IOP Publishing Quantum Sci. Technol. 10 (2025) 035043 https://doi.org/10.1088/2058-9565/addffd

Quantum Science and Technology

OPEN ACCESS

RECEIVED
3 February 2025

REVISED
25 April 2025

ACCEPTED FOR PUBLICATION
3 June 2025

PUBLISHED
20 June 2025

PAPER

Classical shadows with improved median-of-means estimation

Winston Fu[1,2,*], **Dax Enshan Koh**[1,2,3,*], **Siong Thye Goh**[1,2,4] and **Jian Feng Kong**[1,2]

[1] Quantum Innovation Centre (Q.InC), Agency for Science, Technology and Research (A*STAR), 2 Fusionopolis Way, Innovis #08-03, Singapore 138634, Singapore
[2] Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, #16-16 Connexis, Singapore 138632, Singapore
[3] Science, Mathematics and Technology Cluster, Singapore University of Technology and Design, 8 Somapah Road, Singapore 487372, Singapore
[4] Singapore Management University, 81 Victoria St, Singapore 188065, Singapore
* Authors to whom any correspondence should be addressed.

E-mail: Winston_Fu@ihpc.a-star.edu.sg, Dax_Koh@ihpc.a-star.edu.sg, Goh_Siong_Thye@ihpc.a-star.edu.sg and Kong_Jian_Feng@ihpc.a-star.edu.sg

Keywords: classical, shadows, estimation, tomography, quantum

Abstract

The classical shadows protocol, introduced by Huang *et al* (2020 *Nat. Phys.* **16** 1050), makes use of the median-of-means (MoM) estimator to efficiently estimate the expectation values of M observables with failure probability δ using only $\mathcal{O}(\log(M/\delta))$ measurements. In their analysis, Huang *et al* used loose constants in their asymptotic performance bounds for simplicity. However, the specific values of these constants can significantly affect the number of shots used in practical implementations. To address this, we studied a modified MoM estimator proposed by Minsker (2023 *Proc. 36th Conf. on Learning Theory* (PMLR) 195 5925) that uses optimal constants and involves a U-statistic over the data set. For efficient estimation, we implemented two types of incomplete U-statistics estimators, the first based on random sampling and the second based on cyclically permuted sampling. We compared the performance of the original and modified estimators when used with the classical shadows protocol with single-qubit Clifford unitaries (Pauli measurements) for an Ising spin chain, and global Clifford unitaries (Clifford measurements) for the Greenberger–Horne–Zeilinger state. While the original estimator outperformed the modified estimators for Pauli measurements, the modified estimators showed improved performance over the original estimator for Clifford measurements. Our findings highlight the importance of tailoring estimators to specific measurement settings to optimize the performance of the classical shadows protocol in practical applications.

1. Introduction

Classical shadows is a protocol recently proposed by Huang, Kueng, and Preskill [1] that builds on ideas from Aaronson's shadow tomography [2] and quantum state tomography [3, 4] to efficiently estimate properties of an unknown quantum state ρ, without requiring a complete description of ρ to be learned. The protocol tackles the fundamental scaling problem of quantum systems, where learning a full description of the quantum system requires a number of measurements that scales exponentially with the number of qubits [5, 6]. By instead constructing an approximate classical description of the system, the expectation values of M observables can be predicted using only $\mathcal{O}(\log M)$ measurements, independent of system size. Since its introduction, the protocol and its variants [7–42] have found applications across diverse domains, including chemistry [43–47], machine learning [48–50], error mitigation [51–53], entanglement characterization [54–56], and variational quantum algorithms [57–60], among others [61–66]. Moreover, noise-robust adaptations of the protocol have been introduced, allowing it to sustain efficient performance under realistic, noisy conditions [67–75].

To estimate expectation values of observables using classical shadows, the median-of-means (MoM) estimator [76, 77] is commonly employed, offering favorable scaling of sample complexity with respect to the

failure probability of the protocol. In the original construction, Huang *et al* chose conservative numerical constants for estimator bounds, which, although not affecting theoretical analyses, play a critical role in practical implementations [78–81]. Specifically, these constants directly affect the number of measurements required to achieve a desired level of accuracy in the estimation process. While theoretical analyses typically focus on large sample sizes where constants become negligible, they become critical in practical scenarios with limited measurements or resource constraints. Tighter constants in the MoM estimator allow for more precise estimates with fewer measurements, reducing overhead.

Recently, Minsker [82, 83] presented optimal constants for the MoM estimator, and also proposed a modified version of the estimator with even tighter constants. In this study, we apply Minsker's results to classical shadows. Using ideas from incomplete U-statistics [84], we introduce practical implementations of Minsker's estimator which would otherwise be impossible to run for large datasets. Through numerical simulations, we benchmarked the performance of the original and modified estimators using an Ising spin chain and the Greenberger–Horne–Zeilinger (GHZ) state [85] to measure their performance, using the single-qubit Clifford unitary ensemble and global Clifford unitary ensemble respectively. We found that while the original MoM estimator performed better in the former case, the modified version showed improvements in the latter case.

The rest of our manuscript is structured as follows. In section 2, we review the basics of the classical shadows protocol. In section 3, we present the original MoM estimator used in the protocol, followed by Minsker's modified estimator in section 4. Next, in section 5, we present two practical implementations of Minsker's estimator using incomplete U-statistics in the context of classical shadows. We then benchmark the estimators for the Ising spin chain in section 6.1 and GHZ state in section 6.2. Finally, in section 7, we provide a discussion of the results.

2. Preliminaries

The purpose of the classical shadows protocol is to efficiently predict functions of a density matrix ρ. Of particular importance are linear functions, such as expectation values $\{o_i\}$ of a set of M observables $\{O_i\}$, which can be expressed as:

$$o_i(\rho) = \mathrm{tr}(O_i \rho) \quad 1 \leqslant i \leqslant M. \tag{1}$$

The protocol works by applying a transformation $\rho \mapsto U\rho U^\dagger$, where U is randomly selected from an ensemble of unitaries. We consider two commonly used ensembles: random global Clifford unitaries, and tensor products of random single-qubit Clifford unitaries. After the transformation, we perform a computational-basis measurement to obtain $|\hat{b}\rangle \in \{0,1\}^r$, where r is the number of qubits. Using this, we define the shadow channel $\mathcal{M}$, defined as

$$\mathcal{M}(\rho) = \mathbb{E}\left[U^\dagger |\hat{b}\rangle\langle\hat{b}| U\right]. \tag{2}$$

We typically choose ensembles for which the object $U^\dagger |\hat{b}\rangle\langle\hat{b}| U$ can be efficiently stored classically. Assuming that the inverse map $\mathcal{M}^{-1}$ exists, it can be shown that the classical snapshot $\hat{\rho} = \mathcal{M}^{-1}(U^\dagger |\hat{b}\rangle\langle\hat{b}| U)$ is an unbiased estimator of ρ, i.e.

$$\rho = \mathbb{E}[\hat{\rho}] = \mathbb{E}\left[\mathcal{M}^{-1}\left(U^\dagger |\hat{b}\rangle\langle\hat{b}| U\right)\right]. \tag{3}$$

While the inverse map $\mathcal{M}^{-1}$ is not necessarily a valid quantum channel, it can still be applied during classical post-processing. Repeatedly sampling U from the ensemble of unitaries N times results in a classical shadow S of length N:

$$S(\rho; N) = \{\hat{\rho}_1, \ldots, \hat{\rho}_N\}. \tag{4}$$

3. MoM estimator

Given M observables O_i, which return $o_i = \mathrm{tr}(O_i\rho)$ when acting on the system ρ, we can construct an unbiased estimator $\hat{o}_i$ of o_i using the set of classical shadows $S(\rho; N)$:

$$\hat{o}_i(N, 1) = \frac{1}{N}\sum_{j=1}^{N} \mathrm{tr}\left(O_i \hat{\rho}_j\right). \tag{5}$$

Algorithm 1. Median-of-means estimator for a list of data (MoM).

Input:
- $x_1, \ldots, x_N$: Array of data.
- k: The number of disjoint subsets.

Output: Estimated result.

```
function MoM (x_1, ..., x_N, k)
    A ← {}
    for i ← 0, ⌊N/k⌋ − 1 do
        A ← A ∪ {(1/k) Σ_{p=ik+1}^{k(i+1)} x_p}
    end for
    return median(A)
end function
```

From Chebyshev's inequality,

$$P\left(\left|\hat{o}_i(N,1) - \mathrm{tr}\left(O_i\rho\right)\right| \geqslant \varepsilon\right) \leqslant \frac{\mathrm{Var}\left[\hat{o}_i\right]}{\varepsilon^2} \tag{6}$$

$$= \frac{\mathrm{Var}\left[\mathrm{tr}\left(O_i\hat{\rho}_1\right)\right]}{N\varepsilon^2} \tag{7}$$

$$\equiv \frac{\delta}{M}. \tag{8}$$

This means that to achieve a failure probability below δ, the number of samples required scales as $N \sim 1/\delta$. To circumvent this $1/\delta$ dependence, the MoM estimator is often used:

$$\hat{o}_i(N,k) = \mathrm{median}\left\{\hat{o}_i^{(1)}\left(\left\lfloor\frac{N}{k}\right\rfloor, 1\right), \ldots, \hat{o}_i^{(k)}\left(\left\lfloor\frac{N}{k}\right\rfloor, 1\right)\right\}. \tag{9}$$

This involves splitting the N snapshots into groups of size $\lfloor N/k\rfloor$, and finding the mean of each group. Then, we take the median of these groups. This procedure is shown in algorithm 1.

Let $\hat{\mu}_{\mathrm{MoM}}$ be an MoM estimator of N random variables X_i, so

$$\hat{\mu}_{\mathrm{MoM}} = \mathrm{median}\left(\bar{X}_1, \ldots, \bar{X}_k\right), \tag{10}$$

where $\bar{X}_i$ denotes the mean of the group with size $\lfloor N/k\rfloor$. The estimator $\hat{\mu}_{\mathrm{MoM}}$ obeys the inequality

$$P\left(\left|\hat{\mu}_{\mathrm{MoM}} - \mu\right| \geqslant C\sigma\sqrt{\frac{t}{N}}\right) \leqslant 2\mathrm{e}^{-t}, \tag{11}$$

where C can be taken to be $8e^2$ (see appendix A). Using $N = 34\sigma^2 k/\varepsilon^2$, $t = k/2$, Huang *et al* [1] obtain the bound

$$P\left(\left|\hat{\mu}_{\mathrm{MoM}} - \mu\right| \geqslant \varepsilon\right) \leqslant 2\mathrm{e}^{-k/2}. \tag{12}$$

However, Minsker [83, 86] showed that a tighter bound is possible.

Theorem 1 (Theorem 2.1 of Minsker [83].).

$$P\left(\left|\hat{\mu}_{\mathrm{MoM}} - \mu\right| \geqslant \sigma\sqrt{\frac{t}{N}}\right) \leqslant 2\exp\left(-\frac{t}{\pi}\left(1 + o\left(1\right)\right)\right), \tag{13}$$

where $o(1)$ is a function that goes to 0 as $k, N/k \to \infty$, uniformly over $t \in [l_{k,N}, u_{k,N}]$ for any sequences $l_{k,N} \gg kg^2(N/k)$ and $u_{k,N} \ll k$, where g satisfies the following inequality:

$$g(m) \leqslant C\,\mathbb{E}\left|\frac{X_1 - \mu}{\sigma}\right|^q m^{-(q-2)/2} \tag{14}$$

whenever $\mathbb{E}|X_1 - \mu|^q < \infty$ for some $q \in (2, 3]$.

This corresponds to $C = \sqrt{\pi} + o(1)$ in equation (11). In the context of quantum measurements, the condition $\mathbb{E}|X_1 - \mu|^q < \infty$ is trivially satisfied. Next, if $\sqrt{kg}(N/k) \to 0$ as $k, N \to \infty$, then the range of t becomes $1 \leqslant t \ll k$. For classical shadows, we can choose k such that $N \gg k$ and $o(1) \ll 1$. Hence, we will ignore this factor in our subsequent analysis.

Given M quantities we want to estimate, we construct the estimators $\hat{\mu}_1, \ldots, \hat{\mu}_M$. We want to find a value for t such that:

$$P(|\hat{\mu}_i - \mu_i| \geqslant \varepsilon \; \forall i) \leqslant \delta. \tag{15}$$

Starting from theorem 1, we make the substitution $t \mapsto \pi t$, $\varepsilon = \sigma\sqrt{\pi t/N}$:

$$P(|\hat{\mu}_i - \mu_i| \geqslant \varepsilon) \leqslant 2\mathrm{e}^{-t} = \frac{\delta}{M}. \tag{16}$$

From the union bound,

$$P\left(\bigcup_{i=1}^{M} |\hat{\mu}_i - \mu_i| \geqslant \varepsilon\right) \leqslant \sum_{i=1}^{M} P(|\hat{\mu}_i - \mu_i| \geqslant \varepsilon) \tag{17}$$

$$\leqslant \sum_{i=1}^{M} \frac{\delta}{M} \tag{18}$$

$$= \delta. \tag{19}$$

Therefore, $t = \log(2M/\delta)$. To satisfy the condition $t \ll k$, we will choose $t = k/\log(k)$.

4. Modified MoM for classical shadows

It was shown by Minsker [82] that a modified, permutation-invariant version of Mom can give even tighter bounds. Similar to before, we first split our set of N data points $\{X_1, \ldots, X_N\}$ into disjoint subsets of size $n = kl$ and take the mean of each subset:

$$G_1 \cup \ldots \cup G_n \subseteq [N], \quad Z_j = \bar{X}_j = \frac{1}{|G_j|} \sum_{i \in G_j} X_i, \tag{20}$$

where $[N] = \{1, \ldots, N\}$. For some set of integers $J \subseteq [n]$ of cardinality $|J| = l$, define $\bar{Z}_J = \frac{1}{l}\sum_{j \in J} Z_j$. Then, to construct our estimator $\hat{\mu}_{\text{comb}}$, we take the median of all possible $\bar{Z}_J$:

$$\hat{\mu}_{\text{comb}} \equiv \text{median}\left(\bar{Z}_J, J \in \mathcal{A}_n^{(l)}\right), \tag{21}$$

$$\mathcal{A}_n^{(l)} = \{J \subset [n] : |J| = l\}. \tag{22}$$

This process is shown in algorithm 2.

Making appropriate substitutions to Minsker's results, we get a bound corresponding to $C = \sqrt{2} + o(1)$ in equation (11)—a tighter constant compared to Mom.

Theorem 2 (Theorem 1 of Minsker [82].). *Assume that $\mathbb{E}|X_1 - \mu|^{2+a} < \infty$ for some $a > 0$. Suppose that $l = o(m^a)$ and let $L(n,l)$ and $M(n,l)$ be any sequences such that $L(n,l) \gg \frac{n}{l}g^2(m)$ and $M(n,l) \ll \frac{n}{l^2}$. Then for all $L(n,l) \leqslant t \leqslant M(n,l)$,*

$$P\left(|\hat{\mu}_{\text{MoM}} - \mu| \geqslant \sigma\sqrt{\frac{t}{N}}\right) \leqslant 3\exp\left(-\frac{t}{2(1+o(1))}\right), \tag{23}$$

where $m = \lfloor N/k \rfloor$, $o(1) \to 0$ as $l, k \to \infty$ uniformly over all $t \in [L(n,l), M(n,l)]$.

A full treatment can be found in Minsker's [82] work, which we summarize below. Using the same analysis as equation (15), we choose $t = \log(3M/\delta)$. Additionally, we will use the values $l = \log(N/k)$, $t = n/l^2\log(l)$.

The MoM estimator can be written as an M-estimator:

$$\hat{\mu}_{\text{MoM}} \in \underset{z \in \mathbb{R}}{\text{argmin}} \sum_{J \in \mathcal{A}_n^{(l)}} |\sqrt{m}\left(\bar{Z}_J - z\right)|. \tag{24}$$

Algorithm 2. Median-of-means estimator with combinations (MomComb).

Input:
- $x_1, \ldots, x_N$: Array of data.
- k: Changes the number of initial disjoint subsets, where kl is the number of subsets.
- l: Size of sets of sample averages.

Output: Estimated result.

function MomComb $(x_1, \ldots, x_N, k, l)$
  $A \leftarrow \{\}$
  for $i \leftarrow 0, \lfloor \frac{N}{kl} \rfloor - 1$ **do** ▷ Split data into disjoint subsets, like in Mom.
    $A \leftarrow A \cup \{\frac{1}{kl} \sum_{j=ikl+1}^{kl(i+1)} x_j\}$
  end for

  $B \leftarrow$ all possible subsets of size l of A ▷ B contains $\binom{kl}{l}$ sets of size l.
  $C \leftarrow \{\}$
  for all $J \in B$ **do**
    $C \leftarrow C \cup \{\frac{1}{l} \sum_{p=1}^{l} J_p\}$
  end for
  return median(C)
end function

The left derivative $|\cdot|'_-$ of the absolute value function $|\cdot|$ is $\leqslant 0$. Using the fact that $|\cdot|'_-$ is a non-decreasing function, it can be shown that the performance of the MomComb estimator is bounded by that of the U-statistic:

$$U_{n,l}(|\cdot|'_-) = \binom{n}{l}^{-1} \sum_{J \in \mathcal{A}_n^{(l)}} \left(\left(|\sqrt{m}(\bar{Z}_J - \mu - \sqrt{t/N})|'_- - \mathbb{E}|\cdot|'_- \right) \geqslant -\sqrt{k}\mathbb{E}|\cdot|'_- \right), \tag{25}$$

where

$$\mathbb{E}|\cdot|'_- \equiv \mathbb{E}\left| \left(\sqrt{m} \left(\bar{Z}_J - \mu - \sqrt{t/N} \right) \right) \right|'_- . \tag{26}$$

One downside of this approach is its computational complexity of $\mathcal{O}(k^l \log k)$, which can quickly become untenable. A practical implementation therefore requires sampling a subset $D \subseteq \mathcal{A}_n^{(l)}$ of m elements. In the proceeding sections, we will draw heavily on analysis by Lee [84] to present and benchmark two possible methods of sampling: randomly, and using cyclic permutations.

5. Incomplete modified MoM

An incomplete U-statistic $U_n^{(0)}$ involves considering a subset $\mathcal{D}$ of size m out of the $\binom{n}{l}$ terms in MomComb, where $\mathcal{D}$ is commonly called the design of $U_n^{(0)}$. For a given m, we want to choose $\mathcal{D}$ to minimize the variance of $U_n^{(0)}$ to maximize the efficiency of the estimator.

A simple and common way of choosing the design is to randomly sample from $\mathcal{A}_n^{(l)}$ either with or without replacement. For simplicity, we will consider the former. This approach is shown in algorithm 3. It can be shown that [84]

$$\operatorname{Var} U_n^{(0)} = \frac{\sigma_l^2}{m} + \left(1 - m^{-1}\right) \operatorname{Var} U_n, \tag{27}$$

where for some U-statistic $U_n = \binom{n}{l}^{-1} \sum_{(n,l)} \psi(X_{i_1}, \ldots, X_{i_l})$ and the sum is taken over all permutations $(i_1, \ldots, i_l)$ of $\{1, 2, \ldots, l\}$,

$$\sigma_c^2 = \operatorname{Var}\left[\psi_c(X_1, \ldots, X_c)\right], \tag{28}$$

where

$$\psi_c(x_1, \ldots, x_c) = \mathbb{E}\left[\psi(X_1, \ldots, X_k) | X_1 = x_1, \ldots, X_c = x_c\right]. \tag{29}$$

Algorithm 3. Median-of-means estimator with random sampling (MomRand).

Input:
- $x_1, \ldots, x_N$: Array of data.
- k: Changes the number of initial disjoint subsets, where kl is the number of subsets.
- l: Size of sets of sample averages.
- m: Number of randomly sampled sets of sample averages.

Output: Estimated result.

function MomRand $(x_1, \ldots, x_N, k, l, m)$
  $A \leftarrow \{\}$
  for $i \leftarrow 0, \lfloor \frac{N}{kl} \rfloor - 1$ **do** ▷ Split data into disjoint subsets, like in Mom.
    $A \leftarrow A \cup \{\frac{1}{kl} \sum_{p=ikl+1}^{kl(i+1)} x_p\}$
  end for

  $B \leftarrow \{\}$
  for $i \leftarrow 0, m-1$ **do** ▷ Randomly sample m subsets of A.
    $B \leftarrow B \cup \{J \subseteq A : |J| = l\}$ where J is a random subset of A.
  end for

  $C \leftarrow \{\}$
  for all $J \in B$ **do**
    $C \leftarrow C \cup \{\frac{1}{l} \sum_{p=1}^{l} J_p\}$
  end for
  return median(C)
end function

Let $m = Kn$ for some constant K. Comparing the variance with the complete estimator U_n, we find the asymptotic relative efficiency (ARE)

$$\mathrm{ARE} = \lim_{n\to\infty} \frac{\mathrm{Var}\, U_n}{\mathrm{Var}\, U_n^{(0)}} = \frac{Kl^2\sigma_1^2}{l^2\sigma_1^2 + \sigma_l^2}. \tag{30}$$

We can choose $\mathcal{D}$ in a manner that is theoretically more efficient than simple random sampling. One possible minimum variance design employs cyclic permutations of our data set. To our data set of n elements Z_j, we apply the cyclic permutations

$$\begin{pmatrix} 1 & 2 & \ldots & n \\ d_\nu \oplus 1 & d_\nu \oplus 2 & \ldots & d_\nu \oplus n \end{pmatrix} \tag{31}$$

for $\nu = 1, 2, \ldots, l$. The symbol $\oplus$ denotes addition (mod n), and d_ν's are chosen such that $(d_\nu - d_{\nu'})$ (mod n) are distinct for any pair (ν, ν') with $\nu \neq \nu'$. In other words, the chosen d_ν must form a modified Golomb ruler [87], which can be found in $\mathcal{O}(l^2)$ time. Furthermore, the values of d_ν can be precomputed and stored in a lookup table for reference during the execution of the estimator. Hence, this step has minimal impact on the performance of the algorithm.

After applying each cyclic permutation with offset d_ν, the elements are sequentially grouped into sets of size l, with their average taken to give $\bar{Z}_J$. Repeating the process K times with a new set of d_ν, where all pairs d_ν used have distinct differences, gives a total of $m = Kn$ values for $\bar{Z}_J$. This is summarized in algorithm 4.

This estimator MomCyc has variance

$$\mathrm{Var}\, U_n^{(0)} = m^{-1}\left(l(lK-1)\sigma_1^2 + \sigma_l^2\right) \tag{32}$$

and

$$\mathrm{ARE} = \frac{Kl^2\sigma_1^2}{l(lK-1)\sigma_1^2 + \sigma_l^2}, \tag{33}$$

which is more efficient than MomRand for the same m.

Algorithm 4. Median-of-means estimator with cyclic permutations (MomCyc).

Input:
- $x_1, \ldots, x_N$: Array of data.
- k: Changes the number of initial disjoint subsets, where kl is the number of subsets.
- l: Size of sets of sample averages.
- m: Number of rounds.

Output: Estimated result.

$\Rightarrow$ CycPerm(A, d): cyclically permutes all elements in A by offset d.

$\Rightarrow$ GolRul(l, D): generates a Golomb ruler (mod n) of length l, while also accounting for the differences stored in D.

function MomCyc $(x_1, \ldots, x_N, k, l, m)$
  $A \leftarrow \{\}$
  for $i \leftarrow 0, \lfloor \frac{N}{kl} \rfloor - 1$ **do** ▷ Split data into disjoint subsets, like in Mom.
    $A \leftarrow A \cup \{\frac{1}{kl} \sum_{p=ikl+1}^{kl(i+1)} x_p\}$
  end for

  $B \leftarrow \{\}$
  $D \leftarrow \{\}$
  for $i \leftarrow 0, m$ **do**
    $C \leftarrow$ GolRul (l, D)
    $D \leftarrow$ differences between any two elements in C.
    for all $d \in C$ **do**
      CycPerm (A, d)
      for $j \leftarrow 0, \lfloor \frac{N}{kl^2} \rfloor - 1$ **do** ▷ Split data into disjoint subsets of length l and store the mean of each subset.
        $B \leftarrow B \cup \{\frac{1}{l} \sum_{p=j+1}^{j+l} A_p\}$
      end for
    end for
  end for
  return median(B)
end function

6. Numerical experiments

6.1. Benchmarking with Ising Chain

We first used classical shadows with random single-qubit Cliffords (hereby referred to as 'Pauli measurements') to estimate the two-point correlator function $\langle \sigma_1^z \sigma_{i+1}^z \rangle$ of a transverse field Ising model, with Hamiltonian

$$H = J \sum_i \sigma_i^Z \sigma_{i+1}^Z + h \sum_i \sigma_i^X. \tag{34}$$

We found the correlator for 50 qubits with between 1000 and 50 000 Pauli measurements, using tensor network techniques described in section 9.

Figure 1 shows the results of the simulations using 4 different kinds of estimators. Figure 2 shows the difference between the predicted and exact values, together with the 'Mean bound' (equation (8)), 'Huang-Kueng-Preskill (HKP) bound' (equation (12)), 'Original bound' (theorem 1), and 'New bound' (theorem 2). The short dotted lines in the figure correspond to 3.3σ values, which have a 0.1% probability of occurring. This allows for direct comparison between numerical results and the theoretical bounds (long dashed lines) with failure probability $\delta = 0.1\%$.

6.2. Benchmarking with noisy GHZ state

Next, we examined the fidelity between a noisy GHZ state and a pure GHZ state. We define

$$|\mathrm{GHZ}_r^{\pm}\rangle = \frac{1}{\sqrt{2}} \left(|0\rangle^{\otimes r} \pm |1\rangle^{\otimes r} \right) \tag{35}$$

and introduce a GHZ source which has a phase error with probability p

$$\rho_p = (1-p)\, |\mathrm{GHZ}_r^+\rangle\langle \mathrm{GHZ}_r^+| + p |\mathrm{GHZ}_r^-\rangle\langle \mathrm{GHZ}_r^-|. \tag{36}$$

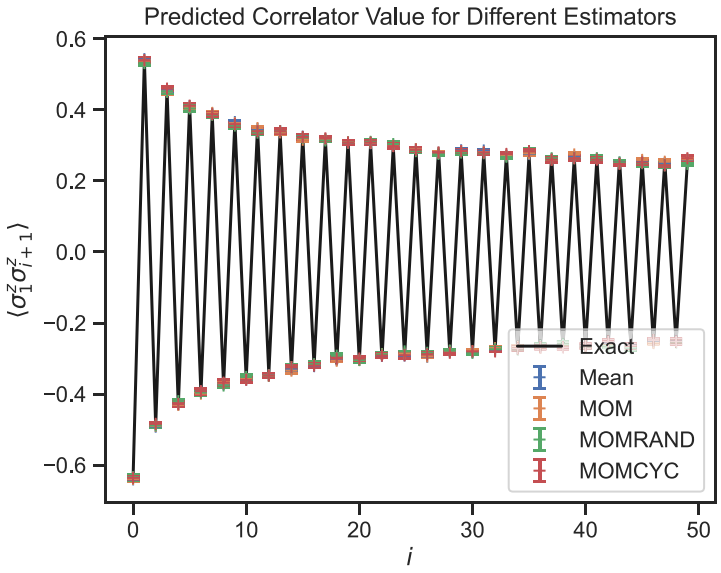

Figure 1. Predicted and exact values for the 2-point correlator function over 50 qubits, using Pauli measurements.

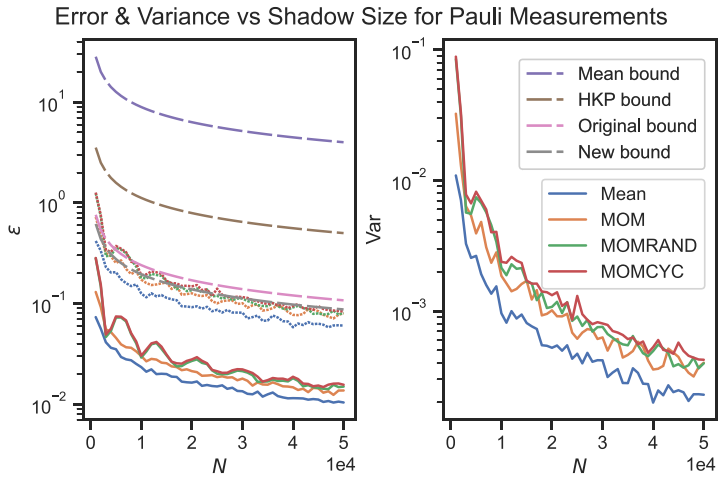

Figure 2. (Left) Average error of each correlator over 10 runs against shadow size for the mean, Mom, MomRand, and MomCyc. Short dotted lines of the corresponding color plot the 3.3σ boundary for each estimator. Long dashed lines show each of the bounds in equation (8), theorems 1, 2, and equation (12). (Right) Variance for the same runs.

Using random global Clifford unitaries (hereby referred to as 'Clifford measurements'), we found the classical shadow of ρ_p and calculated the fidelity with the pure GHZ state $\rho_{\text{pure}} = |\text{GHZ}_r^+\rangle\langle\text{GHZ}_r^+|$,

$$F\left(\rho_p, \rho_{\text{pure}}\right) = \text{tr}\left(\rho_p \rho_{\text{pure}}\right). \tag{37}$$

Figure 3 shows the exact and predicted values of $F(\rho_p, \rho_{\text{pure}})$ for difference estimators and $r = 2$. Figure 4 shows the average error and variance.

Huang *et al* [1] showed that the performance of the classical shadows protocol is independent of the number of qubits in the system. To verify this for the new estimators, we estimated the fidelity of the pure $p = 0$ GHZ state for $r = 5, 10, 15, 20$ and the average. This is shown in figure 5. In this case, the 3.3σ value of Mom exceeds the bounds of the modified estimator (theorem 2). Next, the shadow size N required to attain a fidelity of 0.98 was found, as shown in figure 6, showing independence between accuracy and shadow size regardless of estimator.

Finally, we tested the estimators on quadratic functions of ρ. Instead of median of means, Huang *et al* employ a median of U-statistics to estimate quadratic functions. The details of these modified estimators can be found in section 9.

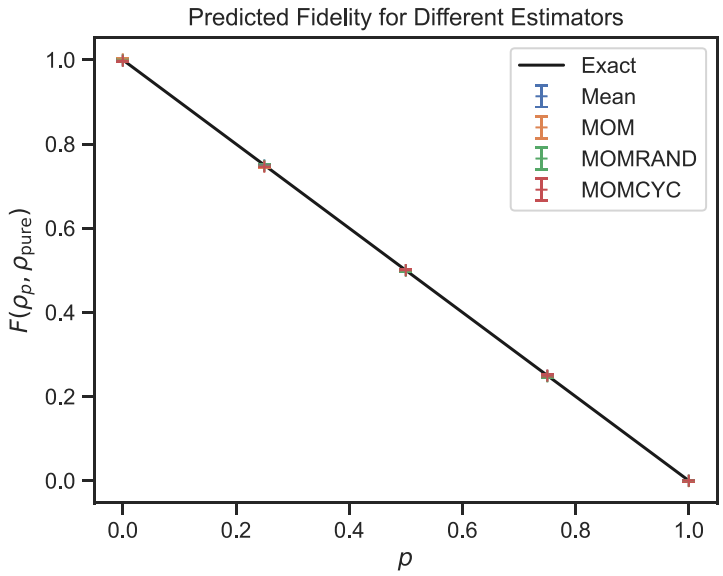

Figure 3. Predicted and exact values of fidelity for different phase error probability p using Clifford measurements.

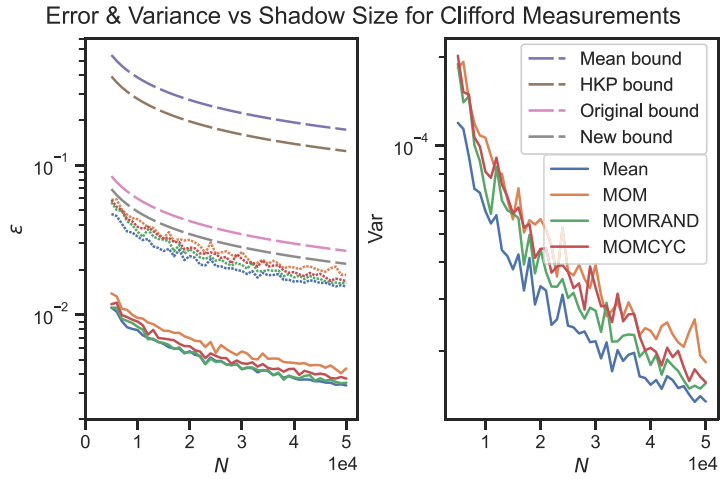

Figure 4. (Left) Average error for the fidelity over 100 runs against shadow size for the mean, Mom, MomComb, MomRand, and MomCyc for the $n = 2$-qubit noisy GHZ state. Short dotted lines of the corresponding color plot the 3.3σ boundary for each estimator. Long dashed lines show the bounds. (Right) Variance for the same runs.

Figure 7 shows the predicted purity for different values of p over 10 independent runs for $p = 0, 0.25, 0.5, 0.75, 1$ and figure 8 shows the error and variances of each estimator.

7. Discussion

Despite having the worst bound, taking the mean of the data had the lowest average error out of all the estimators. This agrees with prior experiments [78–80], as the distribution of the estimator $\hat{o}_i(N, 1)$ tends towards a normal distribution as $N \to \infty$ and gives a tighter bound than theorem 2. Our results strengthen the case for the normality of $\hat{o}_i(N, 1)$ for the number of samples $N \geqslant 1000$ examined (appendix B).

When estimating Pauli observables using Pauli measurements, Mom had the best performance out of the remaining estimators despite having worse bounds than MomRand and MomCyc. MomRand had error that was in between Mom and MomCyc. This result is further discussed in appendix C. The 3.3σ error of MomRand and MomCyc exceeded the tightened bounds by Minsker (theorem 2), indicating that these estimators may not be suitable for use with Pauli measurements.

When using Clifford measurements to predict linear functions, the practical performance of the MoM estimators followed the tightened bounds closely, indicating the validity of the assumptions made. Mom now had higher error and variance than MomRand and MomCyc. This shows that Minsker's estimator offers an

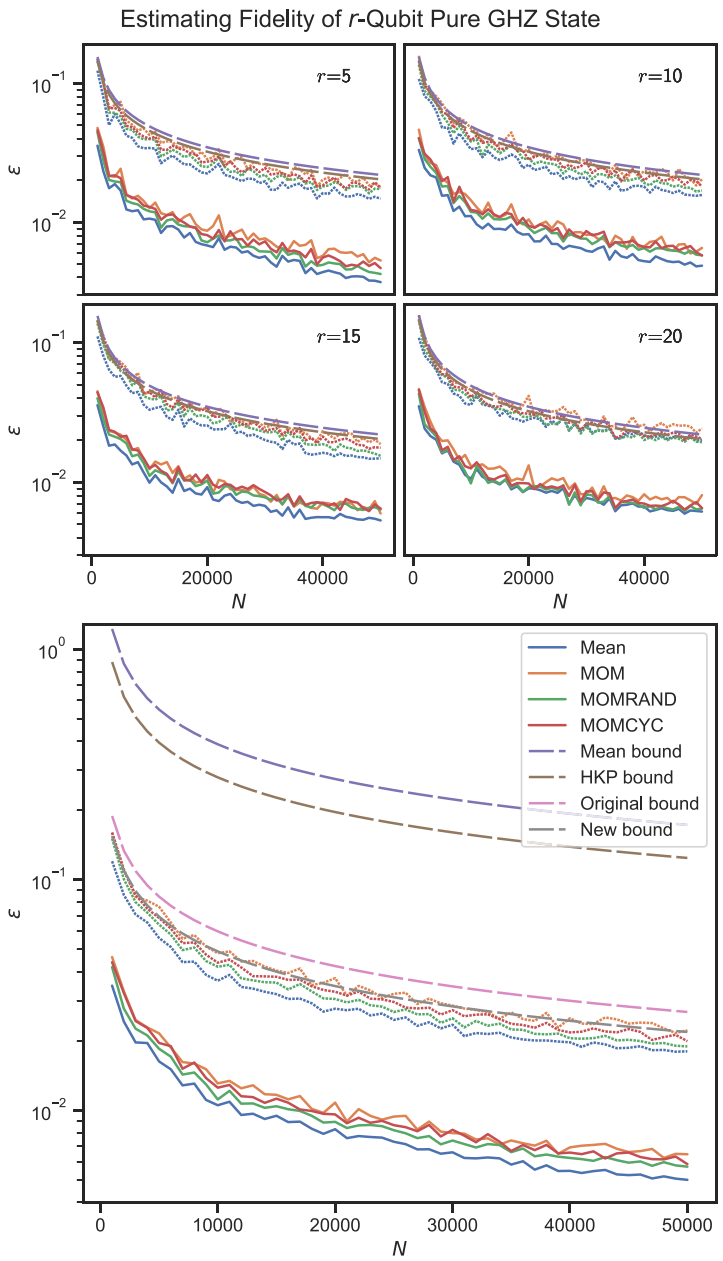

Figure 5. (Top half): Error in fidelity with the r-Qubit pure GHZ state with $r = 5, 10, 15, 20$ over 10 independent runs. (Bottom half): Average error over all four values of n.

advantage over the traditional MoM protocol for Clifford measurements. In this case, the mean and MomRand performed similarly, while MomCyc performed slightly worse. We have also demonstrated the tightness of Minsker's bound (theorem 2) in figure 5, as the bound was exceeded by Mom but not MomRand and MomCyc.

When predicting quadratic functions such as purity, the mean continued to perform the best. The MomRand and MomCyc estimators had worse performance than Mom, in contrast to the improved performance for linear functions.

Despite the higher ARE of MomCyc compared to MomRand, the latter showed better performance for Clifford measurements.

While the mean showed the best performance in all the simulations, MomRand had similar results when using Clifford measurements, with the added benefit that the MoM estimators are backed by performance bounds under weaker assumptions. Using the bound in theorem 2, the number of shots needed can be greatly reduced. As a simple illustration, the minimum number of shots required is summarized in table 1 for

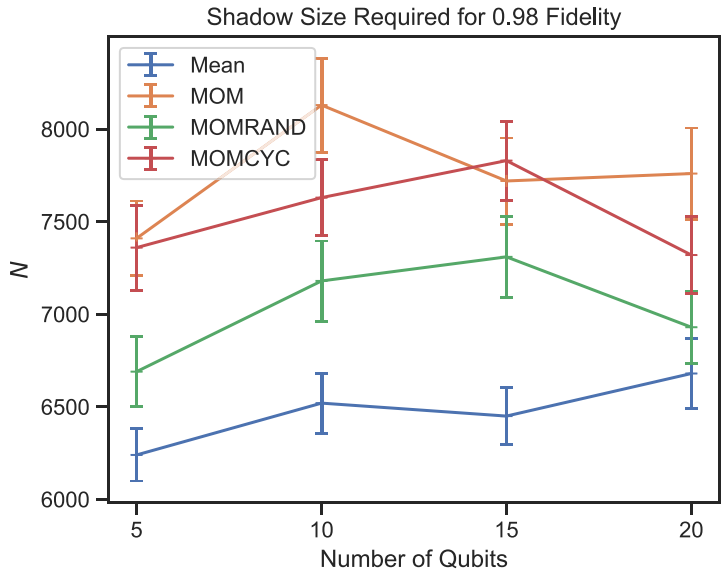

Figure 6. Number of snapshots needed to achieve 0.98 fidelity ($\varepsilon < 0.02$) for increasing system size, using various estimators, Averaged over 10 independent runs.

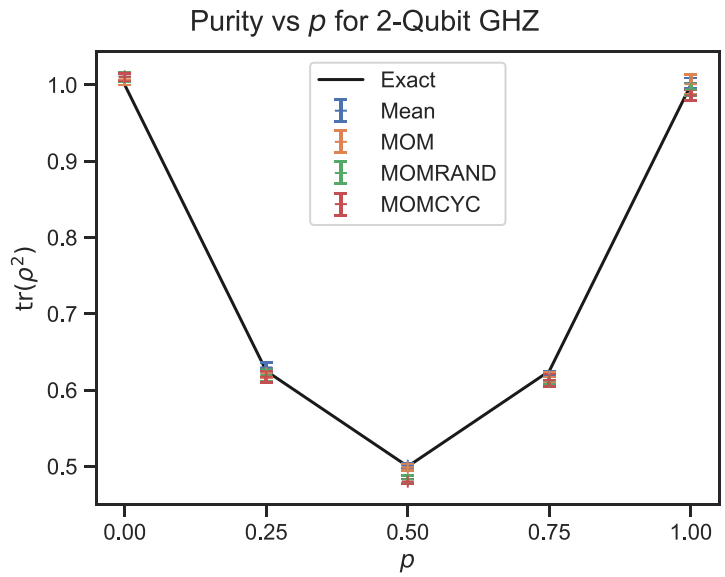

Figure 7. Predicted purity for the 2-qubits GHZ state using mean, Mom, MomRand, and MomCyc, for $p = 0, 0.25, 0.5, 0.75, 1$ and $N = 10\,000$ shadows.

Table 1. Number of samples required to achieve an average error of $\varepsilon = 0.1$, using the different bounds.

Bound	N
Mean	80×10^6
HKP	1.3×10^6
Original	58×10^3
New	38×10^3

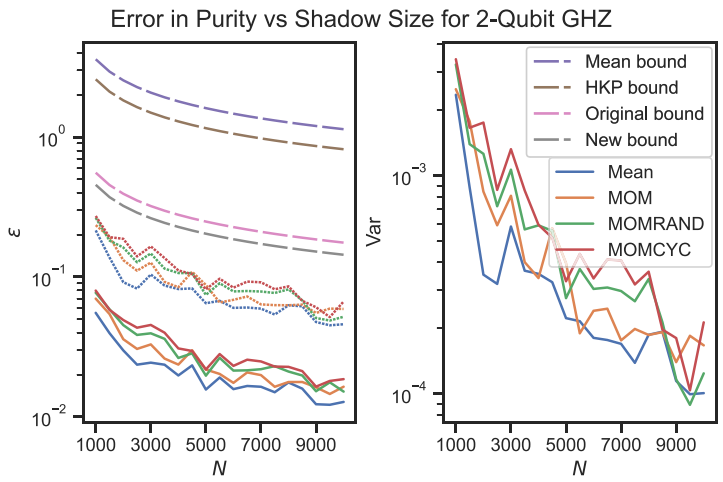

Figure 8. Error and variance of predictions with varying shadow size N using different estimators and their associated bounds, averaged over $p = 0, 0.25, 0.5, 0.75, 1$ and 10 independent runs.

an average error $M = 50,\ \varepsilon < 0.1,\ \delta = 0.001$. The loose bound used by Huang *et al* requires several orders of magnitude more shots than the tighter bounds by Minsker.

8. Conclusions

Since its inception, the classical shadows protocol has seen widespread adoption in areas such as the study of open quantum systems [88, 89] and variational quantum algorithms [57]. The high cost of running quantum circuits on noisy intermediate-scale quantum computers relative to classical compute power incentivizes the minimization of quantum sample complexity using more efficient classical post-processing techniques, as proposed in this paper.

In this paper, we have shown using numerical experiments that different estimators can improve the accuracy of estimated values, depending on the ensemble employed in the classical shadows protocol. Previous studies have likewise shown that estimator performance can vary depending on both the ensemble and the physical system under investigation [23]. Future work includes a more detailed investigation into how different estimators used in classical post-processing affect sample complexity. It would also be valuable to explore alternative variants of classical shadows, such as derandomized shadows [8], decision diagram-based methods [90], and shallow shadows [39].

Since our proposed changes pertain solely to the classical post-processing stage of the classical shadows protocol, our findings are immediately applicable to any experiment using the protocol, without requiring changes to the experimental setup, and can be readily applied to existing datasets.

9. Methods

9.1. Pauli measurements

The ground state of the Ising Hamiltonian was found using the infinite time-evolving block decimation algorithm [91, 92], in four steps of bond dimensions $\chi = 16, 32, 50, 100$ and time-steps $\tau = 0.1, 0.01, 0.001, 0.0001$ respectively. This was treated as the ground truth, and the matrix product state tensors were used to find the classical shadow. The bit-string $\hat{b} = (\hat{b}_1, \ldots, \hat{b}_r) \in \{0,1\}^r$ was found using methods proposed by Ferris and Vidal [93]. 50 000 snapshots were taken for 10 independent experiments. Finally, the various estimators were used on the shadows with $M = 50,\ \delta = 0.001$. For MomRand and MomCyc, the number of samples was chosen to be $m = 10\,\mathrm{kl}$.

To generate shadows of size between 1000 and 50 000, the appropriate number of snapshots was randomly sampled from the full shadow. The single-qubit Clifford unitary U_j applied to each qubit was stored and used to reconstruct each snapshot

$$\hat{\rho} = \bigotimes_{j=1}^{r} \left(3U_j^\dagger |\hat{b}_j\rangle\langle\hat{b}_j| U_j - \mathbb{1}\right). \tag{38}$$

9.2. Clifford measurements

The GHZ state was simulated efficiently using the stabilizer formalism, implemented using the Qiskit library. Using the global Clifford ensemble, it can be shown that the snapshots have the form

$$\hat{\rho} = (2^r + 1)\, U^\dagger |\hat{b}\rangle\langle\hat{b}| U - \mathbb{1}. \tag{39}$$

When estimating fidelity we can use this form to show:

$$F(\rho_0, \hat{\rho}) = (2^r + 1)\, |\langle\hat{b}|U|\mathrm{GHZ}^+(r)\rangle|^2 - 1. \tag{40}$$

The shadows of the 2-qubit GHZ state were found for $p = 0, 0.25, 0.5, 0.75, 1$ for 100 independent runs and shadow size $N = 50\,000$. To estimate the fidelity for $1000 \leqslant N \leqslant 50\,000$, the appropriate number of snapshots was sampled randomly from the full shadow.

Next, 10 shadows of the $r = 5, 10, 15, 20$ GHZ states were found for $p = 0$. For each shadow, 10 fidelity estimates were obtained for each value of $1000 \leqslant N \leqslant 50\,000$, sampled randomly, using each of the estimators, to obtain 100 runs. To find the minimum number of snapshots for 0.98 fidelity, a rolling average over 5 values of N was used to smooth out the data, then the minimum N was found for each run, before taking the average.

To estimate purity,

$$\mathrm{tr}\left(U_{\mathrm{swap}}\hat{\rho}_1 \otimes \hat{\rho}_2\right) = (2^r + 1)^2\, |\langle\hat{b}_1|U_1 U_2^\dagger|\hat{b}_2\rangle|^2 - 2\,(2^r + 1) + 2^r, \tag{41}$$

where U_{swap} is the SWAP operator. Instead of taking the MoM of disjoint subsets, we instead used each subset to find the U-statistic:

$$\hat{o}_i^{(q)}\left(\left\lfloor\frac{N}{k}\right\rfloor, 1\right) = \frac{1}{\left\lfloor\frac{N}{k}\right\rfloor\left(\left\lfloor\frac{N}{k}\right\rfloor - 1\right)} \sum_{\substack{u \neq v \\ u,v \in \left\{N\left(1-\frac{1}{q}\right)+1,\ldots,N\right\}}} \mathrm{tr}\left(O_i\, \hat{\rho}_u \otimes \hat{\rho}_v\right) \tag{42}$$

for $1 \leqslant q \leqslant k$. Then, we took the median of this set of U-statistics:

$$\hat{o}_i(N, k) = \mathrm{median}\left\{\hat{o}_i^{(1)}\left(\left\lfloor\frac{N}{k}\right\rfloor, 1\right), \ldots, o_i^{(k)}\left(\left\lfloor\frac{N}{k}\right\rfloor, 1\right)\right\}. \tag{43}$$

We define the 'mean' estimator to be the mean of the set of $\hat{o}_i^{(p)}(N, 1)$ from which the median is found for Mom.

To extend this estimator to MomRand and MomCyc, we treated this U-statistic as the mean in the first step of the algorithm (line 4 of algorithms 3 and 4), then grouped them to form $\bar{Z}_J$. This was used to estimate the purity $\mathrm{tr}(\rho^2) = \mathrm{tr}(U_{\mathrm{swap}}\rho \otimes \rho)$ of the $n = 2$ noisy GHZ state for $p = 0, 0.25, 0.5, 0.75, 1$ from 10 shadows for $1000 \leqslant N \leqslant 10\,000$.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgment

The authors want to thank Hsin-Yuan Huang for sharing his code. This research is supported by the National Research Foundation, Singapore, and the Agency for Science, Technology and Research (A*STAR), Singapore, under its Quantum Engineering Programme (NRF2021-QEP2-02-P03); A*STAR C230917003; and A*STAR under the Central Research Fund (CRF) Award for Use-Inspired Basic Research (UIBR) and the Quantum Innovation Centre (Q.InC) Strategic Research and Translational Thrust.

Conflict of interests

All authors declare no financial or non-financial competing interests.

Author contributions

D E K and S T G conceptualized the project. Analyses and numerical experiments were led by W F and discussed with D E K, S T G and J F K The manuscript was written by W F, D E K, S T G, and J FK.

Appendix A. MoM bound

We present a proof for the bound

$$P[|\hat{\mu}_N - \mathbb{E}(X)| \geqslant \varepsilon] \leqslant 2\mathrm{e}^{-k/2}, \tag{A1}$$

where $\varepsilon = \sqrt{4e^2\sigma^2 k/N}$ and N is the total number of data points for an MoM estimator $\hat{\mu}_N$, adapted from Devroye *et al* [76].

Given random variables $X_1, \ldots, X_N$, we split them into k disjoint sets of size $b = \lfloor N/k \rfloor$, and find their mean. If $|\hat{\mu}_N - \mathbb{E}(X)| \geqslant \varepsilon$, this means more than half of the means $\bar{X}_i^{(b)} = b^{-1}\sum_{j\in J_i} X_j$, $|J_i| = b$ deviates from $\mathbb{E}(X)$ by more than ε. Let $Y_i^{(b)} = \bar{X}_i^{(b)} - \mathbb{E}(X)$:

$$\sum_{i=1}^{k} I\left\{|Y_i^{(b)}| \geqslant \varepsilon\right\} \geqslant \frac{k}{2}. \tag{A2}$$

As $Y_i^{(b)}$ are i.i.d. let $\mathrm{Var}[X_i] = \sigma^2 < \infty$ and $\mathrm{Var}\left[Y_i^{(b)}\right] = k\sigma^2/N$ so that by Chebyshev's inequality,

$$P\left(|Y_i^{(b)}| \geqslant \varepsilon\right) \leqslant \frac{k\sigma^2}{N\varepsilon^2}. \tag{A3}$$

$|Y^{(l)}| \geqslant \varepsilon$ is a Bernoulli event with probability $\leqslant k\sigma^2/N\varepsilon^2$. This implies

$$P\left(\sum_{i=1}^{k} I\left\{|Y_i^{(b)}| \geqslant \varepsilon\right\} = r\right) \leqslant \binom{k}{r}\left(\frac{k\sigma^2}{N\varepsilon^2}\right)^r\left(1 - \frac{k\sigma^2}{N\varepsilon^2}\right)^{k-r}. \tag{A4}$$

Putting everything together,

$$\begin{aligned} P(|\hat{\mu}_N - \mathbb{E}(X)| \geqslant \varepsilon) &\leqslant P\left(\sum_{i=1}^{k} I\left\{|Y_i^{(b)}| \geqslant \varepsilon\right\} \geqslant \frac{k}{2}\right) && \text{(A5)}\\ &\leqslant \sum_{r=\lceil k/2\rceil}^{k}\binom{k}{r}\left(\frac{k\sigma^2}{N\varepsilon^2}\right)^r\left(1 - \frac{k\sigma^2}{N\varepsilon^2}\right)^{k-r} && \text{(A6)}\\ &\leqslant \left(\frac{k\sigma^2}{N\varepsilon^2}\right)^{\lceil k/2\rceil}\sum_{r=\lceil k/2\rceil}^{k}\binom{k}{r} && \text{(A7)}\\ &\leqslant \left(\frac{k\sigma^2}{N\varepsilon^2}\right)^{\lceil k/2\rceil}\sum_{r=0}^{k}\binom{k}{r} && \text{(A8)}\\ &= \left(\frac{k\sigma^2}{N\varepsilon^2}\right)^{\lceil k/2\rceil} 2^k. && \text{(A9)} \end{aligned}$$

Choosing $N = 4e^2\sigma^2 k/\varepsilon^2$ ensures that

$$\begin{aligned} P(|\hat{\mu}_N - \mathbb{E}(X)| \geqslant \varepsilon) &\leqslant \mathrm{e}^{-k} && \text{(A10)}\\ &\leqslant 2\mathrm{e}^{-k/2}, && \text{(A11)} \end{aligned}$$

thereby obtaining the bound given in equation (A1).

Appendix B. Bounds for Gaussian distribution

In the simulations, the mean performed better despite having worse bounds than the MoM estimators when using Chebyshev's inequality (equation (8)). However, for large shadow size, the distribution of the estimator $\hat{o}_i(N,1)$ tends towards a Gaussian distribution, with $P(|\hat{\mu}-\mu| \geqslant \varepsilon) \leqslant \exp\left(-\frac{N\varepsilon^2}{2\sigma^2}\right) \equiv \frac{\delta}{M}$, giving an explanation for the better-than-expected performance of the mean. Here, we include comparisons between this tighter bound under this normality assumption, and the bound in theorem 2.

This 'Mean Gaussian bound' is tighter than theorem 2 and obeyed by the mean for the Ising model (figure 9), 2-qubit GHZ (figure 10), 5, 10, 15, 20-qubit GHZ (figure 11), and GHZ purity (figure 12). In figure 11, the tighter bound is obeyed by the mean and MomRand, but not Mom and MomCyc. These results demonstrate the validity of the normality assumption for classical shadows.

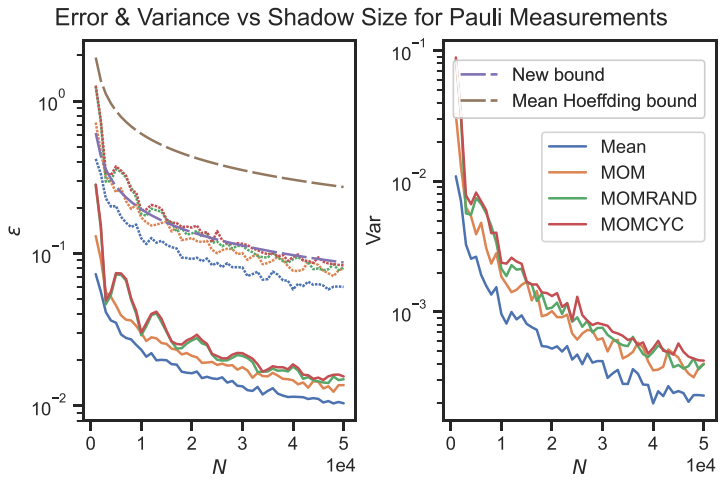

Figure 9. Error and variance vs shadow size for the Ising model using Pauli measurements when estimating the correlator, as in figure 2. We compare the bounds under the assumption that the distribution of the mean of the shadow is Gaussian, and the bound by Minsker in theorem 2.

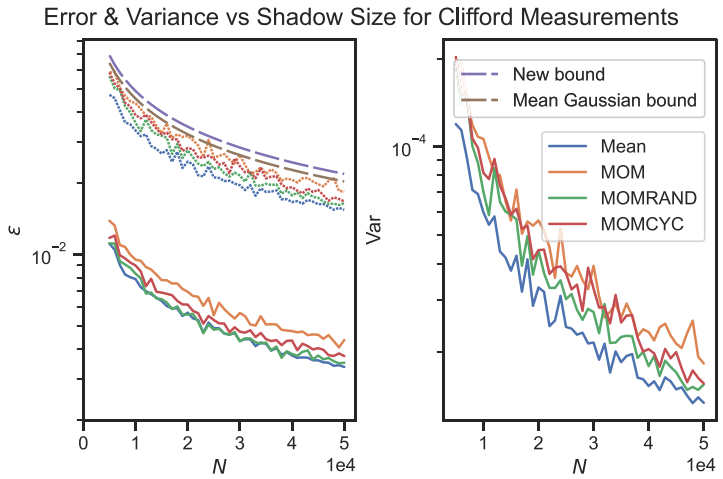

Figure 10. Error and variance vs shadow size for the 2-qubit GHZ state using Clifford measurements when estimating fidelity, as in figure 4, with comparison of bounds under the normality assumption.

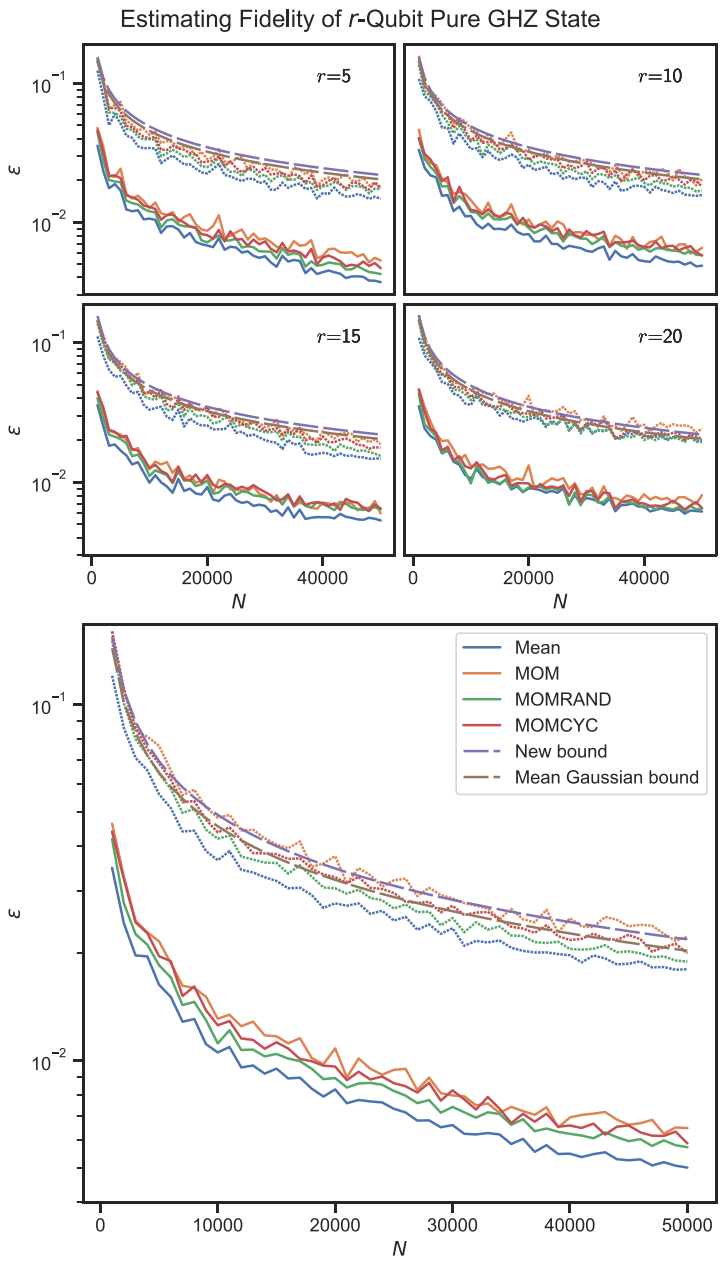

Figure 11. Error and variance vs shadow size for the 5, 10, 15, 20-qubit GHZ states using Clifford measurements when estimating fidelity, as in figure 5, with comparison of bounds under the normality assumption.

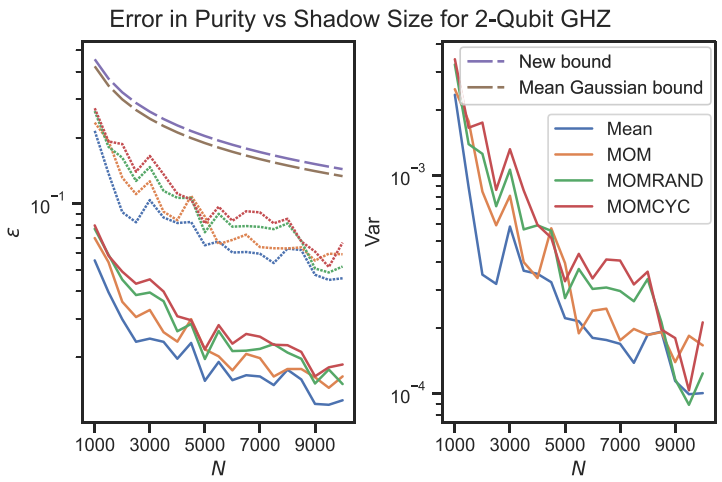

Figure 12. Error and variance vs shadow size for the 2-qubit GHZ state using Clifford measurements when estimating the purity, as in figure 8, with comparison of bounds under the normality assumption.

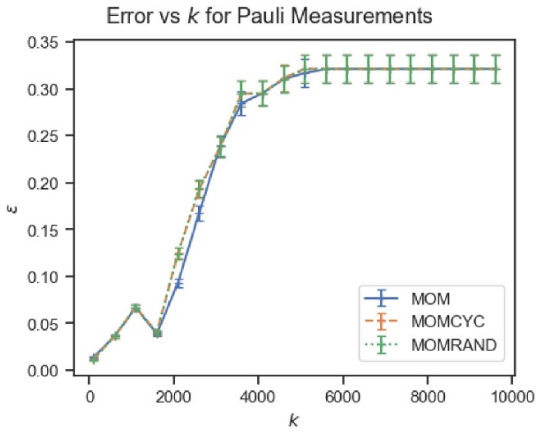

Figure 13. Error vs group size k for predicting the 2-point correlator for 50 qubits, using different estimators, averaged over 10 independent runs.

Appendix C. Performance of estimators for Pauli observables

The performance of the estimators may be explained by the different values of k used for each, as suggested by analysis by Huang *et al* [8]. The mean is equivalent to Mom with $k = 1$. For Mom, $k = 43$. For MomRand and MomCyc, $k = [65, 219]$ when $N = [1000, 50\,000]$. k changes the number of data points $\mathrm{tr}(O_i \hat{\rho}_j)$ that the final estimate is averaged over, corresponding to $\lfloor N/k \rfloor$.

Figure 13 shows that, rather than the type of estimator used, the error depends most heavily on the number of groups k. The fewer data points per group and the higher k used, the higher the error and variance of the estimator.

In the aforementioned paper, Huang *et al* showed that the failure probability of the estimator for Pauli observables and measurements is exponentially suppressed by the 'hit count' of the estimator. For a system of r qubits, we perform N Pauli measurements $p_i \in \{X, Y, Z\}^r$, to estimate M observables O_j, with each measurement corresponding to a snapshot $\hat{\rho}_i$. For Pauli observables, equation (38) implies that $\mathrm{tr}(\hat{\rho}_i O_j)$ is non-zero if the measurement 'hits' the observable, $O_j \triangleright p_i$. That is, by substituting $\{X, Y, Z\}$ in p_i with I, we can obtain O_j if $O_j \triangleright p_i$. Huang *et al* showed that the performance of the estimator improves exponentially with the number of hits. By splitting the data set into groups, the hit count in each group is reduced, exponentially increasing the failure probability of each $\bar{X}_i$ for Mom and $\bar{Z}_J$ for MomRand and MomCyc.

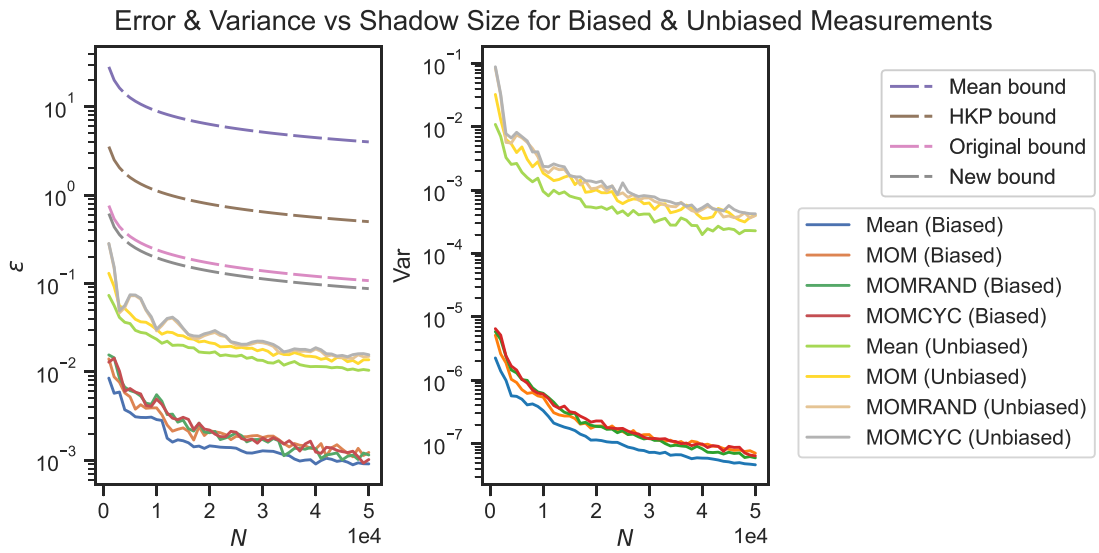

Figure 14. Error and variance vs shadow size for biased and unbiased classical shadows when estimating the 2-point correlator for 50 qubits.

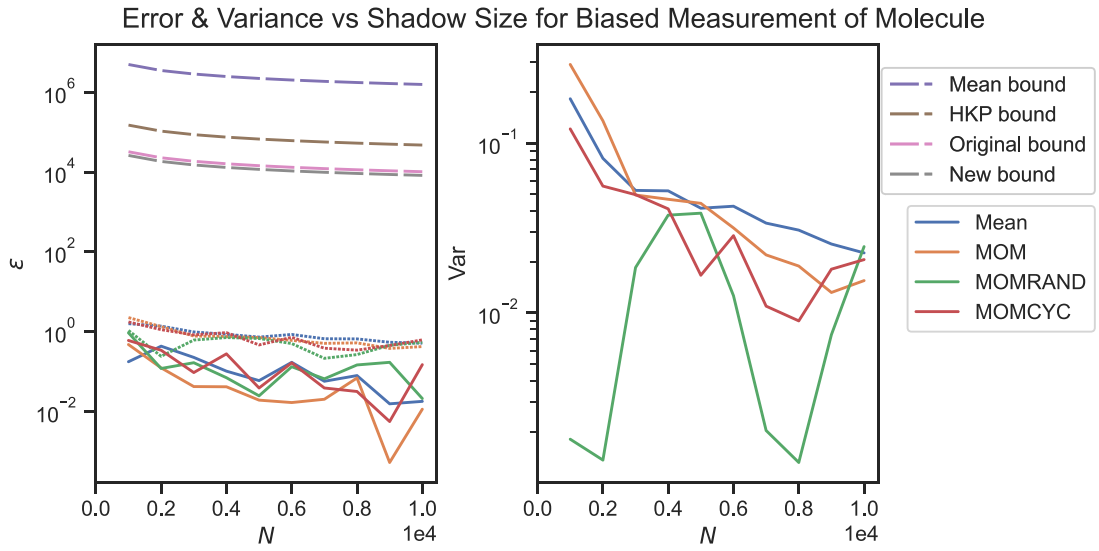

Figure 15. Error and variance vs shadow size for biased measurements of a H_2O molecule.

Appendix D. Application to locally biased classical shadows (LBCS)

We repeated the Ising experiment using LBCS [11]. In this case, as the observables were composed solely of '*Z*' Paulis, the optimized single-qubit Clifford ensemble comprised only the '*I*' operator. The results are shown in figure 14. The relative performance of the different estimators matched that of the unbiased case.

LBCS was also applied to a H_2O molecule. From the Jordan–Wigner transformation, the ground-state energy was estimated from the resulting 14-qubit system, and the various estimators were applied. The results are shown in figure 15.

Appendix E. Time comparison

For the Ising and GHZ systems, we compared the time taken to run the various estimators. These are shown in figures 16 and 17 respectively. As MomCyc and MomRand sample a subset of possible values, the number of samples can be varied. Here, we used the same values as in the rest of the experiments. The modified estimators have a higher classical computational cost compared to 'Mean' and Mom. Nevertheless, quantum operations dominate the total resource budget while the estimator is executed entirely on classical hardware.

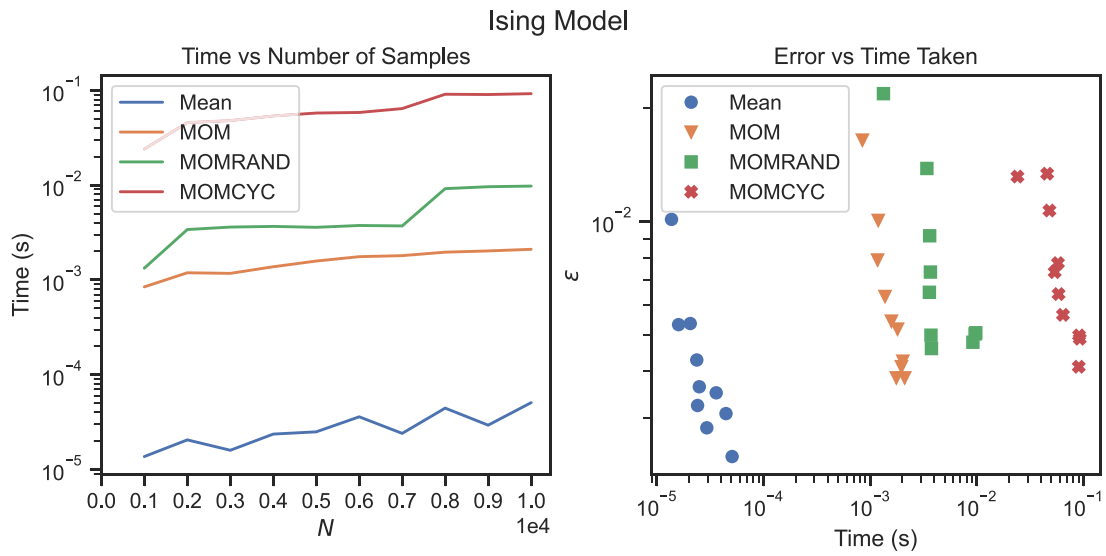

Figure 16. Time vs shadow size, and error vs time for various estimators applied to the Ising model.

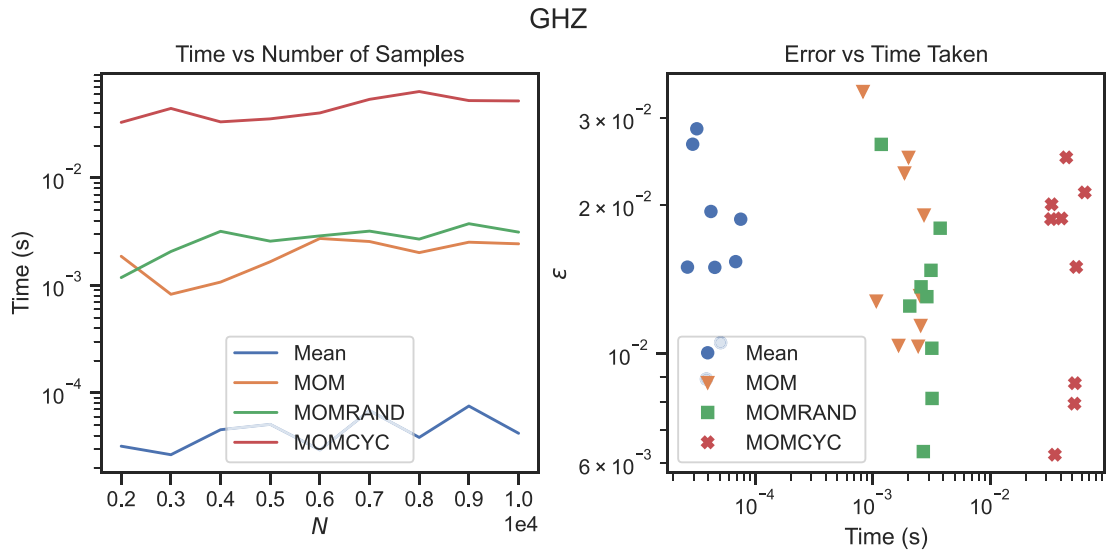

Figure 17. Time vs shadow size, and error vs time for various estimators applied to the GHZ system.

Hence, the runtime of the estimators remains a secondary consideration as long as it remains within reasonable bounds.

ORCID iDs

Winston Fu https://orcid.org/0009-0000-1383-6454
Dax Enshan Koh https://orcid.org/0000-0002-8968-591X
Siong Thye Goh https://orcid.org/0000-0001-7563-0961
Jian Feng Kong https://orcid.org/0000-0001-5980-4140

References

[1] Huang H-Y, Kueng R and Preskill J 2020 Predicting many properties of a quantum system from very few measurements *Nat. Phys.* **16** 1050
[2] Aaronson S 2018 Shadow tomography of quantum states *Proc. 50th Annual ACM SIGACT Symp. on Theory of Computing* (Association for Computing Machinery) STOC 2018, pp 325–38
[3] Guţă M, Kahn J, Kueng R and Tropp J A 2020 Fast state tomography with optimal error bounds *J. Phys. A: Math. Theor.* **53** 204001
[4] Sugiyama T, Turner P S and Murao M 2013 Precision-guaranteed quantum tomography *Phys. Rev. Lett.* **111** 160406

[5] Haah J, Harrow A, Ji Z, Wu X and Yu N 2017 Sample-optimal tomography of quantum states *IEEE Trans. Inf. Theory* **63** 5628
[6] O'Donnell R and Wright J 2016 Efficient quantum tomography *Proc. 48th Annual ACM Symp. on Theory of Computing* pp 899–912
[7] Lukens J M, Law K J and Bennink R S 2021 A Bayesian analysis of classical shadows *npj Quantum Inf.* **7** 113
[8] Huang H-Y, Kueng R and Preskill J 2021 Efficient estimation of Pauli observables by derandomization *Phys. Rev. Lett.* **127** 030503
[9] Hadfield C 2021 Adaptive Pauli shadows for energy estimation (arXiv:2105.12207)
[10] Zhao A, Rubin N C and Miyake A 2021 Fermionic partial tomography via classical shadows *Phys. Rev. Lett.* **127** 110504
[11] Hadfield C, Bravyi S, Raymond R and Mezzacapo A 2022 Measurements of quantum Hamiltonians with locally-biased classical shadows *Commun. Math. Phys.* **391** 951
[12] Zhou Y and Liu Z 2024 A hybrid framework for estimating nonlinear functions of quantum states *npj Quantum Inf.* **10** 62
[13] Akhtar A A, Hu H-Y and You Y-Z 2023 Scalable and flexible classical shadow tomography with tensor networks *Quantum* **7** 1026
[14] Hu H-Y, Choi S and You Y-Z 2023 Classical shadow tomography with locally scrambled quantum dynamics *Phys. Rev. Res.* **5** 023027
[15] Shivam S, von Keyserlingk C W and Sondhi S L 2023 On classical and hybrid shadows of quantum states *SciPost Phys.* **14** 094
[16] Wan K, Huggins W J, Lee J and Babbush R 2023 Matchgate shadows for fermionic quantum simulation *Commun. Math. Phys.* **404** 629
[17] Innocenti L, Lorenzo S, Palmisano I, Albarelli F, Ferraro A, Paternostro M and Palma G M 2023 Shadow tomography on general measurement frames *PRX Quantum* **4** 040328
[18] Ippoliti M, Li Y, Rakovszky T and Khemani V 2023 Operator relaxation and the optimal depth of classical shadows *Phys. Rev. Lett.* **130** 230403
[19] Helsen J and Walter M 2023 Thrifty shadow estimation: reusing quantum circuits and bounding tails *Phys. Rev. Lett.* **131** 240602
[20] Grier D, Pashayan H and Schaeffer L 2024 Principal eigenstate classical shadows *Proc. 37th Conf. on Learning Theory* (*Proc. of Machine Learning Research*) vol 247, ed S Agrawal and A Roth (PMLR) pp 2122–65
[21] Sauvage F and Larocca M 2024 Classical shadows with symmetries (arXiv:2408.05279)
[22] Bu K, Koh D E, Garcia R J and Jaffe A 2024 Classical shadows with Pauli-invariant unitary ensembles *npj Quantum Inf.* **10** 6
[23] Grier D, Pashayan H and Schaeffer L 2024 Sample-optimal classical shadows for pure states *Quantum* **8** 1373
[24] West M, Mele A A, Larocca M and Cerezo M 2024 Real classical shadows *J. Phys. A: Math. Theor.* **58** 245304
[25] Ippoliti M 2024 Classical shadows based on locally-entangled measurements *Quantum* **8** 1293
[26] King R, Gosset D, Kothari R and Babbush R 2025 Triply efficient shadow tomography *PRX Quantum* **6** 010336
[27] De Palma G, Klein T and Pastorello D 2024 Classical shadows meet quantum optimal mass transport *J. Math. Phys.* **65** 092201
[28] Becker S, Datta N, Lami L and Rouze C 2024 Classical shadow tomography for continuous variables quantum systems *IEEE Trans. Inf. Theory* **70** 3427
[29] Hearth S N, Flynn M O, Chandran A and Laumann C R 2024 Efficient local classical shadow tomography with number conservation *Phys. Rev. Lett.* **133** 060802
[30] Cai Z, Chapman A, Jnane H and Koczor B 2025 Biased estimator channels for classical shadows *Phys. Rev.* A **111** L030402
[31] Chen D T S, Saleem Z H and Perlin M A 2024 Quantum circuit cutting for classical shadows *ACM Trans. Quantum Comput.* **5(2)** 13
[32] Heyraud V, Chomet H and Tilly J 2025 Unified framework for matchgate classical shadows *npj Quantum Inf.* **11** 65
[33] Akhtar A A, Anand N, Marshall J and You Y-Z 2024 Dual-unitary classical shadow tomography (arXiv:2404.01068)
[34] Wang Y Quantum advantage via efficient post-processing on qudit shadow tomography (arXiv:2408.16244)
[35] Wu Y, Wang C, Yao J, Zhai H, You Y-Z and Zhang P 2024 Contractive unitary and classical shadow tomography (arXiv:2412.01850)
[36] Zhou T-G and Zhang P 2024 Efficient classical shadow tomography through many-body localization dynamics *Quantum* **8** 1467
[37] Cioli R, Ercolessi E, Ippoliti M, Turkeshi X and Piroli L 2024 Approximate inverse measurement channel for shallow shadows (arXiv:2407.11813)
[38] Liu Q, Li Z, Yuan X, Zhu H and Zhou Y 2024 Auxiliary-free replica shadow estimation (arXiv:2407.20865)
[39] Bertoni C, Haferkamp J, Hinsche M, Ioannou M, Eisert J and Pashayan H 2024 Shallow shadows: expectation estimation using low-depth random Clifford circuits *Phys. Rev. Lett.* **133** 020602
[40] Anselmetti G L R, Degroote M, Moll N, Santagati R and Streif M 2024 Classical optimisation of reduced density matrix estimations with classical shadows using N-representability conditions under shot noise considerations (arXiv:2411.18430)
[41] Mao C, Yi C and Zhu H 2024 Qudit shadow estimation based on the Clifford group and the power of a single magic gate *Phys. Rev. Lett.* **134** 160801
[42] Caprotti A, Morris J and Dakić B 2024 Optimizing quantum tomography via shadow inversion *Phys. Rev. Res.* **6** 033301
[43] Zhang Z-J, Nakaji K, Choi M and Aspuru-Guzik A 2023 A composite measurement scheme for efficient quantum observable estimation (arXiv:2305.02439)
[44] Huang B, Chen Y-T, Gupt B, Suchara M, Tran A, McArdle S and Galli G 2024 Evaluating a quantum-classical quantum Monte Carlo algorithm with matchgate shadows *Phys. Rev. Res.* **6** 043063
[45] Avdic I and Mazziotti D A 2024 Enhanced shadow tomography of molecular excited states via the enforcement of *N*-representability conditions by semidefinite programming *Phys. Rev. A* **110** 052407
[46] Wang Y, Avdic I and Mazziotti D A 2024 Shadow ansatz for the many-fermion wave function in scalable molecular simulations on quantum computing devices (arXiv:2408.11026)
[47] Avdic I and Mazziotti D A 2024 Fewer measurements from shadow tomography with *N*-representability conditions *Phys. Rev. Lett.* **132** 220802
[48] Huang H-Y, Kueng R, Torlai G, Albert V V and Preskill J 2022 Provably efficient machine learning for quantum many-body problems *Science* **377** eabk3333
[49] Lewis L, Huang H-Y, Tran V T, Lehner S, Kueng R and Preskill J 2024 Improved machine learning algorithm for predicting ground state properties *Nat. Commun.* **15** 895
[50] Jerbi S, Gyurik C, Marshall S C, Molteni R and Dunjko V 2024 Shadows of quantum machine learning *Nat. Commun.* **15** 5676
[51] Zhao A and Miyake A 2024 Group-theoretic error mitigation enabled by classical shadows and symmetries *npj Quantum Inform.* **10** 57
[52] Seif A, Cian Z-P, Zhou S, Chen S and Jiang L 2023 Shadow distillation: quantum error mitigation with classical shadows for near-term quantum processors *PRX Quantum* **4** 010303
[53] Chen Y, Yu Z, Zhu C and Wang X 2023 Efficient information recovery from Pauli noise via classical shadow (arXiv:2305.04148)
[54] Elben A *et al* 2020 Mixed-state entanglement from local randomized measurements *Phys. Rev. Lett.* **125** 200501
[55] Neven A *et al* 2021 Symmetry-resolved entanglement detection using partial transpose moments *npj Quantum Inf.* **7** 152
[56] Rath A, Branciard C, Minguzzi A and Vermersch B 2021 Quantum Fisher information from randomized measurements *Phys. Rev. Lett.* **127** 260501

[57] Boyd G and Koczor B 2022 Training variational quantum circuits with CoVaR: covariance root finding with classical shadows *Phys. Rev. X* **12** 041022
[58] Sack S H, Medina R A, Michailidis A A, Kueng R and Serbyn M 2022 Avoiding barren plateaus using classical shadows *PRX Quantum* **3** 020365
[59] Basheer A, Feng Y, Ferrie C and Li S 2023 Alternating layered variational quantum circuits can be classically optimized efficiently using classical shadows *Proc. AAAI Conf. on Artificial Intelligence* vol 37 pp 6770–8
[60] Nakaji K, Endo S, Matsuzaki Y and Hakoshima H 2023 Measurement optimization of variational quantum simulation by classical shadow and derandomization *Quantum* **7** 995
[61] Garcia R J, Zhou Y and Jaffe A 2021 Quantum scrambling with classical shadows *Phys. Rev. Res.* **3** 033155
[62] Helsen J, Ioannou M, Kitzinger J, Onorati E, Werner A, Eisert J and Roth I 2023 Shadow estimation of gate-set properties from random sequences *Nat. Commun.* **14** 5039
[63] White G A L, Modi K and Hill C D 2023 Filtering crosstalk from bath non-Markovianity via spacetime classical shadows *Phys. Rev. Lett.* **130** 160401
[64] Ippoliti M and Khemani V 2024 Learnability transitions in monitored quantum dynamics via eavesdropper's classical shadows *PRX Quantum* **5** 020304
[65] Ruiz Guzman E A and Lacroix D 2024 Restoring symmetries in quantum computing using classical shadows *Eur. Phys. J. A* **60** 112
[66] Conrad J, Eisert J and Flammia S T 2024 Chasing shadows with Gottesman-Kitaev-Preskill codes (arXiv:2411.00235)
[67] Chen S, Yu W, Zeng P and Flammia S T 2021 Robust shadow estimation *PRX Quantum* **2** 030348
[68] Koh D E and Grewal S 2022 Classical shadows with noise *Quantum* **6** 776
[69] Brieger R, Heinrich M, Roth I and Kliesch M 2025 Stability of classical shadows under gate-dependent noise *Phys. Rev. Lett.* **134** 090801
[70] Nguyen H C 2023 Shadow tomography with noisy readouts (arXiv:2310.17328)
[71] Wu B and Koh D E 2024 Error-mitigated fermionic classical shadows on noisy quantum devices *npj Quantum Inf.* **10** 39
[72] Rozon P-G, Bao N and Agarwal K 2024 Optimal twirling depth for classical shadows in the presence of noise *Phys. Rev. Lett.* **133** 130803
[73] Jnane H, Steinberg J, Cai Z, Nguyen H C and Koczor B 2024 Quantum error mitigated classical shadows *PRX Quantum* **5** 010324
[74] Farias R M S, Peddinti R D, Roth I and Aolita L 2025 Robust ultra-shallow shadows *Quantum Sci. Technol.* **10** 025044
[75] Onorati E, Kitzinger J, Helsen J, Ioannou M, Werner A, Roth I and Eisert J 2024 Noise-mitigated randomized measurements and self-calibrating shadow estimation (arXiv:2403.04751)
[76] Devroye L, Lerasle M, Lugosi G and Oliveira R I 2016 Sub-Gaussian mean estimators *Ann. Stat.* **44** 2695
[77] Lugosi G and Mendelson S 2019 Mean estimation and regression under heavy-tailed distributions: a survey *Found. Comput. Math.* **19** 1145
[78] Struchalin G, Zagorovskii Y A, Kovlakov E, Straupe S and Kulik S 2021 Experimental estimation of quantum state properties from classical shadows *PRX Quantum* **2** 010307
[79] Zhang T, Sun J, Fang X-X, Zhang X-M, Yuan X and Lu H 2021 Experimental quantum state measurement with classical shadows *Phys. Rev. Lett.* **127** 200501
[80] Levy R, Luo D and Clark B K 2024 Classical shadows for quantum process tomography on near-term quantum computers *Phys. Rev. Res.* **6** 013029
[81] Dutt A, Kirby W, Raymond R, Hadfield C, Sheldon S, Chuang I L and Mezzacapo A 2023 Practical benchmarking of randomized measurement methods for quantum chemistry Hamiltonians (arXiv:2312.07497)
[82] Minsker S 2023 Efficient median of means estimator *Proc. 36th Conf. on Learning Theory* (*Proc. Machine Learning Research*) vol 195, ed G Neu and L Rosasco (PMLR) pp 5925–33
[83] Minsker S 2023 U-statistics of growing order and sub-Gaussian mean estimators with sharp constants *Math. Stat. Learn.* **7** 1
[84] Lee A J 2019 *U-Statistics: Theory and Practice* (Routledge)
[85] Greenberger D M, Horne M A and Zeilinger A 1989 Going beyond Bell's theorem *Bell's Theorem, Quantum Theory and Conceptions of the Universe* ed M M Kafatos (Springer) pp 69–72
[86] Minsker S 2019 Distributed statistical estimation and rates of convergence in normal approximation *Electron. J. Stat.* **13** 5213
[87] Drakakis K 2009 A review of the available construction methods for Golomb rulers *Adv. Math. Commun.* **3** 235
[88] Myerson-Jain N, Hughes T L and Xu C 2025 Decoherence through Ancilla Anyon Reservoirs *Phys. Rev. Lett.* **134** 096503
[89] Lessa L A, Ma R, Zhang J-H, Bi Z, Cheng M and Wang C 2025 Strong-to-weak spontaneous symmetry breaking in mixed quantum states *PRX Quantum* **6** 010344
[90] Hillmich S, Hadfield C, Raymond R, Mezzacapo A and Wille R 2021 Decision diagrams for quantum measurements with shallow circuits *2021 IEEE Int. Conf. on Quantum Computing and Engineering (QCE)* pp 24–34
[91] Vidal G 2004 Efficient simulation of one-dimensional quantum many-body systems *Phys. Rev. Lett.* **93** 040502
[92] Vidal G 2007 Classical simulation of infinite-size quantum lattice systems in one spatial dimension *Phys. Rev. Lett.* **98** 070201
[93] Ferris A J and Vidal G 2012 Perfect sampling with unitary tensor networks *Phys. Rev. B* **85** 165146