

Efficient sampling for Pauli-measurement-based shadow tomography in direct fidelity estimation

Hyunho Cha^{✉*} and Jungwoo Lee^{✉†}

NextQuantum and Department of Electrical and Computer Engineering, Seoul National University, Seoul 08826, Republic of Korea



(Received 11 February 2025; accepted 13 August 2025; published 15 September 2025)

A constant number of random Clifford measurements allows the classical shadow protocol to perform direct fidelity estimation (DFE) with high precision. However, estimating properties of an unknown quantum state is expected to be more feasible with random Pauli measurements than with random Clifford measurements in the near future. Inspired by the importance sampling technique applied to sampling Pauli measurements for DFE, we show that similar strategies can be derived from classical shadows. Specifically, we describe efficient methods using only local Pauli measurements to perform DFE with Greenberger-Horne-Zeilinger (GHZ), W , and Dicke states, establishing tighter bounds (by factors of 14.22 and 16 for GHZ and W /Dicke, respectively) on the number of measurements required for desired precision. These protocols are derived by adjusting the distribution of observables. Notably, they require no preprocessing steps other than the sampling algorithms.

DOI: [10.1103/4pzk-lznf](https://doi.org/10.1103/4pzk-lznf)

I. INTRODUCTION

Advances in the preparation of many-body entangled quantum states have marked significant progress in the field of quantum information [1–4]. A critical step in these experiments is verifying that the system's state aligns with the desired one [5–8]. One application of this verification step is in the verifier-prover paradigm of cloud quantum computing [9,10]. Here, an untrusted server (prover) prepares an entangled many-body state, and the client (verifier) must certify that the returned state matches the target. This is accomplished by the prover supplying multiple copies of the state, which the verifier then measures to validate the preparation. Traditional methods for assessing quantum state fidelity include quantum state tomography (QST), which reconstructs the full density matrix of a quantum state, offering detailed information [11–19]. However, QST is resource intensive and scales poorly with system size, posing significant challenges for large-scale quantum systems [20]. Recent approaches have introduced various techniques aimed at simplifying the processes of verification [10,21–23] and fidelity estimation [24–26]. Among these, direct fidelity estimation (DFE) has gained attention for its potential to provide accurate fidelity measurements with fewer resources compared to full state tomography. Unlike full tomography, DFE focuses on directly estimating the fidelity without requiring a complete reconstruction of the quantum state, thus significantly reducing the measurement overhead. For some classes of states, efficient

methods using Pauli measurement [24,25,27,28], some relying on *importance sampling* [29,30], are known.

Meanwhile, the *classical shadow* estimation technique is known for its pioneering approach to efficiently predicting numerous properties of unknown quantum states [31], and recent studies have begun to explore its variants [32,33]. Specifically, it can estimate any quantity related to a quantum state ρ , expressed as $\text{tr}(\rho O)$ for some linear operator O . The fidelity between ρ and a particular pure state is in this form, which implies that classical shadows can be used for DFE.

A classical snapshot of ρ is generated by performing a random unitary operation U on ρ followed by a computational basis measurement. The efficiency of classical shadow tomography is determined by the distribution from which U is chosen. Theoretically, random Clifford measurements [34–37] are preferable in the sense that a constant number of measurements suffices to estimate the fidelity with *any* pure target state. Detailed analyses of the performance of this measurement have been carried out in the context of DFE. Unfortunately, performing random Clifford measurements in a laboratory setting is currently impractical [38]. More feasible alternatives are random Pauli measurements [37]. They are much more resource efficient and need precise control over only single-qubit operations. The significant drawback of random Pauli measurements is that, in the worst case, exponentially many samples with respect to the locality of an observable are necessary for precise estimation [31]. However, for observables associated with the Greenberger-Horne-Zeilinger (GHZ) state [39] and the W /Dicke state [40–42], which have maximum locality, the number of measurements required is only polynomial in the number of qubits. These entangled states play a central role as computational resources in various quantum protocols [43–47].

In this work, we show that for some classes of target states, random Pauli measurements can be efficiently synthesized with classical shadows for DFE. Our protocols are designed to improve previous methods (based on characteristic functions)

*Contact author: aiden132435@cml.snu.ac.kr

†Contact author: junglee@snu.ac.kr

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

in the nonadaptive local Pauli measurement setting [24]. The central aspect is that we employ a target-specific distribution of Pauli measurements rather than the default uniform distribution. This is analogous to sampling measurements from the characteristic function $\chi_\rho(k)$, which depends on the target ρ . Accordingly, the shadow inversion step is adapted to align with the modified distribution of measurements. As a result, the bound on the number of measurements necessary to achieve a specified precision is tightened by a factor between 14.22 and 28.44, and between 16 and 32 for the GHZ state and the W /Dicke state, respectively. The proposed estimator is more *direct* than Ref. [24] (as will be explained in the discussion), resulting in more accurate estimations with the same number of measurements. Moreover, it does not require any preprocessing steps or memory to store the distribution. The proposed method is analyzed in a more general setting in Sec. VI, and it can also be extended to other scenarios as in [48,49].

II. DFE WITH IMPORTANCE SAMPLING

The fidelity between an unknown state and a target state can be estimated from a few Pauli measurements using importance sampling techniques [24]. It is carried out in two stages: first by sampling a Pauli observable, followed by estimating its corresponding characteristic function. The fidelity is estimated to an additive error of 2ϵ with probability $\geq 1 - 2\delta$ using

$$N \leq l + 1 + \frac{2 \ln(2/\delta)}{\epsilon^2} \alpha$$

measurements, where we may choose

$$(l, \alpha) = \left(\frac{2 \ln(2/\delta)}{\epsilon^2}, 1 \right) \quad \text{and} \quad (l, \alpha) = \left(\frac{1}{\epsilon^2 \delta}, n^2 \right) \quad (1)$$

for the GHZ state

$$|\text{GHZ}\rangle = \frac{|0\rangle^{\otimes n} + |1\rangle^{\otimes n}}{\sqrt{2}}$$

and the W state

$$|W\rangle = \frac{|100 \cdots 0\rangle + |010 \cdots 0\rangle + \cdots + |00 \cdots 01\rangle}{\sqrt{n}},$$

respectively. Replacing (ϵ, δ) with $(\epsilon/2, \delta/2)$, we can estimate the fidelity to an additive error of ϵ with probability $\geq 1 - \delta$ using less than

$$\frac{16 \ln(4/\delta)}{\epsilon^2} \quad \text{and} \quad \frac{8 \ln(4/\delta)}{\epsilon^2} n^2 \quad (2)$$

measurements for the GHZ state and the W state, respectively. The choices for l in (1) and the approximations in (2) are given because asymptotics with respect to (ϵ, δ) are considered for the GHZ state, whereas for the W state, asymptotics with respect to n are considered while keeping (ϵ, δ) constant.

III. CLASSICAL SHADOWS

Many properties of a quantum state can be predicted without its full characterization. Such processes are referred to as *shadow tomography*. Throughout this work, ρ denotes an unknown quantum state in $d = 2^n$ dimensions, associated

with an n -qubit system. Let $\mathcal{U}$ be a distribution over unitary operators in a d -dimensional Hilbert space. Repeating the following steps generates random classical snapshots of ρ :

- (1) Sample $U \sim \mathcal{U}$.
- (2) Apply U to ρ ($\rho \mapsto \rho' = U\rho U^\dagger$).
- (3) Measure ρ' in the computational basis.
- (4) Upon receiving the measurement outcome $\hat{b} \in \{0, 1\}^n$, store an efficient description of $U^\dagger |\hat{b}\rangle \langle \hat{b}| U$ in classical memory.

This snapshot, averaged over U and $\hat{b}$, can be viewed as a quantum channel [37,50,51]

$$\mathcal{E}(\rho) \equiv \mathbb{E}[U^\dagger |\hat{b}\rangle \langle \hat{b}| U],$$

where $\mathcal{E}$ depends on $\mathcal{U}$. If $\mathcal{U}$ defines a tomographically complete set of measurements, then $\mathcal{E}$ has an inverse and the snapshot of ρ is defined as

$$\hat{\rho} = \mathcal{E}^{-1}[U^\dagger |\hat{b}\rangle \langle \hat{b}| U],$$

where $\mathbb{E}[\hat{\rho}] = \rho$.

In many cases, a useful property of ρ is given in the form of the expectation value of an observable:

$$\langle O \rangle = \text{tr}(\rho O). \quad (3)$$

A snapshot of this quantity is $\text{tr}(\hat{\rho} O)$ and its average equals $\langle O \rangle$. One important application of this technique is in DFE. We restrict our attention to estimating the fidelity with a *pure* target state

$$\sigma = |\psi\rangle \langle \psi|.$$

The most straightforward way to estimate the fidelity

$$F(\rho, \sigma) = \langle \psi | \rho | \psi \rangle = \text{tr}(\rho \sigma)$$

is to substitute O with σ for the snapshot, i.e., calculate the average of $\text{tr}(\hat{\rho} \sigma)$ as shown in Fig. 1(a). However, this approach can be highly inefficient, and our objective is to derive target-dependent modifications of this process as in Fig. 1(b).

IV. RANDOM MEASUREMENTS

Although classical shadows can be defined for any distribution $\mathcal{U}$, the following two distributions are mathematically convenient and widely discussed in literature.

A. Random n -qubit Clifford circuits

The n -qubit Clifford group $\text{Cl}(2^n)$ is a group of n -qubit unitary operators generated by the gates $\{H, S, \text{cnot}\}$. A circuit U is chosen uniformly from this group and the associated snapshot is

$$\hat{\rho} = (2^n + 1)U^\dagger |\hat{b}\rangle \langle \hat{b}| U - \mathbb{I},$$

where $\mathbb{I}$ denotes the 2^n -dimensional identity operator. A classical shadow of size independent n is sufficient to accurately estimate the fidelity with any pure target state [31, Theorem 1].

B. Tensor products of random single-qubit Clifford circuits

In laboratory settings, random circuits in $\text{Cl}(2^n)$ are challenging to implement in the near future due to the presence of entangling gates such as cnot. Practical circuits are those that

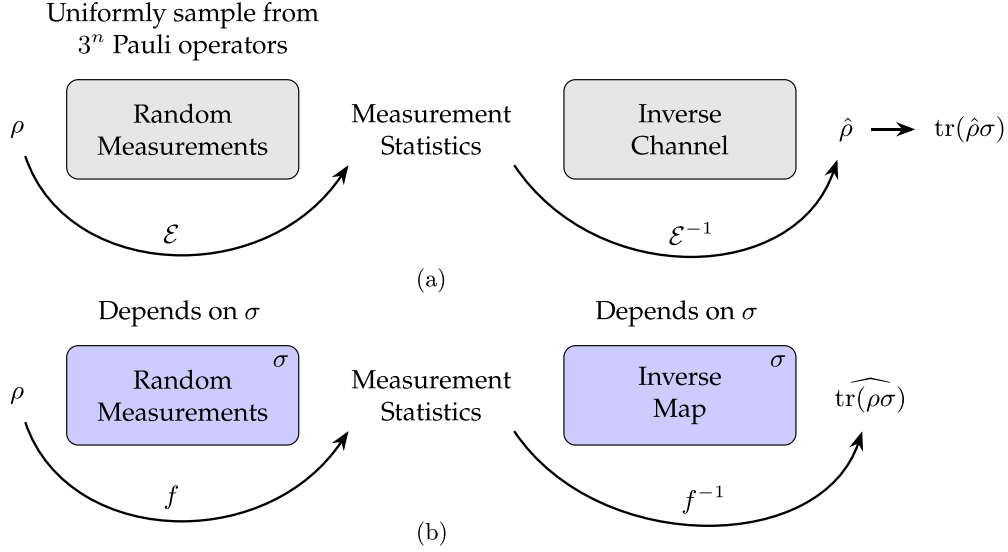


FIG. 1. (a) Vanilla implementation of DFE with classical shadows. (b) Sample-efficient implementation of DFE derived from classical shadows.

consist only of local (single-qubit) unitaries. Throughout the rest of this paper, we focus exclusively on these circuits. For each qubit i , a single-qubit unitary U_i is chosen uniformly and independently from $\text{Cl}(2)$. Instead, we can simply measure each qubit in a random Pauli basis chosen uniformly from

$$S = \{X, Y, Z\},$$

which is equivalent to choosing U_i uniformly from

$$T = \{I, H, HS^\dagger\}.$$

In either case, the n -qubit unitary is $U = \bigotimes_{i=1}^n U_i$ and the associated snapshot is

$$\hat{\rho} = \bigotimes_{i=1}^n (3U_i^\dagger |\hat{b}_i\rangle\langle \hat{b}_i| U_i - \mathbb{I}). \quad (4)$$

The worst-case upper bound on the number of measurements to estimate (3) scales exponentially with the locality of O [31, Proposition 3] or the stabilizer Rényi entropy [28, Theorem 1]. However, many efficient methods exist for DFE with some well-conditioned states (as defined in Ref. [24]), despite the locality of O being n .

V. FIDELITY ESTIMATION FROM PAULI MEASUREMENTS

The remainder of this work will be devoted to describing efficient methods to estimate the fidelity $F(\rho, \sigma)$ using random Pauli measurements and classical shadows, where σ is either the GHZ state or the W state. For these states, classical shadows allow us to exclude unnecessary circuits by exploiting underlying symmetries.

A. Fidelity with computational basis states

We begin with the trivial case of estimating the fidelity between ρ and a computational basis state, interpreted in the context of classical shadows. Suppose we want to use classical shadows to estimate $F(\rho, |b\rangle\langle b|)$, where $b \in \{0, 1\}^n$.

A naive approach is to sample a Pauli string uniformly from S^n for each measurement. The associated circuit $\bigotimes_{i=1}^n U_i$ is an element of $T^{\otimes n}$. From (4), the estimator of $F(\rho, |b\rangle\langle b|)$ is calculated as

$$\begin{aligned} \langle b|\hat{\rho}|b\rangle &= \langle b|\bigotimes_{i=1}^n (3U_i^\dagger |\hat{b}_i\rangle\langle \hat{b}_i| U_i - \mathbb{I})|b\rangle \\ &= \prod_{i=1}^n (3\langle b_i|U_i^\dagger |\hat{b}_i\rangle\langle \hat{b}_i| U_i |b_i\rangle - \langle b_i|b_i\rangle) \\ &= \prod_{i=1}^n (3|\langle \hat{b}_i|U_i|b_i\rangle|^2 - 1). \end{aligned} \quad (5)$$

More generally,

$$\langle b_1|\hat{\rho}|b_2\rangle = \prod_{i=1}^n (3\langle b_{1,i}|U_i^\dagger |\hat{b}_i\rangle\langle \hat{b}_i| U_i |b_{2,i}\rangle - \delta_{b_{1,i}, b_{2,i}}). \quad (6)$$

Note that $|\langle \hat{b}_i|U_i|b_i\rangle| = 1/\sqrt{2}$ whenever $U_i \in T \setminus \{\mathbb{I}\}$, which can be verified with simple calculations. That is, if we measure the i th qubit in the X or Y basis, then $3|\langle \hat{b}_i|U_i|b_i\rangle|^2 - 1$ is always $1/2$ (regardless of $\hat{b}_i$). This allows us to obtain a sample for each of the 3^n Pauli measurements just from a single $ZZ \cdots Z$ measurement outcome $|\hat{b}\rangle$, because we can replace some of the Z measurements with X or Y measurements for which the multiplier is $1/2$ with certainty. Let $m = |\{i|\hat{b}_i = b_i\}|$. Substituting U_i with $\mathbb{I}$ for all i in (5) gives

$$\langle b|\hat{\rho}|b\rangle = 2^m (-1)^{n-m}.$$

Suppose we replace a out of the measurements for $\{i|\hat{b}_i = b_i\}$ and b out of the measurements for $\{i|\hat{b}_i = \bar{b}_i\}$ with X or Y . Using the same $|\hat{b}\rangle$ gives

$$\langle b|\hat{\rho}|b\rangle = 2^{m-a} (-1)^{n-m-b} (1/2)^{a+b}.$$

TABLE I. Values of local traces for each measurement basis and measurement outcome.

Measurement basis	X	Y	Z
U_i	H	$HS^\dagger$	$\mathbb{I}$
$\langle 0 U_i^\dagger 0\rangle\langle 0 U_i 1\rangle$	$1/2$	$-i/2$	0
$\langle 0 U_i^\dagger 1\rangle\langle 1 U_i 1\rangle$	$-1/2$	$i/2$	0
$\langle 1 U_i^\dagger 0\rangle\langle 0 U_i 0\rangle$	$1/2$	$i/2$	0
$\langle 1 U_i^\dagger 1\rangle\langle 1 U_i 0\rangle$	$-1/2$	$-i/2$	0

Since we have 2^{a+b} choices for replacement, the average of all 3^n estimator values evaluates to

$$\begin{aligned} \text{avg} &= \frac{1}{3^n} \sum_{a=0}^m \sum_{b=0}^{n-m} \frac{\binom{m}{a} \binom{n-m}{b} 2^{a+b} 2^{m-a} (-1)^{n-m-b}}{2^{a+b}} \\ &= \frac{1}{3^n} \sum_{a=0}^m \binom{m}{a} 2^{m-a} \sum_{b=0}^{n-m} \binom{n-m}{b} (-1)^{n-m-b} \\ &= \frac{1}{3^n} (1+2)^m (1-1)^{n-m} \quad (\text{define } 0^0 = 1) \\ &= \delta_{b,b}. \end{aligned}$$

As expected, estimating the fidelity with the computational basis state $|b\rangle$ just amounts to repeating a $ZZ \cdots Z$ measurement and counting the number of the measurement outcomes equal to $|b\rangle$. This interpretation will support our further analysis to derive efficient methods to estimate the fidelity with the GHZ state or the W state using Pauli measurements and classical shadows.

Throughout the derivation of our methods for the GHZ state and the W state, we will frequently refer to Table I.

B. Fidelity with GHZ state

The density matrix of the GHZ state is

$$\sigma = \frac{1}{2}(|0\rangle\langle 0|^{\otimes n} + |1\rangle\langle 1|^{\otimes n} + |0\rangle\langle 1|^{\otimes n} + |1\rangle\langle 0|^{\otimes n}).$$

We estimate the diagonal part

$$d = \text{tr}(\rho \frac{1}{2}(|0\rangle\langle 0|^{\otimes n} + |1\rangle\langle 1|^{\otimes n}))$$

and the off-diagonal part

$$o = \text{tr}(\rho \frac{1}{2}(|0\rangle\langle 1|^{\otimes n} + |1\rangle\langle 0|^{\otimes n}))$$

separately, where $F(\rho, \sigma) = d + o$.

To estimate d , we repeat a $ZZ \cdots Z$ measurement with the estimator defined as

$$\hat{d} = \begin{cases} 1/2 & \text{if } \delta_{b,0} + \delta_{b,1} = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Also, we can estimate $\text{tr}(\rho|0\rangle\langle 1|^{\otimes n})$ by choosing a random Pauli string from $\{X, Y, Z\}^n$ and applying (3):

$$\begin{aligned} \text{tr}(\rho|0\rangle\langle 1|^{\otimes n}) &= \mathbb{E}_{(U_i), \delta}[\langle \mathbf{1} | \hat{\rho} | \mathbf{0} \rangle] \\ &= \sum_{(U_i) \in T^n} \frac{1}{3^n} \mathbb{E}_{\hat{\rho}}[\langle \mathbf{1} | \hat{\rho} | \mathbf{0} \rangle] \end{aligned}$$

$$\begin{aligned} &= \sum_{(U_i) \in T^n} \frac{1}{3^n} \mathbb{E}_{\hat{\rho}} \left[3^n \prod_{i=1}^n \langle 1|U_i^\dagger|\hat{b}_i\rangle\langle \hat{b}_i|U_i|0\rangle \right] \\ &= \sum_{(U_i) \in T^n} \mathbb{E}_{\hat{\rho}} \left[\prod_{i=1}^n \langle 1|U_i^\dagger|\hat{b}_i\rangle\langle \hat{b}_i|U_i|0\rangle \right]. \end{aligned} \quad (7)$$

However, for any i , if we measure the i th qubit in the Z basis, i.e., $U_i = \mathbb{I}$, then Table I shows that the whole product in (7) evaluates to zero. Therefore, it suffices to measure each qubit in the X or Y basis, and we replace (7) with

$$\sum_{(U_i) \in (T \setminus \{\mathbb{I}\})^n} \mathbb{E}_{\hat{\rho}} \left[\prod_{i=1}^n \langle 1|U_i^\dagger|\hat{b}_i\rangle\langle \hat{b}_i|U_i|0\rangle \right]. \quad (8)$$

From Table I, the term inside the square brackets in (8) is either real if

$$|\{i|U_i = HS^\dagger\}| \equiv 0 \pmod{2} \quad (9)$$

or pure imaginary otherwise. In either case, it has an absolute value of $1/2^n$. Meanwhile, $\text{tr}(\rho|1\rangle\langle 0|^{\otimes n})$ is the complex conjugate of (8): the sample

$$V = \prod_{i=1}^n \langle 1|U_i^\dagger|\hat{b}_i\rangle\langle \hat{b}_i|U_i|0\rangle + \prod_{i=1}^n \langle 0|U_i^\dagger|\hat{b}_i\rangle\langle \hat{b}_i|U_i|1\rangle \quad (10)$$

evaluates to $\pm 1/2^{n-1}$ if (9) holds and zero otherwise. This symmetry allows us to further reduce the sample space of (U_i) as

$$o = \sum_{\substack{(U_i) \in (T \setminus \{\mathbb{I}\})^n, \\ |\{i|U_i = HS^\dagger\}| \equiv 0 \pmod{2}}} \mathbb{E}_{\hat{\rho}} \left[\frac{V}{2} \right].$$

Let T_{GHZ} and $\mathcal{U}_{\text{GHZ}}$ denote this space and the uniform distribution over this space, respectively, where $|T_{\text{GHZ}}| = 2^{n-1}$. It is not hard to show that

$$\frac{V}{2} = \frac{(-1)^{|\{i|U_i = HS^\dagger\}|/2 + |\{i|\hat{b}_i = 1\}|}}{2^n}. \quad (11)$$

Considering that each (U_i) is chosen with probability $1/2^{n-1}$, we may write

$$o = \sum_{(U_i) \in T_{\text{GHZ}}} \mathbb{E}_{\hat{\rho}} \left[\frac{1}{2^{n-1}} \hat{o} \right] = \mathbb{E}_{(U_i) \sim \mathcal{U}_{\text{GHZ}}, \hat{\rho}} \hat{o},$$

where $\hat{o} \equiv 2^{n-2}V$. Therefore, we have $0 \leq \hat{d} \leq 1/2$ and $-1/2 \leq \hat{o} \leq 1/2$.

Finally, $\hat{F}$ is defined as a sample generated by sampling $3(\hat{d} - 1/4)$ with probability $1/3$ and $3\hat{o}/2$ with probability $2/3$. Then $|\hat{F}| \leq 3/4$ and

$$\begin{aligned} \mathbb{E}[\hat{F}] + \frac{1}{4} &= \frac{1}{3} \mathbb{E}[3(\hat{d} - 1/4)] + \frac{2}{3} \mathbb{E}[3\hat{o}/2] + \frac{1}{4} \\ &= d + o \\ &= F(\rho, \sigma). \end{aligned}$$

We generate

$$N = \left\lceil \frac{9 \ln(2/\delta)}{8\epsilon^2} \right\rceil \quad (12)$$

ALGORITHM 1. DFE for n -qubit GHZ state.

Input: Unknown state ρ , error parameters (ϵ, δ)
Output: Estimate of the fidelity $F(\rho, \sigma)$, where σ is the n -qubit GHZ state

```

 $N \leftarrow \lceil 2 \ln(2/\delta)/\epsilon^2 \rceil$ 
 $\text{sum} \leftarrow 0$ 
for  $i$  in range( $N$ ):
    if  $1/3 > X \sim \mathcal{U}_{(0,1)}$ :
         $\hat{b} = \text{measureComputationalBasis}(\rho)$ 
         $\text{sum} += 3(\delta_{\hat{b},0} + \delta_{\hat{b},1})/2 - 3/4$ 
    else
        Sample  $U_i \in \{H, HS^\dagger\}$  uniformly and independently for  $1 \leq i \leq n-1$ .
        Choose  $U_n \in \{H, HS^\dagger\}$  such that  $|\{i|U_i = HS^\dagger\}| \bmod 2 = 0$ .
         $U \leftarrow \bigotimes_{i=1}^n U_i$ 
         $\rho' \leftarrow U\rho U^\dagger$ 
         $\hat{b} = \text{measureComputationalBasis}(\rho')$ 
         $\text{sum} += 3 \cdot (-1)^{|\{i|U_i = HS^\dagger\}|/2 + |\{i|\hat{b}_i = 1\}|} / 4$ 
    end if
end for
return  $\tilde{F} = \text{sum}/N + 1/4$ 

```

samples of $\tilde{F}$. From Hoeffding's inequality [52,53],

$$P\left(\left|\tilde{F} + \frac{1}{4} - F\right| \geq \epsilon\right) \leq \delta.$$

The total number of measurements is about $9 \ln(2/\delta)/8\epsilon^2$, which tightens the bound (2) by a factor of $128(1 + \ln_{2/\delta} 2)/9$, a constant between 128/9 and 256/9 (closer to 128/9 for small δ). Algorithm 1 provides a complete pseudocode for the GHZ state. It should be noted that the measurement settings for fidelity estimation in Algorithm 1 also appear in some verification protocols [54,55], but in our context these settings are further associated with a fidelity estimator defined by the specific measurement basis and outcome.

C. Fidelity with W state

The density matrix of the W state is

$$\sigma = \frac{1}{n} \left(\sum_{i=1}^n |e_i\rangle\langle e_i| + \sum_{i \neq j} |e_i\rangle\langle e_j| \right),$$

where

$$|e_i\rangle \equiv \underbrace{|00 \dots 0 1}_{\times(i-1)} \underbrace{|00 \dots 0\rangle}_{\times(n-i)}.$$

We estimate the diagonal part

$$d = \text{tr} \left(\rho \frac{1}{n} \sum_{i=1}^n |e_i\rangle\langle e_i| \right)$$

and the off-diagonal part

$$o = \text{tr} \left(\rho \frac{1}{n} \sum_{i \neq j} |e_i\rangle\langle e_j| \right)$$

separately, where $F(\rho, \sigma) = d + o$.

To estimate d , we repeat a $ZZ \dots Z$ measurement with the estimator defined as

$$\hat{d} = \begin{cases} 1/n & \text{if } \sum_{i=1}^n \delta_{\hat{b}, e_i} = 1 \\ 0 & \text{otherwise.} \end{cases}$$

The off-diagonal estimator $\hat{o}$ can be found in Appendix A. We have $0 \leq \hat{d} \leq 1/n$ and $-(n-1)/2 \leq \hat{o} \leq (n-1)/2$.

Finally, $\hat{F}$ is defined as a sample generated by sampling $(n^2 - n + 1)(\hat{d} - 1/2n)$ with probability $1/(n^2 - n + 1)$ and $(n^2 - n + 1)\hat{o}/n(n-1)$ with probability $n(n-1)/(n^2 - n + 1)$. Then $|\hat{F}| \leq (n^2 - n + 1)/2n$ and

$$\begin{aligned} \mathbb{E}[\hat{F}] + \frac{1}{2n} &= \frac{\mathbb{E}[(n^2 - n + 1)(\hat{d} - 1/2n)]}{n^2 - n + 1} \\ &\quad + \frac{n(n-1)\mathbb{E}[(n^2 - n + 1)\hat{o}/n(n-1)]}{n^2 - n + 1} + \frac{1}{2n} \\ &= d + o \\ &= F(\rho, \sigma). \end{aligned}$$

We generate

$$N = \left\lceil \frac{\ln(2/\delta)(n^2 - n + 1)^2}{2\epsilon^2 n^2} \right\rceil \quad (13)$$

samples of $\hat{F}$. From Hoeffding's inequality,

$$P\left(\left|\tilde{F} + \frac{1}{2n} - F\right| \geq \epsilon\right) \leq \delta.$$

The total number of measurements is about $\ln(2/\delta)n^2/2\epsilon^2$, which tightens the bound (2) by a factor of $16(1 + \ln_{2/\delta} 2)$, a constant between 16 and 32 (closer to 16 for small δ). Algorithm 2 provides a complete pseudocode for the W state. For relevant definitions, see Appendix A.

ALGORITHM 2. DFE for n -qubit W state.

Input: Unknown state ρ , error parameters (ϵ, δ)
Output: Estimate of the fidelity $F(\rho, \sigma)$, where σ is the n -qubit W state

```

 $N \leftarrow \lceil 2 \ln(2/\delta)(n-1)^2/\epsilon^2 \rceil$ 
 $\text{sum} \leftarrow 0$ 
for in range( $N$ ):
    if  $1/(n^2 - n + 1) > X \sim \mathcal{U}_{(0,1)}$ :
         $\hat{b} = \text{measureComputationalBasis}(\rho)$ 
         $\text{sum} += (n^2 - n + 1)(2 \sum_{i=1}^n \delta_{\hat{b}, e_i} - 1)/2n$ 
    else
        Sample  $\{i, j\} \sim f$  and  $U' \sim g$ .
         $U_i, U_j \leftarrow U'$ 
         $U_k \leftarrow \mathbb{I} \quad \forall k \notin \{i, j\}$ 
         $U \leftarrow \bigotimes_{m=1}^n U_m$ 
         $\rho' \leftarrow U\rho U^\dagger$ 
         $\hat{b} = \text{measureComputationalBasis}(\rho')$ 
         $\text{sum} += (n^2 - n + 1)(2\delta_{\hat{b}_i, \hat{b}_j} - 1)\delta_{\hat{b}_{[n] \setminus \{i, j\}}, 0}/2n$ 
    end if
end for
return  $\tilde{F} = \text{sum}/N + 1/2n$ 

```

D. Fidelity with Dicke state

The density matrix of the Dicke state with k excitations is

$$\sigma = \binom{n}{k}^{-1} \left(\sum_{|\mathbf{i}|=k} |\mathbf{i}\rangle\langle\mathbf{i}| + \sum_{\substack{|\mathbf{i}|=|\mathbf{j}|=k, \\ |\mathbf{i}| \neq |\mathbf{j}|}} |\mathbf{i}\rangle\langle\mathbf{j}| \right),$$

where $|\mathbf{i}|$ denotes the Hamming weight of $\mathbf{i} \in \{0, 1\}^n$. We estimate the diagonal part

$$d = \text{tr} \left(\rho \binom{n}{k}^{-1} \sum_{|\mathbf{i}|=k} |\mathbf{i}\rangle\langle\mathbf{i}| \right)$$

and the off-diagonal part

$$o = \text{tr} \left(\rho \binom{n}{k}^{-1} \sum_{\substack{|\mathbf{i}|=|\mathbf{j}|=k, \\ |\mathbf{i}| \neq |\mathbf{j}|}} |\mathbf{i}\rangle\langle\mathbf{j}| \right)$$

separately, where $F(\rho, \sigma) = d + o$.

To estimate d , we repeat a $ZZ \cdots Z$ measurement with the estimator defined as

$$\hat{d} = \begin{cases} \binom{n}{k}^{-1} & \text{if } \sum_{|\mathbf{i}|=k} \delta_{\hat{b}, \mathbf{i}} = 1 \\ 0 & \text{otherwise.} \end{cases}$$

We have $0 \leq \hat{d} \leq \binom{n}{k}^{-1}$ and $-\binom{n}{k}^{-1} c_l \leq \hat{o}_l \leq \binom{n}{k}^{-1} c_l$, where the definition of c_l and the off-diagonal estimator $\hat{o}_l$ can be found in Appendix B. Moreover, let

$$S \equiv \frac{1}{2} + \sum_{l=\max(0, 2k-n)}^{k-1} c_l.$$

Finally, $\hat{F}$ is defined as a sample generated by sampling $2S[\hat{d} - \frac{1}{2}\binom{n}{k}^{-1}]$ with probability $1/2S$ and $S\hat{o}_l/c_l$ with probability c_l/S . Then $|\hat{F}| \leq \binom{n}{k}^{-1} S$ and

$$\begin{aligned} \mathbb{E}[\hat{F}] + \frac{1}{2} \binom{n}{k}^{-1} &= \frac{2S\mathbb{E}[\hat{d} - \frac{1}{2}\binom{n}{k}^{-1}]}{2S} \\ &+ \sum_{l=\max(0, 2k-n)}^{k-1} \frac{c_l \mathbb{E}[S\hat{o}_l/c_l]}{S} + \frac{1}{2} \binom{n}{k}^{-1} \\ &= d + \sum_{l=\max(0, 2k-n)}^{k-1} o_l \\ &= d + o \\ &= F(\rho, \sigma). \end{aligned}$$

We generate

$$N = \left\lceil \frac{2 \ln(2/\delta) S^2}{\epsilon^2 \binom{n}{k}^2} \right\rceil$$

samples of $\hat{F}$. From Hoeffding's inequality,

$$P\left(\left|\bar{F} + \frac{1}{2} \binom{n}{k}^{-1} - F\right| \geq \epsilon\right) \leq \delta.$$

The total number of measurements is $O(n^{2k})$, where the asymptotics are with respect to n , while k , ϵ , and δ remain constant. As a detailed comparison, it follows from [24] that estimating the fidelity with an n -qubit Dicke state σ with k excitations to additive error ϵ and failure probability $\leq \delta$ needs roughly

$$\frac{8 \ln(4/\delta)}{\epsilon^2} \binom{n}{k}^2 \approx \frac{8 \ln(4/\delta)}{\epsilon^2} \frac{n^{2k}}{(k!)^2} \quad (14)$$

Pauli measurements. This follows from the fact that for every Pauli string $W \in \{I, X, Y, Z\}^{\otimes n}$,

$$|\text{tr}(\sigma W)| = 0 \quad \text{or} \quad |\text{tr}(\sigma W)| \geq \binom{n}{k}^{-1}.$$

Our protocol leverages Dicke state symmetries to reduce the sample space. For large n (with fixed k) we have

$$S \approx c_0 = \binom{n}{2k} \binom{2k}{0} \binom{2k-1}{k-1} \approx \frac{n^{2k}}{2(k!)^2}.$$

Consequently, the required number of samples becomes

$$\begin{aligned} \frac{2 \ln(2/\delta)}{\epsilon^2} \frac{S^2}{\binom{n}{k}^2} &\approx \frac{2 \ln(2/\delta)}{\epsilon^2} \frac{n^{4k}}{4(k!)^4} \frac{(k!)^2}{n^{2k}} \\ &= \frac{\ln(2/\delta)}{2\epsilon^2} \frac{n^{2k}}{(k!)^2}. \end{aligned} \quad (15)$$

Equation (15) improves upon (14) by the factor $16(1 + \ln_{2/\delta} 2)$, which recovers the W state gain when $k = 1$. Algorithm 3 provides a complete pseudocode for the Dicke state with k excitations. For relevant definitions, see Appendix B.

ALGORITHM 3. DFE for n -qubit Dicke state with k excitations.

Input: Unknown state ρ , error parameters (ϵ, δ)
Output: Estimate of the fidelity $F(\rho, \sigma)$, where σ is the n -qubit Dicke state with k excitations
Define c_l as in (B3).
 $N \leftarrow \lceil 2 \ln(2/\delta) S^2 / \epsilon^2 \binom{n}{k}^2 \rceil$, $S \leftarrow \frac{1}{2} + \sum_{l=\max(0, 2k-n)}^{k-1} c_l$
 $\text{sum} \leftarrow 0$
for $\text{in range}(N)$:
 Sample l with probability c_l/S for $\max(0, 2k-n) \leq l \leq k-1$
 and $1/2S$ for $l = k$
 if $l = k$:
 $\hat{b} = \text{measureComputationalBasis}(\rho)$
 $\text{sum} += 2S \binom{n}{k}^{-1} (\sum_{|\mathbf{i}|=k} \delta_{\hat{b}, \mathbf{i}} - \frac{1}{2})$
 else
 Sample $\{\mathbf{i}, \mathbf{j}\} \sim f_l$ and $U'_{m \in s(\mathbf{i}) \Delta s(\mathbf{j}) - \max[s(\mathbf{i}) \Delta s(\mathbf{j})]} \sim g$ independently.
 $U_{m \in s(\mathbf{i}) \Delta s(\mathbf{j}) - \max[s(\mathbf{i}) \Delta s(\mathbf{j})]} \leftarrow U'_m$, $U_{\max[s(\mathbf{i}) \Delta s(\mathbf{j})]} \leftarrow \mathcal{F}(\{U'_m\})$
 $U_{m \notin s(\mathbf{i}) \Delta s(\mathbf{j})} \leftarrow \mathbb{I}$
 $U \leftarrow \bigotimes_{m=1}^n U_m$
 $\rho' \leftarrow U \rho U^\dagger$
 $\hat{b} = \text{measureComputationalBasis}(\rho')$
 $x \leftarrow |\{m \in s(\mathbf{i}) \Delta s(\mathbf{j}) | U_m = HS^\dagger\}|/2 + |\{m \in s(\mathbf{i}) - s(\mathbf{j}) | (U_m, \hat{b}_m) \in (H, 1), (HS^\dagger, 0)\}|$
 $+ |\{m \in s(\mathbf{j}) - s(\mathbf{i}) | \hat{b}_m = 1\}|$
 $\text{sum} += (-1)^x S \delta_{\hat{b}_{m \in s(\mathbf{i}) \cap s(\mathbf{j})}, 1} \delta_{\hat{b}_{m \notin s(\mathbf{i}) \cup s(\mathbf{j})}, 0} / \binom{n}{k}$
 end if
end for
return $\bar{F} = \text{sum}/N + 1/2 \binom{n}{k}$

VI. A GENERIC OPTIMIZATION RULE

The protocols developed for the GHZ, W , and Dicke states can be subsumed under a more general optimization framework. Consider a real-valued observable

$$O = \sum_{b_1, b_2 \in \{0,1\}^n} c_{b_1, b_2} |b_1\rangle\langle b_2|, \quad c_{b_1, b_2} \in \mathbb{R}.$$

For a fixed pair (b_1, b_2) define

$$\begin{aligned} \text{eq}(b_1, b_2) &= \{i | b_{1,i} = b_{2,i}\}, \\ \text{neq}(b_1, b_2) &= \{i | b_{1,i} = 1 - b_{2,i}\}. \end{aligned}$$

To estimate the symmetric combination $\langle b_1 | \rho | b_2 \rangle + \langle b_2 | \rho | b_1 \rangle$, it suffices to perform a Pauli measurement that meets the following local basis rules:

- (1) qubits in $\text{eq}(b_1, b_2)$ are measured in the Z basis.
- (3) qubits in $\text{neq}(b_1, b_2)$ are measured in the X or Y basis, with even number of Y basis across $\text{neq}(b_1, b_2)$.

Let $\mathcal{P}^{(b_1, b_2)}$ be the set of Pauli strings obeying this rule. Because O has real coefficients (so $c_{b_1, b_2} = c_{b_2, b_1}$ when $b_1 \neq b_2$), the expectation value can be reconstructed solely from

$$\bigcup_{\substack{b_1, b_2 \in \{0,1\}^n, \\ c_{b_1, b_2} \neq 0}} \mathcal{P}^{(b_1, b_2)}. \quad (16)$$

After restricting the sampling domain to (16) we reweight and renormalize the estimators and the sampling probabilities [which is itself a nontrivial process that depends on both the set (16) and the coefficients c_{b_1, b_2}], replacing the original shadow estimation using all 3^n Pauli operators. An immediate corollary is that Pauli strings containing an odd number of Y are never required when $O \in \mathbb{R}^{2^n \times 2^n}$.

Suppose the number of nonzero off-diagonal, right triangular elements of O is r . Then, according to the measurement rule described above, we obtain a very rough upper bound on the number of measurement settings required to estimate $\text{tr}(\rho O)$, which is $1 + rd/2$. If $r = \text{poly}(n)$, this already yields an exponential reduction compared to the total number of local Pauli measurement settings, which is $3^n = (\frac{3}{2})^n d$.

Whenever the target state factorizes as a tensor product of independent subsystems—such as a collection of Bell pairs used as essential resources in distributed quantum computing [56–58]—one incurs no loss of optimality by simply performing the optimal measurement protocol on each subsystem in parallel. Moreover, to estimate every subsystem fidelity to within an additive error ϵ with confidence at least $1 - \delta$, the total number of measurements need only grow logarithmically with the number of subsystems.

In quantum machine learning, one must estimate the quantity

$$\text{tr}(|\psi(\mathbf{x}; \boldsymbol{\theta})\rangle\langle\psi(\mathbf{x}; \boldsymbol{\theta})|O),$$

where

$$|\psi(\mathbf{x}; \boldsymbol{\theta})\rangle = \prod_{l=1}^L V^{(l)}(\boldsymbol{\theta}) U^{(l)}(\mathbf{x}) |0\rangle^{\otimes n}$$

for parametrized unitaries $V^{(l)}(\boldsymbol{\theta})$ and data-encoding unitaries $U^{(l)}(\mathbf{x})$. There is growing interest in applying shadow

tomography to observable estimation in variational quantum neural networks [59,60]. As our approach can target a wide range of observables, it naturally extends to such applications.

VII. NUMERICAL SIMULATIONS

Using the previous importance sampling technique, the average number of samples falls strictly below the bound (2) for many configurations. Furthermore, the bound (2) itself may be quite rough, as it involves two layers of estimation: one for importance sampling and another for characteristic function estimation. Thus, it is essential to experimentally demonstrate the improvement of our method over the importance sampling approach.

In practice, the precision of estimates in Algorithms 1 and 2 is not directly controllable, but the number of samples can be specified. Therefore, we compare Ref. [24] with Algorithms 1 and 2 as follows: First, Ref. [24] with parameters (ϵ, δ) is used to estimate the fidelity. Next, Algorithms 1 and 2 are executed using the same number of samples as Ref. [24]. Finally, the mean-squared error (MSE) of the estimates is evaluated as $\mathbb{E}[\{\tilde{F} - F(\rho, \sigma)\}^2]$, where $\tilde{F}$ is the output of our estimation protocol.

The states were randomly generated with fidelities ranging from 0 to 1 in increments of 0.01 (i.e., 0, 0.01, 0.02, ..., 1), as described in Algorithm 4. Each state is generated by first creating a random mixed state and projecting it onto the subspace orthogonal to the target state (GHZ or W). The output state with the desired fidelity is then obtained as a linear combination of the target state and the projected state. Figure 2 shows the comparison of MSE for up to 7 qubits. Algorithms 1 and 2 significantly improve estimation accuracy over Ref. [24], although the gain in this direction is less pronounced than the reduction implied by (2) vs (12) and (13). The simulations were performed using pennylane's classical simulation framework [62].

VIII. DISCUSSION

We have outlined efficient methods to estimate the fidelity between an unknown laboratory state and a pure target state using random Pauli measurement and classical shadows, where the target is either the GHZ state, the W state, or the Dicke state. The upper bounds on the required number of measurements to achieve a certain accuracy are tightened by constant factors. The improvement can be explained as follows.

- (1) Instead of following the two-step procedure of sampling a Pauli operator and then estimating its characteristic

ALGORITHM 4. Generating a random state with a specified fidelity relative to a target state.

Input: Target state σ , fidelity f

Output: Random state ρ such that $F(\rho, \sigma) = f$

$\rho = \text{randomDensityMatrix}()$

$P \leftarrow I - \sigma$

$\rho \leftarrow P \rho P$

$\rho \leftarrow (1 - f)/\text{tr}(\rho)$

return $\rho + f\sigma$

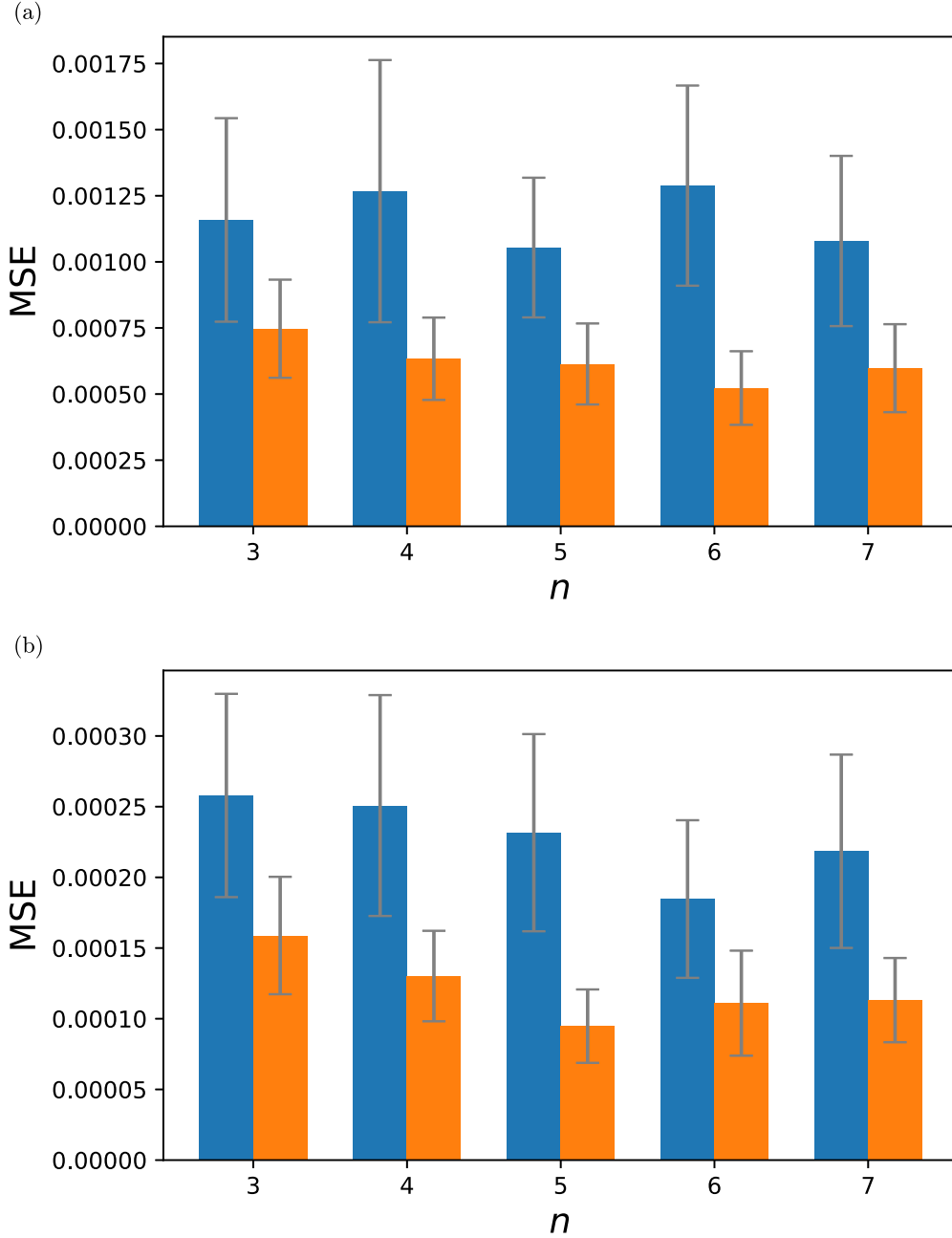


FIG. 2. MSE comparison for (a) GHZ state with $(\epsilon, \delta) = (0.1, 0.1)$, (b) GHZ state with $(\epsilon, \delta) = (0.05, 0.05)$, (c) W state with $(\epsilon, \delta) = (0.1, 0.1)$, and (d) W state with $(\epsilon, \delta) = (0.05, 0.05)$. Blue corresponds to the algorithm in Ref. [24], which is based on Pauli-Liouville coefficients [61], and orange to our approach. Error bars represent 95% confidence intervals.

function, we adopt a somewhat more direct way of estimating the fidelity based on a modified form of shadow tomography.

(2) Local observables are sampled from $\{X, Y, Z\}^{\otimes n}$ rather than $\{\mathbb{I}, X, Y, Z\}^{\otimes n}$ (sampling the local observable $\mathbb{I}$ might be considered redundant since it always has an expectation value of 1). Moreover, deterministic and canceled-out terms are carefully removed to further increase sample efficiency.

Optimizing the estimation of fidelities with states other than those considered in this work based on similar approaches is a promising direction for future research.

Additionally, integrating recent techniques that employ strategies other than nonadaptive local Pauli measurements could yield further improvement.

Practical measurements suffer from noise, and the robust shadow estimation protocol was proposed to specifically mitigate this source of error [61]. Its calibration procedure for uniformly random Pauli measurements requires a number of samples that grows exponentially with the locality of the observable. Since every observable considered in this work is n local, a direct application of that result would incur prohibitive overhead. However, our protocol departs from uniform Pauli

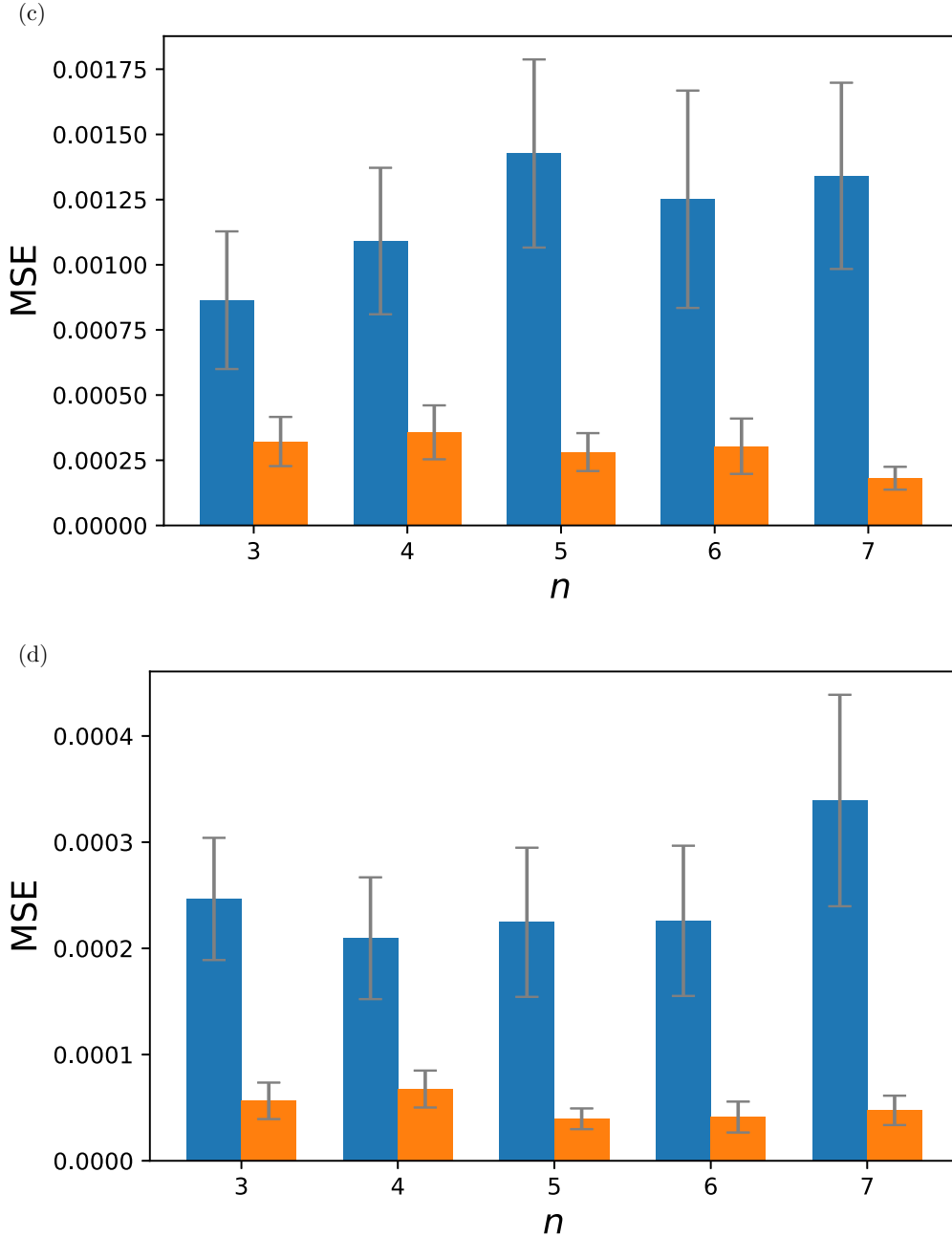


FIG. 2. (Continued).

sampling by tailoring the measurement ensemble to each target observable, and we believe that one can similarly devise bespoke calibration schemes whose sample complexity does not scale exponentially in n . We therefore leave the development of such observable-specific calibration procedures as a key avenue for future work.

ACKNOWLEDGMENTS

This work is in part supported by the National Research Foundation of Korea (NRF; Grants No. RS-2024-00451435 and No. RS-2024-00413957), Institute of Information and Communications Technology Planning and Evaluation (IITP;

Grant No. 2021-0-01059), a grant funded by the Ministry of Science and ICT (MSIT), Institute of New Media and Communications (INMAC), and the Brain Korea 21 FOUR program of the Education and Research Program for Future ICT Pioneers.

H.C. conceptualized the study and conducted the experiments. J.L. supervised the project and provided feedback on the manuscript.

DATA AVAILABILITY

The data that support the findings of this article are openly available [63].

APPENDIX A: OFF-DIAGONAL TERMS FOR W STATE

For the off-diagonal part, consider the term

$$\begin{aligned}
 \text{tr}(\rho|e_i\rangle\langle e_j|) &= \mathbb{E}_{(U_k), \hat{b}}[\langle e_j|\hat{\rho}|e_i\rangle] \\
 &= \sum_{(U_k) \in T^n} \frac{1}{3^n} \mathbb{E}_{\hat{b}}[\langle e_j|\hat{\rho}|e_i\rangle] \\
 &= \sum_{\substack{U_i, U_j \in T \setminus \{\mathbb{I}\}, \\ U_{k \notin \{i, j\}} = \mathbb{I}}} \frac{1}{3^2} \mathbb{E}_{\hat{b}}[3^2 \langle 0|U_i^\dagger|\hat{b}_i\rangle\langle\hat{b}_i|U_i|1\rangle\langle 1|U_j^\dagger|\hat{b}_j\rangle\langle\hat{b}_j|U_j|0\rangle\delta_{\hat{b}_{[n] \setminus \{i, j\}}, \mathbf{0}}] \\
 &= \sum_{\substack{U_i, U_j \in T \setminus \{\mathbb{I}\}, \\ U_{k \notin \{i, j\}} = \mathbb{I}}} \mathbb{E}_{\hat{b}}[\langle 0|U_i^\dagger|\hat{b}_i\rangle\langle\hat{b}_i|U_i|1\rangle\langle 1|U_j^\dagger|\hat{b}_j\rangle\langle\hat{b}_j|U_j|0\rangle\delta_{\hat{b}_{[n] \setminus \{i, j\}}, \mathbf{0}}], \tag{A1}
 \end{aligned}$$

where $[n] \equiv \{1, 2, \dots, n\}$. The third equality follows from previous analyses to reduce the sample space.

From Table I, we see that the term inside the square brackets in (A1) is zero if $\delta_{\hat{b}_{[n] \setminus \{i, j\}}, \mathbf{0}} = 0, \pm 1/4$ if $\delta_{\hat{b}_{[n] \setminus \{i, j\}}, \mathbf{0}} = 1$ and $U_i = U_j$, and $\pm i/4$ if $\delta_{\hat{b}_{[n] \setminus \{i, j\}}, \mathbf{0}} = 1$ and $U_i \neq U_j$. Meanwhile, $\text{tr}(\rho|e_j\rangle\langle e_i|)$ is the complex conjugate of (A1), and the sample

$$V_W = (\langle 0|U_i^\dagger|\hat{b}_i\rangle\langle\hat{b}_i|U_i|1\rangle\langle 1|U_j^\dagger|\hat{b}_j\rangle\langle\hat{b}_j|U_j|0\rangle + \langle 0|U_j^\dagger|\hat{b}_j\rangle\langle\hat{b}_j|U_j|1\rangle\langle 1|U_i^\dagger|\hat{b}_i\rangle\langle\hat{b}_i|U_i|0\rangle)\delta_{\hat{b}_{[n] \setminus \{i, j\}}, \mathbf{0}} \tag{A2}$$

evaluates to $\pm 1/2$ if $\delta_{\hat{b}_{[n] \setminus \{i, j\}}, \mathbf{0}} = 1$ and $U_i = U_j$. Otherwise, it is always zero. This symmetry allows us to further reduce the sample space of (U_i) :

$$o = \sum_{i < j} \sum_{\substack{U_i = U_j \in T \setminus \{\mathbb{I}\}, \\ U_{k \notin \{i, j\}} = \mathbb{I}}} \mathbb{E}_{\hat{b}}\left[\frac{V_W}{n}\right].$$

Let f denote the uniform distribution over all 2-subsets of $[n]$ and g denote the uniform distribution over $T \setminus \{\mathbb{I}\}$. Then

$$\begin{aligned}
 o &= \sum_{i < j} \frac{2}{n(n-1)} \sum_{U' \in T \setminus \{\mathbb{I}\}} \frac{1}{2} \mathbb{E}_{\hat{b}}[\hat{o}|U_i = U_j = U', U_{k \notin \{i, j\}} = \mathbb{I}] \\
 &= \mathbb{E}_{\{i, j\} \sim f, U' \sim g, \hat{b}}[\hat{o}|U_i = U_j = U', U_{k \notin \{i, j\}} = \mathbb{I}],
 \end{aligned}$$

where $\hat{o} \equiv (n-1)V_W$. Therefore, we have $-(n-1)/2 \leq \hat{o} \leq (n-1)/2$.

APPENDIX B: OFF-DIAGONAL TERMS FOR DICKE STATE

For the off-diagonal part, consider $\mathbf{i}, \mathbf{j} \in \{0, 1\}^n$ such that $|\mathbf{i}| = |\mathbf{j}| = k$ and $|\mathbf{s}(\mathbf{i}) \cap \mathbf{s}(\mathbf{j})| = l$, where $\mathbf{s}(\mathbf{i}) \equiv \{m|\mathbf{i}_m = 1\}$ denotes the support of $\mathbf{i}$ and $l \in [\max(0, 2k-n), k-1]$. For example, if

$$\begin{aligned}
 \mathbf{i} &= (1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0), \\
 \mathbf{j} &= (0 \quad 0 \quad 0 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 0),
 \end{aligned}$$

we have $n = 10, k = 5$, and $l = 2$. Then

$$\begin{aligned}
 \text{tr}(\rho|\mathbf{i}\rangle\langle\mathbf{j}|) &= \mathbb{E}_{(U_m), \hat{b}}[\langle\mathbf{j}|\hat{\rho}|\mathbf{i}\rangle] \\
 &= \sum_{(U_m) \in T^n} \frac{1}{3^n} \mathbb{E}_{\hat{b}}[\langle\mathbf{j}|\hat{\rho}|\mathbf{i}\rangle] \\
 &= \sum_{\substack{U_{m \in \mathbf{s}(\mathbf{i}) \Delta \mathbf{s}(\mathbf{j})} \in T \setminus \{\mathbb{I}\}, \\ U_{m \notin \mathbf{s}(\mathbf{i}) \Delta \mathbf{s}(\mathbf{j})} = \mathbb{I}}} \frac{1}{3^{2k-2l}} \mathbb{E}_{\hat{b}} \left[3^{2k-2l} \prod_{m \in \mathbf{s}(\mathbf{i}) - \mathbf{s}(\mathbf{j})} \langle 0|U_m^\dagger|\hat{b}_m\rangle\langle\hat{b}_m|U_m|1\rangle \prod_{m \in \mathbf{s}(\mathbf{j}) - \mathbf{s}(\mathbf{i})} \langle 1|U_m^\dagger|\hat{b}_m\rangle\langle\hat{b}_m|U_m|0\rangle \delta_{\hat{b}_{m \in \mathbf{s}(\mathbf{i}) \cap \mathbf{s}(\mathbf{j})}, \mathbf{1}} \delta_{\hat{b}_{m \notin \mathbf{s}(\mathbf{i}) \cup \mathbf{s}(\mathbf{j})}, \mathbf{0}} \right] \\
 &= \sum_{\substack{U_{m \in \mathbf{s}(\mathbf{i}) \Delta \mathbf{s}(\mathbf{j})} \in T \setminus \{\mathbb{I}\}, \\ U_{m \notin \mathbf{s}(\mathbf{i}) \Delta \mathbf{s}(\mathbf{j})} = \mathbb{I}}} \mathbb{E}_{\hat{b}} \left[\prod_{m \in \mathbf{s}(\mathbf{i}) - \mathbf{s}(\mathbf{j})} \langle 0|U_m^\dagger|\hat{b}_m\rangle\langle\hat{b}_m|U_m|1\rangle \prod_{m \in \mathbf{s}(\mathbf{j}) - \mathbf{s}(\mathbf{i})} \langle 1|U_m^\dagger|\hat{b}_m\rangle\langle\hat{b}_m|U_m|0\rangle \delta_{\hat{b}_{m \in \mathbf{s}(\mathbf{i}) \cap \mathbf{s}(\mathbf{j})}, \mathbf{1}} \delta_{\hat{b}_{m \notin \mathbf{s}(\mathbf{i}) \cup \mathbf{s}(\mathbf{j})}, \mathbf{0}} \right], \tag{B1}
 \end{aligned}$$

where $A \Delta B$ denotes the symmetric difference of the sets A and B . The third equality follows from previous analyses to reduce the sample space.

From Table I, we see that the term inside the square brackets in (B1) is zero if $\delta_{\hat{b}_{m \in s(\mathbf{i}) \cap s(\mathbf{j})}, \mathbf{1}} = 0$ or $\delta_{\hat{b}_{m \in s(\mathbf{i}) \cup s(\mathbf{j})}, \mathbf{0}} = 0, \pm 1/2^{2k-2l}$ if $\delta_{\hat{b}_{m \in s(\mathbf{i}) \cap s(\mathbf{j})}, \mathbf{1}} \delta_{\hat{b}_{m \in s(\mathbf{i}) \cup s(\mathbf{j})}, \mathbf{0}} = 1$ and $|\{m \in s(\mathbf{i}) \Delta s(\mathbf{j}) | U_m = HS^\dagger\}| \equiv 0 \pmod{2}$, and $\pm i/2^{2k-2l}$ if $\delta_{\hat{b}_{m \in s(\mathbf{i}) \cap s(\mathbf{j})}, \mathbf{1}} \delta_{\hat{b}_{m \in s(\mathbf{i}) \cup s(\mathbf{j})}, \mathbf{0}} = 1$ and $|\{m \in s(\mathbf{i}) \Delta s(\mathbf{j}) | U_m = HS^\dagger\}| \equiv 1 \pmod{2}$. Meanwhile, $\text{tr}(\rho | \mathbf{j} \rangle \langle \mathbf{i} |)$ is the complex conjugate of (B1), and the sample

$$V_D = \left(\prod_{m \in s(\mathbf{i}) - s(\mathbf{j})} \langle 0 | U_m^\dagger | \hat{b}_m \rangle \langle \hat{b}_m | U_m | 1 \rangle \prod_{m \in s(\mathbf{j}) - s(\mathbf{i})} \langle 1 | U_m^\dagger | \hat{b}_m \rangle \langle \hat{b}_m | U_m | 0 \rangle \right. \\ \left. + \prod_{m \in s(\mathbf{j}) - s(\mathbf{i})} \langle 0 | U_m^\dagger | \hat{b}_m \rangle \langle \hat{b}_m | U_m | 1 \rangle \prod_{m \in s(\mathbf{i}) - s(\mathbf{j})} \langle 1 | U_m^\dagger | \hat{b}_m \rangle \langle \hat{b}_m | U_m | 0 \rangle \right) \delta_{\hat{b}_{m \in s(\mathbf{i}) \cap s(\mathbf{j})}, \mathbf{1}} \delta_{\hat{b}_{m \in s(\mathbf{i}) \cup s(\mathbf{j})}, \mathbf{0}} \quad (\text{B2})$$

evaluates to $\pm i/2^{2k-2l-1}$ if $\delta_{\hat{b}_{m \in s(\mathbf{i}) \cap s(\mathbf{j})}, \mathbf{1}} \delta_{\hat{b}_{m \in s(\mathbf{i}) \cup s(\mathbf{j})}, \mathbf{0}} = 1$ and $|\{m \in s(\mathbf{i}) \Delta s(\mathbf{j}) | U_m = HS^\dagger\}| \equiv 0 \pmod{2}$. Otherwise, it is always zero. This symmetry allows us to further reduce the sample space of (U_m) :

$$o_l = \sum_{\substack{\mathbf{i} < \mathbf{j}, \\ |\mathbf{i}| = |\mathbf{j}| = k, \\ |s(\mathbf{i}) \cap s(\mathbf{j})| = l}} \sum_{\substack{U_{m \in s(\mathbf{i}) \Delta s(\mathbf{j})} \in T \setminus \{\mathbb{I}\}, \\ U_{m \in s(\mathbf{i}) \Delta s(\mathbf{j})} = \mathbb{I}, \\ |\{m \in s(\mathbf{i}) \Delta s(\mathbf{j}) | U_m = HS^\dagger\}| \equiv 0 \pmod{2}}} \mathbb{E}_{\hat{b}} \left[V_D \binom{n}{k}^{-1} \right].$$

Let f_l denote the uniform distribution over all $\{\mathbf{i}, \mathbf{j}\}$ such that $|\mathbf{i}| = |\mathbf{j}| = k$ and $|s(\mathbf{i}) \cap s(\mathbf{j})| = l$. There are

$$c_l \equiv \binom{n}{2k-l} \binom{2k-l}{l} \binom{2k-2l-1}{k-l-1} \quad (\text{B3})$$

such sets. Then

$$o_l = \sum_{\substack{\mathbf{i} < \mathbf{j}, \\ |\mathbf{i}| = |\mathbf{j}| = k, \\ |s(\mathbf{i}) \cap s(\mathbf{j})| = l}} \frac{1}{c_l} \sum_{U'_{m \in s(\mathbf{i}) \Delta s(\mathbf{j})} - \max[s(\mathbf{i}) \Delta s(\mathbf{j})] \in T \setminus \{\mathbb{I}\}} \frac{1}{2^{2k-2l-1}} \mathbb{E}_{\hat{b}} \left[\hat{o}_l \left| \begin{array}{l} U_{m \in s(\mathbf{i}) \Delta s(\mathbf{j}) - \max[s(\mathbf{i}) \Delta s(\mathbf{j})]} = U'_m, \\ U_{\max[s(\mathbf{i}) \Delta s(\mathbf{j})]} = \mathcal{F}(\{U'_m\}), \\ U_{m \notin s(\mathbf{i}) \Delta s(\mathbf{j})} = \mathbb{I} \end{array} \right. \right] \\ = \mathbb{E}_{\{\mathbf{i}, \mathbf{j}\} \sim f_l, U'_{m \in s(\mathbf{i}) \Delta s(\mathbf{j})} - \max[s(\mathbf{i}) \Delta s(\mathbf{j})] \sim g, \hat{b}} \left[\hat{o}_l \left| \begin{array}{l} U_{m \in s(\mathbf{i}) \Delta s(\mathbf{j}) - \max[s(\mathbf{i}) \Delta s(\mathbf{j})]} = U'_m, \\ U_{\max[s(\mathbf{i}) \Delta s(\mathbf{j})]} = \mathcal{F}(\{U'_m\}), \\ U_{m \notin s(\mathbf{i}) \Delta s(\mathbf{j})} = \mathbb{I} \end{array} \right. \right],$$

where $\hat{o}_l \equiv V_D \binom{n}{k}^{-1} c_l 2^{2k-2l-1}$ and $\mathcal{F}(\{U'_m\}) \in T \setminus \{\mathbb{I}\}$ such that $|\{U \in \{U'_m\} \cup \mathcal{F}(\{U'_m\}) | U = HS^\dagger\}| \equiv 0 \pmod{2}$. Therefore, we have $-\binom{n}{k}^{-1} c_l \leq \hat{o}_l \leq \binom{n}{k}^{-1} c_l$.

-
- [1] D. Lacroix, Symmetry-assisted preparation of entangled many-body states on a quantum computer, *Phys. Rev. Lett.* **125**, 230502 (2020).
 - [2] M. Roghani and H. Weimer, Dissipative preparation of entangled many-body states with Rydberg atoms, *Quantum Sci. Technol.* **3**, 035002 (2018).
 - [3] F. Reiter, D. Reeb, and A. S. Sørensen, Scalable dissipative preparation of many-body entanglement, *Phys. Rev. Lett.* **117**, 040501 (2016).
 - [4] B. Kraus, H. P. Büchler, S. Diehl, A. Kantian, A. Micheli, and P. Zoller, Preparation of entangled states by quantum Markov processes, *Phys. Rev. A* **78**, 042307 (2008).
 - [5] R. Jozsa, Fidelity for mixed quantum states, *J. Mod. Opt.* **41**, 2315 (1994).
 - [6] M. M. Wilde, *Quantum Information Theory* (Cambridge university press, Cambridge, 2013).
 - [7] Y.-C. Liang, Y.-H. Yeh, P. E. Mendonça, R. Y. Teh, M. D. Reid, and P. D. Drummond, Quantum fidelity measures for mixed states, *Rep. Prog. Phys.* **82**, 076001 (2019).
 - [8] A. J. Baldwin and J. A. Jones, Efficiently computing the Uhlmann fidelity for density matrices, *Phys. Rev. A* **107**, 012427 (2023).
 - [9] T. Morimae and K. Fujii, Blind quantum computation protocol in which Alice only makes measurements, *Phys. Rev. A* **87**, 050301(R) (2013).
 - [10] Y. Takeuchi and T. Morimae, Verification of many-qubit states, *Phys. Rev. X* **8**, 021060 (2018).
 - [11] D. T. Smithey, M. Beck, M. G. Raymer, and A. Faridani, Measurement of the wigner distribution and the density matrix of a light mode using optical homodyne tomography: Application to squeezed states and the vacuum, *Phys. Rev. Lett.* **70**, 1244 (1993).
 - [12] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, Measurement of qubits, *Phys. Rev. A* **64**, 052312 (2001).
 - [13] R. T. Thew, K. Nemoto, A. G. White, and W. J. Munro, Qudit quantum-state tomography, *Phys. Rev. A* **66**, 012303 (2002).
 - [14] H. Häffner, W. Hänsel, C. Roos, J. Benhelm, D. Chek-al Kar, M. Chwalla, T. Körber, U. Rapol, M. Riebe, P. Schmidt *et al.*,

- Scalable multiparticle entanglement of trapped ions, *Nature (London)* **438**, 643 (2005).
- [15] D. Leibfried, E. Knill, S. Seidelin, J. Britton, R. B. Blakestad, J. Chiaverini, D. B. Hume, W. M. Itano, J. D. Jost, C. Langer, R. Ozeri, R. Reichle, and D. J. Wineland, Creation of a six-atom “Schrödinger cat” state, *Nature (London)* **438**, 639 (2005).
 - [16] A. I. Lvovsky and M. G. Raymer, Continuous-variable optical quantum-state tomography, *Rev. Mod. Phys.* **81**, 299 (2009).
 - [17] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, Quantum state tomography via compressed sensing, *Phys. Rev. Lett.* **105**, 150401 (2010).
 - [18] G. Tóth, W. Wieczorek, D. Gross, R. Krischek, C. Schwemmer, and H. Weinfurter, Permutationally invariant quantum tomography, *Phys. Rev. Lett.* **105**, 250403 (2010).
 - [19] J. Shang, Z. Zhang, and H. K. Ng, Superfast maximum-likelihood reconstruction for quantum tomography, *Phys. Rev. A* **95**, 062336 (2017).
 - [20] M. Cramer, M. B. Plenio, S. T. Flammia, R. Somma, D. Gross, S. D. Bartlett, O. Landon-Cardinal, D. Poulin, and Y.-K. Liu, Efficient quantum state tomography, *Nat. Commun.* **1**, 149 (2010).
 - [21] H. Zhu and M. Hayashi, Efficient verification of hypergraph states, *Phys. Rev. Appl.* **12**, 054047 (2019).
 - [22] N. Dangniam, Y.-G. Han, and H. Zhu, Optimal verification of stabilizer states, *Phys. Rev. Res.* **2**, 043323 (2020).
 - [23] H.-Y. Huang, J. Preskill, and M. Soleimanifar, Certifying almost all quantum states with few single-qubit measurements, in *2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, New York, 2024), pp. 1202–1206.
 - [24] S. T. Flammia and Y.-K. Liu, Direct fidelity estimation from few Pauli measurements, *Phys. Rev. Lett.* **106**, 230501 (2011).
 - [25] X. Zhang, M. Luo, Z. Wen, Q. Feng, S. Pang, W. Luo, and X. Zhou, Direct fidelity estimation of quantum states using machine learning, *Phys. Rev. Lett.* **127**, 130503 (2021).
 - [26] A. Seshadri, M. Ringbauer, J. Spainhour, R. Blatt, T. Monz, and S. Becker, Versatile fidelity estimation with confidence, *Phys. Rev. Lett.* **133**, 020402 (2024).
 - [27] M. P. da Silva, O. Landon-Cardinal, and D. Poulin, Practical characterization of quantum devices without tomography, *Phys. Rev. Lett.* **107**, 210404 (2011).
 - [28] L. Leone, S. F. E. Oliviero, and A. Hama, Nonstabilizerness determining the hardness of direct fidelity estimation, *Phys. Rev. A* **107**, 022429 (2023).
 - [29] H. Kahn and T. E. Harris, Estimation of particle transmission by random sampling, *National Bureau of Standards, Applied Mathematics Series* (BibSonomy, 1951), Vol. 12, p. 27.
 - [30] T. Kloek and H. K. Van Dijk, Bayesian estimates of equation system parameters: an application of integration by Monte Carlo, *Econometrica* **46**, 1 (1978).
 - [31] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, *Nat. Phys.* **16**, 1050 (2020).
 - [32] A. A. Akhtar, H.-Y. Hu, and Y.-Z. You, Scalable and flexible classical shadow tomography with tensor networks, *Quantum* **7**, 1026 (2023).
 - [33] H.-Y. Hu, S. Choi, and Y.-Z. You, Classical shadow tomography with locally scrambled quantum dynamics, *Phys. Rev. Res.* **5**, 023027 (2023).
 - [34] A. R. Calderbank, E. M. Rains, P. M. Shor, and N. J. Sloane, Quantum error correction via codes over GF(4), *IEEE Trans. Inf. Theory* **44**, 1369 (1998).
 - [35] S. Aaronson and D. Gottesman, Improved simulation of stabilizer circuits, *Phys. Rev. A* **70**, 052328 (2004).
 - [36] K. N. Patel, I. L. Markov, and J. P. Hayes, Optimal synthesis of linear reversible circuits, *Quantum Inf. Comput.* **8**, 282 (2008).
 - [37] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge university press, Cambridge, 2010).
 - [38] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke *et al.*, Noisy intermediate-scale quantum algorithms, *Rev. Mod. Phys.* **94**, 015004 (2022).
 - [39] D. M. Greenberger, M. A. Horne, and A. Zeilinger, Going beyond bell’s theorem, *Bell’s Theorem, Quantum Theory and Conceptions of the Universe* (Springer, New York, 1989), pp. 69–72.
 - [40] W. Dür, G. Vidal, and J. I. Cirac, Three qubits can be entangled in two inequivalent ways, *Phys. Rev. A* **62**, 062314 (2000).
 - [41] A. Cabello, Bell’s theorem with and without inequalities for the three-qubit Greenberger-Horne-Zeilinger and W states, *Phys. Rev. A* **65**, 032108 (2002).
 - [42] J. K. Stockton, J. M. Geremia, A. C. Doherty, and H. Mabuchi, Characterizing the entanglement of symmetric many-particle spin-1/2 systems, *Phys. Rev. A* **67**, 022112 (2003).
 - [43] E. D’Hondt and P. Panangaden, The computational power of the w and ghz states, *J. Quantum Inf. Comput.* **6**, 173 (2005).
 - [44] M. Ben-Or and A. Hassidim, Fast quantum byzantine agreement, in *Proceedings of the 37th annual ACM symposium on Theory of computing* (Association for Computing Machinery, 2005), pp. 481–485.
 - [45] P. Faist, S. Nezami, V. V. Albert, G. Salton, F. Pastawski, P. Hayden, and J. Preskill, Continuous symmetries and approximate quantum error correction, *Phys. Rev. X* **10**, 041018 (2020).
 - [46] R. Hayakawa, Quantum algorithm for persistent Betti numbers and topological data analysis, *Quantum* **6**, 873 (2022).
 - [47] D. W. Berry, Y. Su, C. Gyurik, R. King, J. Basso, A. D. T. Barba, A. Rajput, N. Wiebe, V. Dunjko, and R. Babbush, Analyzing prospects for quantum advantage in topological data analysis, *PRX Quantum* **5**, 010319 (2024).
 - [48] Y.-C. Liu, X.-D. Yu, J. Shang, H. Zhu, and X. Zhang, Efficient verification of Dicke states, *Phys. Rev. Appl.* **12**, 044020 (2019).
 - [49] Z. Li, Y.-G. Han, H.-F. Sun, J. Shang, and H. Zhu, Verification of phased Dicke states, *Phys. Rev. A* **103**, 022601 (2021).
 - [50] J. A. Holbrook, D. W. Kribs, and R. Laflamme, Noiseless subsystems and the structure of the commutant in quantum error correction, *Quantum Inf. Process.* **2**, 381 (2003).
 - [51] C. Weedbrook, S. Pirandola, R. García-Patrón, N. J. Cerf, T. C. Ralph, J. H. Shapiro, and S. Lloyd, Gaussian quantum information, *Rev. Mod. Phys.* **84**, 621 (2012).
 - [52] W. Hoeffding, Probability inequalities for sums of bounded random variables, *The Collected Works of Wassily Hoeffding* (Springer Science & Business Media, Berlin, 1994), p. 409.
 - [53] J. Kahane, Local properties of functions with random Fourier series, *Stud. Math.* **19**, 1 (1960).
 - [54] S. Pallister, N. Linden, and A. Montanaro, Optimal verification of entangled states with local measurements, *Phys. Rev. Lett.* **120**, 170502 (2018).

- [55] Z. Li, Y.-G. Han, and H. Zhu, Optimal verification of Greenberger-Horne-Zeilinger states, *Phys. Rev. Appl.* **13**, 054002 (2020).
- [56] H. Buhrman and H. Röhrig, Distributed quantum computing, *International Symposium on Mathematical Foundations of Computer Science* (Springer, New York, 2003), pp. 1–20.
- [57] D. Cuomo, M. Caleffi, and A. S. Cacciapuoti, Towards a distributed quantum computing ecosystem, *IET Quantum Commun.* **1**, 3 (2020).
- [58] M. Caleffi, M. Amoretti, D. Ferrari, J. Illiano, A. Manzalini, and A. S. Cacciapuoti, Distributed quantum computing: A survey, *Comput. Networks* **254**, 110672 (2024).
- [59] P.-W. Huang and P. Rebentrost, Post-variational quantum neural networks, [arXiv:2307.10560](https://arxiv.org/abs/2307.10560).
- [60] S. Jerbi, C. Gyurik, S. C. Marshall, R. Molteni, and V. Dunjko, Shadows of quantum machine learning, *Nat. Commun.* **15**, 5676 (2024).
- [61] S. Chen, W. Yu, P. Zeng, and S. T. Flammia, Robust shadow estimation, *PRX Quantum* **2**, 030348 (2021).
- [62] V. Bergholm, J. Izaac, M. Schuld, C. Gogolin, S. Ahmed, V. Ajith, M. S. Alam, G. Alonso-Linaje, B. AkashNarayanan, A. Asadi *et al.*, PennyLane: Automatic differentiation of hybrid quantum-classical computations, [arXiv:1811.04968](https://arxiv.org/abs/1811.04968).
- [63] <https://github.com/HyunHoCha/DFEPauli>.