Unified approach to data-driven quantum error mitigation

Angus Lowe,[1,*] Max Hunter Gordon,[2,*] Piotr Czarnik [ORCID],[3] Andrew Arrasmith,[3] Patrick J. Coles,[3,4] and Lukasz Cincio[3,4]

[1]*Department of Combinatorics and Optimization and Institute for Quantum Computing, University of Waterloo,*
Waterloo, Ontario, Canada N2L 3G1
[2]*Instituto de Física Teórica, UAM/CSIC, Universidad Autónoma de Madrid, 28049 Madrid, Spain*
[3]*Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*
[4]*Quantum Science Center, Oak Ridge, Tennessee 37931, USA*

Achieving near-term quantum advantage will require effective methods for mitigating hardware noise. Data-driven approaches to error mitigation are promising, with popular examples including zero-noise extrapolation (ZNE) and Clifford data regression (CDR). Here, we propose a scalable error mitigation method that conceptually unifies ZNE and CDR. Our approach, called variable-noise Clifford data regression (vnCDR), significantly outperforms these individual methods in numerical benchmarks. vnCDR generates training data first via near-Clifford circuits (which are classically simulable) and second by varying the noise levels in these circuits. We employ a noise model obtained from IBM's Ourense quantum computer to benchmark our method. For the problem of estimating the energy of an 8-qubit Ising model system, vnCDR improves the absolute energy error by a factor of 33 over the unmitigated results and by factors of 20 and 1.8 over ZNE and CDR, respectively. For the problem of correcting observables from random quantum circuits with 64 qubits, vnCDR improves the error by factors of 2.7 and 1.5 over ZNE and CDR, respectively.

I. INTRODUCTION

Quantum computers are approaching the important milestone of having a demonstrable advantage over classical computers for practical applications, such as chemistry and materials science [1]. Such a quantum advantage is expected to be demonstrated with near-term devices that do not have the number of qubits or the gate fidelities required to implement full quantum error correction [2]. Nevertheless, the noise of such devices remains a serious obstacle to practical applications [3]. While near-term devices will not be able to completely remove errors caused by device noise, it is often possible to mitigate them.

Such so-called error mitigation (EM) techniques are sure to be an essential part of demonstrating the utility of quantum technologies, for example, for achieving chemical accuracy in chemistry applications. To this end, many distinct EM methods have been proposed [4,5]. One approach is to optimize quantum circuits using compiling and machine learning [6–8], while another employs variational quantum algorithms [9–11] to reduce circuit depth and potentially remove the effects of incoherent noise [12–17]. More recently, quantum phase estimation has been employed for error mitigation [18].

Zero-noise extrapolation (ZNE) is a classical postprocessing approach to EM that has received a significant amount of attention [4]. ZNE combines observables evaluated at several controlled noise levels through stretching gate times or inserting identities [19–23], enabling extrapolation to the zero-noise limit. Despite much success [24], this method is not without its limitations. Due to the uncertainty of the extrapolation, performance guarantees are difficult in general. In particular, ZNE struggles when a low-degree polynomial fit to the noisy expectation values fails to match the behavior in the zero-noise limit. For simple noise models or very low depth circuits this extrapolation can be well behaved. However, in real devices using less trivial circuits, the lowest error points available are often too noisy for such fits to be helpful.

Recently, alternative mitigation methods have been developed that make use of learning from data sets constructed using Clifford quantum circuit data [25,26]. These methods are attractive based on their relative simplicity and scalability due to the classically simulable nature of quantum circuits composed mainly of Clifford gates (gates that map Pauli operators to Pauli operators).

For example, the Clifford data regression (CDR) method [26] first chooses a training set of near-Clifford quantum circuits related to the circuit of interest. A scalable classical simulator of near-Clifford circuits and a noisy quantum computer are used to compute the noise-free and noisy data, respectively. Finally, the trained ansatz is used to predict the noise-free observable for the quantum circuit of interest.

Both ZNE and CDR are data-driven approaches to error mitigation, but they use different types of data. ZNE uses variable-noise data, while CDR uses variable Clifford
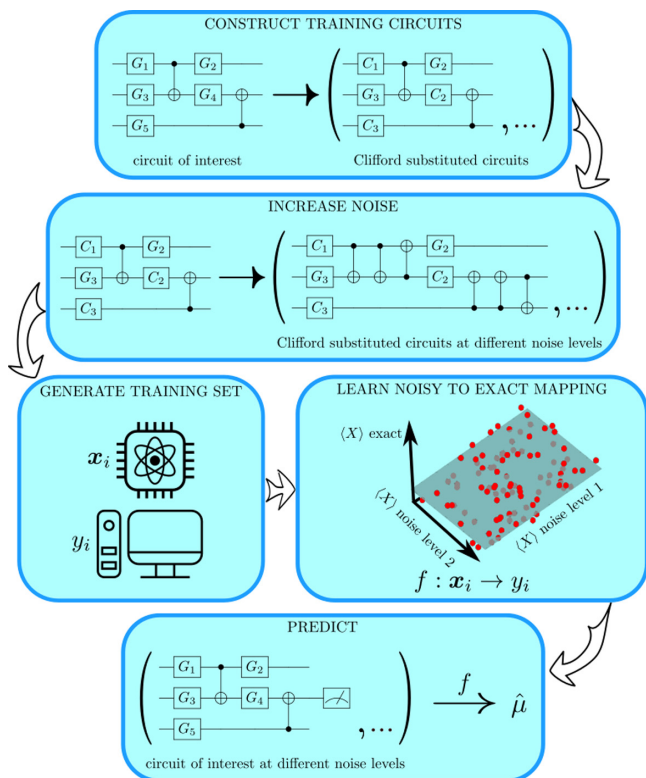

FIG. 1. The variable-noise Clifford data Rregression (vnCDR) method. The first step constructs a set of near-Clifford training circuits that are close, in some sense, to the circuit of interest. The second step increases the size of the training set by adding variable amounts of noise to the circuits generated in the first step. The third step involves both classical simulation and quantum evaluation of the training circuits to generate the noise-free and noisy training data, respectively. The fourth step trains the parameters of an ansatz, which we take as a hyperplane, to fit the training data. Finally, one uses this fitted ansatz to predict the desired observable for the circuit of interest.

circuit data. A natural question is whether combining these approaches could lead to a unified technique that is more powerful than the individual ones. In this paper, we propose a method that answers this in the affirmative.

Our approach is called variable-noise Clifford data regression (vnCDR). vnCDR considers a collection of near-Clifford training circuits like CDR, each evaluated at multiple noise levels as in ZNE. One can think of this process either as informing the extrapolation in ZNE about the zero-noise limit for similar circuits or as adding relevant features to the regression model in CDR. In the latter view, this is philosophically similar to data augmentation techniques that introduce artificial noise in machine learning [27]. The ansatz employed in vnCDR is motivated by Richardson extrapolation and by noting that it perfectly removes the effects of global depolarizing noise (see Appendix A). We also comment that training on the set of Clifford circuits is sufficient as the Clifford gates span the space of single-qubit unitaries (see Appendix B). Figure 1 gives a schematic illustration of vnCDR.

Below we first provide background on ZNE and CDR, and then we present our unified method. Using a noisy simulator

based on a gate set tomography of IBM's Ourense quantum computer, we compare the performance of ZNE, CDR, and vnCDR for two tasks. The first task is estimating the energy of an 8-qubit transverse Ising model with the quantum alternating operator ansatz (QAOA). Correcting circuits of this form is relevant for both combinatorial optimization problems and condensed matter studies [28,29]. Our second task involves random quantum circuits for large qubit numbers (up to 64 qubits with six CNOT layers) and large circuit depth (up to 16 CNOT layers with 8 qubits). The lack of structure in these random circuits makes them a difficult use case for these EM methods, and they give us a notion of these methods' utility in more general settings.

For both use cases, vnCDR outperforms ZNE and CDR. For the QAOA task we analyze the absolute energy error and obtain with vnCDR a factor of 20 improvement over ZNE and a factor of 1.8 improvement over CDR. For the random circuit task we obtain, in the case of 64 qubits, factors of 2.7 and 1.5 improvement over ZNE and CDR, respectively, while for the case of 16 layers we obtain factors of 2.3 and 1.3 improvement over those methods.

II. BACKGROUND

A. Zero-noise extrapolation

ZNE [4] involves varying the noise level of a quantum circuit to infer the noise-free behavior. Assuming a dependence on noise parameter ϵ, the correction is performed by taking linear combinations of the noisy expectation values in such a way that errors attributable to terms of order n or less are canceled, where n is the number of additional noise levels employed.

Following the presentation in Ref. [4], denote the noise-free expectation by μ and consider the task of correcting the expectation value obtained from a noisy quantum device with noise characterized by parameter ϵ. First, one chooses a set of noise levels $\mathcal{C} = \{c_0, c_1, \ldots, c_n | c_0 = 1, c_j < c_{j+1}\}$ and runs the device with amplified noise $c_j \epsilon$ to obtain an estimate $\hat{\mu}_j$ for all noise levels $c_j \in \mathcal{C}$. The final correction $\hat{\mu}$ can then be computed as

$$\hat{\mu} = \sum_{j=0}^{n} \gamma_j \hat{\mu}_j, \tag{1}$$

where the set of coefficients $\{\gamma_j\}$ are chosen to satisfy

$$\sum_{j=0}^{n} \gamma_j = 1, \quad \sum_{j=0}^{n} \gamma_j c_j^k = 0 \quad \forall k \in \{1, \ldots, n\}. \tag{2}$$

This technique, known as Richardson extrapolation [4,30], ensures that the error of the final estimate is of the order $O(\epsilon^{n+1})$.

Though ZNE was originally proposed in a context where one can stretch gate times to achieve the various noise levels c_j, recent work has suggested a hardware-agnostic implementation based on identity insertions [19,20]. For example, inserting two CNOT gates applied one after the other is an identity matrix in the noise-free circuit evaluation but is likely to affect the output in the noisy case. In the fixed identity insertion method (FIIM), the noise levels are taken to

be the number of additional gates added in this manner, so inserting two additional CNOT gates for every CNOT in the original circuit results in noise levels $c_j = 1, 3, 5, \ldots$ being implemented.

As noted in Ref. [20], Richardson extrapolation is equivalent to performing a polynomial interpolation on the various noisy expectations, treating the noise levels c_j as the independent variable. To see this, note that for any solution $\{b_k\}_{k=0}^n$ to the system of equations

$$\hat{\mu}_j = b_0 + \sum_{k=1}^n b_k c_j^k, \quad j = 0, 1, \ldots, n, \tag{3}$$

it holds that

$$\begin{aligned}
\hat{\mu} &= \sum_{j=0}^n \gamma_j \left(b_0 + \sum_{k=1}^n b_k c_j^k \right) \\
&= b_0 \left(\sum_{j=0}^n \gamma_j \right) + \sum_{k=1}^n b_k \left(\sum_{j=0}^n \gamma_j c_j^k \right) \\
&= b_0.
\end{aligned} \tag{4}$$

A unique solution exists to (3) for $n + 1$ distinct noise levels c_j when performing an order-n interpolation. (See, for example, Ref. [31]). Alternatively, one can adopt a lower-degree polynomial fit as done in our numerical experiments by computing the least-squares solution to the resulting system of equations. For example, if one wishes to perform a linear fit on the data from $n + 1$ distinct noise levels, one may write

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_0 \\ \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} 1 & c_0 \\ 1 & c_1 \\ \vdots & \vdots \\ 1 & c_n \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} \tag{5}$$

and express the system of equations as

$$\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{b}. \tag{6}$$

Taking the y intercept b_0 of the least-squares solution,

$$\boldsymbol{b} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{\mu}, \tag{7}$$

yields the extrapolated expectation value.

Crucially, the value of the correction is completely determined by a fixed set of noisy expectations for any choice of the noise amplification and extrapolation techniques above. In particular, Eq. (1) enforces an nth-degree polynomial fit when one has data points at $n + 1$ noise levels, which may not provide a good approximation to the behavior near $\epsilon = 0$ when the data points that are experimentally accessible all reflect a fairly high amount of noise. While some authors have also successfully used lower-degree polynomial fits, there is evidence to suggest that the resulting corrections can still be fairly inaccurate [5,19,26]. This motivates our proposal for a method based on learning from efficiently simulable circuits, avoiding some of the drawbacks of ZNE.

B. Clifford data regression

In CDR [26] the expectation values obtained from a quantum device are corrected using a straightforward linear regression based on examples from circuits composed mainly of Clifford gates. These Clifford circuits are efficiently simulable and generated in such a way as to remain similar to the original circuit of interest. Explicitly, the goal is to learn a function which takes noisy expectations to their error-mitigated values:

$$f(\hat{\mu}_0) = a_1 \hat{\mu}_0 + a_2, \tag{8}$$

where $\hat{\mu}_0$ is the noisy expectation and a_1, a_2 are parameters chosen optimally by least-squares regression on the Clifford circuit data set, i.e., for a training set of m noisy Clifford circuit expectations $\{x_i\}$ and corresponding targets $\{y_i\}$ obtained via classical simulation, one computes

$$(a_1, a_2) = \underset{(a_1, a_2)}{\operatorname{argmin}} \sum_{i=1}^m [y_i - (a_1 x_i + a_2)]^2. \tag{9}$$

The form of the ansatz can be physically motivated using a simplified noise model. Let ρ be the density matrix for the state of a device which has undergone some noise-free evolution and consider a global depolarizing noise channel $\mathcal{E}$ which acts on this state before a measurement of the observable X. It then holds that

$$\operatorname{Tr}(\mathcal{E}(\rho)X) = (1 - \epsilon)\operatorname{Tr}(\rho X) + \frac{\epsilon \operatorname{Tr}(X)}{d}, \tag{10}$$

where d is the dimension of the system and ϵ is a parameter characterizing the noise. Identifying $\hat{\mu}_0 = \operatorname{Tr}(\mathcal{E}(\rho)X)$ and

$$a_1 = 1/(1 - \epsilon), \quad a_2 = -\frac{\epsilon}{d(1 - \epsilon)}\operatorname{Tr}(X), \tag{11}$$

we see that the desired quantity $\operatorname{Tr}(\rho X)$ can be recovered using Eq. (8).

When applied to a more realistic noise model obtained from an IBM quantum device, empirical results suggest that CDR yields significantly more scalable corrections than those from ZNE [26], at least in the plausible setting of being limited to coarse-grained noise amplification. In the following sections, we improve upon the CDR method by incorporating data obtained at variable noise rates, which leads to more accurate predictions of the noise-free expectation values.

III. THE vnCDR METHOD

Let U be a quantum circuit, $|0\rangle$ be its initial state, and X be an observable of interest. Consider the task of estimating the expectation value $\mu = \langle 0| U^\dagger X U |0\rangle$ from measurements of a noisy quantum device. The variable-noise Clifford data regression (vnCDR) method is performed with the following steps.

(1) Clifford data. Choose a set of circuits $\mathcal{S} = \{V_i\}_{i=1}^m$ based on U, which will be used to form the training set $\mathcal{T}$ in step (3). The circuits in $\mathcal{S}$ must be efficient to simulate classically, which is ensured by constructing them primarily from Clifford gates. The number of non-Clifford gates used is denoted by N. Note that N is assumed to be a constant parameter here, so the simulations are classically tractable.

(2) Noise data. Choose a set of noise levels $\mathcal{C} = \{c_0, c_1, \ldots, c_n\}$, where $1 = c_0 < c_1 < \cdots < c_n$, which will be used to form the training set $\mathcal{T}$ in step (3). If the noise is characterized by a parameter ϵ, then running the device with noise level c_j means that the new parameter is $c_j \epsilon$.

(3) Training set. For each of the m circuits V_i in $\mathcal{S}$ and $n + 1$ noise levels $c_j \in \mathcal{C}$, produce an estimate of the observable expectation called $x_{i,j}$. Also, for each of the m circuits, compute $y_i = \langle 0| V_i^\dagger X V_i |0 \rangle$ using a classical simulation. The training set $\mathcal{T}$ is then defined as $\mathcal{T} = \{(\boldsymbol{x}_i, y_i)\}$, where $\boldsymbol{x}_i = (x_{i,0}, \dots, x_{i,n})$ is the vector of noisy estimates originating from the ith circuit.

(4) Learning. Learn a function $f : \mathbb{R}^n \to \mathbb{R}$ that takes a set of noisy estimates at the $n + 1$ different noise levels and outputs an estimate for the noise-free value. Specifically, we take the linear ansatz $g : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$,

$$g(\boldsymbol{x}; \boldsymbol{a}) = \boldsymbol{a} \cdot \boldsymbol{x}. \tag{12}$$

We use least-squares regression on the data set $\mathcal{T}$ to pick optimal parameters $\boldsymbol{a}^*$, i.e.,

$$\boldsymbol{a}^* = \underset{\boldsymbol{a}}{\mathrm{argmin}} \sum_{i=1}^{m} [y_i - g(\boldsymbol{x}_i; \boldsymbol{a})]^2, \tag{13}$$

so that we expect $f(\boldsymbol{x}) = g(\boldsymbol{x}; \boldsymbol{a}^*)$ to output a good estimate for the noise-free expected value given a vector of noisy ones.

(5) Correction. Use the estimate $\hat{\mu} = \boldsymbol{a}^* \cdot \boldsymbol{\mu}$, where $\boldsymbol{\mu} = (\hat{\mu}_0, \dots, \hat{\mu}_n)$ is composed of the $n + 1$ noisy expectations for the original circuit.

Our method shares common features with both ZNE and CDR. Specifically, the functional form of the ansatz we choose resembles a Richardson extrapolation on the noisy expected values as shown in Sec. II A. However, the method differs in its approach to relating the noisy values to the final estimate. Namely, in ZNE, the output is a fixed function of the various $\hat{\mu}_j$, whereas vnCDR attempts to learn the best candidate from a family of functions parametrized by $\boldsymbol{a}$. In a certain sense, vnCDR is thus choosing the best possible extrapolation of the noisy data from the original circuit using examples from Clifford circuits which are similar in structure.

The method can also be viewed as adding relevant variables to the CDR method, before performing a multiple linear regression on the new data set. This description comes with the caveat that—in contrast to the CDR ansatz [see Eq. (8)]—the new parametrization $g(\boldsymbol{x}; \boldsymbol{a})$ is a linear mapping without a constant term. There are two motivations for this. Firstly, such a parametrization corresponds well with the linear combination of noisy expectations that is utilized in the ZNE method [Eq. (1)]. Secondly, restricting the class of functions we are searching over to be linear awards us an intuitively desirable property: If the function we arrive at achieves zero error on all circuits composed of Clifford gates, then it will predict the expectation values of arbitrary circuits with zero error. This result boils down to the observation that Clifford gates span the space of single-qubit unitaries. (See Appendix B for the proof of this statement).

The form of the ansatz can be further motivated by considering the action of a global depolarizing channel [see Eq. (10)]. We note that this simple model was proposed recently to effectively describe dominant effects of the noise in real devices [32,33]. The vnCDR ansatz can be shown to completely mitigate the effect of such a channel on some observable of interest (see Appendix A), similar to CDR.

IV. NUMERICAL RESULTS

A. Transverse-field Ising model

First, we consider a task of variational simulation of the ground state of a one-dimensional (1D) transverse-field Ising model using parametrized quantum circuits. The Hamiltonian of the system is given by

$$H = -g \sum_j \sigma_X^j - \sum_{\langle j, j' \rangle} \sigma_Z^j \sigma_Z^{j'}, \tag{14}$$

where σ_X, σ_Z are Pauli matrices and $\langle j, j' \rangle$ are nearest-neighbor sites on the lattice. We assume here open boundary conditions. We consider the case of $g = 2$ corresponding to a paramagnetic phase. We use the QAOA [28,29]

$$|\psi(\boldsymbol{\beta}, \boldsymbol{\gamma})\rangle = \prod_{j=p,p-1,\dots,1} e^{-i\beta_j H_2} e^{-i\gamma_j H_1} (|+\rangle)^{\otimes Q}, \tag{15}$$

where $\boldsymbol{\beta}, \boldsymbol{\gamma}$ are the rotation angles to be optimized, $H_1 = \sum_{\langle j, j' \rangle} \sigma_Z^j \sigma_Z^{j'}$, $H_2 = \sum_j \sigma_X^j$, $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, and Q is the number of qubits. A decomposition of (15) into a quantum circuit is described in Appendix C.

We perform the optimization for $Q = 8$ qubits using a circuit depth $p = 4$. We minimize the energy evaluated with a noisy simulator using a MATLAB implementation of quasi-Newton gradient descent. The noise model we employ is obtained by gate set tomography of IBM's Ourense quantum computer and described in detail in Ref. [8]. Furthermore, we assume perfect measurement as measurement errors can be mitigated by specialized techniques [34,35]. To carry out the benchmark of our method, we run 27 instances of the optimization and correct the resulting observable expectations using the ZNE, CDR, and vnCDR methods. The corrections are realized on each of the 1- and 2-qubit terms which make up the Hamiltonian from which we then estimate the ground-state energy. The results are summarized in Fig. 2, showing that vnCDR outperforms ZNE and CDR with a factor of 33 improvement of the mean absolute energy error while ZNE and CDR give a factor of 1.7 and 19 improvement, respectively.

In the case of CDR and vnCDR, for each of the circuits we construct training sets with 80 classically simulable circuits, setting the number of non-Clifford gates to $N = 16$. We remark that there are 60 non-Clifford gates in total for the circuit of interest. For further information regarding the construction of our training sets, see Sec. V B.

For the vnCDR and ZNE corrections, we computed the expectation values using the set of noise levels $\mathcal{C} = \{1, 3, 5\}$ and the fixed identity insertion noise amplification method [20], which we elaborate upon in Sec. V A. The noise level is defined as the ratio of CNOT gates in the modified circuits compared with the original one. Note that this is a fairly coarse-grained set of noise levels, which may explain why the ZNE performance is quite poor. We also found that including in $\mathcal{C}$ noise levels higher than 5 did not improve the performance of the methods. For further details, see Secs. V A and V C.

B. Random quantum circuits

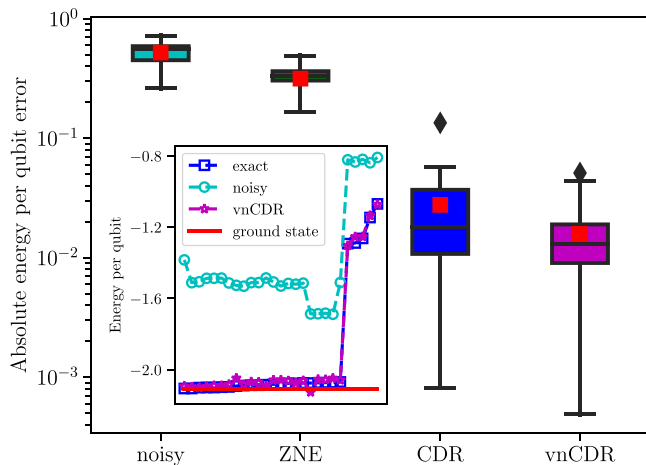Next we consider an implementation of the IBMQ-hardware-efficient ansatz with random parameters; see

FIG. 2. Correcting 27 local minima of the quantum Ising model (14) energy minimization for $Q = 8$ qubits. As a variational ansatz we use the QAOA (15) with $p = 4$ layers. We compare absolute errors of the noisy and corrected energies for different error mitigation methods. The results obtained are shown as a box plot, where boxes show the interval between the first and the third quartile. The red squares denote the mean values, while the central lines represent the median values. The whiskers show the range of the data outside the quartiles, and diamonds indicate outliers. The inset displays the energies per qubit of the minima calculated by the QAOA energy minimization.

Fig. 3. The ansatz consists of layers of alternating nearest-neighbor CNOT gates decorated with general 1-qubit unitaries $U(\alpha, \beta, \gamma)$. We compute 1- and 2-qubit observables for 30 random instances and correct them with ZNE, CDR, and vnCDR methods; see details in the caption of Fig. 4. We analyze the scaling of the observables' absolute error with increasing $Q = 8, 16, 32, 64$ for $p = 6$ and the scaling with increasing $p = 4, 8, 12, 16$ for $Q = 8$. To simulate large-Q systems, we employ a matrix product operator (MPO) [36] noisy simulator with the same noise model as in the case of the Ising QAOA simulations. We discuss the simulator in
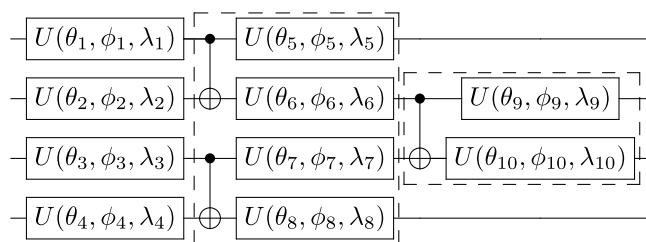

FIG. 3. An example of the IBMQ-hardware-efficient ansatz with $p = 2$ layers for $Q = 4$ qubits. The layers, represented by gates within the dashed contours, act on a random product state created by general single-qubit unitaries U. The general unitary is defined as $U(\theta, \phi, \lambda) = R_Z(\phi + \pi)R_X(\pi/2)R_Z(\theta + \pi)R_X(\pi/2)R_Z(\lambda)$, where $R_Z(\alpha) = e^{-i\alpha/2\sigma_Z}$, $R_X(\alpha) = e^{-i\alpha/2\sigma_X}$. Each layer consists of CNOT gates interleaved with the U gates. The CNOT structure alternates between neighboring layers. We choose angles θ, ϕ, and λ of each U gate randomly creating a random quantum circuit. Note that $R_X(\pi/2)$, CNOT gates, and $R_Z(\alpha)$ are native gates of the IBM computers. Furthermore, CNOT gates and $R_X(\pi/2)$ are Clifford gates.

more detail in Appendix E. Here, to simplify presentation, we show results obtained in the limit of infinite shot number. In Appendix F we show that qualitatively the same results can be obtained using finite shot numbers feasible with current quantum computers.

The results are discussed in detail in Fig. 4. We find that the vnCDR outperforms ZNE and CDR methods for all simulated Q and p values. For the largest system size, $Q = 64$, vnCDR gives a factor of 2.7 improvement of the mean error relative to the noisy results, while ZNE and CDR give factors of 1.0 and 1.8 improvement. For the deepest $p = 16$ the factors are 2.4, 1.0, and 1.8, respectively.

For fixed $p = 6$ we observe that the unmitigated mean absolute error does not grow with increasing Q in the limit of large Q. Such behavior can be explained by the existence of a threshold Q value for which the causal cones of the observables [37] stop increasing with Q. The causal cone is defined here as gates which affect the expectation value of the observable. See Appendix D for an example of causal cone construction. We take the causal cone into account when forming vnCDR and CDR training sets; see details in Sec. V. For such an implementation we find that the vnCDR and CDR mitigated mean errors also do not increase with increasing Q. We remark that our noise model does not include cross talk, which in principle may result in a faster increase in the number of gates in the noisy observables' causal cones. We leave investigation of the scaling in the presence of such noise to future work. With increasing p we find that the quality of the correction decreases for all methods. Nevertheless, even in the case of the deepest circuits, $p = 16$, we obtain a significant improvement when employing vnCDR.

To perform CDR and vnCDR for each observable of interest in each random circuit, we construct a training set using 100 classically simulable circuits with $N = 20$ non-Clifford gates. A detailed discussion of the method used to construct the training sets is given in Sec. V B. We remark that in the case of $p = 6$, $Q = 64$ circuits the largest number of non-Clifford gates within the causal cone of an observable is 60, while for $Q = 8$, $p = 16$ it is 312.

For the vnCDR and ZNE corrections, we increase the noise level by identity insertions as in the case of the QAOA Ising simulations. We find that in both cases it is beneficial to include higher noise levels than in the Ising case: $\mathcal{C} = \{1, 3, 5, 7, 9\}$. We remark that the QAOA circuit having 16 layers of CNOT gates is deeper than most circuits considered here. As ZNE is supposed to correctly capture noise effects for sufficiently small noise this may explain why it is beneficial to use higher noise levels in the case of random quantum circuits. We leave systematic investigation of this effect to future work. For a more detailed description of the ZNE and vnCDR implementations, see Secs. V A and V C.

V. IMPLEMENTATION DETAILS

A. ZNE

We perform the noise amplification in our numerical experiments using identity insertions after each application of a CNOT gate [19,20]. We use the fixed identity insertion method (FIIM) of Ref. [20], which adds pairs of CNOT gates after
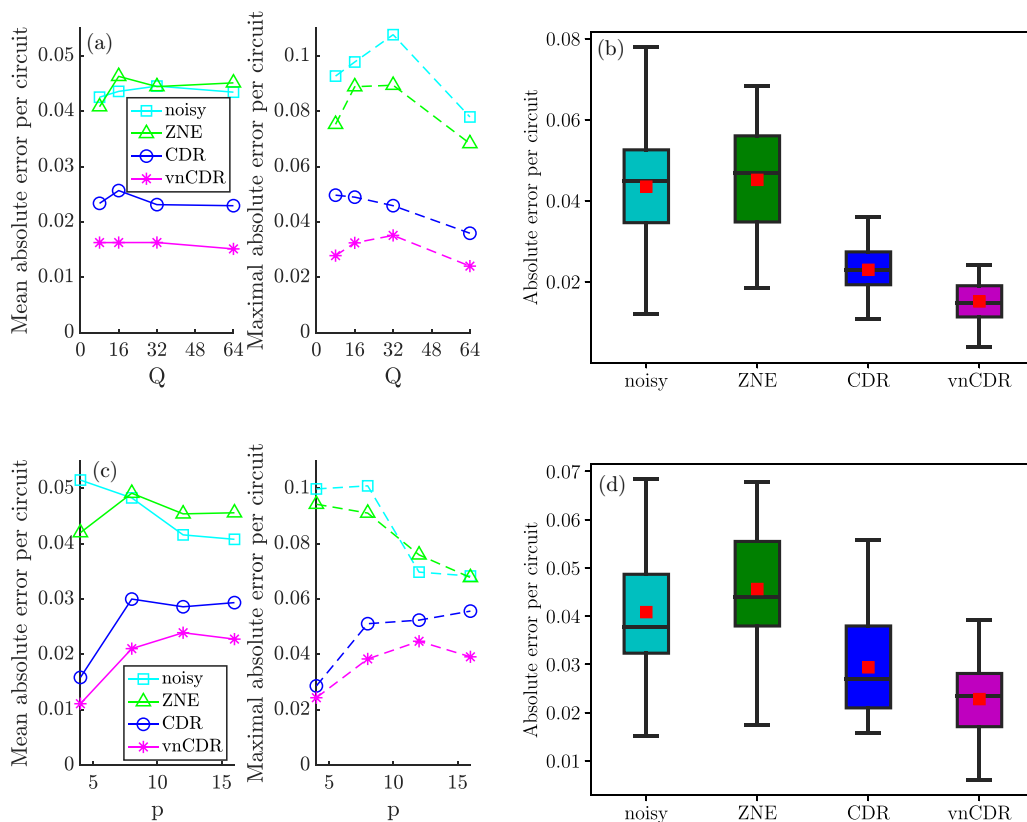
FIG. 4. Correcting the IBMQ-hardware-efficient ansatz with random parameters. For each value of Q and p we analyze 30 random circuits and correct σ_X^1, $\sigma_X^{Q/2}$, $\sigma_Z^1 \sigma_Z^2$, and $\sigma_Z^{Q/2} \sigma_Z^{Q/2+1}$ for each of them. We define an absolute error per circuit as the mean of the observables' absolute errors. The absolute corrected (noisy) observable error is defined as an absolute value of its difference with respect to the exact value. (a) The scaling with Q of the mitigated and unmitigated absolute error per circuit for $p = 6$. In the left panel we show the mean values (solid lines), while the right panel shows the maximal values (dashed lines). (b) Bar plot of the error for $Q = 64$ and $p = 6$. (c) The scaling with increasing p for $Q = 8$. (d) The results for $Q = 8$, $p = 16$.

each CNOT gate of the original circuit. The noise level is defined as the factor by which the number of CNOT gates in the circuit increases. In the first example—the QAOA optimization task—we employ noise levels $\mathcal{C} = \{1, 3, 5\}$, whereas for random circuits we achieved better results with a higher maximum noise level, so we used the set $\mathcal{C} = \{1, 3, 5, 7, 9\}$. We obtained corrected values of the observables of interest by an extrapolation using both a polynomial fit via Eq. (1) and a linear fit to the data, as explained in Sec. II A. In both the Ising case and the case of random quantum circuits, we found that a linear fit performed better than a polynomial regression for extrapolation, so we report those results here.

B. CDR

To construct the training set for a circuit of interest, we substitute most of the non-Clifford gates in the circuit with Clifford gates, with two different substitution strategies, which are explained below. Such a procedure ensures that circuits in the training set are classically simulable and biased towards the circuit of interest. Here, we consider circuits of interest which are compiled for the IBMQ quantum computers. The compiled circuits are built from CNOT gates, $R_X(\pi/2)$

pulses, and general σ_Z rotations $R_Z(\beta) = e^{-i\beta/2 \sigma_Z}$ with $\beta \in [0, 2\pi)$. The pulses and CNOT gates are Clifford gates, while $R_Z(\beta)$ is a Clifford gate only for $\beta = n\pi/2$, where n is an integer. Therefore we substitute most of the R_Z gates by S^n, where $n = 0, 1, 2, 3$ and $S = e^{i\pi/4 \sigma_Z}$ is the phase gate. In both the Ising case and the case of random quantum circuits we find that substantial error reduction can be obtained using training sets built with approximately 100 near-Clifford circuits.

We consider two different substitution strategies. The first one substitutes a randomly chosen non-Clifford $R_Z(\beta)$ by S^n minimizing $d(\beta, n) = ||R_z(\beta) - S^n||$, where $|| \cdot ||$ is the Frobenius norm. This procedure is repeated until N non-Clifford rotations are left in the circuit. We find that this very simple strategy works well for the Ising model enabling us to obtain a factor of 33 improvement in calculating the energy.

In the more general and challenging case of random quantum circuit simulations we find that better results can be obtained with a more sophisticated substitution method. In such a case, to construct a training set, we tailor our choice of classically simulable circuits to an observable of interest, substituting all non-Clifford gates outside its causal cone. By the causal cone definition, such a replacement does not affect its expectation value. See Appendix D for a discussion of the causal cone construction. Taking into account the causal

cone of the observable is especially important in the case of local observables and large-Q shallow circuits because in such a case the causal cone contains only a small fraction of all non-Clifford gates of the circuit of interest. Furthermore, for remaining non-Clifford rotations within the causal cone of the observable of interest, we choose both which gate to replace and what gate to replace it with (S^n) according to a probability distribution $p(\beta_i, n) \propto e^{-d(\beta_i, n)^2/\sigma^2}$. Here, i numbers the remaining non-Clifford rotations in the causal cone. We repeat the procedure until N non-Clifford rotations are left in the causal cone of the observable of interest. Here, we use $\sigma = 0.5$. Such a choice of the probability distribution tends to leave gates which would be most severely distorted by the replacement in the circuit, unchanged. At the same time it produces more diverse training sets than a direct replacement by the closest power of S. We observe that in the case of the random quantum circuits the correction is more challenging as expectation values of the observable of interest become more clustered around 0 with increasing p. Furthermore, we observe that training sets created by the simple substitution method tend to have exact expectation values clustered around 0 more strongly than expectation values of the observable of interest. The more sophisticated procedure generates training sets with more diverse exact expectation values.

C. vnCDR

To construct a vnCDR training set, we choose the same classically simulable circuits which are used for a CDR training set. We also use the same choice of noise levels as used in the ZNE implementation, namely, $\mathcal{C} = \{1, 3, 5\}$ for the Ising and $\mathcal{C} = \{1, 3, 5, 7, 9\}$ for the random quantum circuit mitigation. As in the case of ZNE we observe that including more than five noise levels does not improve results for the Ising case while it is beneficial for the case of random quantum circuits.

VI. CONCLUSIONS

Data-driven error mitigation involves collecting data from multiple different quantum circuits in order to inform the correction of errors in a particular circuit of interest. In this paper, we conceptually unified two distinct, popular methods for data-driven error mitigation: zero-noise extrapolation (ZNE) and Clifford data regression (CDR). Our unified approach, called variable-noise Clifford data regression (vnCDR), appears to be more powerful than the individual methods of ZNE and CDR.

The vnCDR method generates training data from classically simulable near-Clifford circuits, whose noise levels are varied (e.g., by identity insertions). The method then learns how to correct observables on these training circuits. This involves fitting a multidimensional ansatz, which we assume is a hyperplane, to the training data. This enables a guided extrapolation to the noiseless expectation value for the circuit of interest, which dramatically improves the mitigation realized. Rather than doing uninformed extrapolation as in ZNE, the vnCDR method demonstrates that near-Clifford circuits provide an effective guide for the extrapolation process. The fitted

ansatz can be further motivated by considering the effect of a global depolarizing channel on some observable of interest. The effect of such a channel is completely removed using the vnCDR ansatz.

We compared vnCDR with both ZNE and CDR on two tasks: correcting the energy of an Ising transverse spin chain and mitigating local observables of random quantum circuits. For both of them we used a realistic noise model obtained by gate set tomography of IBM's Ourense quantum computer. On each of these tasks, vnCDR outperforms both of these state-of-the-art error mitigation methods. Compared with ZNE, vnCDR was shown to tolerate the relatively high noise levels obtained via fixed identity insertions.

Though preliminary scaling results are promising, further testing on real quantum devices will help determine the number of non-Clifford gates and size of the training sets required to attain accurate predictions. It will also help determine limitations of the method while dealing with large and deep noisy circuits, which are challenging for error mitigation methods. Additionally, it would be interesting to apply the vnCDR method using more sophisticated or fine-grained noise amplification schemes such as random identity insertions or pulse stretching. This may enhance performance for the deep circuits necessary to obtain a quantum advantage. In this regime, we envision that vnCDR could play an important role in yielding quantum advantage for chemistry, materials science, and other applications. Finally, we note that further testing is necessary to determine the potential of our method for quantum computing architectures with gate sets other than IBM's gate set.

APPENDIX A: PERFECT MITIGATION OF GLOBAL DEPOLARIZING NOISE

To motivate the form of the vnCDR ansatz, we consider the action of a global depolarizing channel [see Eq. (10)]. Assuming that this channel acts in our circuit j different times, the final state can be written as

$$\rho_j = (1 - \epsilon)^j \rho + [1 - (1 - \epsilon)^j]\frac{\mathbb{1}}{d}, \qquad \text{(A1)}$$

where d is the dimension of the system and ϵ is a parameter characterizing the noise. Considering the effect of the above

channel on X leads to

$$X_j^{\text{noisy}} = \text{Tr}(\rho_j X) \tag{A2}$$

$$= (1 - \epsilon)^j \mu + [1 - (1 - \epsilon)^j] \frac{\text{Tr}(X)}{d}, \tag{A3}$$

where $\mu = \text{Tr}(\rho X)$. As previously discussed the vnCDR ansatz combines evaluations of the observable of interest at various noise levels:

$$\hat{\mu} = \sum_{j=1}^{n} a_j^* X_j^{\text{noisy}}, \tag{A4}$$

where the parameters a_j^* are chosen by fitting data produced by near-Clifford circuits. The above expression can be expanded:

$$\hat{\mu} = \sum_{j=1}^{n} a_j^* \left((1 - \epsilon)^j \mu + [1 - (1 - \epsilon)^j] \frac{\text{Tr}(X)}{d} \right). \tag{A5}$$

Therefore, for the vnCDR ansatz to completely mitigate the effects of global depolarizing noise, such that $\hat{\mu} = \mu$, we require

$$\sum_{j=1}^{n} a_j^* (1 - \epsilon)^j = 1, \quad \sum_{j=1}^{n} a_j^* [1 - (1 - \epsilon)^j] = 0, \tag{A6}$$

or equivalently,

$$\sum_{j=1}^{n} a_j^* = 1, \quad \sum_{j=1}^{n} a_j^* [1 - (1 - \epsilon)^j] = 0. \tag{A7}$$

The training circuit observables and the observable of interest will behave the same way under such a noise channel. As such, the fitted parameters a_j^* will obey the above relations [Eq. (A7)]. Therefore vnCDR can be seen to perfectly mitigate global depolarizing noise for two or more noise levels.

It is interesting to consider how this contrasts with the ZNE implementation in this paper. We used linear extrapolation to the zero-noise limit and least-squares fitting of the noisy expectation values for the observable of interest. This extrapolation method is not expected to perfectly mitigate the effect of global depolarizing noise. An exponential extrapolation would be required in order to perfectly mitigate the effects of this channel, and polynomial extrapolation is expected to perform better than linear. However, for our simulations we found that in general, a simple linear extrapolation gave better results.

APPENDIX B: SUFFICIENCY OF THE CLIFFORD TRAINING SET

Consider a quantum circuit acting on Q qubits which is represented by a noise-free unitary channel $\mathcal{U}$, and let $\rho_0 \in \mathbb{C}^{d \times d}$ be the initial state, where $d = 2^Q$. Also suppose we have some observable of interest X and a collection of channels $\mathcal{E}_0, \mathcal{E}_1, \ldots, \mathcal{E}_n$ representing $n + 1$ different noise levels. Running the circuit with the jth noise channel returns the value $\text{Tr}((\mathcal{E}_j \circ \mathcal{U})(\rho_0)X)$ in expectation.

Additionally, for all $j \in \{0, \ldots, n\}$, define $\mu_j(\mathcal{V})$ for some unitary channel $\mathcal{V}$ in the following way:

$$\hat{\mu}_j(\mathcal{V}) = \text{Tr}((\mathcal{E}_j \circ \mathcal{V})(\rho_0)X). \tag{B1}$$

Define $\boldsymbol{\mu}(\mathcal{V}) \in \mathbb{R}^{n+1}$ as

$$\boldsymbol{\mu}(\mathcal{V}) = (\hat{\mu}_0(\mathcal{V}), \hat{\mu}_1(\mathcal{V}), \ldots, \hat{\mu}_n(\mathcal{V})). \tag{B2}$$

For a given circuit $\mathcal{V}$, the vnCDR correction is then given by $f(\boldsymbol{\mu}(\mathcal{V})) = \boldsymbol{a} \cdot \boldsymbol{\mu}(\mathcal{V})$, where $\boldsymbol{a}$ is some optimal set of parameters obtained by training the model. Our goal is to show that if the vnCDR estimate $f(\boldsymbol{\mu}(\mathcal{C}))$ is fully accurate for all Clifford circuits $\mathcal{C}$, then the output of $f(\boldsymbol{\mu}(\mathcal{U}))$ is also accurate for estimating the value $\text{Tr}(\mathcal{U}(\rho_0)X)$.

We begin by observing that the action of a non-Clifford rotation gate can be decomposed into Clifford maps since they span the space of single-qubit unitaries. Therefore we can write

$$\mathcal{U} = \sum_{j_1} \alpha_{j_1} \mathcal{C}_{j_1}, \tag{B3}$$

where $\mathcal{C}_{j_1}$ is the unitary map resulting from replacing one of the non-Clifford rotation gates in the circuit by the j_1th Clifford in the basis. Repeating this process recursively for each of the k non-Clifford gates in the circuit, we obtain

$$\mathcal{U} = \sum_{j_1, j_2, \ldots, j_k} \alpha_{j_1} \alpha_{j_2} \cdots \alpha_{j_k} \mathcal{C}_{j_1, \ldots, j_k}. \tag{B4}$$

Each of the unitary maps $\mathcal{C}_{j_1, \ldots, j_k}$ is now composed of Clifford maps only. Furthermore, by the linearity of the trace, noise channels, and function f, it holds that

$$f(\boldsymbol{\mu}(\mathcal{U})) = \sum_{j_1, j_2, \ldots, j_k} \alpha_{j_1} \alpha_{j_2} \cdots \alpha_{j_k} f(\boldsymbol{\mu}(\mathcal{C}_{j_1, \ldots, j_k})) \tag{B5}$$

and

$$\text{Tr}(\mathcal{U}(\rho_0)X) = \sum_{j_1, j_2, \ldots, j_k} \alpha_{j_1} \alpha_{j_2} \cdots \alpha_{j_k} \text{Tr}(\mathcal{C}_{j_1, \ldots, j_k}(\rho_0)X). \tag{B6}$$

Therefore suppose it holds for all Clifford unitaries $\mathcal{C}$ that the loss of our correction is

$$|\delta(\mathcal{C})| := |f(\boldsymbol{\mu}(\mathcal{C})) - \text{Tr}(\mathcal{C}(\rho_0)X)| = 0. \tag{B7}$$

Then, since $\mathcal{C}_{j_1, \ldots, j_k}$ is a Clifford unitary for all $j_1, \ldots, j_k$,

$$|\delta(\mathcal{U})| = |f(\boldsymbol{\mu}(\mathcal{U})) - \text{Tr}(\mathcal{U}(\rho_0)X)|$$

$$= \left| \sum_{j_1, \ldots, j_k} \alpha_{j_1} \cdots \alpha_{j_k} \delta(\mathcal{C}_{j_1, \ldots, j_k}) \right|$$

$$= 0. \tag{B8}$$

In other words, we have achieved zero loss on all arbitrary circuits if we obtain zero loss on training data composed of all possible Clifford circuits. We note that a similar remark is made in Ref. [25] to argue that Clifford circuits suffice for their learning-based approach to quasiprobability representation (QPR) error mitigation. However, unlike in Ref. [25], depending on the channels and initial state involved in the error mitigation, there may not exist a set of parameters which
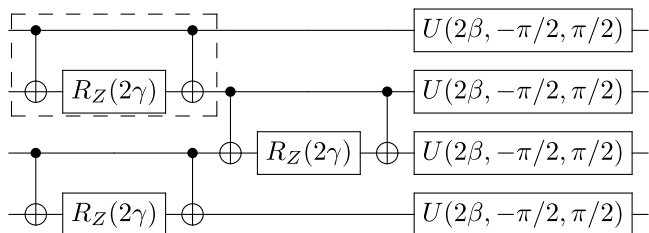
FIG. 5. A layer of the QAOA for a 4-qubit system decomposed into IBMQ natively supported gates. Gates within the dashed contour represent $e^{-i\gamma\sigma_Z^1\sigma_Z^2}$ [38], while a U gate represents $e^{-i\beta\sigma_X}$. To perform the mitigation, we decompose the U gates as explained in Fig. 3.
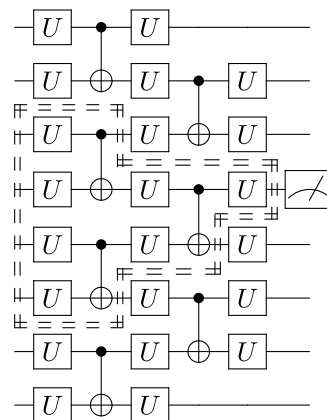

FIG. 6. A causal cone of a single-qubit observable is shown as gates within the double-dashed contour. Here, we consider the case of the hardware-efficient ansatz with $Q = 8$ and $p = 2$. Note that the causal cone is shown for a noisy expectation value while assuming that the Kraus matrices of a noise channel associated with a single- or double-qubit gate act on the same qubits as the gate. This assumption is true for our noise model. The causal cone for a noisy circuit implementation contains the causal cone of the corresponding exact expectation value.

achieves zero loss on all Clifford circuits. Hence the considerations in this Appendix should serve as high-level motivation for the linear form of the ansatz we employ, rather than a rigorous demonstration of the practicality of Clifford-based training sets.

APPENDIX C: THE QAOA DECOMPOSITION INTO A QUANTUM CIRCUIT

In Fig. 5 we show a decomposition of the QAOA (15) into a circuit which was used to perform the simulations in Sec. IV A.

APPENDIX D: CAUSAL CONE

We define the causal cone of an observable of interest as the set of gates which affects its expectation value [37]. In the case of the shallow quantum circuits and 1- or 2-qubit observables considered in Sec. IV B the causal cones contain only a fraction of all gates in the circuit. Therefore, to correct noisy expectation values of these observables, it is beneficial to leave non-Clifford gates only within the causal cone, as we have done when constructing the CDR and vnCDR training sets while correcting observables from random quantum circuits. We show an example of the causal cone construction in Fig. 6.

APPENDIX E: MATRIX PRODUCT OPERATOR SIMULATION OF NOISY STATES

Many-body quantum states can be represented in terms of interconnected tensors called tensor networks. Tensor networks are the basis for many standard numerical and analytical techniques in condensed matter theory [39]. The matrix product operator (MPO) is an example of a tensor network, which corresponds to a 1D array of tensors. In general they can describe any mixed state in the 1D many body Hilbert space, with the dimensionality of the tensors scaling exponentially with the system size. However, states with sufficiently small entanglement can be efficiently represented as MPOs, making them a convenient numerical tool. For a detailed introduction of MPO methods we refer the reader to Ref. [40].

Consider a Q-qubit density matrix,

$$\hat{\rho} = \sum_{\substack{i_1, \ldots, i_Q = 0 \\ i'_1, \ldots, i'_Q = 0}}^{1} c_{(i_1 \cdots i_Q)(i'_1 \cdots i'_Q)} |i_1, \ldots, i_Q\rangle \langle i'_1, \ldots, i'_Q|. \quad \text{(E1)}$$

We can express $c_{(i_1 \cdots i_Q)(i'_1 \cdots i'_Q)}$ as a product of matrices $W^{(1)i_1 i'_1}, W^{(2)i_2 i'_2}, \ldots, W^{(Q)i_Q i'_Q}$,

$$c_{(i_1 \cdots i_Q)(i'_1 \cdots i'_Q)} = \sum_{b_1, \ldots, b_{Q-1} = 0}^{\chi - 1} W_{1, b_1}^{(1)i_1 i'_1} W_{b_1, b_2}^{(2)i_2 i'_2} \cdots W_{b_{Q-1}}^{(Q)i_Q i'_Q}, \quad \text{(E2)}$$

where χ is the bond dimension and $b_1, \ldots, b_{Q-1}$ are bond indices which characterize the entanglement in the state. Above, for simplicity, we assume the same χ for each bond index, but in principle χ can be different for each of them. For a general quantum state represented in this form, one needs to use $\chi = O(2^n)$ [40]. Dropping the summation over the dummy indices, we can write $\hat{\rho}$ as a MPO

$$\hat{\rho} = \sum_{\substack{i_1, \ldots, i_Q \\ i'_1, \ldots, i'_Q}} W^{(1)i_1 i'_1} W^{(2)i_2 i'_2} \cdots$$
$$\cdots W^{(Q)i_Q i'_Q} |i_1, \ldots, i_Q\rangle \langle i'_1, \ldots, i'_Q|. \quad \text{(E3)}$$

In general this is not an efficient description of a quantum state due to the exponential scaling of the bond dimension. However, for a restricted set of states an efficient representation exists, namely, for states with sufficiently small entanglement. Therefore, for this restricted class of states, expectation values can be classically evaluated. To be more specific, let us consider $\langle X_1^{(1)} X_2^{(2)} \cdots X_Q^{(Q)} \rangle$, where $X_1^{(1)}, X_2^{(2)}, \ldots, X_Q^{(Q)}$ are single-qubit observables acting at
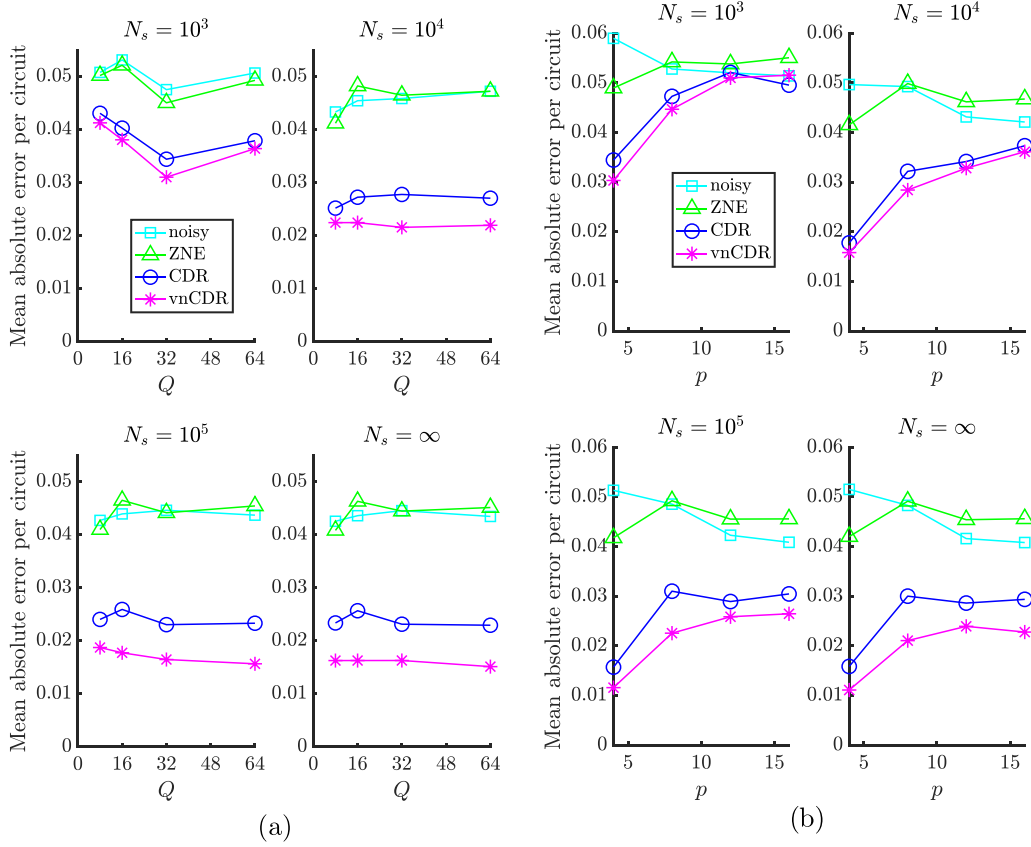
FIG. 7. Correcting the IBMQ-hardware-efficient ansatz with random parameters. Various numbers of shots N_s were used to evaluate every circuit involved in each mitigation method. For each value of Q and p we analyze 30 random circuits and correct $\langle \sigma_X^1 \rangle$, $\langle \sigma_X^{Q/2} \rangle$, $\langle \sigma_Z^1 \sigma_Z^2 \rangle$, and $\langle \sigma_Z^{Q/2} \sigma_Z^{Q/2+1} \rangle$ for each of them. The circuits are the same as those analyzed in Fig. 4. The results from Fig. 4 correspond to the $N_s = \infty$ limit and are shown alongside finite-N_s results. (a) The scaling with Q of the mitigated and unmitigated absolute error per circuit for $p = 6$. In the left panel we show the mean values (solid lines), while the right panel shows the maximal values (dashed lines). (b) The scaling with increasing p for $Q = 8$. The vnCDR method systematically outperforms the ZNE and CDR methods for $N_s \geqslant 10^4$ in the case of the shallow circuits and $N_s \geqslant 10^5$ in the case of the deep circuits making it the method of choice for obtaining high-accuracy results.

qubits $1, 2, \ldots, Q$, respectively. Then

$$\langle X_1^{(1)} X_2^{(2)} \cdots X_Q^{(Q)} \rangle = Y^{(1)} Y^{(2)} \cdots Y^{(Q)}, \qquad \text{(E4)}$$

where $Y^{(1)}, Y^{(2)}, \ldots, Y^{(Q)}$ are matrices obtained as

$$Y_1^{(1)} = \sum_{i_1, i_1'} W^{(1) i_1 i_1'} X_1^{(1) i_1' i_1}, \ldots. \qquad \text{(E5)}$$

Therefore the computational cost scales as $O(\chi^2)$ enabling classical computation for small enough χ.

In the case of our noise model, the initial noisy state corresponding to the exact state $\hat{\rho} = |00 \cdots 0\rangle\langle 00 \cdots 0|$ can be written as a MPO with $\chi = 1$ using standard MPO techniques [40]. Applying these techniques, one can verify that the action of single-qubit gates does not increase χ, while the action of 2-qubit noisy CNOT gates increases χ for bond indices linking 2 qubits at which the CNOT gate acts by at most a factor of 16. Therefore, in the case of our random quantum circuit (RQC) simulation, maximal χ is bounded from above by $16^{p/2}$. In practice we find that MPO representation can be further compressed after each CNOT action using standard MPO compression techniques [40] to discard W elements of the order of numerical precision.

APPENDIX F: RESOURCE SCALING

In the main text we systematically benchmark the performance of the ZNE, CDR, and vnCDR error mitigation methods in the limit of an infinite number of shots. In the case of real-world quantum devices, one is limited to a finite shot number. Here, we analyze the performance of the method in this case. We perform error mitigation of $\langle \sigma_X^1 \rangle$, $\langle \sigma_X^{Q/2} \rangle$, $\langle \sigma_Z^1 \sigma_Z^2 \rangle$, and $\langle \sigma_Z^{Q/2} \sigma_Z^{Q/2+1} \rangle$ for RQC using $N_s = 10^3, 10^4, 10^5$ shots per circuit and compare the results obtained with those evaluated with an infinite number of shots. To enable efficient classical simulation, we consider here independent shots for each observable.

We gather the results in Fig. 7 showing scaling of the error with system size Q and circuit depth p for different shot costs. We find that the improvement of vnCDR over ZNE and CDR grows systematically with increasing N_s. We see that small $N_s = 10^3$ is enough to see a systematic improvement of vnCDR over ZNE for the shallow circuits. With $N_s = 10^4$ we see a systematic improvement of vnCDR over CDR for the shallow circuits and a systematic improvement of vnCDR over ZNE for the deep circuits. With $N_s = 10^5$ shots we also obtain a systematic improvement of vnCDR over CDR for the deep

circuits. We note that for setups in which vnCDR does not outperform other methods, it gives results of a similar quality. We observe that increasing N_s improves performance of the methods only for sufficiently small N_s. In the case of ZNE the results obtained with $N_s = 10^3$ are of quality similar to that of the results obtained in the limit of $N_s = \infty$. For vnCDR and CDR, $N_s = 10^5$ is needed to that end.

To give a full picture of the number of shots needed for the different methods, one also needs to consider the number of circuits required to mitigate the circuit of interest. ZNE requires only the execution of the circuit of interest at various noise levels. Assuming that n noise levels are needed with N_s shots per circuit, the total shot cost for ZNE is $n \times N_s$. Both CDR and vnCDR require the execution of near-Clifford

training circuits as well as the circuit of interest. Assuming a training set consisting of m circuits, each run using N_s shots, the total shot cost for CDR is given as $(m + 1) \times N_s$. vnCDR requires that the Clifford training circuits and the circuit of interest be implemented at various noise levels. With n noise levels and m training circuits, each evaluated using N_s shots, the total shot cost for vnCDR is $(m + 1) \times n \times N_s$. For our RQC results, $n = 5$ and $m = 100$. Therefore the shot cost for ZNE is given as $5 \times N_s$, that for CDR is $101 \times N_s$, and that for vnCDR is $501 \times N_s$. Then, to see systematic improvement over ZNE and CDR, we need 5×10^5 to 5×10^6 and 5×10^6 to 5×10^7 shots in total, respectively. These shot numbers can be obtained with current devices, proving the usefulness of vnCDR.

---

[1] B. Bauer, S. Bravyi, M. Motta, and G. K.-L. Chan, Quantum algorithms for quantum chemistry and quantum materials science, Chem. Rev. (Washington, DC) **120**, 12685 (2020).

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

[3] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, Noise-induced barren plateaus in variational quantum algorithms, arXiv:2007.14384.

[4] K. Temme, S. Bravyi, and J. M. Gambetta, Error Mitigation for Short-Depth Quantum Circuits, Phys. Rev. Lett. **119**, 180509 (2017).

[5] S. Endo, S. C. Benjamin, and Y. Li, Practical Quantum Error Mitigation for Near-Future Applications, Phys. Rev. X **8**, 031027 (2018).

[6] L. Cincio, Y. Subaşı, A. T. Sornborger, and P. J. Coles, Learning the quantum algorithm for state overlap, New J. Phys. **20**, 113022 (2018).

[7] P. Murali, J. M. Baker, A. Javadi-Abhari, F. T. Chong, and M. Martonosi, Noise-adaptive compiler mappings for noisy intermediate-scale quantum computers, in *The 24th ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS'19)* (Association for Computing Machinery, New York, 2019), pp. 1015–1029.

[8] L. Cincio, K. Rudinger, M. Sarovar, and P. J. Coles, Machine learning of noise-resilient quantum circuits, PRX Quantum **2**, 010324 (2021).

[9] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, A variational eigenvalue solver on a photonic quantum processor, Nat. Commun. **5**, 4213 (2014).

[10] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, The theory of variational hybrid quantum-classical algorithms, New J. Phys. **18**, 023023 (2016).

[11] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Variational quantum algorithms, arXiv:2012.09265.

[12] Y. Li and S. C. Benjamin, Efficient Variational Quantum Simulator Incorporating Active Error Minimization, Phys. Rev. X **7**, 021050 (2017).

[13] S. Khatri, R. LaRose, A. Poremba, L. Cincio, A. T. Sornborger, and P. J. Coles, Quantum-assisted quantum compiling, Quantum **3**, 140 (2019).

[14] K. Sharma, S. Khatri, M. Cerezo, and P. J. Coles, Noise resilience of variational quantum compiling, New J. Phys. **22**, 043006 (2020).

[15] R. LaRose, A. Tikku, É. O'Neel-Judy, L. Cincio, and P. J. Coles, Variational quantum state diagonalization, npj Quantum Inf. **5**, 57 (2019).

[16] C. Cirstoiu, Z. Holmes, J. Iosue, L. Cincio, P. J. Coles, and A. Sornborger, Variational fast forwarding for quantum simulation beyond the coherence time, npj Quantum Inf. **6**, 82 (2020).

[17] M. Cerezo, K. Sharma, A. Arrasmith, and P. J. Coles, Variational quantum state eigensolver, arXiv:2004.01372.

[18] T. E. O'Brien, S. Polla, N. C. Rubin, W. J. Huggins, S. McArdle, S. Boixo, J. R. McClean, and R. Babbush, Error Mitigation via Verified Phase Estimation, PRX Quantum **2**, 020317 (2021).

[19] E. F. Dumitrescu, A. J. McCaskey, G. Hagen, G. R. Jansen, T. D. Morris, T. Papenbrock, R. C. Pooser, D. J. Dean, and P. Lougovski, Cloud Quantum Computing of an Atomic Nucleus, Phys. Rev. Lett. **120**, 210501 (2018).

[20] A. He, B. Nachman, W. A. de Jong, and C. W. Bauer, Zero-noise extrapolation for quantum-gate error mitigation with identity insertions, Phys. Rev. A **102**, 012426 (2020).

[21] M. Otten and S. K. Gray, Recovering noise-free quantum observables, Phys. Rev. A **99**, 012338 (2019).

[22] T. Giurgica-Tiron, Y. Hindy, R. LaRose, A. Mari, and W. J. Zeng, Digital zero noise extrapolation for quantum error mitigation, in *2020 IEEE International Conference on Quantum Computing and Engineering (QCE)* (IEEE Computer Society, Los Alamitos, CA, 2020), pp. 306–316.

[23] Z. Cai, Multi-exponential error extrapolation and combining error mitigation techniques for NISQ applications, arXiv:2007.01265.

[24] 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 (London) **567**, 491 (2019).

[25] A. Strikis, D. Qin, Y. Chen, S. C. Benjamin, and Y. Li, Learning-based quantum error mitigation, arXiv:2005.07601.

[26] P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio, Error mitigation with Clifford quantum-circuit data, arXiv:2005.10189.

[27] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge, 2016), http://www.deeplearningbook.org.

[28] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm, arXiv:1411.4028 [quant-ph].

[29] S. Hadfield, Z. Wang, B. O'Gorman, E. G. Rieffel, D. Venturelli, and R. Biswas, From the quantum approximate optimization algorithm to a quantum alternating operator ansatz, Algorithms **12**, 34 (2019).

[30] L. F. Richardson and J. A. Gaunt, VIII. The deferred approach to the limit, Philos. Trans. R. Soc., A **226**, 299 (1927).

[31] G. Szegő, *Orthogonal Polynomials*, American Mathematical Society Colloquium Publications Vol. 23 (American Mathematical Society, Providence, 1975).

[32] J. Vovrosh, K. E. Khosla, S. Greenaway, C. Self, M. Kim, and J. Knolle, Efficient mitigation of depolarizing errors in quantum simulations, arXiv:2101.01690.

[33] M. Urbanek, B. Nachman, V. R. Pascuzzi, A. He, C. W. Bauer, and W. A. de Jong, Mitigating depolarizing noise on quantum computers with noise-estimation circuits, arXiv:2103.08591.

[34] J. M. Chow, L. DiCarlo, J. M. Gambetta, A. Nunnenkamp, Lev S. Bishop, L. Frunzio, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, Detecting highly entangled states with a joint qubit readout, Phys. Rev. A **81**, 062325 (2010).

[35] S. Bravyi, S. Sheldon, A. Kandala, D. C. Mckay, and J. M. Gambetta, Mitigating measurement errors in multiqubit experiments, Phys. Rev. A **103**, 042605 (2021).

[36] M. Fannes, B. Nachtergaele, and R. F. Werner, Finitely correlated states on quantum spin chains, Commun. Math. Phys. **144**, 443 (1992).

[37] G. Vidal, Class of Quantum Many-Body States That Can Be Efficiently Simulated, Phys. Rev. Lett. **101**, 110501 (2008).

[38] J. Abhijith, A. Adedoyin, J. Ambrosiano, P. Anisimov, A. Bärtschi, W. Casper, G. Chennupati, C. Coffrin, H. Djidjev, D. Gunter, S. Karra, N. Lemons, S. Lin, A. Malyzhenkov, D. Mascarenas, S. Mniszewski, B. Nadiga, D. O'Malley, D. Oyen, S. Pakin *et al.*, Quantum algorithm implementations for beginners, arXiv:1804.03719 [cs.ET].

[39] R. Orús, Tensor networks for complex quantum systems, Nat. Rev. Phys. **1**, 538 (2019).

[40] U. Schollwöck, The density-matrix renormalization group in the age of matrix product states, Ann. Phys. (Amsterdam) **326**, 96 (2011).