Deep-learned error mitigation via partially knitted circuits for the variational quantum eigensolver

Simone Cantori[1,2], **Andrea Mari**[1], **David Vitali**[1,2], **Sebastiano Pilati**[1,2]

[1] School of Science and Technology, Physics Division, University of Camerino, I-62032 Camerino (MC), Italy
[2] INFN-Sezione di Perugia, 06123 Perugia, Italy

5 June 2025

Abstract. The variational quantum eigensolver (VQE) is generally regarded as a promising quantum algorithm for near-term noisy quantum computers. However, when implemented with the deep circuits that are in principle required for achieving a satisfactory accuracy, the algorithm is strongly limited by noise. Here, we show how to make VQE functional via a tailored error mitigation technique based on deep learning. Our method employs multilayer perceptrons trained on the fly to predict ideal expectation values from noisy outputs combined with circuit descriptors. Importantly, a circuit knitting technique with partial knitting is adopted to substantially reduce the classical computational cost of creating the training data. We also show that other popular general-purpose quantum error mitigation techniques do not reach comparable accuracies. Our findings highlight the power of deep-learned quantum error mitigation methods tailored to specific circuit families, and of the combined use of variational quantum algorithms and classical deep learning.

1. Introduction

The practical utilization of quantum computing is still challenged by the inherent noise and decoherence in the currently available devices [1, 2, 3, 4]. In fact, noise is expected to be a key challenge also in the near and mid-term future [5, 6]. In this context, variational quantum algorithms [7] have emerged as promising applications that might be less affected by noise. In particular, the variational quantum eigensolver (VQE) algorithm [8] has been proposed for estimating the ground-state energy of quantum systems. However, its performance is, in fact, affected by noise and it is unclear whether the high accuracies required for practical applications can be reached [9, 10, 11, 12, 13]. Indeed, flexible variational ansatzes can be implemented using deep circuits, but the impact of noise rapidly increases with the circuit depth.

For the above reasons, intense research activity is being devoted to the development of error mitigation (EM) strategies [14]. Popular approaches, such as zero-noise extrapolation (ZNE) [15, 16, 17] and probabilistic error cancellation [15, 18, 19, 20], can reduce the impact of noise [21], but often come at the price of considerable

sampling overhead and display limited effectiveness in the context of variational quantum algorithms [22, 23, 24]. Recently, EM techniques based on the integration of machine learning methods with noisy quantum data have been proposed [25, 26, 27, 28, 29, 30, 31, 32]. In particular, deep learning models are expected to succeed in learning the complex mapping between noisy measurements and their corresponding ideal values, effectively filtering out errors. However, a key challenge in using machine learning models is the need for training data with exact expectation values, which demands excessive classical computational resources for generic large circuits. In previous studies, this problem was addressed by the training on easily simulable circuits, such as those representing product states or near-Clifford circuits [26, 28, 29, 33]. However, these circuits differ substantially from those employed in VQE, making them unsuitable for mitigating errors in this task.

In this article, we introduce a deep learning-based quantum error mitigation approach (DL-EM) tailored to the VQE framework. DL-EM leverages artificial neural networks to predict noise-mitigated expectation values by processing hybrid quantum-classical input data composed of unmitigated expectation values (quantum data) and quantum circuit descriptors (classical data). Importantly, the network training is integrated within the VQE process, ensuring that the training circuits are representative of those required in successive stages of the variational optimization. Furthermore, the computational cost for the generation of the training data is drastically reduced by incorporating the circuit knitting technique [34], appropriately adapted with partial knitting. A schematic description of our approach is shown in Fig. 1.

The testbed we consider is the one-dimensional (1D) quantum Ising model with random couplings. Our simulations take into account realistic noise models, validated making comparisons against experiments executed on real IBM quantum devices. We find that while the unmitigated noisy VQE drastically fails to find the ground-state energy, the DL-EM method reaches an accuracy well below the 1% threshold, with a further improvement as the amount of noise is reduced. Instead, general-purpose EM methods, chiefly ZNE and corrections based on damping factors, do not provide comparable improvements. Our combined strategy not only enhances EM but also paves the way for scalable implementations of VQE on current quantum hardware. Our findings highlight the importance of EM methods tailored to specific applications and of the combined use of quantum computers with classical deep learning.

The remainder of this article is organized as follows: In Sec. 2, we describe the target Hamiltonian, the quantum circuits and the noise model, the (partial) circuit knitting technique employed in the generation of the training data, and the VQE process with the iterative DL-EM method. In Sec. 3, a set of numerical experiments is discussed, demonstrating that the DL-EM method substantially improves the accuracy of the noisy VQE algorithm. The performance of other popular general-purpose EM methods is also discussed. Our conclusions are reported in Sec. 4.
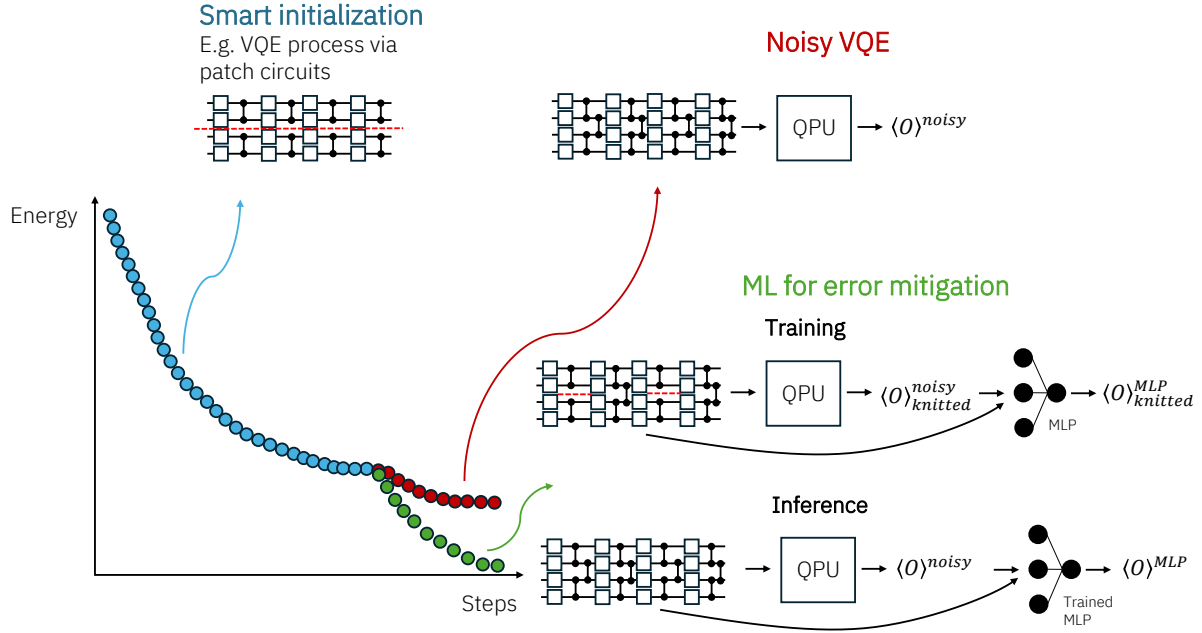
Figure 1: Schematic illustration of the VQE process enhanced by the deep learning-based error mitigation technique (DL-EM). First, a classical smart initialization of the variational parameters is applied performing a classical VQE optimization with simulable patch circuits. Secondly, before proceeding with the final VQE optimization, a neural network, which takes circuit descriptors and noisy expectation values as input data, is trained on partially-knitted quantum circuits having variational parameters close to the optimized configuration. In this training phase, the partially-knitted circuits are executed on the noisy quantum computer and on a classical simulator. The VQE process is then completed using the trained neural network for evaluating error-mitigated expectation values. Note that, in the VQE steps, partial-knitting is not applied and the full variational circuit is executed on the quantum computer.

2. Methods

2.1. Variational quantum eigensolver for the 1D quantum Ising model

Our testbed for the VQE algorithm is the 1D quantum Ising model with nearest-neighbor random couplings. The Hamiltonian reads:

$$H = -\sum_i J_i Z_i Z_{i+1} - h \sum_i X_i \,, \tag{1}$$

where Z_i and X_i are standard Pauli matrices, the couplings are sampled from a uniform distribution in the interval $J_i \in [-1, 1]$, the transverse field is $h = 0.5$, and periodic boundary conditions are adopted, i.e. $Z_{N+1} = Z_1$, with N the number of spins. VQE is a hybrid quantum-classical algorithm designed to approximate the ground-state energy of a quantum system [8]. Leveraging the variational principle, a parameterized quantum

circuit (ansatz) $U(\boldsymbol{\theta})$ with parameters $\boldsymbol{\theta}$ is optimized to minimize the expectation value of the Hamiltonian. A quantum computer estimates the mean energy

$$E(\boldsymbol{\theta}) = \langle \psi(\boldsymbol{\theta})|H|\psi(\boldsymbol{\theta}) \rangle, \tag{2}$$

where $|\psi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})|0\rangle^{\otimes N}$ is the output state of the quantum circuit with input state $|0\rangle^{\otimes N} = |00...0\rangle$. A classical optimizer updates the circuit parameters $\boldsymbol{\theta}$ iteratively, seeking for the minimum. We present results obtained using either gradient-based or gradient-free optimization strategies, specifically, employing the ADAM optimizer [35] with the parameter-shift rule [36, 37], and the COBYLA algorithm [38], respectively. We adopt the circuit architecture described in Ref. [39], which was originally designed for the pure ferromagnetic 1D quantum Ising model, corresponding to the Hamiltonian (1) with $J_i = 1$ for $i = 1, \ldots, N$. Since our model features random couplings, we extend the original circuit ansatz by assigning independent parameters to each gate, allowing different rotation angles for individual qubits across all circuit layers. The resulting ansatz is visualized in Fig. 2a.

A key feature of our DL-EM method is that the neural networks are iteratively trained on the fly during the circuit optimization. The training is repeated when the target circuits differ substantially from those included in the training set. Furthermore, the training circuits are created using the knitting technique, as further explained in Sec. 2.3. A partial knitting procedure allows us to find a balance between the classical computational cost and the similarity between training and target sets. The VQE process starts with a smart parameter initialization, in which a variational optimization is performed with so-called patch circuits [40]. These are formed by removing gates connecting two (or more) patches, allowing efficient classical execution. Clearly, due to the disconnected structure, the patch circuits reach only an approximate representation of the ground state. A schematic representation of the VQE process with our tailored EM method is shown in Fig. 1. In summary, it involves the following steps:

(i) Smart parameter initialization with patch circuits;

(ii) Training of the DL-EM neural network on partially knitted quantum circuits;

(iii) Execution of VQE with DL-EM;

(iv) Re-training of the DL-EM neural network when the circuit parameters deviate significantly from those used in the previous training phase;

(v) Repetition of steps 3 and 4 until convergence is reached.

It is worth mentioning that other smart initialization criteria have been proposed in the literature [41, 42]. However, in our framework, patch circuits are particularly effective since the first steps of the parameter optimization involve small angles in the R_{ZZ} gates connecting the two patches. This feature reduces the computational cost of the circuit knitting technique employed in the training phase.

Even in the absence of noise, the accuracy reachable with VQE depends on the flexibility of the circuit ansatz. In fact, the required circuit depth was shown to scale
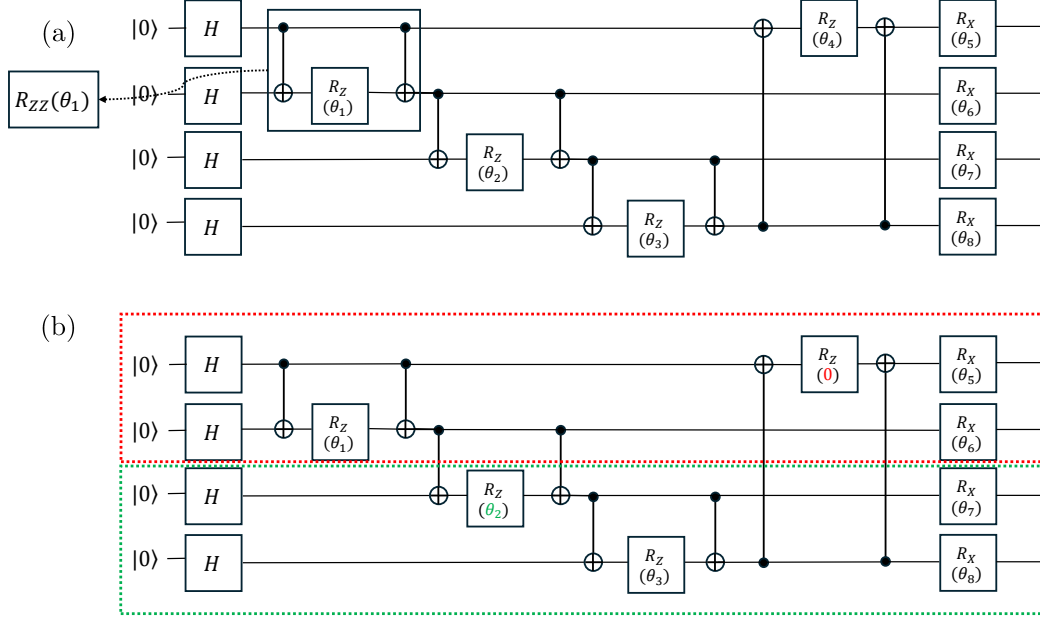
Figure 2: (a) Scheme of the first layer of the VQE ansatz. The same quantum gates, besides the initial superposition obtained with the Hadamard gates, are then repeated P times with different parameters θ_i. The decomposition of the R_{ZZ} gate is also shown. (b) We divide the quantum circuit into 2 sub-circuits with the circuit knitting technique. Some of the parameters of the knitting gates are fixed to 0 (or π) effectively cutting the two partitions. Other parameters are left unchanged ($R_Z(\theta_2)$ in the figure), and a sampling overhead is needed to reconstruct the expectation values of the full circuit starting from the expectation values of the two sub-circuits, as explained in the main text.

with the size of the target system [39, 43]. For this reason, we adopt a relatively deep circuit featuring $P = 8$ layers. Furthermore, in our simulations we adopt a realistic model of noise, as discussed in Sec. 2.2. Therein, its effect is compared to the errors occurring in physical devices available from the IBM Quantum platform. As detailed in Sec. 3, VQE with deep circuits is drastically affected by realistic noise, leading to poor performance. Some previous studies [9, 21, 44] reported relatively good accuracies of unmitigated noisy VQE, but considered shallow circuits suitable for small physical systems.

In Sec. 3, we also perform comparisons with other EM techniques. An important one is based on the observation that, if O is a traceless operator, its noisy expectation value under depolarizing noise is proportional to the exact expectation value $\langle O \rangle_{\text{noisy}} = D \langle O \rangle$, where D is the damping factor [45, 44]. By removing the single-qubit gates, one obtains a quantum circuit whose ideal result is known, because built using only CNOT gates. Hence, measuring the noisy expectation value allows determining D. The estimated damping factor can then be used to mitigate the expectation value of the original circuit
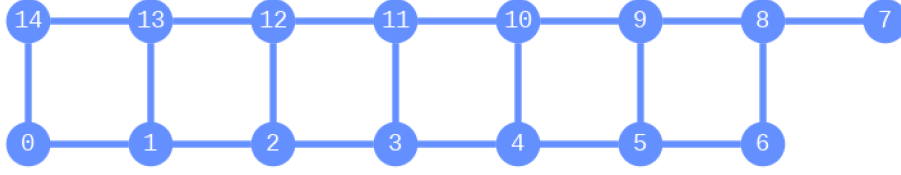
Figure 3: Layout of the IBM Melbourne quantum chip.

by rescaling its noisy output. A similar approach can be used in combination with ZNE [43]. In our analysis, we consider the quantum circuit resulting from the last step of a VQE process, and we execute this error mitigation, dubbed damping-factor method (DM), both with and without ZNE. In both scenarios, the results are far less accurate than those achieved by DL-EM. In Sec. 3.2, we also discuss other machine learning-based EM methods and how they compare with our strategy.

2.2. Noise model

The noisy expectation values are obtained by simulating a realistic quantum device using a noise model imported from Qiskit [46], namely, the FakeMelbourneV2 model. The chip layout corresponding to this device is shown in Fig. 3. Notably, this layout allows preserving periodic boundary conditions across 1D quantum circuits of different sizes, up to $N = 14$. In our simulations, the noisy expectation values $\langle Z_i Z_{i+1} \rangle_{\text{noisy}}$ and $\langle X_i \rangle_{\text{noisy}}$ are obtained performing $S = 10^6$ shots. To thoroughly analyse the impact of noise, we customize the FakeMelbourneV2 model so that the amount of noise can be tuned. To achieve this, we follow the technique originally deployed in Ref. [27]. Specifically, for each gate, the associated error channel is replaced with a noiseless identity with probability $1 - p_{\text{noise}}$. The extreme case $p_{\text{noise}} = 0$ corresponds to a completely noiseless circuit, while the other extreme $p_{\text{noise}} = 1$ represents the original noise model. The readout error is similarly scaled by a factor p_{noise}, meaning that there is no error when $p_{\text{noise}} = 0$, while the error model is unchanged when $p_{\text{noise}} = 1$. To verify that the tunable FakeMelbourneV2 model is indeed realistic, we make a comparison vis-á-vis three actual physical devices available from the IBM Quantum platform. The testbed is a quantum Ising chain with $N = 12$ spins. This size is chosen to allow periodic boundary conditions to be implemented on the native topology of the chips. The test quantum circuit is the final stage of a noise-free VQE process. The error in the energy expectation value compared to the ideal case $p_{\text{noise}} = 0$ is shown in Fig. 4. We find that the three devices perform slightly better than the worst-case scenario $p_{\text{noise}} = 1$, reaching performances comparable to those of the regime $0.2 \lesssim p_{\text{noise}} \lesssim 0.7$. As discussed in Sec. 5, unmitigated VQE would require noise levels reduced by various orders of magnitude compared to this regime.
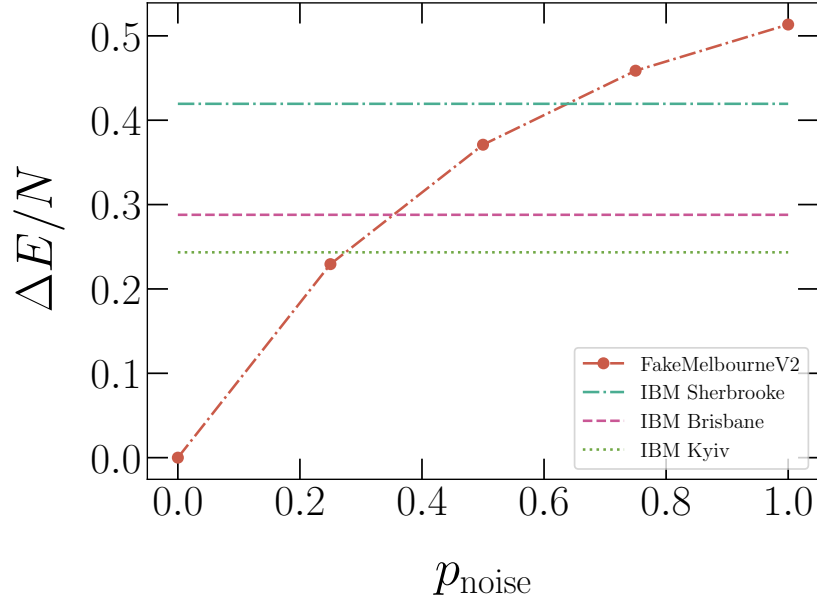
Figure 4: Energy discrepancies $\Delta E/N$ obtained by classically simulating the FakeMelbourneV2 model, as a function of the noise level p_{noise}. These simulations are compared with actual experiments performed on the three quantum devices indicated in the legend, denoted by the horizontal lines. The three devices are accessed via the IBM quantum platform. The testbed is a quantum circuit representing the final step of a VQE process with $N = 12$ qubits. The energy obtained at $p_{\text{noise}} = 0$ is the reference to calculate the discrepancy.

2.3. Classically simulable training circuits via partial circuit knitting

Circuit knitting techniques provide a suitable strategy for computing the output expectation values of quantum circuits with a large number of qubits [34, 47, 48]. Large circuits are decomposed into smaller, more manageable sub-circuits by removing the gates connecting two or more patches. This decomposition is particularly useful for near-term quantum devices, where circuit depth and qubit connectivity are limited by hardware constraints, meaning that removing few gates suffices to form disconnected parts. In our framework, the decomposition leads to small sub-circuits that can be efficiently classically simulated. In detail, we leverage circuit knitting by cutting two rows of R_{ZZ} gates. Taking into account periodic boundary conditions, this partitioning results in two independent sub-circuits, each containing $N/2$ qubits. If necessary, the number of sub-circuits can be increased to reduce their size below the capability of the classical simulator. The ideal expectation values of the entire circuit can be reconstructed by combining (*knitting*) the results from multiple variations of these smaller sub-circuits. However, this reconstruction comes at the cost of a potentially large sampling overhead, depending on the number of cut gates and their rotation parameters.

Quantitatively, the sampling overhead scales as [34]:

$$O = \prod_{\theta_i \in \mathcal{K}} (1 + 2|\sin(\theta_i)|)^2 \,, \tag{3}$$

where $\mathcal{K}$ represents the set of R_{ZZ} gates connecting the two sub-circuits. This overhead can become prohibitively large, making full circuit reconstruction impractical for deep circuits or large qubit numbers, especially if the angles are sizable. Fortunately, in our framework, it is in fact unnecessary to fully reconstruct the whole circuit. Instead, a subset of $\mathcal{K}$ composed of $C \geq 0$ gates can remain unknitted, i.e., their associated parameters θ_i are set to 0 or π. In this way, C cuts are effectively introduced in the original circuit such that, for a sufficiently large C, the computation becomes classically feasible. This partial circuit knitting approach ensures that the generated training data is representative of the target circuit while controlling the sampling cost. A schematic representation of the knitting procedure is shown in Fig. 2b. Each circuit is cut at the qubits q and $(q + N/2) \mod N$, and the random index q is uniformly sampled in the interval $q \in [1, N]$. The layers in which the gates are not knitted are, again, chosen randomly. In Sec. 3.2, we analyse the number of cuts C needed to get accurate results for quantum circuits with different numbers of qubits.

2.4. Training of the neural networks

We employ standard multilayer perceptrons (MLPs) to predict ideal expectation values from the noisy outputs of the quantum device. For a quantum circuit with N qubits and P layers, we train $2 \times N$ independent neural networks to predict $\langle Z_i Z_{i+1} \rangle$ and $\langle X_i \rangle$, for $i = 1, \ldots, N$. These expectation values allow the computation of the ground-state energy E. The inputs of the MLP network are the noisy expectation values and the sine and cosine of the set of angles $\boldsymbol{\theta} = \{\theta_1, \theta_2, \ldots, \theta_{2 \times N \times P}\}$ that characterize each quantum circuit.

The training circuits are designed to resemble those encountered in different stages of the VQE optimization. Starting from the circuit obtained from the previous stage of VQE, random angle variations are applied. First, since the parameter shift rule is used to determine the gradient, the training set is augmented by including circuits that feature one $\pm \frac{\pi}{2}$ angle shift. Specifically, the shift is applied to one of the parametrized gates, randomly selected with a uniform probability, and with the same probability the circuit without shift is considered. Moreover, a small random shift is applied to each angle. This is sampled from a normal distribution with zero mean and standard deviation $\sigma = 0.05$. This additional offset allows using the same neural network for several VQE steps, even when the circuit parameters have sizably changed. The small variance facilitates the training process [49]. In our implementation, the training process is repeated every 9 steps of VQE optimization. This number is chosen because, after this many steps, some gate parameters typically shift more than σ. It is worth mentioning that transfer learning methods [50, 51] might allow reducing the required amount of training data [52, 53]. The training circuits are classically simulated using the circuit

knitting technique with two sub-circuits, as discussed in Sec. 2.3. For large-scale circuits, the sampling technique characteristic of the circuit knitting methods could be adopted. However, in our study, given the small size of the circuits, we employ a statevector simulation [54]. It is important to emphasize that, in our analysis of the DL-EM method, the computational bottleneck is represented by the classical simulations of the noisy training circuits. In fact, these are computationally more demanding than noise-free simulations. Clearly, when deploying the DL-EM technique on actual quantum devices, these simulations are not needed, allowing scaling to larger qubit numbers. The statistical noise induced by finite sampling associated with circuit knitting is analysed in Sec. 3.2.

The training of the MLPs is performed by minimizing the mean-squared-error loss function:

$$\mathcal{L} = \frac{1}{K_{\text{train}}} \sum_{k=1}^{K_{\text{train}}} (y_k - \tilde{y}_k)^2 \, , \tag{4}$$

where K_{train} is the number of training instances, y_k is the target value, and $\tilde{y}_k$ is the corresponding predicted value. The network parameters are optimized using the ADAM algorithm. To evaluate the prediction accuracy, we compute the coefficient of determination

$$R^2 = 1 - \frac{\sum_{k=1}^{K_{\text{test}}} (y_k - \tilde{y}_k)^2}{\sum_{k=1}^{K_{\text{test}}} (y_k - \bar{y})^2}, \tag{5}$$

where $\bar{y}$ is the average of the target values and K_{test} is the number of test instances. The coefficient of determination quantifies how well the model captures the variations in the target data. A normalized error measure is obtained as $1 - R^2$, which corresponds to the ratio of the mean squared error over the data variance.

3. Results

3.1. Testing the VQE algorithm improved by DL-EM

Hereafter, the ground-state energy of random quantum Ising chains featuring $N = 6$ qubits is determined using the VQE algorithm. In particular, the performance of a noisy device model, namely the FakeMelbourneV2 model, is tested using different EM methods. The VQE process for a representative instance of the random Hamiltonian (1) is visualized in Fig. 5. With the chosen circuit depth, the noiseless execution of VQE rapidly approaches the exact ground state energy E_0. Instead, when performed with the noise model, the optimization remains well above E_0. Notably, the DL-EM method restores convergence, as evidenced by the rapid descent after the smart initialization stage. On the other hand, a popular general-purpose EM method, namely, ZNE, provides only a marginal improvement compared to the noisy VQE. The final accuracy is better visualized in Fig. 6, where the energy discrepancy ΔE_0 is plotted as a function of the noise level p_{noise}. Even reducing this parameter to $p_{\text{noise}} = 0.01$, the energy discrepancy of noisy VQE is still above 1%, meaning that VQE performed
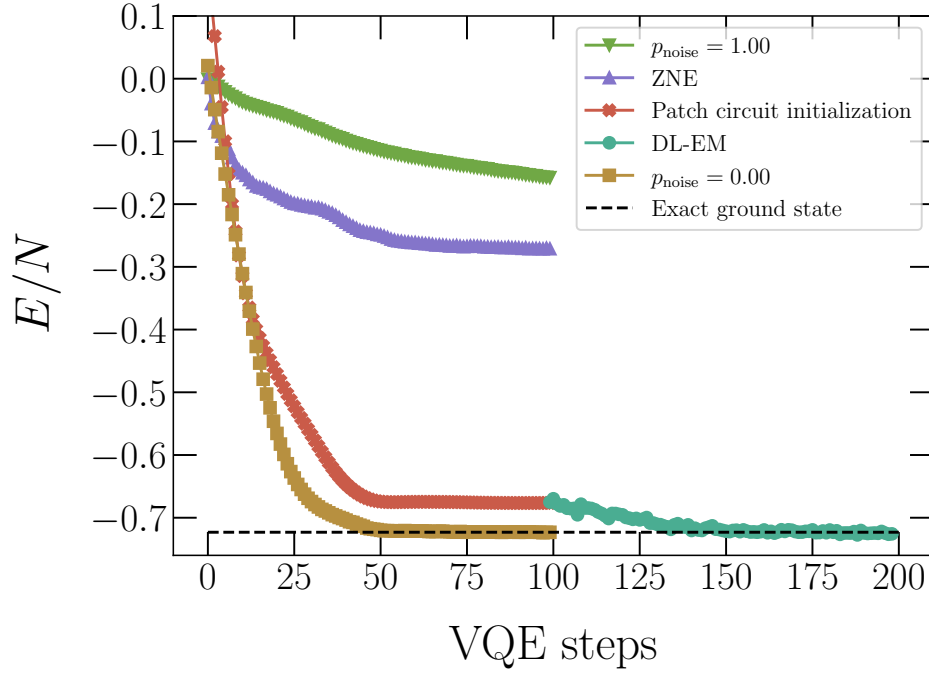
Figure 5: Energy per qubit E/N as a function of the number of VQE optimization steps. We compare the ideal VQE execution ($p_{noise} = 0$), the noisy VQE performed with the FakeMelbourneV2 model ($p_{noise} = 1$), the noisy VQE mitigated using our DL-EM method, and a general-purpose EM method, namely, ZNE. The smart initialization using patch circuits is also shown. The exact ground-state energy is indicated by the horizontal dashed line. VQE is performed using a variational circuit with $P = 8$ layers acting on $N = 6$ qubits. The DL-EM network is trained on partially knitted circuits with $C = 12$ cuts. ZNE is performed using noise scaling factors 1, 3, and 5, corresponding to 0, 1, and 2 repetitions of the quantum circuit and its inverse, with a linear extrapolation method [55].

on near-term quantum devices without EM is impractical for the considered circuit depth. Gradient-based and gradient-free optimization algorithms provide comparable performances. It is worth mentioning that starting the unmitigated VQE from the patch-circuit initialization, as for the DL-EM approach, doesn't improve the results (not shown). The DL-EM method reaches an error below the 1% threshold even when executed with the original noise model, corresponding to $p_{noise} = 1$, and below 0.1% when the noise level is reduced to $p_{noise} = 0.1$. Fig. 6 also analyses the performance of another general-purpose EM strategy, namely, the damping factor method (DM, see Sec. 2.1), executed with or without ZNE. To implement this method, we analyse the quantum circuit resulting from the last step of the VQE process performed with the DL-EM.

To further characterize the performance of the DL-EM method, we repeat the VQE process for four random instances of the quantum Ising chain, at noise level $p_{noise} = 0.1$.
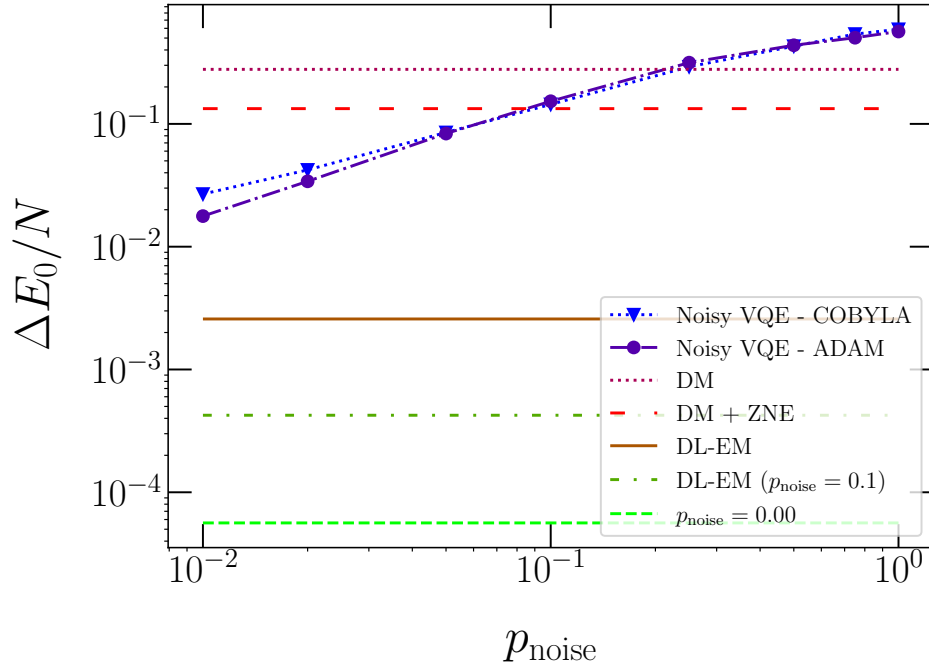
Figure 6: Discrepancy per qubit $\Delta E_0/N$ in the estimate of the ground state energy as a function of the noise level p_{noise}. The two datasets with symbols correspond to noisy VQE performed with COBYLA and with ADAM optimizers. The horizontal lines denote: DM approaches with and without ZNE, noisy VQE boosted by the DL-EM method executed on the FakeMelbourneV2 model and on a reduced noise model ($p_{\text{noise}} = 0.1$), and the noiseless execution of VQE ($p_{\text{noise}} = 0$). For the DM approaches, the discrepancy is computed applying the correction to the last circuit of VQE performed with the DL-EM method. For the other results, the whole VQE process is performed with the ADAM optimizer, unless otherwise specified.

This yields an average error of $\Delta E_0/N \simeq 9.5 \times 10^{-4}$, indicating a systematic convergence of VQE with DL-EM.

3.2. Trade-off between training resources and prediction accuracy in deep learned error mitigation

This section examines how the structural similarity between training and target circuits affects the prediction accuracy. This analysis also demonstrates that the classical computational cost for training the DL-EM can be reduced without compromising its accuracy. One expects that, the more closely the training quantum circuits resemble the target circuit, the better the predictive performance of the model. This expectation is supported by the error measures $1 - R^2$ reported in Fig. 7, as discussed hereafter. The training and test sets analysed in this section are generated from quantum circuits optimized via simulated VQE processes, each executed for 100 steps with random initialization. Then, offsets and cuts are executed as described in Sec. 2. The integer C
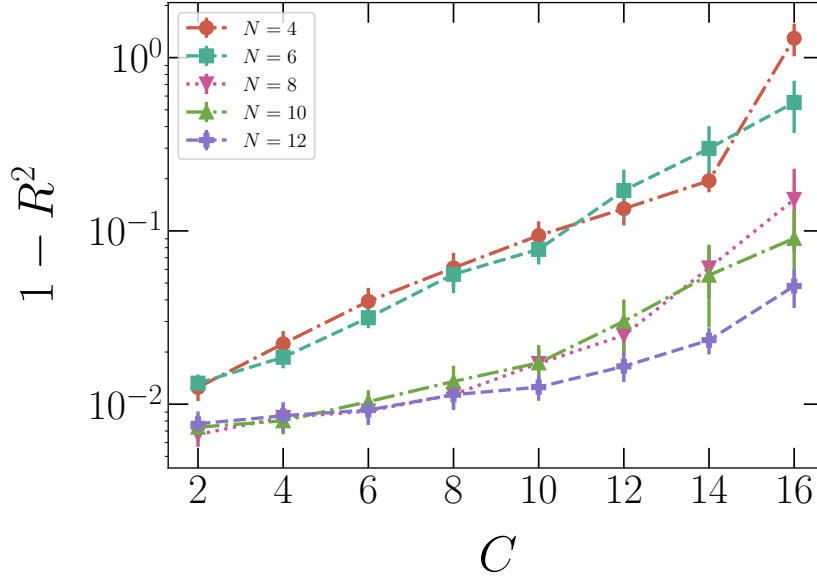
Figure 7: Prediction error $1 - R^2$ as a function of the number of cuts C (i.e, the number of not-knitted R_{ZZ} gates) in the training circuits, for different numbers of qubits N. The training-set includes 10^4 circuits, while the test-set includes 10^3 exactly simulated full quantum circuits, namely, setting $C = 0$.

denotes the number of cuts, i.e., the number of R_{ZZ} gates whose angle parameters θ_i are fixed to 0 or π. Reducing C (i.e., using fewer cuts) results in training circuits that more closely mimic the target circuits, thereby improving prediction accuracy. However, this comes at the cost of increased computational resources during training data generation, due to the larger sampling overhead. As shown in Fig. 7, there is a trade-off between the accuracy of the predictions and the computational cost of simulating the training circuits. Notably, the accuracy of DL-EM does not degrade with the size of the target circuit for a fixed C. On the other hand, if the (partial) knitting is implemented with only two patches, the classical simulation cost scales as $2^{N/2}$.

Some machine learning-based EM methods have already been discussed in the previous literature. In one prominent method, the training is performed on product states [33]. In our framework, product states are obtained by performing $C = 48$ cuts, which is well beyond the regime where the error metric is satisfactory (see Fig. 7). Likewise, approaches that train using near-Clifford circuits [26, 28] are not suitable for our setting. Indeed, these methods require most gates to be Clifford. However, even in the worst-case scenario $C = 16$ of Fig. 7, where the performance is already compromised, 80 of the 96 parameterized gates are non-Clifford, making Clifford-based EM methods inadequate. Other machine learning-based methods [30] that learn how to mimic expectation values mitigated via standard approaches, such as ZNE, are limited by the accuracy of such techniques, and these appear to perform sub-optimally for the analysed circuits.
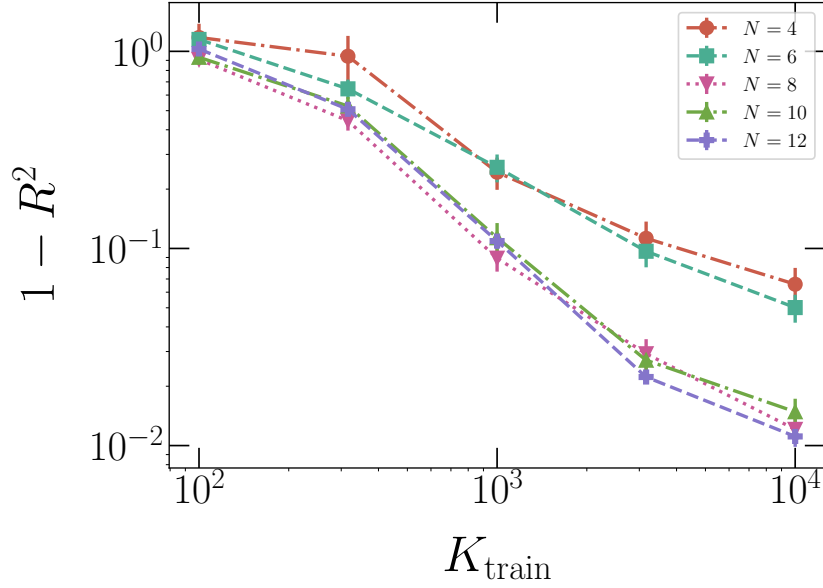
Figure 8: Prediction error $1 - R^2$ as a function of the training-set size K_{train}, for different number of qubits N. The training circuits feature $C = 8$ cuts, while the test-set includes 10^3 full quantum circuits, namely, setting $C = 0$.

Training deep neural networks requires sufficiently large datasets, potentially leading to overwhelming computational costs for their generation. In Fig. 8, we show that the prediction error achieved by our DL-EM model decreases quite rapidly with the training-set size K_{train}. Importantly, we find that the learning speed does not decrease with the qubit number, suggesting that the DL-EM method could be scaled to larger circuits.

Finally, it is important to analyse the possible detrimental effect of shot noise. In fact, shot noise is inherent in the use of quantum devices, since expectation values are estimated with a finite number of shots S, leading to stochastic fluctuations in these estimates. One might expect that shot noise is particularly impactful in our DL-EM method, since the standard circuit knitting technique in principle requires an exponentially scaling shot number, as predicted by Eq. (3). However, it has been found that deep neural networks can often be trained on target values affected by random fluctuations, effectively filtering the noise out [56, 57]. In particular, this phenomenon was observed in the learning of circuit outputs affected by shot noise [52]. Indeed, in Fig. 9 we show that the DL-EM model can be trained on target values derived from partial circuit knitting with a small shot number S. Although noisy circuit outputs do not accurately approximate the exact expectation values, leading to a large error metric $1 - R^2$, the DL-EM model provides more accurate predictions, and these rapidly improve as the size of the training set increases.
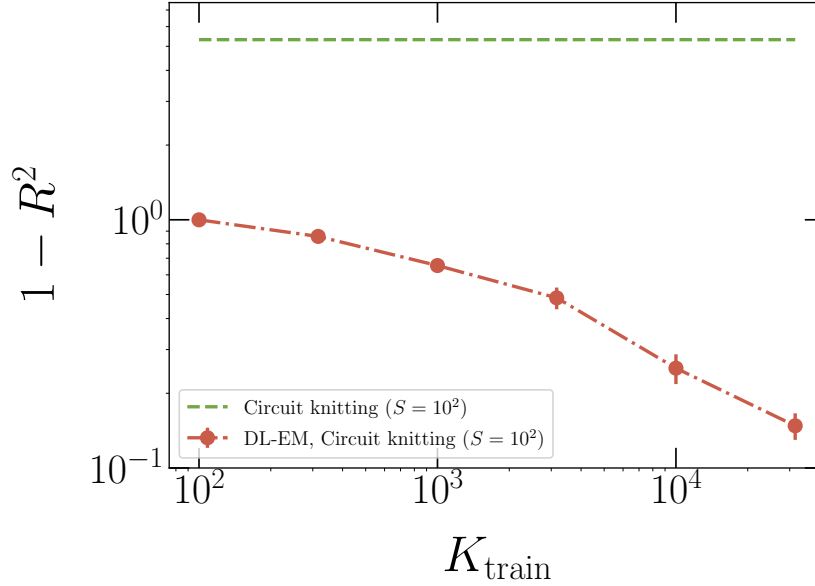
Figure 9: Prediction error $1 - R^2$ as a function of the training-set size K_{train}. The DL-EM is trained on target values obtained with the circuit knitting technique with $C = 14$ cuts (i.e. only two knitted gates) and $S = 10^2$ shots. The model is tested on quantum circuits with the same number of cuts $C = 14$, but simulated using exact statevector techniques (i.e. without shot noise). The horizontal line represents the error of the expectation values obtained with the circuit knitting technique with $S = 10^2$, measured as the deviation from the statevector results.

4. Conclusions

Our work introduces a deep learning-based quantum error mitigation strategy (DL-EM) that significantly improves the performance of variational quantum eigensolvers in the presence of realistic noise. By training deep neural networks on hybrid quantum-classical data, combining noisy expectation values and circuit descriptors, we provide accurate predictions of the ground state energies for the 1D random quantum Ising model. An important aspect is the use of partial circuit knitting, which balances classical computational cost with the need to keep training circuits representative of the target variational circuits, without incurring the exponential overhead of full circuit reconstruction.

Our results demonstrate that DL-EM outperforms general-purpose EM techniques such as ZNE and DM approaches, achieving small energy discrepancies. Notably, the performance further improves as noise is reduced, meaning that DL-EM will strongly benefit from the further development of quantum devices. These findings highlight the critical role of application-specific EM and the synergistic use of quantum resources with classical machine learning.

This work presents DL-EM as a potential approach for improving deep-circuit VQE on near-term quantum hardware by addressing some of the challenges introduced by

noise. Future directions include applying the framework to larger systems and more complex Hamiltonians, implementing transfer learning to reduce training costs, and exploring hybrid methods that combine DL-EM with other EM strategies. Further validation across different quantum architectures and systematic scalability studies will be important to assess its practical impact. Overall, these contributions add to the broader effort to improve the reliability and accuracy of variational quantum algorithms on current hardware.

Acknowledgments

[1] J. Preskill. Quantum Computing in the NISQ era and beyond. *Quantum*, 2:79, August 2018. `doi:10.22331/q-2018-08-06-79`.

[2] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, and A. Aspuru-Guzik. Noisy intermediate-scale quantum algorithms. *Rev. Mod. Phys.*, 94:015004, Feb 2022. URL: `https://link.aps.org/doi/10.1103/RevModPhys.94.015004`, `doi:10.1103/RevModPhys.94.015004`.

[3] D. Stilck França and R. García-Patrón. Limitations of optimization algorithms on noisy quantum devices. *Nature Physics*, 17(11):1221–1227, Nov 2021. `doi:10.1038/s41567-021-01356-3`.

[4] Y. Zhou, E. M. Stoudenmire, and X. Waintal. What limits the simulation of quantum computers? *Phys. Rev. X*, 10:041038, Nov 2020. URL: `https://link.aps.org/doi/10.1103/PhysRevX.10.041038`, `doi:10.1103/PhysRevX.10.041038`.

[5] R. Zhang, G. Wang, and P. Johnson. Computing Ground State Properties with Early Fault-Tolerant Quantum Computers. *Quantum*, 6:761, July 2022. `doi:10.22331/q-2022-07-11-761`.

[6] J. Preskill. Beyond NISQ: The megaquop machine. *ACM Transactions on Quantum Computing*, 6(3):1–7, April 2025. URL: `http://dx.doi.org/10.1145/3723153`, `doi:10.1145/3723153`.

[7] 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. *Nature Reviews Physics*, 3(9):625–644, August 2021. URL: `http://dx.doi.org/10.1038/s42254-021-00348-9`, `doi:10.1038/s42254-021-00348-9`.

[8] J. Tilly, H. Chen, S. Cao, D. Picozzi, K. Setia, Y. Li, E. Grant, L. Wossnig, I. Rungger, G. H. Booth, and J. Tennyson. The variational quantum eigensolver: A review of methods and best practices. *Physics Reports*, 986:1–128, November 2022. URL: `http://dx.doi.org/10.1016/j.physrep.2022.08.003`, `doi:10.1016/j.physrep.2022.08.003`.

[9] A. Bentellis, A. Matic-Flierl, C. B. Mendl, and J. M. Lorenz. Benchmarking the variational quantum eigensolver using different quantum hardware. In *2023 IEEE International Conference*

on Quantum Computing and Engineering (QCE), volume 01, pages 518–523, 2023. `doi: 10.1109/QCE57702.2023.00065`.

[10] E. Fontana, N. Fitzpatrick, D. M. n. Ramo, R. Duncan, and I. Rungger. Evaluating the noise resilience of variational quantum algorithms. *Phys. Rev. A*, 104:022403, Aug 2021. URL: `https://link.aps.org/doi/10.1103/PhysRevA.104.022403`, `doi:10.1103/PhysRevA.104.022403`.

[11] P. Singkanipa and D. A. Lidar. Beyond unital noise in variational quantum algorithms: noise-induced barren plateaus and limit sets. *Quantum*, 9:1617, January 2025. `doi:10.22331/q-2025-01-30-1617`.

[12] W. Saib, P. Wallden, and I. Akhalwaya. The effect of noise on the performance of variational algorithms for quantum chemistry. In *2021 IEEE International Conference on Quantum Computing and Engineering (QCE)*, pages 42–53, 2021. `doi:10.1109/QCE52317.2021.00020`.

[13] S. Brandhofer, S. Devitt, and I. Polian. Error analysis of the variational quantum eigensolver algorithm. In *2021 IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH)*, pages 1–6, 2021. `doi:10.1109/NANOARCH53687.2021.9642249`.

[14] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O'Brien. Quantum error mitigation. *Reviews of Modern Physics*, 95(4), December 2023. URL: `http://dx.doi.org/10.1103/RevModPhys.95.045005`, `doi:10.1103/revmodphys.95.045005`.

[15] K. Temme, S. Bravyi, and J. M. Gambetta. Error mitigation for short-depth quantum circuits. *Phys. Rev. Lett.*, 119:180509, Nov 2017. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.119.180509`, `doi:10.1103/PhysRevLett.119.180509`.

[16] Y. Li and S. C. Benjamin. Efficient variational quantum simulator incorporating active error minimization. *Phys. Rev. X*, 7:021050, Jun 2017. URL: `https://link.aps.org/doi/10.1103/PhysRevX.7.021050`, `doi:10.1103/PhysRevX.7.021050`.

[17] 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(7749):491–495, March 2019. URL: `http://dx.doi.org/10.1038/s41586-019-1040-7`, `doi:10.1038/s41586-019-1040-7`.

[18] S. Endo, S. C. Benjamin, and Y. Li. Practical quantum error mitigation for near-future applications. *Phys. Rev. X*, 8:031027, Jul 2018. URL: `https://link.aps.org/doi/10.1103/PhysRevX.8.031027`, `doi:10.1103/PhysRevX.8.031027`.

[19] S. Zhang, Y. Lu, K. Zhang, W. Chen, Y. Li, J.-N. Zhang, and K. Kim. Error-mitigated quantum gates exceeding physical fidelities in a trapped-ion system. *Nature Communications*, 11(1), January 2020. URL: `http://dx.doi.org/10.1038/s41467-020-14376-z`, `doi:10.1038/s41467-020-14376-z`.

[20] E. van den Berg, Z. K. Minev, A. Kandala, and K. Temme. Probabilistic error cancellation with sparse Pauli–Lindblad models on noisy quantum processors. *Nature Physics*, 19(8):1116–1121, May 2023. URL: `http://dx.doi.org/10.1038/s41567-023-02042-2`, `doi:10.1038/s41567-023-02042-2`.

[21] T. Kurita, H. Qassim, M. Ishii, H. Oshima, S. Sato, and J. Emerson. Synergetic quantum error mitigation by randomized compiling and zero-noise extrapolation for the variational quantum eigensolver. *Quantum*, 7:1184, November 2023. `doi:10.22331/q-2023-11-20-1184`.

[22] S. Wang, P. Czarnik, A. Arrasmith, M. Cerezo, L. Cincio, and P. J. Coles. Can Error Mitigation Improve Trainability of Noisy Variational Quantum Algorithms? *Quantum*, 8:1287, March 2024. `doi:10.22331/q-2024-03-14-1287`.

[23] Y. Quek, D. Stilck França, S. Khatri, J. J. Meyer, and J. Eisert. Exponentially tighter bounds on limitations of quantum error mitigation. *Nature Physics*, 20(10):1648–1658, July 2024. URL: `http://dx.doi.org/10.1038/s41567-024-02536-7`, `doi:10.1038/s41567-024-02536-7`.

[24] R. Takagi, S. Endo, S. Minagawa, and M. Gu. Fundamental limits of quantum error mitigation. *npj Quantum Information*, 8(1), September 2022. URL: `http://dx.doi.org/10.1038/s41534-022-00618-z`, `doi:10.1038/s41534-022-00618-z`.

[25] T. Jiang, J. Rogers, M. S. Frank, O. Christiansen, Y.-X. Yao, and N. Lanatà. Error mitigation in variational quantum eigensolvers using tailored probabilistic machine learning. *Phys. Rev. Res.*, 6:033069, Jul 2024. URL: `https://link.aps.org/doi/10.1103/PhysRevResearch.6.033069`, `doi:10.1103/PhysRevResearch.6.033069`.

[26] P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio. Error mitigation with Clifford quantum-circuit data. *Quantum*, 5:592, November 2021. `doi:10.22331/q-2021-11-26-592`.

[27] S. Cantori, A. Mari, D. Vitali, and S. Pilati. Synergy between noisy quantum computers and scalable classical deep learning for quantum error mitigation. *EPJ Quantum Technology*, 11(1):45, Jul 2024. `doi:10.1140/epjqt/s40507-024-00256-8`.

[28] 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, Jul 2021. URL: `https://link.aps.org/doi/10.1103/PhysRevResearch.3.033098`, `doi:10.1103/PhysRevResearch.3.033098`.

[29] M. Liao, Y. Zhu, G. Chiribella, and Y. Yang. Noise-agnostic quantum error mitigation with data augmented neural models. *npj Quantum Information*, 11(1):8, Jan 2025. `doi:10.1038/s41534-025-00960-y`.

[30] H. Liao, D. S. Wang, I. Sitdikov, C. Salcedo, A. Seif, and Z. K. Minev. Machine learning for practical quantum error mitigation. *Nature Machine Intelligence*, 6(12):1478–1486, Dec 2024. `doi:10.1038/s42256-024-00927-2`.

[31] E. R. Bennewitz, F. Hopfmueller, B. Kulchytskyy, J. Carrasquilla, and P. Ronagh. Neural error mitigation of near-term quantum simulations. *Nature Machine Intelligence*, 4(7):618–624, Jul 2022. `doi:10.1038/s42256-022-00509-0`.

[32] X.-Y. Xu, X. Xue, T. Chen, C. Ding, T. Li, H. Zhou, H.-L. Huang, and W.-S. Bao. Physics-inspired machine learning for quantum error mitigation, 2025. URL: `https://arxiv.org/abs/2501.04558`, `arXiv:2501.04558`.

[33] S. H. Sack and D. J. Egger. Large-scale quantum approximate optimization on nonplanar graphs with machine learning noise mitigation. *Phys. Rev. Res.*, 6:013223, Mar 2024. URL: `https://link.aps.org/doi/10.1103/PhysRevResearch.6.013223`, `doi:10.1103/PhysRevResearch.6.013223`.

[34] K. Mitarai and K. Fujii. Constructing a virtual two-qubit gate by sampling single-qubit operations. *New Journal of Physics*, 23(2):023021, feb 2021. URL: `https://dx.doi.org/10.1088/1367-2630/abd7bc`, `doi:10.1088/1367-2630/abd7bc`.

[35] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. 2017. `arXiv:1412.6980`.

[36] J. Li, X. Yang, X. Peng, and C.-P. Sun. Hybrid quantum-classical approach to quantum optimal control. *Phys. Rev. Lett.*, 118:150503, Apr 2017. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.118.150503`, `doi:10.1103/PhysRevLett.118.150503`.

[37] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii. Quantum circuit learning. *Phys. Rev. A*, 98:032309, Sep 2018. URL: `https://link.aps.org/doi/10.1103/PhysRevA.98.032309`, `doi:10.1103/PhysRevA.98.032309`.

[38] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. `doi:10.1038/s41592-019-0686-2`.

[39] W. W. Ho and T. H. Hsieh. Efficient variational simulation of non-trivial quantum states. *SciPost Phys.*, 6:029, 2019. URL: `https://scipost.org/10.21468/SciPostPhys.6.3.029`, `doi:10.21468/SciPostPhys.6.3.029`.

[40] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney,

A. Dunsworth, E. Farhi, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, K. Guerin, S. Habegger, M. P. Harrigan, M. J. Hartmann, A. Ho, M. Hoffmann, T. Huang, T. S. Humble, S. V. Isakov, E. Jeffrey, Z. Jiang, D. Kafri, K. Kechedzhi, J. Kelly, P. V. Klimov, S. Knysh, A. Korotkov, F. Kostritsa, D. Landhuis, M. Lindmark, E. Lucero, D. Lyakh, S. Mandrà, J. R. McClean, M. McEwen, A. Megrant, X. Mi, K. Michielsen, M. Mohseni, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Y. Niu, E. Ostby, A. Petukhov, J. C. Platt, C. Quintana, E. G. Rieffel, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, K. J. Sung, M. D. Trevithick, A. Vainsencher, B. Villalonga, T. White, Z. J. Yao, P. Yeh, A. Zalcman, H. Neven, and J. M. Martinis. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, Oct 2019. `doi:10.1038/s41586-019-1666-5`.

[41] Y. Wang, B. Qi, C. Ferrie, and D. Dong. Trainability enhancement of parameterized quantum circuits via reduced-domain parameter initialization. *Phys. Rev. Appl.*, 22:054005, Nov 2024. URL: `https://link.aps.org/doi/10.1103/PhysRevApplied.22.054005`, `doi:10.1103/PhysRevApplied.22.054005`.

[42] A. Khan, B. K. Clark, and N. M. Tubman. Pre-optimizing variational quantum eigensolvers with tensor networks, 2023. URL: `https://arxiv.org/abs/2310.12965`, `arXiv:2310.12965`.

[43] H. Yu, Y. Zhao, and T.-C. Wei. Simulating large-size quantum spin chains on cloud-based superconducting quantum computers. *Phys. Rev. Res.*, 5:013183, Mar 2023. URL: `https://link.aps.org/doi/10.1103/PhysRevResearch.5.013183`, `doi:10.1103/PhysRevResearch.5.013183`.

[44] E. Rosenberg, P. Ginsparg, and P. L. McMahon. Experimental error mitigation using linear rescaling for variational quantum eigensolving with up to 20 qubits. *Quantum Science and Technology*, 7(1):015024, jan 2022. URL: `https://dx.doi.org/10.1088/2058-9565/ac3b37`, `doi:10.1088/2058-9565/ac3b37`.

[45] 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. *Phys. Rev. Lett.*, 127:270502, Dec 2021. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.127.270502`, `doi:10.1103/PhysRevLett.127.270502`.

[46] A. Javadi-Abhari, M. Treinish, K. Krsulich, C. J. Wood, J. Lishman, J. Gacon, S. Martiel, P. D. Nation, L. S. Bishop, A. W. Cross, B. R. Johnson, and J. M. Gambetta. Quantum computing with Qiskit, 2024. `arXiv:2405.08810`, `doi:10.48550/arXiv.2405.08810`.

[47] C. Piveteau and D. Sutter. Circuit knitting with classical communication. *IEEE Transactions on Information Theory*, 70(4):2734–2745, 2024. `doi:10.1109/TIT.2023.3310797`.

[48] G. Gentinetta, F. Metz, and G. Carleo. Overhead-constrained circuit knitting for variational quantum dynamics. *Quantum*, 8:1296, March 2024. `doi:10.22331/q-2024-03-21-1296`.

[49] S. Cantori and S. Pilati. Challenges and opportunities in the supervised learning of quantum circuit outputs, 2024. URL: `https://arxiv.org/abs/2402.04992`, `arXiv:2402.04992`.

[50] A. Mari, T. R. Bromley, J. Izaac, M. Schuld, and N. Killoran. Transfer learning in hybrid classical-quantum neural networks. *Quantum*, 4:340, October 2020. `doi:10.22331/q-2020-10-09-340`.

[51] R. Zen, L. My, R. Tan, F. Hébert, M. Gattobigio, C. Miniatura, D. Poletti, and S. Bressan. Transfer learning for scalability of neural-network quantum states. *Physical Review E*, 101(5), May 2020. URL: `http://dx.doi.org/10.1103/PhysRevE.101.053301`, `doi:10.1103/physreve.101.053301`.

[52] S. Cantori, D. Vitali, and S. Pilati. Supervised learning of random quantum circuits via scalable neural networks. *Quantum Science and Technology*, 8(2):025022, mar 2023. URL: `https://dx.doi.org/10.1088/2058-9565/acc4e2`, `doi:10.1088/2058-9565/acc4e2`.

[53] N. Saraceni, S. Cantori, and S. Pilati. Scalable neural networks for the efficient learning of disordered quantum systems. *Phys. Rev. E*, 102:033301, Sep 2020. URL: `https://link.aps.org/doi/10.1103/PhysRevE.102.033301`, `doi:10.1103/PhysRevE.102.033301`.

[54] Q. A. team and collaborators. qsim, September 2020. `doi:10.5281/zenodo.4023103`.

[55] R. LaRose, A. Mari, S. Kaiser, P. J. Karalekas, A. A. Alves, P. Czarnik, M. E. Mandouh, M. H.

Gordon, Y. Hindy, A. Robertson, P. Thakre, M. Wahl, D. Samuel, R. Mistri, M. Tremblay, N. Gardner, N. T. Stemen, N. Shammah, and W. J. Zeng. Mitiq: A software package for error mitigation on noisy quantum computers. *Quantum*, 6:774, Aug 2022. `doi:10.22331/q-2022-08-11-774`.

[56] S. Pilati and P. Pieri. Supervised machine learning of ultracold atoms with speckle disorder. *Scientific Reports*, 9(1), April 2019. URL: `http://dx.doi.org/10.1038/s41598-019-42125-w`, `doi:10.1038/s41598-019-42125-w`.

[57] D. Rolnick, A. Veit, S. Belongie, and N. Shavit. Deep learning is robust to massive label noise, 2018. URL: `https://arxiv.org/abs/1705.10694`, `arXiv:1705.10694`.