Letter

Biased estimator channels for classical shadows

Zhenyu Cai,[1,2] Adrian Chapman,[1,*] Hamza Jnane,[1,2] and Bálint Koczor[3,1,2,†]
[1]Department of Materials, University of Oxford, Parks Road, Oxford OX1 3PH, United Kingdom
[2]Quantum Motion, 9 Sterling Way, London N7 9HJ, United Kingdom
[3]Mathematical Institute, University of Oxford, Woodstock Road, Oxford OX2 6GG, United Kingdom

(Received 1 March 2024; accepted 26 February 2025; published 21 March 2025)

Extracting classical information from quantum systems is of fundamental importance, and classical shadows allow us to extract a large amount of information using relatively few measurements. Conventional shadow estimators are unbiased and thus agree with the true mean in expectation. In this Letter, we consider a biased scheme, intentionally introducing a bias in the expectation value by rescaling the conventional classical-shadow estimators to reduce the error in the finite-sample regime. The approach is straightforward to implement and requires no quantum resources. We analytically prove average-case as well as worst- and best-case scenarios, and rigorously prove that it is, in principle, always worth biasing the estimators. We illustrate our approach in a quantum simulation task of a 12-qubit spin-ring problem and demonstrate how estimating expected values of nonlocal perturbations can be significantly more efficient using our biased scheme.

DOI: 10.1103/PhysRevA.111.L030402

Introduction. We are experiencing rapid progress in the development of quantum hardware but also in theoretical advances [1–6]. Any quantum computational scheme needs to extract classical information from a quantum device. However, this requires multiple repetitions of the experiment due to fundamental limitations posed by quantum mechanics. Each observation of the system collapses its quantum state, preventing one from extracting further information. For this reason, one must extract classical information through the use of statistical estimators [7–10]. It is thus an exciting and fundamentally important challenge to extract classical information, for example, expected values of observables, such that the statistical uncertainty (i.e., shot noise due to having access to only finite samples or circuit repetitions) is minimised. Classical shadows [11] allow one to predict many such properties of quantum states from very few samples (circuit repetitions) and the statistical uncertainty due to shot noise can be rigorously bounded. The approach yields unbiased estimators, i.e., estimators whose expected value agrees with the true value, e.g., the true expected value of an observable.

In the present Letter, we explore the possibility of intentionally introducing a small bias into the estimators, i.e., for a fixed such bias, even an infinite number of samples would not yield the true expected value, but in return the statistical uncertainties are significantly reduced in the finite-sample regime. A significant practical advantage of our biasing scheme is that it is implemented completely in classical post processing: One collects a number of samples as classical shadows using a quantum computer, and in postprocessing predicts many properties of the quantum state. Our approach only slightly modifies this prediction stage whereby the mean estimators are simply scaled down by a factor quantified by a bias parameter ε. We choose this parameter to depend on the number of samples in the experiment such that the resulting estimator remains consistent, converging to the true value in the infinite-sample limit. The approach is thus more general than the standard, unbiased shadow techniques [11–20] which are then contained as a special case at $\varepsilon = 0$.

We comprehensively and rigorously characterize the performance of our biased scheme and find, somewhat surprisingly, that biasing our estimators is always worthwhile assuming that an optimal bias parameter ε that is specific for the particular estimator is known. After briefly recapitulating classical shadows, we start by mathematically deriving the average-case gain of biased shadow tomography when the aim is to predict local density matrices. We then analytically characterize both the worst- and best-case scenarios of our approach when the aim is to predict expected values of observables, and we provide explicit expressions for the optimal bias parameter ε showing it depends only on the theoretical mean value. We then argue that the biased scheme is not specific to classical shadows but can generally be applied to any estimation scheme, e.g., by directly estimating Pauli expected values.

Classical shadows. A classical shadow is a description of a quantum state that can be classically efficiently stored and manipulated, enabling one to bypass the computationally hard task of reconstructing the full density matrix. To construct a classical shadow for an n-qubit quantum state ρ, we repeat the following evolution-measurement process. We sample a random unitary U_i from a suitable distribution $\mathcal{U}$ (Pauli and

*Contact author: adrian.chapman@materials.ox.ac.uk
†Contact author: koczor@maths.ox.ac.uk

Clifford distributions are typical), apply this unitary to ρ, and measure the resulting state $U_i \rho U_i^\dagger$ in the computational basis, yielding a bit string $\mathbf{b} \in \{0, 1\}^n$. We store the index of the unitary and the measurement outcome as a composite index $\ell = (i, \mathbf{b})$.

One considers the *process channel* $\mathcal{M}$ as the average over the previously fixed distribution $\mathcal{U}$ as

$$\mathcal{M}(\rho) = \sum_{\mathbf{b}\in\{0,1\}^n} \int_{\mathcal{U}} dU \, \langle \mathbf{b}| U^\dagger \rho U |\mathbf{b}\rangle \, U|\mathbf{b}\rangle\langle\mathbf{b}|U^\dagger. \quad (1)$$

Each composite index ℓ then identifies a classical snapshot $\hat{\rho}_\ell$ such that

$$\hat{\rho}_\ell = \widetilde{\mathcal{M}}(U_i|\mathbf{b}\rangle\langle\mathbf{b}|U_i^\dagger), \quad (2)$$

which is an unbiased estimator of ρ. Here, $\widetilde{\mathcal{M}}$ is the *estimator channel*, which is implicitly defined by this equation. In practice, one repeats the above procedure N_s times, generating the *classical shadow* of ρ as the collection $S(\rho, N_s) = \{\hat{\rho}_1, \ldots, \hat{\rho}_{N_s}\}$. When discussing averages over ℓ, we drop the corresponding dependence on $\hat{\rho}$ and understand this estimator to be a function of ℓ.

Main result: Biased shadow estimators. A channel $\widetilde{\mathcal{M}}$ that yields an unbiased estimator for the density matrix satisfies the condition $(\widetilde{\mathcal{M}} \circ \mathcal{M})[\rho] = \rho$ for all states ρ, and thus guarantees $\widetilde{\mathcal{M}} = \mathcal{M}^{-1}$ as

$$\rho = \mathbb{E}_{\mathcal{U},\mathbf{b}}[\hat{\rho}] = (\widetilde{\mathcal{M}} \circ \mathcal{M})[\rho], \ \forall \rho. \quad (3)$$

However, in the present work we focus on constructing a biased estimator which does not necessarily satisfy the above property but in return allows us to reduce the variance of the estimator.

To simplify our presentation, we illustrate our results on the simple uniform ensemble over n-qubit product Clifford rotations $\mathcal{U} = \mathcal{C}_1^{\times n}$. This is equivalent to uniformly sampling a local Pauli basis in which to measure each qubit and thus $\hat{\rho}_\ell$ can be decomposed as a tensor product [11],

$$\hat{\rho}_\ell = \bigotimes_{j=1}^{n} \hat{\rho}_\ell^{(j)} = \bigotimes_{j=1}^{n} \left[3\left(U_i^{(j)}\right)^\dagger |b^{(j)}\rangle\langle b^{(j)}|U_i^{(j)} - \mathbb{1}\right], \quad (4)$$

where the superscript (j) indicates that we consider the jth term in the tensor product decomposition. We can then rewrite $\hat{\rho}_\ell$ and define $\widetilde{\mathcal{M}}_{\text{local}}$ as

$$\hat{\rho}_\ell \equiv \bigotimes_{j=1}^{n} \widetilde{\mathcal{M}}_{\text{local}}\left[\left(U_i^{(j)}\right)^\dagger |b^{(j)}\rangle\langle b^{(j)}|U_i^{(j)}\right], \quad (5)$$

in the same spirit as the definition of $\widetilde{\mathcal{M}}$, with

$$\widetilde{\mathcal{M}}_{\text{local}}(\rho) = 3\rho - \mathbb{1}. \quad (6)$$

As the effect of $\mathcal{M}_{\text{local}}$ is to contract the single-qubit Bloch sphere uniformly by a factor of 3, the inverse channel is given by dilating the Bloch sphere by the same factor [11]. Our scheme biases this channel by effectively dilating the Bloch sphere by a smaller factor tuned via a bias parameter ε.

Statement 1 (biased shadow estimators). Given a bias parameter ε we modify the conventional shadow estimator in Eq. (6) and define the biased local estimator of Pauli shadows $\widetilde{\mathcal{M}}_{\text{local}}^{(\varepsilon)}$ as

$$\widetilde{\mathcal{M}}_{\text{local}}^{(\varepsilon)}(\rho) = 3\sqrt{1-\varepsilon}\rho + \tfrac{1}{2}\left(1 - 3\sqrt{1-\varepsilon}\right)\mathbb{1}. \quad (7)$$

This channel dilates the Bloch sphere by a factor $3\sqrt{1-\varepsilon}$ and, indeed, for $\varepsilon = 0$ we recover the unbiased channel of conventional classical shadows.

While the above channel does not converge to the true state in the infinite-sample limit for fixed (independent of N_s) $\varepsilon \neq 0$, it allows us to control shot noise (statistical uncertainty) in the finite-sample regime as the variance is decreased by $(1-\varepsilon)$. As we argue, by choosing the optimal ε as a function of N_s, we obtain a consistent estimator, which converges to the unbiased estimate in the infinite-sample limit.

We now show that this trade-off is on average worthwhile through defining the expected loss as a measure of the *average* performance of the biased scheme as

$$\mathcal{L}_{\mathcal{U}}(\rho, \varepsilon) = \mathbb{E}_{\mathcal{U},\mathbf{b}}\{\text{tr}[(\rho - \hat{\rho})^2]\}. \quad (8)$$

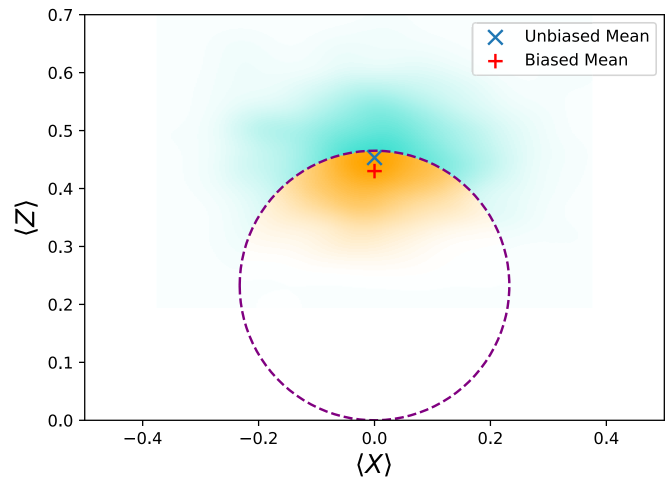

FIG. 1. A geometric illustration of variance-bias tradeoff, displayed via the density of *unbiased* Pauli-X and -Z estimates, where each estimate is obtained as the average over $N_s = 100$ samples. The exterior of the dashed circle corresponds to the region where biasing would hypothetically bring the estimate closer to the true expectation value (with $\langle X\rangle = 0$ and $\langle Z\rangle = 0.45\bar{3}$) for a biasing parameter of $\varepsilon = 0.1$. That is, the color of the density cloud represents the *sign* of the change in loss upon biasing. The radius of the dashed circle is increasing with ε, but this is compensated by the magnitude of the decrease in loss so that biasing improves the estimate for $\varepsilon \leqslant \varepsilon_{\min}$. See details in the Supplemental Material [21].

In Fig. 1, we consider applying the Pauli shadows approach to estimating local properties of a quantum system and consider the reduced density matrix of a single qubit. We fix a number of shots (samples) $N_s = 100$, a biasing parameter of $\varepsilon = 0.1$, and the true state ρ on the Z axis of the Bloch sphere. We calculate the change in loss for a set of averaged classical-shadow estimates of ρ under biasing by ε. The exterior of the dashed circle corresponds to estimates where the corresponding biased estimator is actually more accurate in the particular finite-sample regime than the original, unbiased one. We now concretely state our analytical result that quantifies the expected loss for any particular local density matrix.

Statement 2 (single-shot average-case loss). For a single-qubit local density matrix $\rho = \frac{1}{2}[\mathbb{1} + (\boldsymbol{r}\cdot\boldsymbol{\sigma})]$, where $\mathcal{U}$ forms a 2-design, the expected loss for a single sample is

$$\mathcal{L}_{\mathcal{U}}(\rho, \varepsilon) = \tfrac{1}{2}[\|\boldsymbol{r}\|^2 + 9(1-\varepsilon)] - \sqrt{1-\varepsilon}\|\boldsymbol{r}\|^2. \tag{9}$$

See Supplemental Material (SM) [21] (including Refs. [22–24]) for the derivation. This is a quadratic expression in $\sqrt{1-\varepsilon}$ and thus attains a minimum value at $\sqrt{1-\varepsilon_{\min}} = \frac{1}{9}\|\boldsymbol{r}\|^2$ as

$$\mathcal{L}_{\mathcal{U}}(\rho, \varepsilon_{\min}) = \tfrac{1}{18}\|\boldsymbol{r}\|^2(9 - \|\boldsymbol{r}\|^2). \tag{10}$$

Thus, even if ρ is pure ($\|\boldsymbol{r}\| = 1$), it is worth introducing a strong bias for a single-shot estimator as $\varepsilon_{\min} = \frac{80}{81}$ because it reduces the expected loss to 4/9 from the expected loss 4 of the unbiased scheme. However, in practice the qubit is usually part of an entangled (computational) state ($\|\boldsymbol{r}\| < 1$) which guarantees even more significant gains. In the following sections we explain, however, that the gain is less pronounced as we increase the number of shots.

Analyzing worst- and best-case gain. While in Statement 2 we focused on the average performance of biased shadow tomography, we now analytically predict the performance in extremal scenarios. In particular, we consider the practically pivotal task of predicting expected values of local Pauli observables from the snapshots as $\mathrm{tr}(O\hat{\rho})$ and analyze the worst- and best-case scenarios. We note that we focus on the mean estimator, which is then of course the primary component of the median-of-means estimator [11].

As we will conclude below Eq. (14), our biased estimator yields the least gain in the worst-case scenario that the quantum state ρ is the eigenstate of the Pauli observable via $\mathrm{tr}(O\rho) = +1$ (where O is a Pauli string of weight w_p up to $\pm$ sign). The reason why we still gain even in this worst-case scenario is the following: At each shot our estimator $\mathrm{tr}(O\hat{\rho})$ yields the outcome either $+1$ (when we measure in a compatible Pauli basis) or 0 (when we measure in an incompatible basis). This indeed yields a binomial distribution $B(N_s, p)$ with number of samples N_s and probability of compatible measurements $p = 3^{-w}$. As we illustrate in SM [21], the binomial distribution is not symmetric around the mean but has a tail, and the mean-squared error on the right-hand side is higher than the mean-squared error on the left-hand side—it is thus always worth biasing the estimator by effectively rescaling the estimate. However, as we increase N_s, the binomial distribution quickly tends to a symmetric normal distribution and thus rescaling will not yield an advantage.

In SM [21] we analytically derive the mean error $\langle E_n^{\mathrm{w}}(\varepsilon)\rangle$ in the expected value measurement and plot these relative to the unbiased case in Fig. 2. Indeed, this confirms that (a) the biased scheme is always advantageous as there is always an optimal ε (the minimum of the curves indicated using crosses) for which the relative mean error is smaller than 1. (b) The advantage of the biased scheme grows exponentially as we increase the weight of the Pauli string simply because the contraction of the Bloch sphere is exponential via the factor 3^{w_p} for locality w_p while the tail of the distribution of estimates (via the aforementioned binomial distribution) gets exponentially long, i.e., compare blue, orange, and green lines. (c) The advantage of the biased scheme diminishes as we increase the numbers of shots.

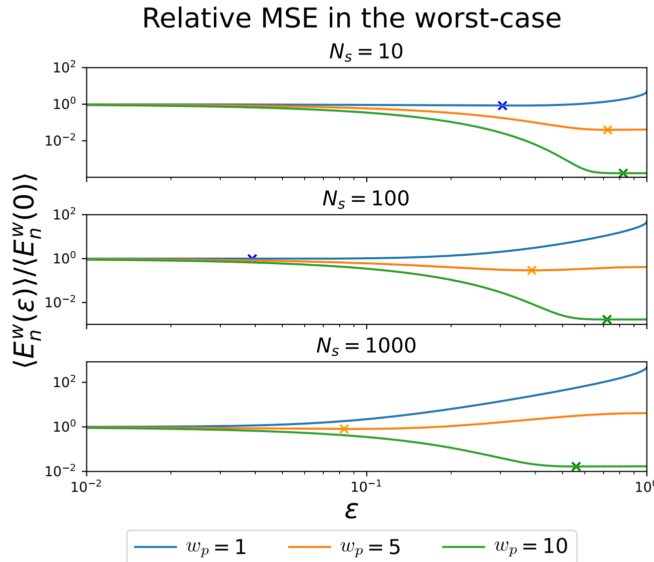

FIG. 2. Mean-squared error $\langle E_n^{\mathrm{w}}(\varepsilon)\rangle$ of the biased scheme relative to the unbiased one when the aim is to estimate Pauli expected values of a weight w_p observable on an n-qubit state $\mathrm{tr}(O\rho)$. Analytical expression plotted for the worst-case scenario (O commutes with ρ) for an increasing number of shots, weight of the Pauli string ($n = w_p$), and bias parameters ε. There is an optimal ε (crosses at the minimum of the curves) for which the relative error is minimized. When $w = 1$ and $N_s = 1000$ the minimum is attained outside the plotted region.

In contrast, the best-case scenario is attained when the quantum state is an eigenstate of an operator that anticommutes with the observable which gives us $\mathrm{Tr}(O\rho) = 0$. It is then clear that biasing, which is equivalent to shrinking the expectation value is always advantageous as it forces the estimate to be closer to 0. In SM [21], we give an expression for the mean error $\langle E_n^{\mathrm{b}}(\varepsilon)\rangle$ which we numerically compute and plot confirming that indeed increasing the bias parameter ε monotonically decreases the expected error.

Relation to variances and quantum mechanical expected values. In the previous section we analyzed the instances when the expected values attain the extremal values $\mathrm{tr}(O\rho) = \pm 1$ and $\mathrm{tr}(O\rho) = 0$, and we now prove that indeed these are the worst- and best-case scenarios, respectively. For this reason we consider an arbitrary sample mean estimator $\overline{R}$ that one obtains from averaging over N_s samples of a single-shot estimator $\hat{R}$, and compare $\overline{R}$ to the corresponding biased estimator $(1-\alpha)\overline{R}$ that one obtains through rescaling with the factor $(1-\alpha)$. In order to derive the optimal biasing point, i.e., the minimum of the curves in Fig. 2 (crosses), we first consider the mean-squared error (MSE) of the unbiased estimator $\overline{R}$ as $\mathrm{MSE}[\overline{R}] = \mathrm{Var}[\hat{R}]/N_s$. We can define and calculate the signal-to-noise ratio (SNR) of $\overline{R}$ as

$$\beta := \frac{\mathbb{E}[\overline{R}]^2}{\mathrm{MSE}[\overline{R}]} = \frac{\mathbb{E}[\hat{R}]^2}{\mathrm{Var}[\hat{R}]/N_s}. \tag{11}$$

The MSE for the biased mean estimator $(1-\alpha)\overline{R}$ is then

$$\mathrm{MSE}[(1-\alpha)\overline{R}] = \underbrace{\alpha^2\mathbb{E}[\hat{R}]^2}_{\text{bias}} + \underbrace{(1-\alpha)^2\mathrm{Var}[\hat{R}]/N_s}_{\text{variance}}, \tag{12}$$

which is minimized at the optimal biasing point $\alpha^* = (1+\beta)^{-1}$ as we derive in SM [21]. Through Eq. (11), we find that the optimal bias parameter approaches $\alpha^* \to 0$ as we increase the number of samples $N_s \to \infty$. As such, our (optimally) biased estimator is actually a consistent estimator, i.e., it asymptotically approaches an unbiased estimator in the infinite-sample limit [25].

We can make the following statement at the optimal biasing point.

Statement 3 (biasing an estimator through rescaling). The SNR of the optimally biased estimator is $\beta_{\text{biased}} = 1 + \beta$ which always guarantees an improved SNR over the unbiased estimator β. The relative SNR gain through biasing is given as

$$\frac{\beta_{\text{biased}}}{\beta} = 1 + \beta^{-1}. \tag{13}$$

As further shown in SM [21], in the case of estimating Pauli expected values for Pauli shadows, the mean and variance of the single-shot estimator are given as $\mathbb{E}[\hat{R}] = \mathrm{tr}[O\rho]$ and $\mathrm{Var}[\hat{R}] = 3^w - \mathrm{tr}[O\rho]^2$. Hence, the SNR of the unbiased estimator is given as $\beta = N_s[3^w\,\mathrm{tr}[O\rho]^{-2} - 1]^{-1}$ and the factor of SNR gain by biasing is given by

$$\frac{\beta_{\text{biased}}}{\beta} = 1 + (3^w\,\mathrm{tr}[O\rho]^{-2} - 1)/N_s. \tag{14}$$

The maximal and minimal gains are then obtained at $\mathrm{tr}[O\rho] = 0$ and $\mathrm{tr}[O\rho] = +1$, respectively, which proves our previous observations on worst- and best-case scenarios.

Let us note that, while the above SNR allows us to rigorously prove that biasing is in principle always advantageous, the mean-squared error analyzed in the previous section remains the more practical measure and we will use it thereafter. It is also worth noting that Statement 3 requires us to know the expectation value $\mathbb{E}[\hat{R}]$ exactly—which may not be possible in practice—in order to predict the optimal biasing point. Nevertheless, we demonstrate in the following section that biasing is still worthwhile even if we can only have approximate knowledge of the optimal bias.

Practical demonstration. We consider a potential practical application whereby one aims to obtain the ground-state energy of a Hamiltonian $\mathcal{H}$ composed of only low-weight (local) Pauli observables. In an experiment one first prepares the ground state and collects a set of Pauli shadows from which the ground-state energy can be predicted in postprocessing. The significant advantage of classical shadows is that they allow us to estimate further Pauli strings beyond the Hamiltonian terms, without repeating the experiment.

For example, one can consider perturbative corrections to the Hamiltonian in the form of high-weight Pauli strings P and predict the expected value of the sum $\mathcal{H} + P$, such as when computing a first-order correction to the energy in perturbation theory. However, the variance in our example is increased exponentially due to the high weight of P; this potentially renders a direct estimation impractical as the increased variance may bring the SNR down below 1, as we detail in SM [21].

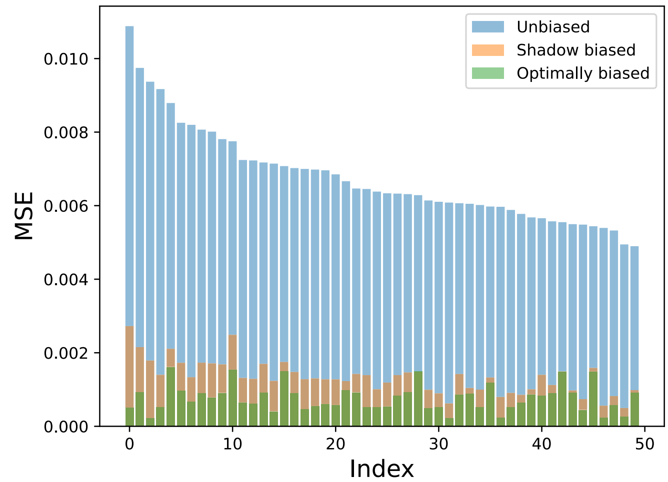

FIG. 3. Illustration of the impact of different biasing strategies. In blue, we plot the MSE of fifty 8-local observables (whose SNR is smaller than 1) sorted by how large the error is. While analytically choosing the optimal bias according to Statement 3 drastically reduces the error (green), this assumes previous knowledge of the exact expectation values, which is not available in practice. Through estimating the optimal bias from experimental data (orange), one still obtains a significant error reduction.

As we demonstrate, biasing then allows us to significantly improve upon this potentially low SNR.

As a concrete example we consider a spin-ring Hamiltonian as

$$\mathcal{H} = \sum_{k\in\text{ring}(N)} \omega_k Z_k^z + J\boldsymbol{\sigma}_k \cdot \boldsymbol{\sigma}_{k+1}, \tag{15}$$

with coupling $J = 0.3$, on-site interaction strengths uniformly randomly generated in the range $-1 \leqslant \omega_k \leqslant 1$ and $\boldsymbol{\sigma}_k = (X_k, Y_k, Z_k)^T$ is a vector of single-qubit Pauli matrices. We prepare the ground state ρ of a 12-qubit Hamiltonian through a variational Hamiltonian ansatz of $l = 5$ layers. We generate a collection of shadows $S(N_s, \rho)$ with $N_s = 10^6$ and estimate the expected value $\mathrm{tr}[(\mathcal{H} + P)\rho]$ with respect to corrections P as $w = 8$ Pauli observables.

In Fig. 3, we plot the mean-squared error obtained by averaging over 10^5 repetitions and consider different 8-local observables with and without biasing (blue). We consider two different biasing strategies. First, we analytically choose the optimal bias parameter α^* according to Statement 3 (green) assuming direct access to the exact expected values (which one does not have access to in practice). Second, we use the experimentally estimated expected values using shadows to estimate the optimal biasing point in Statement 3. While the latter deviates from the exact α^* due to shot noise, Fig. 3 (orange) clearly demonstrates that the MSE is still significantly reduced compared to the unbiased scheme. This confirms robustness against errors in our ability to determine the optimal bias parameter, i.e., the approach can demonstrably be used even when the optimal biasing point can only be estimated from the experimental data.

Discussion and conclusion. In this Letter, we explore the tradeoff between intentionally introducing a bias into classical-shadow estimators which in return allows us to reduce statistical uncertainties due to finite samples—ultimately enabling us to achieve the same precision but using fewer samples. The implementation of the approach is straightforward, and is performed completely in postprocessing, as it is effectively just a rescaling of the conventional shadow estimator.

We obtain rigorous analytical guarantees that biasing is, in principle, always worthwhile given an optimal bias parameter is known: First, optimal biasing improves the relative loss of shadow tomography on average. Second, optimal biasing improves expected value measurements even in the worst-case scenario and may provide significant gains in other scenarios. Third, optimal biasing is guaranteed to increase the signal to noise of any statistical mean estimator. Although the optimal bias parameter is not known *a priori*, we demonstrate in a practically motivated numerical experiment that an approximation directly determined form experimental data is sufficient.

The gain using our biased estimator is more pronounced for a small number of samples or when Pauli strings of high weight are predicted. As such, our approach is thus particularly well suited for practical tasks where only a relatively low number of shots is available, e.g., estimating gradients [26–29] or covariances [30] when training variational circuits or when estimating time-dependent properties [31]. Another particularly interesting application area is quantum error mitigation [32] whereby prior works aimed at recovering an unbiased estimator from noisy measurements. Combinations of QEM and classical shadows have similarly been considered [33,34].

The insights and theoretical results provided in this Letter may prove invaluable in developing further, more advanced biased estimators. Given classical information from quantum systems can only be extracted through statistical estimators, this work is an important step in the crucial task of developing applications of quantum computers that have minimal sample requirements.

Acknowledgments. The numerical modeling involved in this study made use of the Quantum Exact Simulation Toolkit (QuEST) [35], and the recent development QuESTlink [36] which permits the user to use *Mathematica* as the integrated front end, and pyQuEST [37] which allows access to QuEST from PYTHON. We are grateful to those who have contributed to all of these valuable tools. The authors would like to acknowledge the use of the University of Oxford Advanced Research Computing (ARC) facility [38] in carrying out this work and specifically the facilities made available from the EPSRC QCS Hub grant (Agreement No. EP/T001062/1). The authors also acknowledge funding from the EPSRC projects Robust and Reliable Quantum Computing (RoaRQ, EP/W032635/1) and Software Enabling Early Quantum Advantage (SEEQA, EP/Y004655/1). B.K. is thankful for support from UKRI for the Future Leaders Fellowship project titled Theory to Enable Practical Quantum Advantage (MR/Y015843/1). B.K. is thankful for support from the University of Oxford for a Glasstone Research Fellowship and Lady Margaret Hall, Oxford for a Research Fellowship. Z.C. is supported by the Junior Research Fellowship from St John's College, Oxford. A.C. acknowledges support from EU H2020-FETFLAG-03-2018 under grant Agreement No. 820495 (AQTION). This research was funded in whole, or in part, by UKRI (MR/Y015843/1).

The authors of this paper were ordered alphabetically.

[1] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler *et al.*, Quantum supremacy using a programmable superconducting processor, Nature (London) **574**, 505 (2019).

[2] Y. Wu, W.-S. Bao, S. Cao, F. Chen, M.-C. Chen, X. Chen, T.-H. Chung, H. Deng, Y. Du, D. Fan, M. Gong, C. Guo, C. Guo, S. Guo, L. Han, L. Hong, H.-L. Huang, Y.-H. Huo, L. Li, N. Li *et al.*, Strong quantum computational advantage using a superconducting quantum processor, Phys. Rev. Lett. **127**, 180501 (2021).

[3] H.-S. Zhong, Y.-H. Deng, J. Qin, H. Wang, M.-C. Chen, L.-C. Peng, Y.-H. Luo, D. Wu, S.-Q. Gong, H. Su, Y. Hu, P. Hu, X.-Y. Yang, W.-J. Zhang, H. Li, Y. Li, X. Jiang, L. Gan, G. Yang, L. You *et al.*, Phase-programmable Gaussian boson sampling using stimulated squeezed light, Phys. Rev. Lett. **127**, 180502 (2021).

[4] D. Bluvstein, S. J. Evered, A. A. Geim, S. H. Li, H. Zhou, T. Manovitz, S. Ebadi, M. Cain, M. Kalinowski, D. Hangleiter, J. P. B. Ataides, N. Maskara, I. Cong, X. Gao, P. S. Rodriguez, T. Karolyshyn, G. Semeghini, M. J. Gullans, M. Greiner, V. Vuletić *et al.*, Logical quantum processor based on reconfigurable atom arrays, Nature (London) **626**58 (2023).

[5] Y. Kim, A. Eddins, S. Anand, K. X. Wei, E. van den Berg, S. Rosenblatt, H. Nayfeh, Y. Wu, M. Zaletel, K. Temme, and A. Kandala, Evidence for the utility of quantum computing before fault tolerance, Nature (London) **618**, 500 (2023).

[6] R. Acharya, I. Aleiner, R. Allen, T. I. Andersen, M. Ansmann, F. Arute, K. Arya, A. Asfaw, J. Atalaya, R. Babbush, D. Bacon, J. C. Bardin, J. Basso, A. Bengtsson, S. Boixo, G. Bortoli, A. Bourassa, J. Bovaird, L. Brill, M. Broughton *et al.*, Suppressing quantum errors by scaling a surface code logical qubit, Nature (London) **614**, 676 (2023).

[7] T. Sugiyama, P. S. Turner, and M. Murao, Precision-guaranteed quantum tomography, Phys. Rev. Lett. **111**, 160406 (2013).

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

[9] J. J. Meyer, S. Khatri, D. Stilck França, J. Eisert, and P. Faist, Quantum metrology in the finite-sample regime, arXiv:2307.06370.

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

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

[12] H.-Y. Huang, R. Kueng, and J. Preskill, Efficient estimation of Pauli observables by derandomization, Phys. Rev. Lett. **127**, 030503 (2021).
[13] A. Zhao, N. C. Rubin, and A. Miyake, Fermionic partial tomography via classical shadows, Phys. Rev. Lett. **127**, 110504 (2021).
[14] K. Wan, W. J. Huggins, J. Lee, and R. Babbush, Matchgate shadows for fermionic quantum simulation, Commun. Math. Phys. **404**, 629 (2023).
[15] S. Chen, W. Yu, P. Zeng, and S. T. Flammia, Robust shadow estimation, PRX Quantum **2**, 030348 (2021).
[16] H. C. Nguyen, J. L. Bönsel, J. Steinberg, and O. Gühne, Optimizing shadow tomography with generalized measurements, Phys. Rev. Lett. **129**, 220502 (2022).
[17] C. Ferrie and R. Blume-Kohout, Maximum likelihood quantum state tomography is inadmissible, arXiv:1808.01072.
[18] A. A. Akhtar, H.-Y. Hu, and Y.-Z. You, Scalable and flexible classical shadow tomography with tensor networks, Quantum **7**, 1026 (2023).
[19] K. Bu, D. E. Koh, R. J. Garcia, and A. Jaffe, Classical shadows with Pauli-invariant unitary ensembles, npj Quantum Inf. **10**, 6 (2024).
[20] D. E. Koh and S. Grewal, Classical shadows with noise, Quantum **6**, 776 (2022).
[21] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevA.111.L030402 for a comparison with previous literature, derivation of average-, worst-, and best-case errors, a statistical analysis of biased shadows (resulting in the determination of the optimal biasing point), and additional information on the practical applications we explored.
[22] C. Hadfield, S. Bravyi, R. Raymond, and A. Mezzacapo, Measurements of quantum Hamiltonians with locally-biased classical shadows, Commun. Math. Phys. **391**, 951 (2022).
[23] A. Gresch and M. Kliesch, Guaranteed efficient energy estimation of quantum many-body Hamiltonians using ShadowGrouping, Nat. Commun. **16**, 689 (2025).
[24] Z. Zhu, J. M. Lukens, and B. T. Kirby, On the connection between least squares, regularization, and classical shadows, Quantum **8**, 1455 (2024).
[25] A. Scott and C.-F. Wu, On the asymptotic distribution of ratio and regression estimators, J. Am. Stat. Assoc. **76**, 98 (1981).
[26] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, Evaluating analytic gradients on quantum hardware, Phys. Rev. A **99**, 032331 (2019).
[27] B. Koczor and S. C. Benjamin, Quantum analytic descent, Phys. Rev. Res. **4**, 023017 (2022).
[28] D. Wierichs, J. Izaac, C. Wang, and C. Y.-Y. Lin, General parameter-shift rules for quantum gradients, Quantum **6**, 677 (2022).
[29] B. Koczor and S. C. Benjamin, Quantum natural gradient generalized to noisy and nonunitary circuits, Phys. Rev. A **106**, 062416 (2022).
[30] G. Boyd and B. Koczor, Training variational quantum circuits with CoVaR: Covariance root finding with classical shadows, Phys. Rev. X **12**, 041022 (2022).
[31] H. H. S. Chan, R. Meister, M. L. Goh, and B. Koczor, Algorithmic shadow spectroscopy, PRX Quantum **6**, 010352 (2025).
[32] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O'Brien, Quantum error mitigation, Rev. Mod. Phys. **95**, 045005 (2023).
[33] H. Jnane, J. Steinberg, Z. Cai, H. C. Nguyen, and B. Koczor, Quantum error mitigated classical shadows, PRX Quantum **5**, 010324 (2024).
[34] A. Zhao and A. Miyake, Group-theoretic error mitigation enabled by classical shadows and symmetries, npj Quantum Inf. **10**, 57 (2024).
[35] T. Jones, A. Brown, I. Bush, and S. C. Benjamin, QuEST and high performance simulation of quantum computers, Sci. Rep. **9**, 10736 (2019).
[36] T. Jones and S. Benjamin, QuESTlink—mathematica embiggened by a hardware-optimised quantum emulator, Quantum Sci. Technol. **5**, 034012 (2020).
[37] R. Meister, pyQuEST - A Python interface for the Quantum Exact Simulation Toolkit, https://github.com/rrmeister/pyQuEST.
[38] A. Richards, University of Oxford Advanced Research Computing, https://doi.org/10.5281/zenodo.22558.