

[67] W.-Y. Zhang, M.-G. He, H. Sun, Y.-G. Zheng, Y. Liu, A. Luo, H.-Y. Wang, Z.-H. Zhu, P.-Y. Qiu, Y.-C. Shen, *et al.*, Scalable multipartite entanglement created by spin exchange in an optical lattice, Phys. Rev. Lett. **131,** 073401 (2023).

[68] Y. Nakata and M. Murao, Diagonal quantum circuits: Their computational power and applications, Eur. Phys. J. Plus **129,** 152 (2014).

[69] Y. Zhou and A. Hamma, Entanglement of random hypergraph states, Phys. Rev. A **106,** 012410 (2022).

[70] J. Chen, Y. Yan, and Y. Zhou, Magic of quantum hypergraph states, arXiv:2308.01886.

[71] G. Park, Y. S. Teo, and H. Jeong, Resource-efficient shadow tomography using equatorial measurements, arXiv:2311.14622.

[72] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O'Brien, Quantum error mitigation, Rev. Mod. Phys. **95,** 045005 (2023).

[73] S. Chen, W. Yu, P. Zeng, and S. T. Flammia, Robust shadow estimation, PRX Quantum **2,** 030348 (2021).

[74] D. E. Koh and S. Grewal, Classical shadows with noise, Quantum **6,** 776 (2022).

[75] D. H. Mahler, L. A. Rozema, A. Darabi, C. Ferrie, R. Blume-Kohout, and A. M. Steinberg, Adaptive quantum state tomography improves accuracy quadratically, Phys. Rev. Lett. **111,** 183601 (2013).

[76] C. Ferrie, Self-guided quantum tomography, Phys. Rev. Lett. **113,** 190404 (2014).

[77] S. Gandhari, V. V. Albert, T. Gerrits, J. M. Taylor, and M. J. Gullans, Continuous-variable shadow tomography, arXiv:2211.05149.

[78] S. Becker, N. Datta, L. Lami, and C. Rouzé, Classical shadow tomography for continuous variables quantum systems, IEEE Trans. Inf. Theory **70,** 3427 (2024).

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

[80] V. Shoup, New algorithms for finding irreducible polynomials over finite fields, Math. Comput. **54,** 435 (1990).

[81] V. Shoup, Fast construction of irreducible polynomials over finite fields, J. Symb. Comput. **17,** 371 (1994).

[82] R. Sarkar and E. v. d. Berg, On sets of commuting and anticommuting Paulis, arXiv:1909.08123.

[83] V. V. Williams, Y. Xu, Z. Xu, and R. Zhou, in *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)* (SIAM, 2024), p. 3792.

[84] O. H. Ibarra, S. Moran, and R. Hui, A generalization of the fast LUP matrix decomposition algorithm and applications, J. Algorithms **3,** 45 (1982).