# Shadow Distillation: Quantum Error Mitigation with Classical Shadows for Near-Term Quantum Processors

Alireza Seif,[1,*,†] Ze-Pei Cian,[2,3,4,†] Sisi Zhou,[1,5] Senrui Chen,[1] and Liang Jiang[1]

[1] *Pritzker School of Molecular Engineering, University of Chicago, Chicago, Illinois 60637, USA*

[2] *Department of Physics, University of Maryland, College Park, Maryland 20742, USA*

[3] *Joint Quantum Institute, University of Maryland, College Park, Maryland 20742, USA*

[4] *Center for Quantum Information and Computer Science, University of Maryland, College Park, Maryland 20742, USA*

[5] *Institute for Quantum Information and Matter, California Institute of Technology, Pasadena, California 91125, USA*

(Received 28 March 2022; revised 25 September 2022; accepted 22 November 2022; published 9 January 2023)

Mitigating errors in quantum information processing devices is especially important in the absence of fault tolerance. An effective method in suppressing state-preparation errors is using multiple copies to distill the ideal component from a noisy quantum state. Here, we use classical shadows and randomized measurements to circumvent the need for coherent access to multiple copies at an exponential cost. We study the scaling of resources using numerical simulations and find that the overhead is still favorable compared to full state tomography. We optimize measurement resources under realistic experimental constraints and apply our method to an experiment preparing a Greenberger-Horne-Zeilinger state with trapped ions. In addition to improving stabilizer measurements, the analysis of the improved results reveals the nature of errors affecting the experiment. Hence, our results provide a directly applicable method for mitigating errors in near-term quantum computers.

DOI: 10.1103/PRXQuantum.4.010303

## I. INTRODUCTION

One of the main obstacles in operating quantum information processing devices is extreme sensitivity to errors. In principle, these errors can be corrected using error-correcting codes [1]. However, utilizing these codes in a fault-tolerant manner requires a hardware overhead that is pushing the limits of what experiments can achieve today [2]. Therefore, it is interesting to find ways to mitigate the effect of errors and extend the utility of current devices in the absence of fault tolerance. Recently, there have been several proposals for mitigating the effect of errors on estimating expectation values of observables in a quantum circuit [3–9]. These schemes work by acquiring the expectation value of an observable for different noise strengths (e.g., by changing the gate time) and extrapolating them to find the expectation value at the zero-noise limit, or as shown in Refs. [5,6] by learning a correction scheme using circuits that are easy (e.g., Clifford circuits) to simulate and applying the learned correction procedure to general circuits. Additionally, there has been a new endeavor along the ideas of Ref. [10] to extract the state of interest from a noisy mixed state by using multiple copies of the noisy state [7,11–17].

At the same time, quantum devices are growing in size, and that increases the complexity of extracting information from the system. In particular, methods such as quantum state tomography have a complexity that grows exponentially with the system size. Recently, there have been proposals for efficient extraction of certain properties of a quantum system based on randomized measurements and classical shadows [18–20]. Roughly speaking, these methods provide a way for estimating many linear functions of a quantum state with (quantum and classical) resources that scale efficiently with the system size. For nonlinear functions of the state, such as Rényi entropies or topological invariants, protocols based on randomized measurements have an exponential complexity, but are still advantageous compared to full state tomography [18,21–26], making them a useful tool for probing near-term intermediate-scale devices [27].

In this work, we take advantage of the framework of randomized measurements and classical shadows and apply it to the problem of error mitigation. Specifically,

*seif@uchicago.edu

†These authors contributed equally to this work.

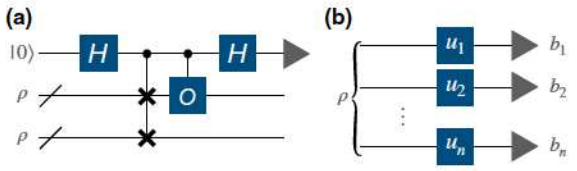


FIG. 1. Schematic representation of error mitigation with multiple copies. (a) Performing an interferometry experiment, where $H$ is the Hadamard gate, with a controlled-SWAP operation on two copies of state $\rho$ together with a controlled-$O$ operation on one of the copies enables measuring $\mathrm{tr}(O\rho^2)$. (b) The procedure in panel (a) can be replaced by randomized single-qubit measurements implemented by $u_i$ and postprocessing the results $b_i$.

we study error mitigation using multiple copies [7,13–16] and study the trade-off between quantum resources (such as two-qubit gates and coherent access to multiple copies of a state) and single-qubit randomized measurements (see Fig. 1). We first explain the error-mitigation framework and show how our protocol incorporates randomized measurements in this framework. We then provide a numerical analysis of the errors and resources and explore the trade-off between the number of measurement settings and the repetitions of each measurement. Finally, using the existing trapped-ion experimental data from Ref. [28], we illustrate the application of our method in optimizing experimental resources for improving the measurements of stabilizers of a five-qubit Greenberger-Horne-Zeilinger (GHZ) state [29]. The success and shortcomings of our protocol, in this case, reveal valuable information about the nature of errors in the experiment.

## II. ERROR MITIGATION USING MULTIPLE COPIES

We first review the scheme using multiple copies for suppressing errors in preparing a quantum state. Let $|\psi\rangle$ denote the ideal state that we are interested in preparing in an experiment. Because of experimental imperfections, we instead end up with $\rho = (1-\varepsilon)|\psi\rangle\langle\psi| + \varepsilon\rho_{\mathrm{error}}$, where $0 < \varepsilon \leq 1$ quantifies the strength of errors. We assume that $\rho_{\mathrm{error}}$ is a density matrix in a subspace orthogonal to $|\psi\rangle$, i.e., $\langle\psi|\rho_{\mathrm{error}}|\psi\rangle = 0$. Realistic noise in an experiment might differ from this model. In Appendix F, we discuss the effectiveness of this scheme for various noise models. Now let us consider the task of estimating the expectation value of an observable $O$. Ideally, we would like to extract $\langle\psi|O|\psi\rangle$. However, because of the errors we obtain $\mathrm{tr}(O\rho)$. To reduce the errors in our estimate, one can instead calculate $\langle O\rangle_{(m)} = \mathrm{tr}(O\rho^m)/\mathrm{tr}(\rho^m)$, where $m$ is an integer, which is referred to as virtual distillation in the literature (see, e.g., Refs. [13,15]). This scheme is effective if $|\psi\rangle$ is the dominant eigenvector of $\rho$, i.e., $1-\varepsilon > \varepsilon p_{\max}$ with $p_{\max}$ being the largest eigenvalue of $\rho_{\mathrm{error}}$, and suppresses the errors exponentially in $m$ [13,15] since

$$\begin{aligned}\frac{\mathrm{tr}(O\rho^m)}{\mathrm{tr}(\rho^m)} &= \frac{\mathrm{tr}\{O[(1-\varepsilon)^m|\psi\rangle\langle\psi| + \varepsilon^m\rho^m_{\mathrm{error}}]\}}{\mathrm{tr}[(1-\varepsilon)^m|\psi\rangle\langle\psi| + \varepsilon^m\rho^m_{\mathrm{error}}]} \\ &= \frac{(1-\varepsilon)^m\langle\psi|O|\psi\rangle + \varepsilon^m\mathrm{tr}(O\rho^m_{\mathrm{error}})}{(1-\varepsilon)^m + \varepsilon^m\mathrm{tr}(\rho^m_{\mathrm{error}})} \\ &\simeq \langle\psi|O|\psi\rangle + f(O,\rho_{\mathrm{error}})\varepsilon^m + \mathcal{O}(\varepsilon^{m+1}), \quad (1)\end{aligned}$$

where $f(O,\rho_{\mathrm{error}}) = \mathrm{tr}(O\rho^m_{\mathrm{error}}) - \langle\psi|O|\psi\rangle\mathrm{tr}(\rho^m_{\mathrm{error}})$. Hence, the access to $\rho^m$ enables suppressing errors exponentially in $m$. Previous works [7,11–16,30] have mostly considered using multiple copies and controlled permutations to prepare $\rho^m$, given access to $m$ copies of $\rho$. This is enabled by using the fact that $\mathrm{tr}(V^{(m)}\rho^{\otimes m}) = \mathrm{tr}(\rho^m)$, where $V^{(m)}$ is a permutation operator acting as $V^{(m)}|\psi_1\rangle|\psi_2\rangle\cdots|\psi_m\rangle = |\psi_m\rangle|\psi_1\rangle\cdots|\psi_{m-1}\rangle$. Such schemes require the use of two-qubit gates between copies of state $\rho$ stored in quantum registers. Note that while $m$ copies of $\rho$ are required for such a procedure, we only need coherent access to two copies at the same time [30]. However, even in this case, preparing two copies of the state in the same quantum register and performing coherent operations on them poses serious difficulties in near-term experiments [31]. In particular, the long-range coupling required for performing the swap operation between copies is challenging in devices with limited connectivity and requires additional gates that introduce further errors in the computation task of interest [32].

Recently, there have been proposals to trade access to copies of the state (circuit width) with circuit depth using a dual-state scheme [11,14]. These methods eliminate the need for quantum operations between different copies of the state, which can be challenging in near-term devices [31]. However, they require knowledge of the unitary operator that prepares the state of interest and its inverse and assume that the noise affecting the state and its dual are similar. In general, implementing the inverse of a known unitary operation is not straightforward, e.g., in an analog simulator without a circuit description or for applications in metrology [17]. The increased depth of the circuit can also be problematic for the latter assumption in the presence of non-Markovian errors [33]. Moreover, these method require midcircuit measurements with feedback, or ancilla-assisted measurements that can again introduce errors in the experiment and require additional error mitigation in the near term.

## III. SHADOW DISTILLATION

In this work, we propose using the framework of randomized measurements and classical shadows to calculate $\mathrm{tr}(O\rho^m)$ and $\mathrm{tr}(\rho^m)$. We refer to our method as shadow

distillation (SD). Our approach trades circuit size with sample complexity. This trade-off can be especially useful in near-term devices where control and circuit depth and width are limited and errors are large. In such cases, the increased depth and width required for error mitigation with multiple copies inevitably introduces additional errors in the computations. These errors limit the applicability of error mitigation as they further bias the results away from their ideal values. In contrast, the errors in our method are mostly statistical and be reduced by collecting, possibly exponentially many, more samples. We come back to this issue when we increase sample complexity in Sec. IV.

Specifically, let $\rho$ denote the state of interest on $n_q$ qubits. To measure the quantum state in $N_U$ random bases, we sample $N_U$ distinct combinations of random single-qubit rotations $U = u_1 \otimes u_2 \otimes \cdots \otimes u_N$ and append them to the circuit that is used to prepare $\rho$. Finally, we perform projective measurements on the computational basis. For each rotation setting $U$, the measurements are repeated $N_S$ shots.

To infer the physical quantities from the randomized measurements, one can convert each measurement outcome to a classical snapshot of the state. For a measurement with a random unitary $U = u_1 \otimes u_2 \otimes \cdots \otimes u_N$ satisfying the three-design property and a measurement outcome $|b\rangle = |b_1, b_2, \ldots, b_N\rangle$, the classical snapshot is of the form

$$\hat{\rho}_{U,b} = \bigotimes_{k=1}^{n_q} (3u_k^\dagger |b_k\rangle\langle b_k| u_k - I), \tag{2}$$

where $I$ is the identity matrix on a single qubit. The collection of these snapshots is referred to as a classical shadow of the state [18]. The density matrix $\rho$ can be inferred from the classical shadow by averaging over $U$ and $b$, i.e., $\rho = \mathbb{E}_{U,b}(\hat{\rho}_{U,b})$. Therefore, one can directly infer the expectation value of an observable $O$ from its expectation value over each snapshot using $\mathrm{tr}(O\rho) = \mathbb{E}_{U,b}[\mathrm{tr}(O\hat{\rho}_{U,b})]$ [18]. Physical quantities that are nonlinear in the density matrix $\rho$, e.g., $\mathrm{tr}(O\rho^2)$, can be calculated through $\mathrm{tr}(O\rho^2) = \mathbb{E}_{U,b,U',b'}[\mathrm{tr}(V^{(2)}(O\hat{\rho}_{U,b}) \otimes \hat{\rho}_{U',b'})]$, where $V^{(2)}$ is the swap operator [18]. For certain choices of measurement bases, such as those corresponding to random global Clifford operations and random Pauli measurements, the shadows can be stored and manipulated efficiently in a time and memory polynomial in $n_q$, $N_U$, and $N_S$ [18].

Here, we focus on second-order error mitigation ($m = 2$) with randomized single-qubit Pauli measurements. Specifically, let $\{U_j\}_{j=1}^{N_U}$ denote the $N_U$ sampled unitary operators from random local Clifford gates, and let $\{|b^{(i_j)}\rangle_{i_j=1}^{N_S}\}$ denote the measurement outcomes of $N_S$ measurements fixing $U = U_j$. We then define $\hat{\rho}_j = (1/N_S)\sum_{i_j=1}^{N_S} \hat{\rho}_{U_j,b^{(i_j)}}$, which corresponds to the average snapshot (2) for a fixed $U$. We denote our estimate of $\mathrm{tr}(O\rho^2)$ by $\hat{o}_2$, given by [34]

$$\hat{o}_2 = \frac{1}{N_U(N_U-1)} \sum_{j \neq j'} \mathrm{tr}(V^{(2)} \hat{\rho}_j \otimes (O\hat{\rho}_{j'})), \tag{3}$$

which is an unbiased estimator (see Appendix A). Note that setting $O = I$ results in an estimate of $\mathrm{tr}(\rho^2)$, which we denote by $\hat{s}_2$.

In this way, $\hat{s}_2$ using $N_U N_S$ snapshots can be calculated in time $\mathcal{O}(\mathrm{poly}(n) N_U^2 N_S^2)$. Moreover, $\hat{o}_2$ for operators $O$ that are products of single-qubit Pauli operators can be obtained with the same complexity [35]. Therefore, using classical shadows enables us to perform error mitigation for such operators using classical computational resources that scale polynomially with the number of samples and the number of qubits $n_q$. However, it should be noted that the number of samples required to achieve a given accuracy can depend on $n_q$. In fact, the sample complexity of estimating quantities nonlinear in state $\rho$ can grow exponentially with system size [18,36]. In the following, we numerically investigate this scaling and show that, for the case of $m = 2$, $\langle O\rangle_{(2)}$ for Pauli observables performs favorably compared to schemes based on full quantum state tomography.

## IV. NUMERICAL INVESTIGATION OF ERROR SCALING

We analyze the scaling of statistical errors in the estimation of $\langle O\rangle_{(2)}$ for Pauli observables with measurement resources, $N_U$ and $N_S$, and $n_q$ qubits using numerical simulations. To study the generic performance of the protocol, we first prepare random pure states under depolarization noise with strength $\varepsilon$:

$$\rho_R = (1-\varepsilon)|\psi_R\rangle\langle\psi_R| + \frac{\varepsilon}{2^{n_q}-1}[I - |\psi_R\rangle\langle\psi_R|]. \tag{4}$$

Here $0 < \varepsilon \leq 1$, $|\psi_R\rangle = U_R|0\rangle$, and $U_R$ is a Haar random unitary operator. We then estimate $\mathrm{tr}(\rho_R^2)$ and $\mathrm{tr}(O\rho_R^2)$, denoted by $\hat{s}_2^{(R)}$ and $\hat{o}_2^{(R)}$, respectively, using Eq. (3), by sampling $N_U$ random bases and $N_S$ shots. Let

$$\Delta_R^2 = \left(\frac{\mathrm{tr}(O\rho_R^2)}{\mathrm{tr}(\rho_R^2)} - \frac{\hat{o}_2^{(R)}}{\hat{s}_2^{(R)}}\right)^2 \tag{5}$$

denote the squared error of estimating $\langle O\rangle_{(2)}$ for the particular state $\rho_R$. In our simulations, we examine the mean squared error (MSE) $\Delta^2 = (1/N_R)\sum_R \overline{\Delta_R^2}$, over $N_R = 100$ random choices of $U_R$. The overbar denotes the average taken over different realizations of measurements for each $U_R$ obtained by bootstrap sampling over 250 instances; see Appendix B for more information on the bootstrap resampling techniques. We emphasize that $\Delta$ only captures errors of our SD scheme for estimating $\langle O\rangle_{(2)}$ and

does not include the errors that are not corrected using this error-mitigation procedure. The effectiveness of the error-mitigation scheme has been studied in other works; see, e.g., Refs. [11,13–16]. We discuss that aspect in the discussion of our results for the trapped-ion experiment.

Figures 2(a) and 2(b) show the scaling of the statistical error as a function of $N_U$ and $N_S$ for various observables $O$ for $n_q = 4$ and $\varepsilon = 0.1$. We observe that $\Delta^2$ scales as $1/N_U$. Moreover, for a fixed value of $N_U$, it converges as $1/N_S$ to a constant determined by $N_U$. We also observe a fast convergence of $\Delta^2$ to a constant value determined by $N_U$ and $N_S$ as a function of purity, $\mathrm{tr}(\rho^2)$, as shown in Fig. 2(c). Note that our estimator $\hat{o}_2/\hat{s}_2$ is, in general, a biased estimator for $\langle O\rangle_{(2)}$. Moreover, there is no closed-form formula for the variance of $\hat{o}_2/\hat{s}_2$. In Appendix C, we derive analytical bounds for $\mathrm{Var}(\hat{o}_2)$ and $\mathrm{Var}(\hat{s}_2)$. While these bounds do not directly translate to a bound on $\mathrm{Var}(\hat{o}_2/\hat{s}_2)$, they can still provide an intuition on the behavior of the errors and help us find empirical expressions for the scaling of errors. In fact, the scaling that we observe in Figs. 2(a)–2(b) agrees with our bound for the variance of the numerator. Additionally, errors in $\hat{s}_2$ can lead to large errors in estimating the ratio $\hat{o}_2/\hat{s}_2$, especially in the small-$N_U$ regime. In these cases it might be beneficial to incorporate prior knowledge about the value of the purity $\mathrm{tr}(\rho^2)$ to reduce the errors. We further explore this idea in Appendix E. We show that, given a measurement of purity $s_2$, a prior guess for the value of purity $\mu_0$, and a hyperparameter $\alpha$ that quantifies the confidence in our guess, a modified estimator of the form $(s_2 + \lambda\mu_0)/(1+\lambda)$ with $\lambda = \alpha/N_U$ can be obtained using Bayes' rule.

Finally, in Fig. 2(d) we investigate the number of basis measurements $N_U$ required to reach a certain value of $\Delta^2$ as a function of the number of qubits $n_q$ with $\varepsilon = 0.1$. We find that, although $N_U$ scales exponentially with $n_q$, i.e., $N_U \sim 2^{\gamma n_q}$, the exponent $\gamma \approx 0.82$, which is favorable compared to full quantum state tomography with $N_U \sim 3^{n_q}$ [37]. Therefore, the scheme is favorable for the near-term regime, where we are pushing the boundaries of the classical simulability of quantum systems.

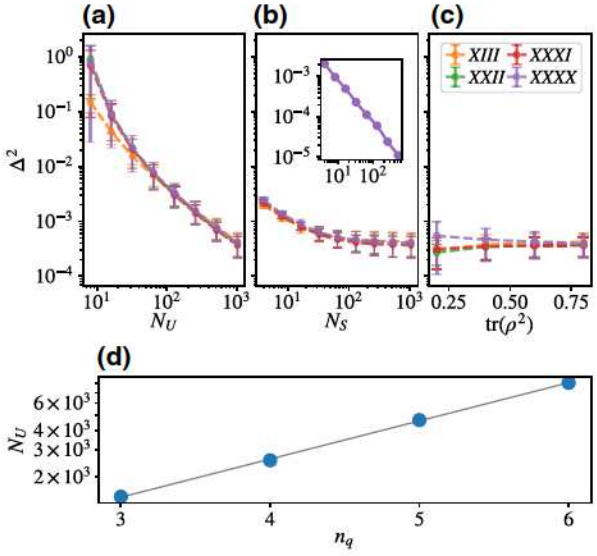


FIG. 2. Scaling of the mean squared error $\Delta^2$ for $n_q = 4$ qubits and error strength $\varepsilon = 0.1$: (a) the number of unitaries $N_U$ with a fixed number of shots $N_S = 1024$; (b) the number of shots with a fixed $N_U = 1024$. The inset shows the convergence to the final value with $N_S$. (c) The scaling of $\Delta^2$ with purity, $\mathrm{tr}(\rho^2)$, for $N_S = N_U = 1024$. The legend indicates the choices for $O$ [see Eq. (5)] in panels (a)–(c). Error bars are standard deviations of $\Delta^2$ over 100 random states. (d) The number of bases $N_U$ in order to reach mean squared error $\Delta^2$ versus the number of qubits $n_q$ for $N_S = 1$. The solid lines are fitting curves $N_U = c2^{\gamma n_q}$. For $\Delta = 0.01$, $\gamma = 0.82$.

## V. TRAPPED-ION EXPERIMENT

We illustrate the utility of our proposed SD method, by applying it to the existing data from an experiment with trapped-ion qubits [28]; see also Appendix F for more information on the experimental device.

In the experiment, a five-qubit GHZ state, i.e., $|\psi_{\mathrm{GHZ}}\rangle = (|0\rangle^{\otimes 5} + |1\rangle^{\otimes 5})/\sqrt{2}$ is prepared. This is a stabilizer state with generators $\mathcal{G} = \{Z_1Z_2, Z_2Z_3, Z_3Z_4, Z_4Z_5, \prod_i X_i\}$, where we use $\prod_i X_i$ to denote $X_1X_2X_3X_4X_5$ [35]. Ideally, for this state, $\langle O\rangle = 1$ for all $O \in \mathcal{G}$. Because of experimental errors, the actual state $\tilde{\rho}_{\mathrm{GHZ}}$ differs from the ideal state and $\langle O\rangle \leq 1$. Here, we investigate how our proposed error-mitigation technique can improve estimates of these expectation values. Note that these expectation values can then be used to estimate the fidelity of the GHZ state [38–40]. A practical consideration in this experiment is that performing measurements in different bases takes roughly 1000 times longer than repeating measurements in a fixed basis. Therefore, it is interesting to explore the possibility of a trade-off between the $N_U$ and $N_S$ for a fixed measurement time.

To optimize resources, we first repeat our simulations by fixing state $|\psi_R\rangle$ in Eq. (4) to be a five-qubit GHZ state and setting $\varepsilon = 0.1$. This allows us to extract the scaling of errors with resources for this particular state. By examining the simulation data we empirically find that the MSE scales as $\Delta^2_{\mathrm{GHZ}} = (3384/N_U^2)(1 + 22/N_S^2)$, which is better than the average scaling that we observed in Fig. 2 for random states (see Appendix D). In Fig. 3 we compare our empirical fit with the numerically obtained contour and find good agreement between the two. Next, we model the experiment time by $T = N_U(1000 + N_S)$ to capture the trade-off between changing the measurement basis and repeating the measurements in the same basis. Finally, for a fixed $T$, we find the optimal choices of $N_S$ and $N_U$ that give us the lowest error (see Fig. 3). We note that the optimal choices of $N_S$ and $N_U$ obtained in our simulation may not be the optimal choices for the experiment, as their values may depend

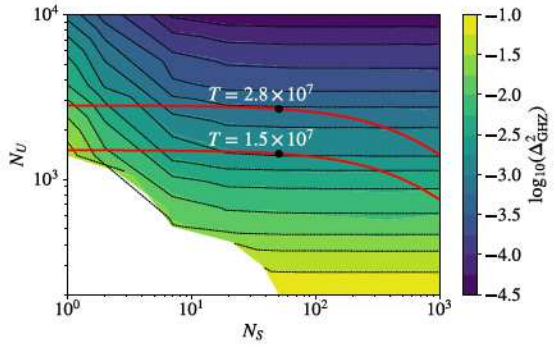


FIG. 3. Contour plot of the simulated error mitigation for the GHZ state for varying $N_S$ and $N_U$. The color bar indicates the logarithm of the mean squared error $\Delta^2_{\mathrm{GHZ}}$. The dashed lines indicate the contours obtained from the empirical fit $\Delta^2_{\mathrm{GHZ}} = (3384/N_U^2)(1 + 22/N_S^2)$. The red lines show the contours of fixed $T = N_U(1000 + N_S)$ for $T = 1.5 \times 10^7$ and $T = 2.8 \times 10^7$, with the circles indicating the optimal choices of $N_U$ and $N_S$. In the white region of the plot $\Delta^2_{\mathrm{GHZ}} > 0.1$.

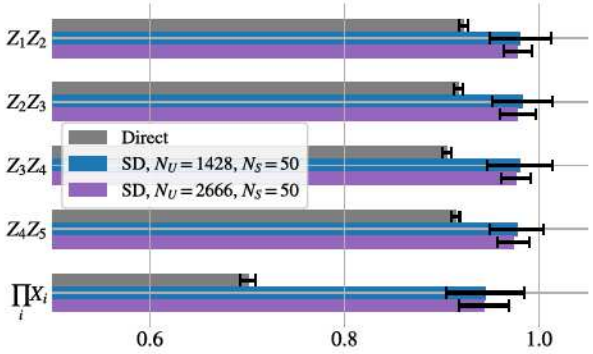


FIG. 4. We compare the experimental results of directly measured expectation values $\mathrm{tr}(\rho O)$ using $N_S = 2000$ with mitigated values $\langle O \rangle_2$ using $N_U = 1428, 2666$ and $N_S = 50$ measurements. The labels on the $y$ axis indicate the choices for $O$ and the $x$ axis shows the expectation value. Increasing $N_U$ reduces the error on the estimate. Error bars are standard deviations obtained from bootstrap sampling in SD and from normal approximation in the direct method.

on the specific error channel and the purity of the experimental state. Moreover, for a large-scale experiment, such simulations may not be feasible. However, as shown in Appendix D, optimizing the ratio of $N_S$ and $N_U$ for smaller experiments can guide us to better allocate resources in larger experiments.

After finding the optimal choices of $N_U$ and $N_S$ we resample the experimental measurement data of Ref. [28] and use our error-mitigation scheme to recover the expectation values of the stabilizers. Specifically, in Fig. 4 we observe that $\prod_i X_i$, which is the operator that is most severely affected by the errors and benefits the most from the SD scheme. In Appendix F, we simulate and analyze possible sources of errors in the experiment and, based on the performance of SD, identify detection errors and dephasing as major sources of noise in the system. Moreover, by increasing $N_U$ from 1428 to 2666, corresponding to the optimal choices for $T = 1.5 \times 10^7$ and $2.8 \times 10^7$, respectively (shown in Fig. 3), we observe that the error bars (standard deviations obtained by bootstrap resampling) in the mitigated values decrease (see Fig. 4). Note that while in this comparison the total number of measurements ($N_U N_S$) in the SD scheme is larger compared to the direct unmitigated method, increasing the number of measurements in the latter only reduces the variance and cannot help with reducing the bias errors from physical noise processes in the experiment.

## VI. DISCUSSION

In this work, we considered an alternative to error mitigation with multiple copies [13,41] and dual state schemes [11,14] that trades circuit volume with sample complexity. While these methods have great potential, the hardware requirement for implementing them can be demanding. In cases where the control and quantum resources required for implementing them is unavailable or limited, it can be useful to trade quantum resources for classical ones and use our proposed method.

We have shown that it is possible to mitigate state-preparation errors using classical shadows and provided numerical evidence of a better sample complexity of this approach compared to full state tomography. We also discussed the possibility of incorporating prior knowledge in our estimates and presented a scheme for optimizing measurement resources given experimental constraints. It is interesting to further develop these heuristics to enhance the capabilities of quantum devices in the near term.

Another aspect of the resource analysis, in addition to the sample complexity, is classical postprocessing. As mentioned earlier, the complexity of evaluating the mitigated expectation values using $M = N_U \times N_S$ snapshots scales as $\mathcal{O}(M^2)$. If our numerical error scaling persists [see Fig. 2(d)], we expect $M \sim 2^{0.82 n_q}$. Note that the second-order mitigation ($m = 2$) has its limitations and even with infinitely many measurements one cannot completely eliminate the errors. One can obtain the full density matrix by taking the average of the measurement snapshots, which allows mitigation with an arbitrary $m$. Therefore, the ultimate mitigation ($m \to \infty$) can be achieved by obtaining the dominant eigenvector of $\rho$ [41], which takes time $\mathcal{O}(2^{3n_q})$. However, taking the latter approach has a higher complexity than simulating the full quantum system and is unlikely to be useful beyond a proof-of-concept illustration. Therefore, we believe that the application of our proposed SD method is at the limit where storing and manipulating the full density matrix is out of reach, but storing the shadows and processing them is possible.

Finally, we note that the data collected for SD do not have to come from a single experimental platform. Combining data from different experiments might help with turning coherent errors into incoherent ones that can be mitigated using this scheme. Such a parallel approach helps mitigate errors when multiple experimental systems are available, but performing coherent operations between those systems is not possible.

## ACKNOWLEDGMENTS

We thank Andreas Elben, Hsin-Yuan Huang, and Benoît Vermersch for helpful discussions. We thank Norbert Linke for helpful comments and for sharing data from Ref. [28] for this work. We gratefully acknowledge Y. Zhu, A. M. Green, C. Huerta Alderete, and N. H. Nguyen who took the measurements. We acknowledge support from the ARO (Grants No. W911NF-18-1-0020, No. W911NF-18-1-0212), ARO MURI (Grants No. W911NF-16-1-0349, No. W911NF-21-1-0325), AFOSR MURI (Grants No. FA9550-19-1-0399, No. FA9550-21-1-0209), AFRL (Grant No. FA8649-21-P-0781), DoE Q-NEXT, NSF (Grants No. OMA-1936118, No. EEC-1941583, No. OMA-2137642), NTT Research, and the Packard Foundation (Grant No. 2020-71479). S.Z. acknowledges funding provided by the Institute for Quantum Information and Matter, an NSF Physics Frontiers Center (NSF Grant No. PHY-1733907). A.S. is supported by a Chicago Prize Postdoctoral Fellowship in Theoretical Quantum Science. Z.P. is supported by AFOSR Grant No. FA9550-19-1-0399, ARO Grants No. W911NF2010232 and No. W911NF-15-1-0397, and the NSF Physics Frontier Center at the Joint Quantum Institute.

*Note added.* Recently, we became aware of a related work [42] that uses similar techniques for error mitigation.

## APPENDIX A: AN UNBIASED ESTIMATOR FOR $\mathrm{tr}(O\rho^2)$

In this section, we show that our estimator in Eq. (3) is unbiased. We first remind the reader that $\{U_j\}_{j=1}^{N_U}$ denotes the $N_U$ sampled unitary operators from random local Clifford gates, and that $\{|b^{(i_j)}\rangle\}_{i_j=1}^{N_S}\}$ denotes the measurement outcomes of $N_S$ measurements fixing $U = U_j$. We can then expand $\hat{\rho}_j = (1/N_S)\sum_{i_j=1}^{N_S}\hat{\rho}_{U_j,b^{(i_j)}}$ in Eq. (3) and calculate its expectation value

$$\begin{aligned}
\mathbb{E}\hat{o}_2 &= \frac{1}{N_U(N_U-1)}\frac{1}{N_S^2}\mathbb{E}\sum_{\substack{i,i',j,j'\\ i\neq i'}}\mathrm{tr}[V^{(2)}(O\hat{\rho}_{U_j,b^{(i_j)}})\otimes\hat{\rho}_{U_{j'},b^{(i'_{j'})}}]\\
&= \frac{1}{N_U(N_U-1)N_S^2}[\underbrace{N_U(N_U-1)N_S(N_S-1)\mathrm{tr}(O\rho^2)}_{i\neq i',j\neq j'}+\underbrace{N_U(N_U-1)N_S\mathrm{tr}(O\rho^2)}_{i\neq i',j=j'}]\\
&= \frac{N_U(N_U-1)N_S^2}{N_U(N_U-1)N_S^2}\mathrm{tr}(O\rho^2)\\
&= \mathrm{tr}(O\rho^2),
\end{aligned} \tag{A1}$$

where we have used the identity $\mathrm{tr}(O\rho^2) = \mathbb{E}_{U,b,U',b'}[\mathrm{tr}(V^{(2)}(O\hat{\rho}_{U,b})\otimes\hat{\rho}_{U',b'})]$ in the second line.

## APPENDIX B: DETAILS OF THE NUMERICAL SIMULATIONS

In this section, we provide the details of the numerical simulations performed for the scaling of the errors $\Delta^2$ with $N_U$, $N_S$, and the purity $\mathrm{tr}(\rho^2)$. We first generate a random mixed state defined in Eq. (4) by sampling a random unitary operator from the Haar distribution.

In order to generate the mixed state with certain purity $\mathrm{tr}(\rho^2)$, we note that the purity is solely determined by the parameter $\varepsilon$ and

$$\mathrm{tr}(\rho_R^2) = (1-\varepsilon)^2 + \left(\frac{\varepsilon}{2^{n_q}-1}\right)^2. \tag{B1}$$

One can therefore vary the parameter $\varepsilon$ to tune the purity of the mixed state.

To estimate the squared error $\Delta_R^2$ defined in Eq. (5) as a function of $N_U$, $N_S$, and $\mathrm{tr}(\rho^2)$, we use the bootstrap resampling technique. We perform randomized measurements for 10 000 different random bases, each with 10 000 shots. These data form the empirical distribution of the classical shadow for a given $\rho_R$.

For a given pair of $(N_U, N_S)$, we sample the classical shadow for $N_U$ random bases and $N_S$ shots from the empirical distribution and estimate $\langle O\rangle_{(2)}$ using Eq. (3). The squared error of the estimation $\Delta_R$ is defined as the squared difference between the estimation and the exact value $\langle O\rangle_{(2)}$ as defined in Eq. (5). We perform the resampling 250 times to obtain the average of $\Delta_R^2$, denoted $\overline{\Delta_R^2}$.

Finally, we average over the random mixed states $\rho_R$ by generating $N_R = 100$ different random mixed states $\rho_R$ and calculate $\Delta^2 = (1/N_R)\sum_R \overline{\Delta_R^2}$. The standard deviation used for plotting the error bars is given by $\mathrm{std}(\Delta^2) = \sqrt{\sum_R (1/N_R)(\Delta^2 - \overline{\Delta_R^2})^2}$.

## APPENDIX C: ANALYTICAL UPPER BOUNDS ON THE ESTIMATION VARIANCE AS A FUNCTION OF $N_S$ AND $N_U$

### 1. Definition

Given an $n$-qubit ($n = n_q$ in the main text) quantum state $\rho$, we perform a random local Clifford unitary $U$ operation on $\rho$ and then perform the computational basis measurement $N_S$ times. Suppose that $\{|b^{(i)}\rangle\}_{i=1}^{N_S}$ are the measurement outcomes (note that here the $b^{(i)}$ are $n$-bit strings); then in this section, we define the following unbiased estimator of $\rho$:

$$\hat{\rho} = \frac{1}{N_S}\sum_{i=1}^{N_S} \hat{\rho}_{U,b^{(i)}} = \frac{1}{N_S}\sum_{i=1}^{N_S} \mathcal{M}^{-1}(U^\dagger |b^{(i)}\rangle \langle b^{(i)}| U)$$

with

$$\begin{aligned}
\mathcal{M}(\rho) &= \frac{1}{N_S}\sum_{i=1}^{N_S} \mathbb{E}[U^\dagger |b^{(i)}\rangle \langle b^{(i)}| U] \\
&= \frac{1}{N_S}\sum_{i=1}^{N_S} \mathbb{E}_{U\sim\mathcal{U}} \sum_{\{b^{(i)}\}} [U^\dagger |b^{(i)}\rangle \langle b^{(i)}|U\rho U^\dagger|b^{(i)}\rangle \langle b^{(i)}| U] \\
&= \mathcal{D}_{1/3}^{\otimes n}(\rho), \\
\mathcal{M}^{-1}(\rho) &= (\mathcal{D}_{1/3}^{-1})^{\otimes n}(\rho).
\end{aligned}$$

Here $\mathcal{D}_{1/3}(\rho) = \frac{1}{3}\rho + \frac{1}{3}\mathrm{tr}(\rho)I$, $\mathcal{D}_{1/3}^{-1}(\rho) = 3\rho - \mathrm{tr}(\rho)I$, and $\mathcal{U}$ denotes the uniform distribution of local Clifford operations on $n$ qubits.

### 2. Variance of estimating $\mathrm{tr}(O\rho)$

Clearly, $\mathrm{tr}(O\hat{\rho})$ is an unbiased estimator of $\mathrm{tr}(O\rho)$. Now we compute its variance:

$$\begin{aligned}
\mathrm{Var}[\mathrm{tr}(O\hat{\rho})] &= \mathbb{E}_{U\sim\mathcal{U}} \sum_{\{b^{(i)}\}} \left(\prod_{i=1}^{N_S} \langle b^{(i)}|U\rho U^\dagger|b^{(i)}\rangle\right)\left(\frac{1}{N_S}\sum_{i=1}^{N_S} \langle b^{(i)}|U\mathcal{M}^{-1}(O)U^\dagger|b^{(i)}\rangle\right)^2 - \mathrm{tr}(\rho O)^2 \\
&= \frac{1}{N_S^2}\mathbb{E}_{U\sim\mathcal{U}} \sum_{\{b^{(i)}\}} \left(\prod_{i=1}^{N_S} \langle b^{(i)}|U\rho U^\dagger|b^{(i)}\rangle\right) \sum_{i,i'=1}^{N_S} \langle b^{(i)}|U\mathcal{M}^{-1}(O)U^\dagger|b^{(i)}\rangle \langle b^{(i')}|U\mathcal{M}^{-1}(O)U^\dagger|b^{(i')}\rangle - \mathrm{tr}(\rho O)^2 \\
&= \frac{1}{N_S}\left(\mathbb{E}_{U\sim\mathcal{U}} \sum_{b} \langle b|U\rho U^\dagger|b\rangle \langle b|U\mathcal{M}^{-1}(O)U^\dagger|b\rangle^2 - \mathrm{tr}(\rho O)^2\right) \\
&\quad + \frac{N_S - 1}{N_S}\left(\mathbb{E}_{U\sim\mathcal{U}} \sum_{b,b'} \langle b|U\rho U^\dagger|b\rangle \langle b'|U\rho U^\dagger|b'\rangle \langle b|U\mathcal{M}^{-1}(O)U^\dagger|b\rangle \langle b'|U\mathcal{M}^{-1}(O)U^\dagger|b'\rangle - \mathrm{tr}(\rho O)^2\right).
\end{aligned}$$

It is known from Proposition S3 of Ref. [18] that, when $O$ is a weight-$k$ operator and has a Pauli decomposition $O = \sum_{\mathbf{p}} \alpha_{\mathbf{p}} P_{\mathbf{p}}$, $\mathbf{p} \in \{I, X, Y, Z\}^k$ (note that we use $O$ as shorthand for $O \otimes I^{\otimes n-k}$ acting on $n$ qubits), the first term is equal to

$$\begin{aligned}&\frac{1}{N_S}\mathbb{E}_{U\sim\mathcal{U}} \sum_b \langle b|U\rho U^\dagger|b\rangle \, \langle b|U\mathcal{M}^{-1}(O)U^\dagger|b\rangle^2 \\ &\quad - \frac{1}{N_S}\mathrm{tr}(\rho O)^2 \\ &\quad = \frac{1}{N_S}\left(\frac{1}{3^k}\sum_{\mathbf{s}\in\{X,Y,Z\}^k} \mathrm{tr}(\rho O_{\mathbf{s}}^2) - \mathrm{tr}(\rho O)^2\right),\end{aligned}$$

where $O_{\mathbf{s}} = \sum_{\mathbf{q}\triangleright\mathbf{s}} 3^{|\mathbf{q}|}\alpha_{\mathbf{q}} P_{\mathbf{q}}$, and $\mathbf{q} \triangleright \mathbf{s}$ means that $q_i$ is equal to either $s_i$ or $I$ for all $i$.

Now we compute the second term. We first compute

$$\mathbb{E}_{U\sim\mathcal{U}} \sum_{b,b'} (U\otimes U)\,|bb'\rangle\langle bb'|\,(U^\dagger\otimes U^\dagger)\langle b|U^\dagger\mathcal{M}^{-1}(P_{\mathbf{p}})U|b\rangle$$

$$\langle b'|U^\dagger\mathcal{M}^{-1}(P_{\mathbf{q}})U|b'\rangle = \bigotimes_{i=1}^n F(p_i, q_i),$$

where $P_{\mathbf{p}}$, $P_{\mathbf{q}}$ are Pauli operators.

$$\begin{aligned}F(p_i, q_i) = \mathbb{E}_{U_1\in\mathcal{U}_1} \sum_{x_1,x_2=0}^{1} &(U_1\otimes U_1)\,|x_1x_2\rangle\,\langle x_1x_2|\,(U_1^\dagger\otimes U_1^\dagger) \\ &\langle x_1|U_1^\dagger P_{p_i}U_1|x_1\rangle\langle x_2|U_1^\dagger P_{q_i}U_1|x_2\rangle\,,\end{aligned}$$

where $\mathcal{U}_1$ is the uniform distribution of Clifford gates on one qubit. After a few calculations, we obtain

$$F(p_i, q_i) = \begin{cases} I\otimes I, & p_i = q_i = 0, \\ \frac{1}{3}P_{p_i}\otimes P_{q_i}, & p_i = q_i \neq 0,\ p_i = 0, q_i \neq 0,\ \text{or } p_i \neq 0, q_i = 0, \\ 0, & \text{otherwise.}\end{cases}$$

Let

$$f(\mathbf{p},\mathbf{q}) = \begin{cases} 0, & \exists i \text{ such that } p_i \neq q_i \text{ and } p_i, q_i \neq I, \\ 3^s, & s = |\{i : p_i = q_i, p_i \neq I\}|; \end{cases}$$

then the second term is equal to

$$\begin{aligned}&\frac{N_S-1}{N_S}\mathbb{E}_{U\sim\mathcal{U}} \sum_{b,b'} \langle b|U\rho U^\dagger|b\rangle \, \langle b'|U\rho U^\dagger|b'\rangle \\ &\quad \langle b|U\mathcal{M}^{-1}(O)U^\dagger|b\rangle \, \langle b'|U\mathcal{M}^{-1}(O)U^\dagger|b'\rangle - \mathrm{tr}(\rho O)^2 \\ &\quad = \mathrm{tr}\left((\rho\otimes\rho)\sum_{\mathbf{pq}}\alpha_{\mathbf{p}}\alpha_{\mathbf{q}}\bigotimes_{i=1}^n F(p_i,q_i)\right) - \mathrm{tr}(\rho O)^2 \\ &\quad = \sum_{\mathbf{pq}}\alpha_{\mathbf{p}}\alpha_{\mathbf{q}} f(\mathbf{p},\mathbf{q})\mathrm{tr}(\rho P_{\mathbf{p}})\mathrm{tr}(\rho P_{\mathbf{q}}) - \mathrm{tr}(\rho O)^2 \\ &\quad = \frac{1}{3^k}\sum_{\mathbf{s}\in\{X,Y,Z\}^k} \mathrm{tr}(\rho O_{\mathbf{s}})^2 - \mathrm{tr}(\rho O)^2,\end{aligned}$$

where the last equality can be verified by comparing the coefficients of $\alpha_{\mathbf{p}}\alpha_{\mathbf{q}}$.

Therefore, we have

$$\begin{aligned}\mathrm{Var}[\mathrm{tr}(O\hat{\rho})] &= \frac{1}{N_S}\left(\frac{1}{3^k}\sum_{\mathbf{s}\in\{X,Y,Z\}^k}\mathrm{tr}(\rho O_{\mathbf{s}}^2) - \mathrm{tr}(\rho O)^2\right) \\ &\quad + \frac{N_S-1}{N_S}\left(\frac{1}{3^k}\sum_{\mathbf{s}\in\{X,Y,Z\}^k}\mathrm{tr}(\rho O_{\mathbf{s}})^2 - \mathrm{tr}(\rho O)^2\right) \\ &=: u_0(O,\rho) + \frac{1}{N_S}u_1(O,\rho), \qquad \text{(C1)}\end{aligned}$$

where

$$\begin{aligned}u_0(O,\rho) &:= \frac{1}{3^k}\sum_{\mathbf{s}\in\{X,Y,Z\}^k}\mathrm{tr}(\rho O_{\mathbf{s}})^2 - \mathrm{tr}(\rho O)^2 \\ \text{and}\quad u_1(O,\rho) &:= \frac{1}{3^k}\sum_{\mathbf{s}\in\{X,Y,Z\}^k}\mathrm{tr}(\rho O_{\mathbf{s}}^2) - \mathrm{tr}(\rho O_{\mathbf{s}})^2 \end{aligned} \qquad \text{(C2)}$$

are all non-negative functions of $O$ and $\rho$. It is known from Ref. [18] that

$$u_0(O,\rho) + u_1(O,\rho) \leq 2^k\mathrm{tr}(O^2).$$

Clearly, when $u_0(O,\rho)$ is significantly smaller than $u_1(O,\rho)$, a large $N_S$ is desirable.

### 3. Variance of estimating $\mathrm{tr}(O\rho^2)$

Let $\hat{o}_2 = [1/N_U(N_U-1)]\sum_{j\neq j'}\mathrm{tr}(V^{(2)}\hat{\rho}_j\otimes(O\hat{\rho}_{j'}))$. This is an unbiased estimator of $o_2=\mathrm{tr}(O\rho^2)$. Now we compute its variance. Let $O^{(2)} := \frac{1}{2}(V^{(2)}(I\otimes O)+(I\otimes O)V^{(2)})$. We have

$$\hat{o}_2 = \binom{N_U}{2}^{-1}\sum_{j<j'}\mathrm{tr}(O^{(2)}\hat{\rho}_j\otimes\hat{\rho}_{j'}),$$

where $\hat{\rho}_j = (1/N_S)\sum_{i_j=1}^{N_S}\mathcal{M}^{-1}(U_j^\dagger|b^{(i_j)}\rangle\langle b^{(i_j)}|U_j)$, $\{U_j\}_{j=1}^{N_U}$ is sampled from random local Clifford gates, and $\{|b^{(i_j)}\rangle_{i_j=1}^{N_S}\}$ are measurement outcomes of $N_S$ measurements fixing $U=U_j$. A total $N_SN_U$ number of measurements are performed.

In order to derive an upper bound of $\mathrm{Var}[\hat{o}_2]$, we first note from our discussion above that

$$\mathrm{Var}[\mathrm{tr}(A\hat{\rho})]\le u_0(A,\rho)+\frac{1}{N_S}u_1(A,\rho)$$

for an arbitrary Hermitian operator $A$, and, for $j\neq j'$,

$$\mathrm{Var}[\mathrm{tr}(O^{(2)}\hat{\rho}_j\otimes\hat{\rho}_{j'})]\le u_0(O^{(2)},\rho\otimes\rho)+\frac{1}{N_S}u_1(O^{(2)},\rho\otimes\rho).$$

Consider

$$(\hat{o}_2)^2 = \binom{N_U}{2}^{-2}\sum_{j<j'}\sum_{k<k'}\mathrm{tr}(O^{(2)}\hat{\rho}_j\otimes\hat{\rho}_{j'})\mathrm{tr}(O^{(2)}\hat{\rho}_k\otimes\hat{\rho}_{k'}).$$

There are $\binom{N_U}{2}^2$ terms in total. For terms where all indices are distinct [$\binom{N_U}{2}\binom{N_U-2}{2}$ terms in all],

$$\mathbb{E}[\mathrm{tr}(O^{(2)}\hat{\rho}_j\otimes\hat{\rho}_{j'})\mathrm{tr}(O^{(2)}\hat{\rho}_k\otimes\hat{\rho}_{k'})]=\mathrm{tr}(\rho^2O)^2;$$

for terms where two of the indices coincide [$2\binom{N_U}{2}(N_U-2)$ terms in all],

$$\mathbb{E}[\mathrm{tr}(O^{(2)}\hat{\rho}_j\otimes\hat{\rho}_{j'})\mathrm{tr}(O^{(2)}\hat{\rho}_k\otimes\hat{\rho}_{k'})]=\mathbb{E}[\mathrm{tr}((\hat{\rho}\otimes\rho)O^{(2)})^2]=\mathbb{E}[\mathrm{tr}(\hat{\rho}A)^2],$$

where $A:=\frac{1}{2}(\rho O+O\rho)$; for terms where $(j,j')$ coincides with $(k,k')$ [$\binom{N_U}{2}$ terms in all],

$$\mathbb{E}[\mathrm{tr}(O^{(2)}\hat{\rho}_j\otimes\hat{\rho}_{j'})\mathrm{tr}(O^{(2)}\hat{\rho}_k\otimes\hat{\rho}_{k'})]=\mathbb{E}[\mathrm{tr}(O^{(2)}\hat{\rho}_j\otimes\hat{\rho}_{j'})^2].$$

Then we have

$$\begin{aligned}\mathrm{Var}[\hat{o}_2] &= \binom{N_U}{2}^{-1}\left(2(N_U-2)\left(u_0(A,\rho)+\frac{1}{N_S}u_1(A,\rho)\right)+\left(u_0(O^{(2)},\rho\otimes\rho)+\frac{1}{N_S}u_1(O^{(2)},\rho\otimes\rho)\right)\right)\\ &\le \frac{4}{N_U}\left(u_0(A,\rho)+\frac{1}{N_S}u_1(A,\rho)\right)+\frac{4}{N_U^2}\left(u_0(O^{(2)},\rho\otimes\rho)+\frac{1}{N_S}u_1(O^{(2)},\rho\otimes\rho)\right).\end{aligned}\tag{C3}$$

In particular, when $O=I$, we have

$$\begin{aligned}\mathrm{Var}[\hat{s}_2] &= \binom{N_U}{2}^{-1}\left(2(N_U-2)\left(u_0(\rho,\rho)+\frac{1}{N_S}u_1(\rho,\rho)\right)+\left(u_0(V^{(2)},\rho\otimes\rho)+\frac{1}{N_S}u_1(V^{(2)},\rho\otimes\rho)\right)\right)\\ &\le \frac{4}{N_U}\left(u_0(\rho,\rho)+\frac{1}{N_S}u_1(\rho,\rho)\right)+\frac{4}{N_U^2}\left(u_0(V^{(2)},\rho\otimes\rho)+\frac{1}{N_S}u_1(V^{(2)},\rho\otimes\rho)\right).\end{aligned}\tag{C4}$$

## APPENDIX D: NUMERICAL SIMULATIONS OF THE GHZ STATE

In this section, we provide more information about the simulations of the GHZ state used for producing Fig. 3. Additionally, we investigate how the optimal choices of $N_U$ and $N_S$ vary as we change the number of qubits.

We generate a five-qubit GHZ state with $\varepsilon=0.1$ and, for each value of $N_S$ and $N_U$, simulate the randomized measurement protocol 1000 times. We then calculate the mean squared error $\Delta^2_{\mathrm{GHZ}}$ using these samples; see Fig. 5. Based on the observed scaling for large $N_S$ and $N_U$, we use the expression $\Delta^2_{\mathrm{GHZ}}=(c_1/N_U^2)(1+c_2/N_S^2)$ to fit the data and find that $c_1=3384$ and $c_2=22$. Since the values $\Delta^2_{\mathrm{GHZ}}$ span orders of magnitudes, we use $\log_{10}(\Delta^2_{\mathrm{GHZ}})$ to fit the data and capture the correct behavior across a large range of values.

We now repeat these simulations for $n_q=3,4,6$ using the same parameters and number of samples. We then obtain the empirical fit, and similar to the procedure used in Fig. 3 we obtain the optimal choices of $N_S$ and $N_U$ for given measurement budgets. The results shown in Fig. 6 illustrate that the optimal distribution of resources for smaller systems can guide us in choosing the appropriate $N_S$ and $N_U$ for larger systems that may eventually be out of reach for a classical computer. The results also reinforce the

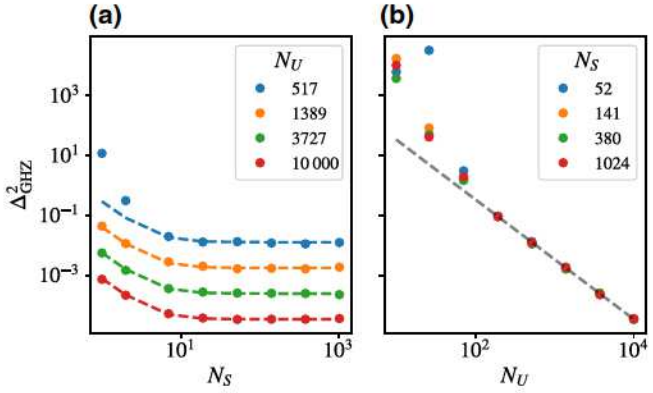


FIG. 5. Scaling of the mean squared error $\Delta^2_{\mathrm{GHZ}}$ for a five-qubit GHZ state for varying (a) $N_S$ and (b) $N_U$. The dashed lines indicate the empirical fits $\Delta^2_{\mathrm{GHZ}} = (3384/N_U^2)(1 + 22/N_S^2)$.

observation that increasing $N_S$ is only helpful to a certain degree and increasing $N_S$ beyond this point has a diminishing return in reducing the errors.

## APPENDIX E: A BIASED ESTIMATOR FOR PURITY

As noted in the main text, it might be beneficial to incorporate prior knowledge about the value of purity $\mathrm{tr}(\rho^2)$ to reduce the errors. We now show one approach to incorporating prior knowledge using a Gaussian prior and Bayes' rule.

Let $\mu$ denote the true value of $\mathrm{tr}(\rho^2)$ and assume that we have a prior belief that $\mu \sim N(\mu_0, \sigma_0^2)$, i.e., a normal distribution with mean $\mu_0$ and variance $\sigma_0^2$. Next, assume that after performing an experiment we estimate the purity to be $s_2$. We also assume that this observation is normally distributed with the variance, $\sigma^2$, that is known. Therefore, based on our measurements and assumptions we have

$$\Pr(s_2|\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(s_2-\mu)^2}{2\sigma^2}\right]. \tag{E1}$$

Moreover, our prior is

$$\Pr(\mu) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left[-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right]. \tag{E2}$$

Then, using Bayes' rule $\Pr(\mu|s_2) = \Pr(\mu)\Pr(s_2|\mu)/\Pr(s_2)$, we find the posterior

$$\begin{aligned}\Pr(\mu|s_2) &\propto \Pr(\mu)\Pr(s_2|\mu) \\ &= \exp\left[-\frac{(s_2-\mu)^2}{2\sigma^2} - \frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right] \\ &\propto \exp\left[-\frac{(\mu-\mu')^2}{2\sigma'^2}\right],\end{aligned} \tag{E3}$$

where our updated mean and variance are

$$\mu' = \frac{\mu_0\sigma^2 + s_2\sigma_0^2}{\sigma^2 + \sigma_0^2}, \tag{E4}$$

$$\sigma'^2 = \frac{\sigma^2\sigma_0^2}{\sigma^2 + \sigma_0^2}. \tag{E5}$$

We now use $\mu'$ as our estimator for purity. We assume that $\sigma^2 \propto 1/N_U$, and define a parameter $\alpha$ such that $\sigma^2/\sigma_0^2 = \alpha/N_U$. We then have

$$\mu' = \frac{s_2 + \alpha\mu_0/N_U}{1 + \alpha/N_U}. \tag{E6}$$

We can then treat $\alpha$ as a hyperparameter that quantifies our confidence in our initial guess. Large values of $\alpha$ indicate our high confidence in $\mu_0$.

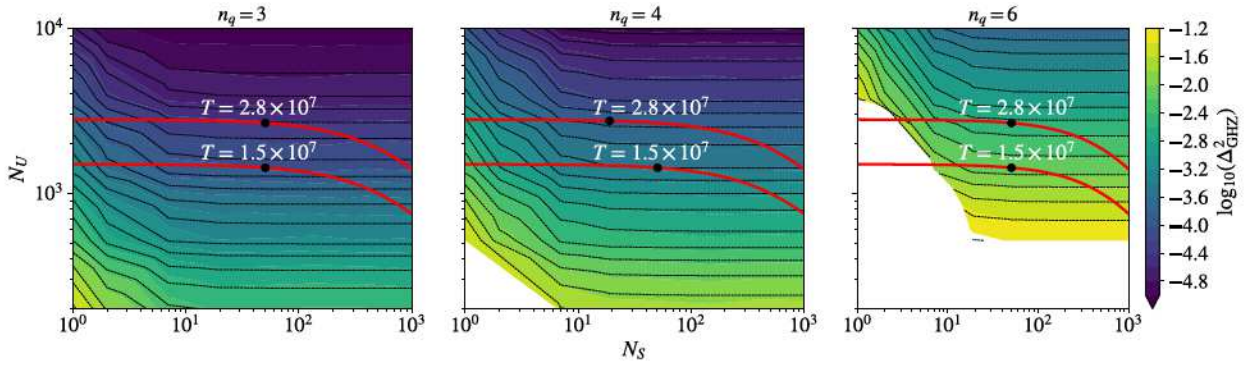


FIG. 6. Optimal choices of $N_U$ and $N_S$ for a given measurement budget for different numbers of qubits $n_q = 3, 4, 6$. The color bar indicates the logarithm of the mean squared error $\Delta^2_{\mathrm{GHZ}}$. The dashed lines indicate the contours obtained from the empirical fit $\Delta^2_{\mathrm{GHZ}} = (c_1/N_U^2)(1 + c_2/N_S^2)$. The red lines show the contours of fixed $T = N_U(1000 + N_S)$ for $T = 1.5 \times 10^7$ and $T = 2.8 \times 10^7$, with the circles indicating the optimal choices of $N_U$ and $N_S$. In the white region of the plot $\Delta^2_{\mathrm{GHZ}} > 0.1$. We observe that, in this problem and measurement budget regime, the optimal choices of $N_U$ and $N_S$ remain similar as we change the number of qubits.

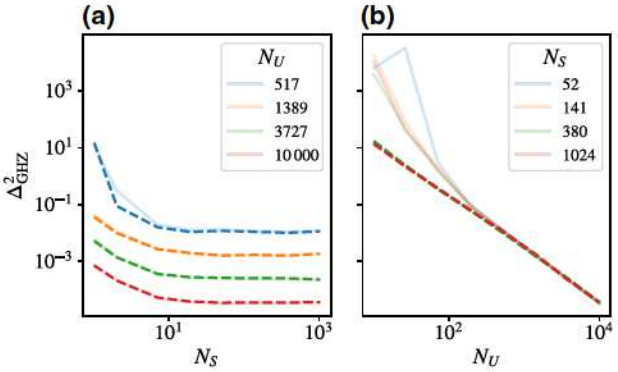


FIG. 7. Using the biased estimator for purity in Eq. (E6) on the data from Fig. 5 reduces the errors for smaller values of $N_U$. The dashed lines show the errors calculated using the biased estimator; the solid lines are the original data from Fig. 5.

This method is particularly useful if we have a good guess about the purity of the state in our experiment. To illustrate, we apply this modified estimator to our data in Fig. 5, with $\mu_0 = 0.9$ and $\alpha = 100$. The true value of purity in this case is 0.81. The results in Fig. 7 show that, even with more than 10% error in the prior, using this biased estimator improves the errors for smaller values of $N_U$.

## APPENDIX F: DETAILS OF THE EXPERIMENT AND ERRORS

### 1. Experimental setup

The trapped-ion experiment is performed on a quantum computer consisting of a chain of nine $^{171}Yb^+$ ions confined in a Paul trap with blade electrodes. Typical single- and two-qubit gate fidelities are 99.5(2)% and 98%–99%. Detailed performance of the system is described in Ref. [43]. The GHZ state in the experiment is prepared by running the circuit show in Fig. 8 on five qubits. The circuit utilizes the two-qubit gate $R_{XX}(\theta) = \exp(-i\theta XX/2)$, and the single-qubit rotations $R_\alpha(\theta) = \exp(-i\sigma_\alpha\theta/2)$ with $\alpha = x, y, z$.

### 2. Error channels and simulation

In this section, we describe the detailed implementations for the simulation of the error channels. We simulate the circuit in Fig. 8 on a classical computer.

To simulate coherent errors, we replace the $R_{XX}(\theta)$ gate by $R_{XX}(\theta(1+\delta_{\mathrm{coh}}))$, where $\delta_{\mathrm{coh}}$ is the over-rotation rate. The dephasing error is simulated by applying the following noise channel at the end of the simulation:

$$\mathcal{C}(\rho) = (1 - p_{\mathrm{deph}})\rho + \frac{p_{\mathrm{deph}}}{n_q}\sum_{i=1}^{n_q} Z_i \rho Z_i \tag{F1}$$

with $p_{\mathrm{deph}}$ the dephasing rate.

To simulate detection errors in the measurements, we first rotate the density matrix to the basis that the measurement will be performed. For example, to measure the $\prod_{i=1}^{n_q} X_i$ operator, we perform Hadamard rotation for all the qubits. After the rotation, we take the diagonal part of the density matrix, $P$. It corresponds to the probability distribution of the measurement outcomes. We then apply the detection error matrix, $M$, to the probability distribution $P$. In this work, we focus on uncorrelated detection errors. Matrix $M$ in this case is given by

$$M = \bigotimes_{i=1}^{n_q} A_i \tag{F2}$$

with

$$A_i = \begin{pmatrix} 1-p_0 & p_1 \\ p_0 & 1-p_1 \end{pmatrix}. \tag{F3}$$

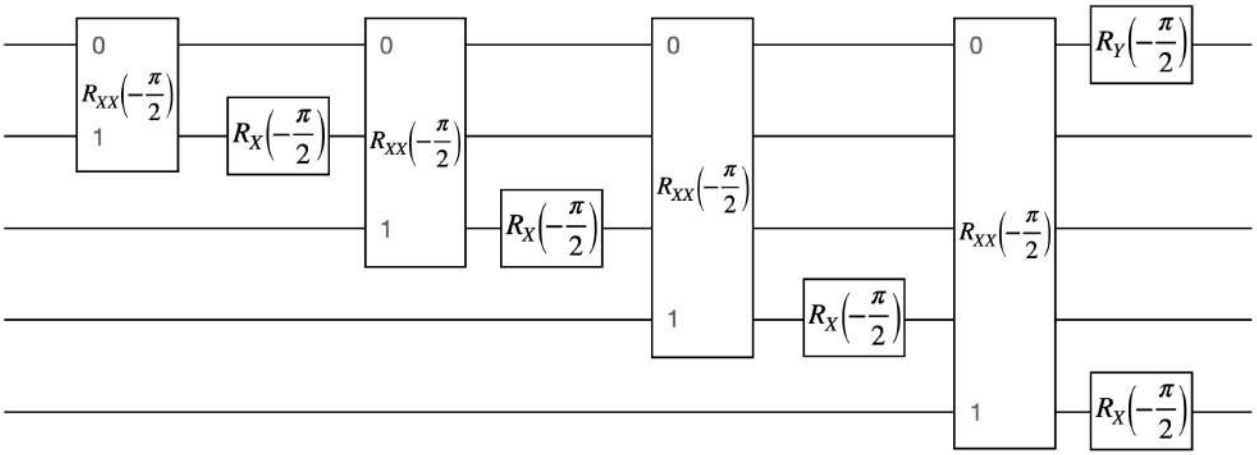


FIG. 8. The circuit used for preparing the GHZ state in the trapped-ion quantum computer, where $R_{XX}(\theta) = \exp(-i\theta XX/2)$, $R_\alpha(\theta) = \exp(-i\sigma_\alpha\theta/2)$ and $\alpha = x, y, z$.

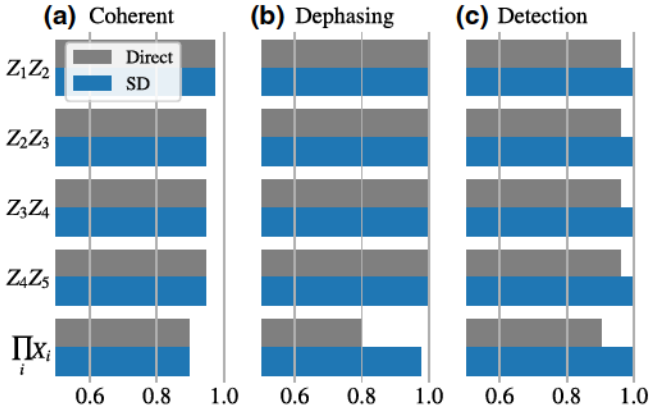


FIG. 9. Simulated effect of errors on second-order mitigation. We compare the direct approach $\mathrm{tr}(\rho O)$ (Direct) against the second-order mitigation $\langle O\rangle_{(2)}$ (SD) for (a) coherent error with $\delta_{\mathrm{coh}} = 0.15$ relative over-rotation, (b) single-qubit dephasing error with $p_{\mathrm{deph}} = 0.1$, and (c) detection errors with $p_{\mathrm{det}} = 0.01$.

Here $p_0$ ($p_1$) denotes the probability that the detector gives outcome 1 (0) where the true outcome should be 0 (1). We assume that $p_0 = p_1 = p_{\mathrm{det}}$ for simplicity. After the application of $M$, we calculate the expectation value of the observables based on the modified probability distribution.

To simulate the measurement of second-order mitigation $\langle O\rangle_{(2)}$ with detection errors, we first simulate the measurement of all $4^n$ Pauli strings with detection errors using the method described in the previous section. We then define the reconstructed density matrix as $\rho = (1/2^{nq})\sum_{k=0}^{4^{nq}-1} c_k P_k$, where $P_k$ is the $k$th Pauli string operator and $c_k$ is the simulated measurement result of $P_k$ with detection errors. Finally, the second-order mitigation is computed as $\langle O\rangle_{(2)} = \mathrm{tr}(O\rho^2)/\mathrm{tr}(\rho^2)$.

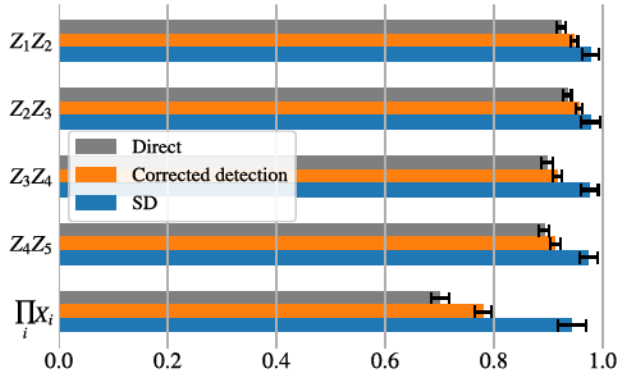


FIG. 10. Comparison of the experimental data (Direct), corrected data using detection calibration (Calibrated detection), and shadows distillation (SD) with $N_U = 2666$ and $N_S = 50$ measurements. The gap between the corrected detection and SD data in $\langle \prod_i X_i\rangle$ and its absence in the $\langle ZZ\rangle$ can be explained by the existence of dephasing error in the experiment.

### 3. Analysis of errors

In addition to correcting the expectation values, our method also reveals some facts about the nature of errors in the system. We first note that static coherent errors do not benefit from SD [see Fig. 9(a)]. This is because these errors change the eigenstates of $\rho$ while leaving the eigenvalues unaffected. From the experimental results in Fig. 4, we can see that $\prod_i X_i$ is the operator that is most affected by the errors. However, the fact that it benefits considerably from the error-mitigation protocol suggests that the errors are mostly incoherent. These observations are further validated by the numerical simulation of coherent errors [Fig. 9(a)], dephasing errors [Fig. 9(b)], and detection errors [Fig. 9(c)]. We see that, unlike coherent errors, the latter two benefit from SD. The contrast between the $ZZ$ and $\prod_i X_i$ can be due to either dephasing or detection errors. However, in the next section, we provide a detailed analysis using a different error-mitigation technique [44] that only mitigates detection errors, and show that it is unlikely that detection errors are the only source of errors in this experiment. The residual errors in Fig. 4 either correspond to higher-order incoherent errors, incoherent errors that modify the eigenvectors of $\rho$ (also known as the coherent mismatch [13,15,41]), or coherent errors originating from under(over)-rotation in two-qubit gates, which is a known source of error in trapped-ion systems [45].

### 4. Correcting detection errors

It is also possible to correct detection errors by first calibrating matrix $M$ in Eq. (F2) in the experiment and applying $M^{-1}$ to the vector of outcome probabilities obtained from the measurements [44]. We apply this procedure to the experimentally obtained expectation values and show the results in Fig. 10. We observe that the corrected expectation values are still lower than those obtained from SD, which indicates that SD is mitigating errors beyond just those in the detection process.

[1] D. A. Lidar and T. A. Brun, *Quantum Error Correction* (Cambridge university press, Cambridge, 2013).
[2] L. Egan, D. M. Debroy, C. Noel, A. Risinger, D. Zhu, D. Biswas, M. Newman, M. Li, K. R. Brown, and M. Cetina, *et al.*, Nature (2021), ISSN 1476-4687.
[3] K. Temme, S. Bravyi, and J. M. Gambetta, Error Mitigation for Short-Depth Quantum Circuits, Phys. Rev. Lett. **119,** 180509 (2017).
[4] S. Endo, S. C. Benjamin, and Y. Li, Practical Quantum Error Mitigation for Near-Future Applications, Phys. Rev. X **8,** 031027 (2018).
[5] P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio, Error mitigation with Clifford quantum-circuit data, Quantum **5,** 592 (2021).

[6] A. Strikis, D. Qin, Y. Chen, S. C. Benjamin, and Y. Li, Learning-Based Quantum Error Mitigation, PRX Quantum **2**, 040330 (2021).

[7] A. Lowe, M. H. Gordon, P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio, Unified approach to data-driven quantum error mitigation, Phys. Rev. Res. **3**, 033098 (2021).

[8] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, Error mitigation extends the computational reach of a noisy quantum processor, Nature **567**, 491 (2019).

[9] J. Vovrosh, K. E. Khosla, S. Greenaway, C. Self, M. S. Kim, and J. Knolle, Simple mitigation of global depolarizing errors in quantum simulations, Phys. Rev. E **104**, 035309 (2021).

[10] A. Peres, Error symmetrization in quantum computers, Int. J. Theor. Phys. **38**, 799 (1999).

[11] Z. Cai, Resource-efficient Purification-based Quantum Error Mitigation, (2021), arXiv preprint ArXiv:2107.07279.

[12] J. Cotler, S. Choi, A. Lukin, H. Gharibyan, T. Grover, M. E. Tai, M. Rispoli, R. Schittko, P. M. Preiss, and A. M. Kaufman, *et al.*, Quantum Virtual Cooling, Phys. Rev. X **9**, 031013 (2019).

[13] W. J. Huggins, S. McArdle, T. E. O'Brien, J. Lee, N. C. Rubin, S. Boixo, K. B. Whaley, R. Babbush, and J. R. McClean, Virtual Distillation for Quantum Error Mitigation, Phys. Rev. X **11**, 041036 (2021).

[14] M. Huo and Y. Li, Dual-state purification for practical quantum error mitigation, Phys. Rev. A **105**, 022427 (2022).

[15] B. Koczor, Exponential Error Suppression for Near-Term Quantum Devices, Phys. Rev. X **11**, 031057 (2021).

[16] Y. Xiong, S. X. Ng, and L. Hanzo, Quantum error mitigation relying on permutation filtering, IEEE Trans. Commun. **70**, 1927 (2022).

[17] K. Yamamoto, S. Endo, H. Hakoshima, Y. Matsuzaki, and Y. Tokunaga, Error-Mitigated Quantum Metrology via Virtual Purification, Phys. Rev. Lett. **129**, 250503 (2022).

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

[19] M. Paini, A. Kalev, D. Padilha, and B. Ruck, Estimating expectation values using approximate quantum states, Quantum **5**, 413 (2021).

[20] S. Chen, W. Yu, P. Zeng, and S. T. Flammia, Robust Shadow Estimation, PRX Quantum **2**, 030348 (2021).

[21] A. Elben, B. Vermersch, C. F. Roos, and P. Zoller, Statistical correlations between locally randomized measurements: A toolbox for probing entanglement in many-body quantum states, Phys. Rev. A **99**, 052323 (2019).

[22] A. Elben, B. Vermersch, R. van Bijnen, C. Kokail, T. Brydges, C. Maier, M. K. Joshi, R. Blatt, C. F. Roos, and P. Zoller, Cross-Platform Verification of Intermediate Scale Quantum Devices, Phys. Rev. Lett. **124**, 010504 (2020).

[23] A. Rath, R. van Bijnen, A. Elben, P. Zoller, and B. Vermersch, Importance Sampling of Randomized Measurements for Probing Entanglement, Phys. Rev. Lett. **127**, 200503 (2021).

[24] A. Elben, J. Yu, G. Zhu, M. Hafezi, F. Pollmann, P. Zoller, and B. Vermersch, Many-body topological invariants from randomized measurements in synthetic quantum matter, Sci. Adv. **6**, eaaz3666 (2020).

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

[26] Z.-P. Cian, H. Dehghani, A. Elben, B. Vermersch, G. Zhu, M. Barkeshli, P. Zoller, and M. Hafezi, Many-Body Chern Number from Statistical Correlations of Randomized Measurements, Phys. Rev. Lett. **126**, 050501 (2021).

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

[28] D. Zhu, Z.-P. Cian, C. Noel, A. Risinger, D. Biswas, L. Egan, Y. Zhu, A. M. Green, C. H. Alderete, N. H. Nguyen, *et al.*, Cross-platform comparison of arbitrary quantum states, Nat. Commun. **13**, 1 (2022).

[29] D. M. Greenberger, M. A. Horne, and A. Zeilinger, in *Bell's theorem, quantum theory and conceptions of the universe* (Springer, 1989), p. 69.

[30] P. Czarnik, A. Arrasmith, L. Cincio, and P. J. Coles, Qubit-efficient exponential suppression of errors, (2021), arXiv preprint ArXiv:2102.06056.

[31] N. M. Linke, S. Johri, C. Figgatt, K. A. Landsman, A. Y. Matsuura, and C. Monroe, Measuring the Rényi entropy of a two-site Fermi-Hubbard model on a trapped ion quantum computer, Phys. Rev. A **98**, 052334 (2018).

[32] N. M. Linke, D. Maslov, M. Roetteler, S. Debnath, C. Figgatt, K. A. Landsman, K. Wright, and C. Monroe, Experimental comparison of two quantum computing architectures, Proc. Nat. Acad. Sci. **114**, 3305 (2017).

[33] H. Hakoshima, Y. Matsuzaki, and S. Endo, Relationship between costs for quantum error mitigation and non-Markovian measures, Phys. Rev. A **103**, 012611 (2021).

[34] A. Elben, R. Kueng, H.-Y. R. Huang, R. van Bijnen, C. Kokail, M. Dalmonte, P. Calabrese, B. Kraus, J. Preskill, and P. Zoller, *et al.*, Mixed-State Entanglement from Local Randomized Measurements, Phys. Rev. Lett. **125**, 200501 (2020).

[35] D. Gottesman, The Heisenberg Representation of Quantum Computers, (1998), arXiv preprint ArXiv:quant-ph/9807006.

[36] S. Chen, J. Cotler, H.-Y. Huang, and J. Li, A Hierarchy for Replica Quantum Advantage, (2021), arXiv preprint ArXiv:2111.05874.

[37] R. O'Donnell and J. Wright, in *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing* (2016), p. 899.

[38] S. T. Flammia and Y.-K. Liu, Direct Fidelity Estimation from Few Pauli Measurements, Phys. Rev. Lett. **106**, 230501 (2011).

[39] M. P. da Silva, O. Landon-Cardinal, and D. Poulin, Practical Characterization of Quantum Devices without Tomography, Phys. Rev. Lett. **107**, 210404 (2011).

[40] A. Kalev, A. Kyrillidis, and N. M. Linke, Validating and certifying stabilizer states, Phys. Rev. A **99**, 042337 (2019).

[41] B. Koczor, The dominant eigenvector of a noisy quantum state, New J. Phys. **23**, 123047 (2021).

[42] H.-Y. Hu, R. LaRose, Y.-Z. You, E. Rieffel, and Z. Wang, Logical shadow tomography: Efficient estimation of error-mitigated observables, (2022), arXiv preprint ArXiv:2203.07263.

[43] K. A. Landsman, C. Figgatt, T. Schuster, N. M. Linke, B. Yoshida, N. Y. Yao, and C. Monroe, Verified quantum information scrambling, Nature **567**, 61 (2019).

[44] C. Shen and L. Duan, Correcting detection errors in quantum state engineering through data processing, New J. Phys. **14**, 053053 (2012).

[45] A. O. Maksymov, J. Nguyen, V. Chaplin, Y. Nam, and I. L. Markov, Detecting Qubit-coupling Faults in Ion-trap Quantum Computers, (2021), arXiv preprint ArXiv:2108.03708.