



Synergy between noisy quantum computers and scalable classical deep learning for quantum error mitigation

Simone Cantori^{1,2*}, Andrea Mari¹, David Vitali^{1,2} and Sebastiano Pilati^{1,2}

*Correspondence:

simone.cantori@unicam.it

¹School of Science and Technology,
Physics Division, Università di
Camerino, Via Madonna delle
Carceri, Camerino (MC), 62032, Italy

²Sezione di Perugia, INFN, Perugia,
06123, Italy

Abstract

We investigate the potential of combining the computational power of noisy quantum computers and of classical scalable convolutional neural networks (CNNs). The goal is to accurately predict exact expectation values of parameterized quantum circuits representing the Trotter-decomposed dynamics of quantum Ising models. By incorporating (simulated) noisy expectation values alongside circuit structure information, our CNNs effectively capture the underlying relationships between circuit architecture and output behaviour, enabling, via transfer learning, also predictions for circuits with more qubits than those included in the training set. Notably, thanks to the quantum information, our CNNs succeed even when supervised learning based only on classical descriptors fails. Furthermore, they outperform a popular error mitigation scheme, namely, zero-noise extrapolation, demonstrating that the synergy between quantum and classical computational tools leads to higher accuracy compared with quantum-only or classical-only approaches. By tuning the noise strength, we explore the crossover from a computationally powerful classical CNN assisted by quantum noisy data, towards rather precise quantum computations, further error-mitigated via classical deep learning.

Keywords: Quantum computing; Quantum circuits; Supervised learning; Deep neural networks; Quantum error mitigation; Transfer learning

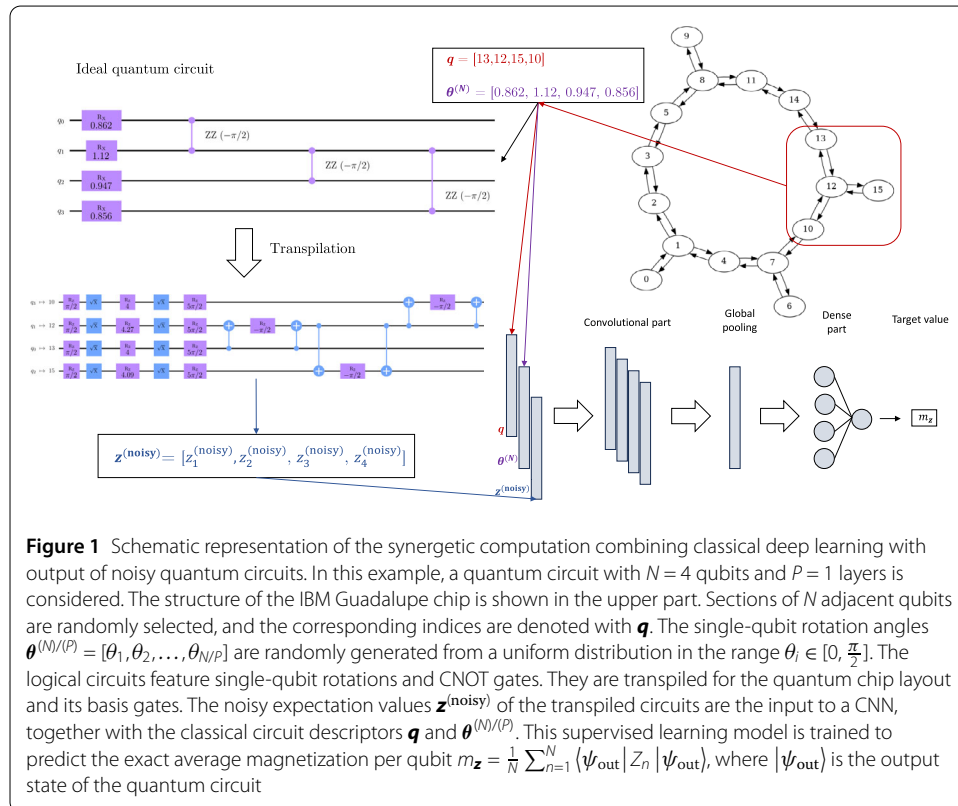
1 Introduction

Quantum computers promise to solve computational problems that are intractable on classical machines [1, 2]. However, efforts to exploit the full power of quantum computing are currently limited by hardware errors. To address this issue, quantum error mitigation techniques have been developed to minimize noise and obtain potentially useful results [3–8]. While error mitigation methods reduce noise in expectation values of observables, they may display limited accuracy or suffer from prohibitive sampling overheads [9–11]. In this scenario, classical machine learning emerges as a suitable tool for post-processing noisy quantum measurements, achieving accurate expectation values at a potentially lower computational cost [12, 13]. In fact, supervised machine learning has been successfully applied to various challenging computational tasks within quantum many-body physics [14–18] and quantum computing [12, 19–27]. Moreover, scalable supervised

© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

learning models allow generalizing beyond the size of the training quantum systems, potentially reaching system sizes out of reach for direct classical simulations [28–32]. On the other hand, classically supervised learning was shown to fail in emulating certain relevant quantum circuits [25], e.g., circuits featuring random inter-layer variations [26].

In this work, we investigate the computational synergy between noisy quantum computers and classical deep learning. Specifically, our focus is on the task of predicting expectation values of large quantum circuits representing the Trotter-decomposed dynamics of an Ising Hamiltonian [4, 12, 33]. These circuits are simulated taking into account the connectivity of an actual quantum chip and considering a realistic model of hardware errors. Our approach involves incorporating noisy quantum expectation values alongside information about the circuit architecture, to be used as input features for classical neural networks. A schematic representation is shown in Fig. 1. Leveraging scalable network generalization, our method shows remarkable performance in emulating quantum circuits with more qubits than those included in the training set. Extrapolation to deeper circuits is also possible, depending on the noise level. In this way, our approach also performs accurate quantum error mitigation, but circumventing the need of explicit target values for large circuits. Thus, it departs from the requirement of error-mitigated expectation values as training data [12]. On the other hand, our investigation improves upon the practice of relying only on circuit-structure information for predicting expectation values [24–26, 34]. Notably, this allows us to emulate circuits that are otherwise intractable for purely classical supervised learning. This approach underlines the potential of combining the outputs provided by quantum computers and classical deep learning methods. The synergy between these two strategies promises results that surpass the individual capabilities of each.



The rest of the article is organized as follows: in Sect. 2 we describe the quantum circuits we address and the structure of the quantum chip on which they can be implemented. We also introduce the error model used to simulate the noisy expectation values, as well as the technique we implement to tune the noise level. The CNNs and the training protocol are described in the final part of the section. The scalability of the CNNs on larger quantum circuits is analysed in Sect. 3. Here, we compare the accuracy of the predictions for different quantum circuit configurations, different numbers of qubits, and different levels of noise. Notably, comparison is made also against a prominent error-mitigation technique, namely zero-noise extrapolation (ZNE) [6, 35, 36]. In Sect. 4 we report our conclusions. Further details on how we tune the noise model and how we implement ZNE are available in Appendix A and Appendix B, respectively.

2 Methods

2.1 Quantum circuits and qubit arrangement

We consider quantum circuits composed of N qubits and P layers of gates. In each layer, a parameterized single-qubit gate R_X is applied to each qubit, and two-qubit gates R_{ZZ} are applied to chosen qubit pairs. The matrix representations of these gates are:

$$R_X(\theta) = \begin{bmatrix} \cos(\frac{\theta}{2}) & -i \sin(\frac{\theta}{2}) \\ -i \sin(\frac{\theta}{2}) & \cos(\frac{\theta}{2}) \end{bmatrix}, \quad R_{ZZ}(\phi) = \begin{bmatrix} e^{-i\frac{\phi}{2}} & 0 & 0 & 0 \\ 0 & e^{i\frac{\phi}{2}} & 0 & 0 \\ 0 & 0 & e^{i\frac{\phi}{2}} & 0 \\ 0 & 0 & 0 & e^{-i\frac{\phi}{2}} \end{bmatrix}. \quad (1)$$

This type of quantum circuit can be used to simulate the time dynamics of a many-body quantum system described by the transverse-field Ising Hamiltonian, which is defined as:

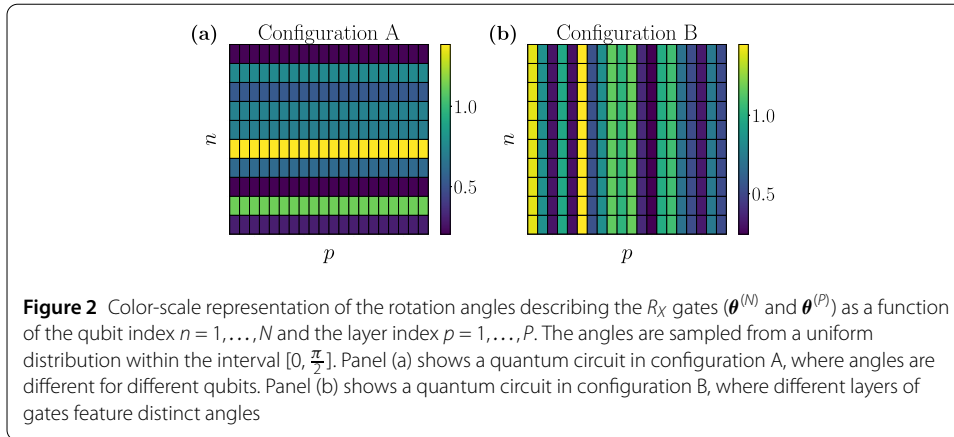
$$H(t) = H_{ZZ} + H_X = -J \sum_{\langle ij \rangle} Z_i Z_j + \sum_i h_i(t) X_i, \quad (2)$$

where X_i and Z_i are Pauli operators, J is the coupling between nearest-neighbour spins on the chosen graph, and $h_i(t)$ is the time-dependent transverse field acting on qubit i . Indeed, from the first-order Trotter decomposition of the time-evolution operator, we get

$$e^{-iH_{ZZ}\delta t} = \prod_{\langle ij \rangle} e^{iJ\delta t Z_i Z_j} = \prod_{\langle ij \rangle} R_{Z_i Z_j}(-2J\delta t) \quad (3)$$

$$e^{-iH_X\delta t} = \prod_i e^{-ih(t)\delta t X_i} = \prod_i R_{X_i}(2h(t)\delta t), \quad (4)$$

where the total evolution time T is discretized into $\frac{T}{\delta t}$ Trotter steps, $-2J\delta t = \phi$, and $2h(t)\delta t = \theta$. We set $\phi = -\frac{\pi}{2}$, following the approach of Ref. [4], despite employing a different circuit transpilation method. The angles θ for the R_X gates are randomly selected from a uniform distribution within the interval $[0, \frac{\pi}{2}]$. As shown in Fig. 2, we consider two distinct circuit configurations: A and B. In configuration A, the angles are randomly assigned to each qubit, but the same angle set is used across the P layers of gates. Instead, in configuration B the single-qubit gates feature different angles for different layers, but the angles are consistent across qubits within a specific layer.



The qubit pairs connected by the R_{ZZ} gates consist exclusively of the nearest neighbours on the graph of the IBM Guadalupe chip. This is illustrated in Fig. 1. Specifically, different portions of the chip are considered in different random realizations of the parametrized circuit. We consider all the possible connections of the quantum chip except the one between the physical qubit 4 and the physical qubit 1 (open boundary conditions). Therefore, each realization is uniquely determined by two arrays. The first, indicated with \mathbf{q} , includes the indices labelling the physical qubits selected in the considered circuit realization (see Fig. 1). This information is important for identifying the connections among qubits. The second array is the set of angles $\theta^{(N)} = \{\theta_1, \theta_2, \dots, \theta_N\}$ for configuration A, or $\theta^{(P)} = \{\theta_1, \theta_2, \dots, \theta_P\}$ for configuration B. To accurately model the noise characteristics of the IBM Guadalupe chip, we need to transpile our ideal quantum circuit into a form that can be executed on the quantum device, using the available gate set. This process is performed by Qiskit and is visualized in Fig. 1. While the arrays θ and \mathbf{q} uniquely identify each circuit realization and, hence, are suitable to perform purely classical supervised learning, we augment the circuit description with the set of noisy expectation values that would be produced by noisy quantum circuits, as discussed hereafter.

2.2 Noisy expectation values

The target value our CNNs shall predict is the average magnetization per qubit:

$$m_{\mathbf{z}} = \frac{1}{N} \sum_{n=1}^N z_n \equiv \frac{1}{N} \sum_{n=1}^N \langle \psi_{\text{out}} | Z_n | \psi_{\text{out}} \rangle ; \quad (5)$$

$|\psi_{\text{out}}\rangle$ is the output state of the quantum circuit after the application of P layers of gates on the input state $|\psi_{\text{in}}\rangle = |0\rangle^{\otimes N}$. For each circuit, the target value is exactly determined via state-vector simulations, which provide numerically exact expectation values of ideal, error-free, circuits. We also numerically emulate the execution of a noisy quantum computer. For this, we adopt the noise model encoded in the virtual backend *FakeGuadalupe* available in the Qiskit library [37]. This model replicates the noise characteristics of the original IBM Guadalupe quantum chip. In this case, the expectation values are averaged over a finite number of shots, namely, 10^4 . This number is large enough to suppress the effect of shot noise for the considered circuit sizes. This choice is motivated by our goal of addressing the effect of hardware errors only. The noisy quantities corresponding to

the exact single-qubit expectation values z_n , for $n = 1, \dots, N$, will be collectively denoted as $\mathbf{z}^{(\text{noisy})} = \{z_1^{(\text{noisy})}, z_2^{(\text{noisy})}, \dots, z_N^{(\text{noisy})}\}$. These noisy expectation values might help the network to predict the corresponding ground-truth results. Hence, we provide them as a further input to the CNNs, in addition to the classical circuit descriptors θ and \mathbf{q} . This combination of (here, simulated) quantum data and classical circuit-features allows overcoming previous approaches that either used classical descriptors only, or error-mitigated same-size circuit outputs, without exploiting scalable classical networks. Clearly, with our approach we aim to obtain predictions that at least outperform the accuracy of the trivial estimation:

$$m_{\mathbf{z}^{(\text{noisy})}} = \frac{1}{N} \sum_{n=1}^N z_n^{(\text{noisy})}. \quad (6)$$

In the following, it will be useful to tune the amount of noise in the circuit outputs. Specifically, we choose to focus on the errors associated with the CNOT gates, which are dominant compared to other errors, e.g. those associated with the single-qubit rotations or with readout operations. Our procedure to tune the noise level is described in Appendix A. In short, we introduce the parameter p_{noise} , with $1 \geq p_{\text{noise}} \geq 0$, which determines the noise strength associated to the CNOT gates. The value $p_{\text{noise}} = 1$ corresponds to the standard noise model of the quantum chip, while $p_{\text{noise}} = 0$ indicates the total cancellation of the noise related to the CNOT gates. Notice that, while somewhat less effective, other errors such as the ones on the other gates and readout errors are still allowed.

2.3 Convolutional neural networks

As discussed in the previous sections, we train deep CNNs to predict expectation values $m_{\mathbf{z}}$ of different quantum circuits. For quantum circuits in configuration A, the network input is one dimensional and it features three channels, resulting in the input shape $(N, 3)$. The first channel includes the qubit indices \mathbf{q} ,¹ the second one includes the angles $\theta^{(N)}$, while the third channel includes the noisy expectation values $\mathbf{z}^{(\text{noisy})}$. These three channels allow the CNNs combining classical circuit descriptors with noisy quantum data. For circuits in configuration B, we implement a two-dimensional CNN with input shape $(N, P, 3)$. To fit this shape, the length- P array $\theta^{(P)}$ is repeated N times. Both \mathbf{q} and $\mathbf{z}^{(\text{noisy})}$ are repeated P times for the same reason. We compare the performance of these CNNs with analogous networks that process only the classical circuit descriptors, namely $\theta^{(N)/(P)}$ (for configuration A/B) and \mathbf{q} . In these cases, the networks have two input channels. To distinguish the above models, we respectively indicate the network with hybrid classical-quantum inputs with $\text{CNN}(\theta^{(N)/(P)}, \mathbf{q}, \mathbf{z}^{(\text{noisy})})$, and the one with only classical descriptors with $\text{CNN}(\theta^{(N)/(P)}, \mathbf{q})$.

Our final goal is to predict expectation values of quantum circuits larger than those included in the training set. To adapt the network to the different circuit sizes, a scalable architecture is crucial. Conventional CNNs featuring convolutional filters followed by dense layers are not entirely scalable. Indeed, while convolutional layers can handle variable-sized inputs, dense layers necessitate a fixed input size. To overcome this constraint, we incorporate a global pooling operation after the last convolutional layer, emu-

¹The actual descriptors are normalized as $\mathbf{q}' = \mathbf{q}/10$, so that values in different channels are of the same order of magnitude. With more qubits, a higher normalization factor might be appropriate.

lating the strategy employed in Refs. [24, 29]. This enhancement transforms the architecture into a fully scalable framework. Moreover, consistently training the neural network on a fixed set of physical qubits poses a challenge. Indeed, when tested on larger circuits, the network would encounter configurations involving connections among qubits that were not part of its training data, making scalability impractical. To address this limitation, the CNN is trained on circuits implemented on randomly-selected consecutive portions of the Guadalupe chip, as illustrated in Fig. 1. In other words, the CNN is trained with varying combinations of \mathbf{q} .

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

$$\mathcal{L} = \frac{1}{K_{\text{train}}} \sum_{k=1}^{K_{\text{train}}} (y_k - \tilde{y}_k)^2, \quad (7)$$

where K_{train} is the number of instances included in the training set, $y_k = m_{\mathbf{z},k}$ is the target value, and \tilde{y}_k is the corresponding predicted value. The network parameters are optimized via a widely used form of stochastic gradient descent, namely, the ADAM algorithm [38].

To assess the prediction accuracy, we evaluate the coefficient of determination

$$R^2 = 1 - \frac{\sum_{k=1}^{K_{\text{test}}} (y_k - \tilde{y}_k)^2}{\sum_{k=1}^{K_{\text{test}}} (y_k - \bar{y})^2}, \quad (8)$$

where \bar{y} is the average of the target values and K_{test} is the number of instances in the test set. The metric R^2 quantifies how accurately the variations of the target values are predicted by the regression model. Notice that a constant model with the correct average corresponds to the score $R^2 = 0$, and that in fact R^2 can be negative. Another useful metric is the difference $1 - R^2$. It coincides with the ratio of the mean squared error over the data variance, thus representing a normalized error measure. In the following, it will be useful to estimate the correlation between noisy expectation values and the exact ones. For this, we determine the Pearson correlation coefficient:

$$\rho = \frac{\sum_i (m_{\mathbf{z}}^{(i)} - \overline{m_{\mathbf{z}}}) (m_{\mathbf{z}(\text{noisy})}^{(i)} - \overline{m_{\mathbf{z}(\text{noisy})}})}{\sqrt{\sum_i (m_{\mathbf{z}}^{(i)} - \overline{m_{\mathbf{z}}})^2 \sum_i (m_{\mathbf{z}(\text{noisy})}^{(i)} - \overline{m_{\mathbf{z}(\text{noisy})}})^2}}. \quad (9)$$

In Eq. (9), $\overline{m_{\mathbf{z}}}$ and $\overline{m_{\mathbf{z}(\text{noisy})}}$ represent the average of $m_{\mathbf{z}}$ and $m_{\mathbf{z}(\text{noisy})}$ across the selected sample of quantum circuits.

3 Results and discussion

3.1 Quantum circuits in configuration A

The first test we discuss is on quantum circuits of depth $P = 20$ in configuration A. In this scenario, the CNN is trained using quantum circuits with $N \in \{6, \dots, 10\}$ qubits. Next, the network is tested on quantum circuits featuring up to $N = 16$ qubits. Figure 3 shows the prediction accuracy as a function of the number of qubits in the test circuits. Here and for the remaining results, the error bars represent the estimated standard deviation of the average over three repetitions of the training process. We observe that the network which processes only classical circuit descriptors, namely, $\text{CNN}(\theta^{(N)}, \mathbf{q})$, achieves satisfactory

Figure 3 Prediction error $1 - R^2$ as a function of the number of qubits N of the quantum circuits in the test set. We compare the network processing only classical descriptors, namely, $\text{CNN}(\theta^{(N)}, \mathbf{q})$, the one processing also (simulated) noisy quantum outputs $\text{CNN}(\theta^{(N)}, \mathbf{q}, \mathbf{z}^{(\text{noisy})})$, as well as the noisy expectation values $m_{\mathbf{z}^{(\text{noisy})}}$ and the corresponding results after zero-noise extrapolation (ZNE). For the latter two, the coefficient of determination R^2 w.r.t. the exact expectation values $m_{\mathbf{z}}$ is computed via eq. (8). The CNNs are trained on $K_{\text{train}} \simeq 6 \times 10^5$ quantum circuits with $N \leq 10$ qubits (see vertical dotted line). The depth of the quantum circuits is $P = 20$

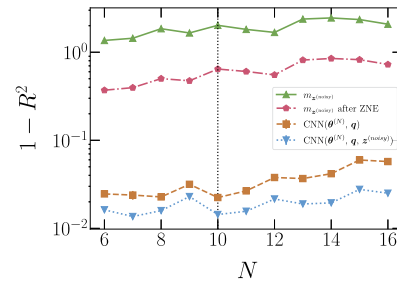
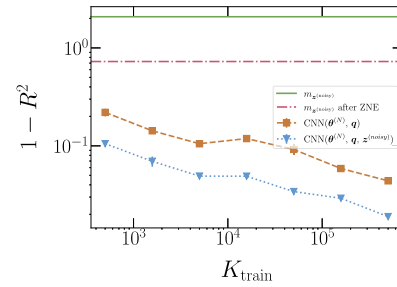


Figure 4 Prediction error $1 - R^2$ as a function of the number of instances in the training set K_{train} . The CNNs are trained on quantum circuits in configuration A, $N \leq 10$ qubits and $P = 20$ layers of gates. The test is performed on quantum circuits with $N = 16$ qubits. The different datasets are defined as in Fig. 3



accuracies. Analogous findings have been previously reported in Ref. [26] for a similar circuit structure. However, the hybrid network $\text{CNN}(\theta^{(N)/(P)}, \mathbf{q}, \mathbf{z}^{(\text{noisy})})$, which processes also the noisy quantum expectation values $\mathbf{z}^{(\text{noisy})}$, consistently reaches superior performances. Importantly, we observe that both CNNs outperform the output of the simulated quantum computer, even when the noise is mitigated through ZNE. In Fig. 4, we show the performance of the CNNs, tested on the qubit number $N = 16$, as a function of the number of instances in the training set K_{train} . Notably, the accuracy of the CNNs are better than the ones obtained with the simulated quantum chip even for training sets as small as $K_{\text{train}} \simeq 500$.

It is worth emphasizing that, in the approach envisioned here, there is no sampling overhead during the prediction phase. In other words, once the network has been trained, for each testing circuit we use the same number of measurements (and even the same noisy results) that are required for the trivial direct estimation of the average magnetization. Moreover, apart from the negligible classical computing cost of computing the output of the CNN, no classical simulation of test circuits is required. Furthermore, during the training phase, only small-scale circuits must be classically simulated, meaning that large-scale simulations at the size of the test circuits are never required.

3.2 Quantum circuits in configuration B

It was recently shown that classical neural networks trained via supervised learning fail to emulate quantum circuits featuring rapid random inter-layer angle fluctuations [26]. This failure is replicated here for quantum circuits in configuration B, as shown in Fig. 5. Indeed, the network $\text{CNN}(\theta, \mathbf{q})$, which processes only classical inputs, fails to reach reasonable accuracies $R^2 \simeq 1$, even with as many as $K_{\text{train}} \simeq 10^6$ training circuits (training sizes $N = 6, \dots, 10$, testing size $N = 16$). In this test, the advantage of including noisy expectation values $\mathbf{z}^{(\text{noisy})}$ is extreme. Indeed, we find that the network with hybrid inputs,

Figure 5 Prediction error $1 - R^2$ as a function of the number of instances in the training set K_{train} . The CNNs are trained on quantum circuits in configuration B, featuring $N \leq 10$ qubits and $P = 20$ layers of gates. The test is performed on quantum circuits with $N = 16$ qubits. The different datasets are defined as in Fig. 3

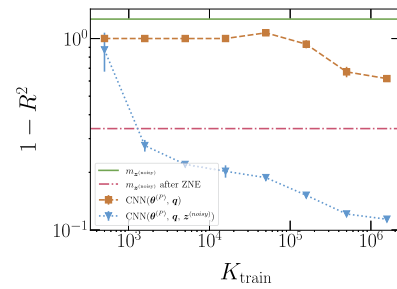
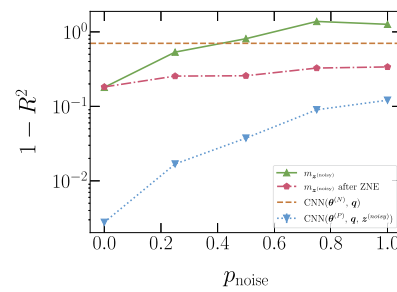
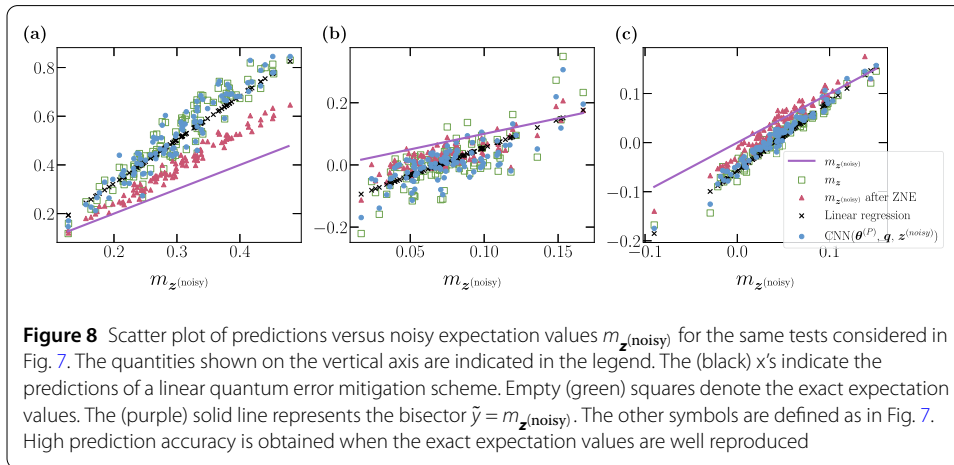
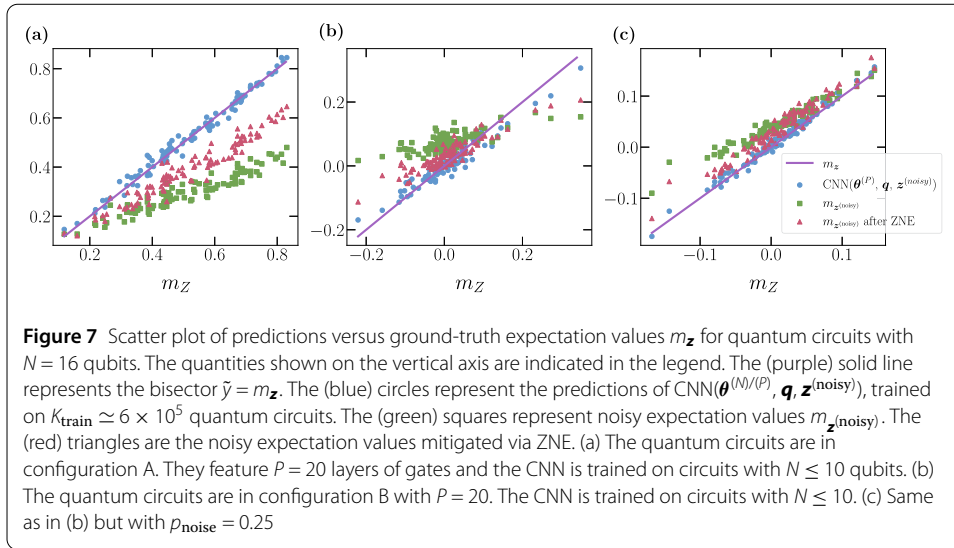


Figure 6 Prediction error $1 - R^2$ as a function of the level of noise p_{noise} associated with the CNOT gates. The CNNs are trained on $K_{\text{train}} \simeq 6 \times 10^5$ quantum circuits in configuration B, featuring $N \leq 10$ qubits and $P = 20$ layers of gates. The test is performed on quantum circuits with $N = 16$ qubits. The different datasets are defined as in Fig. 3



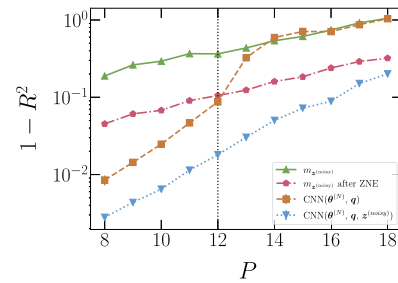
namely, $\text{CNN}(\theta^{(P)}, \mathbf{z}^{(\text{noisy})}, \mathbf{q})$, produces results with acceptable accuracies. In fact, it outperforms the accuracy of ZNE already with $K_{\text{train}} \gtrsim 10^3$ training circuits. Still, the accuracy is inferior to the one obtained for configuration A. This might be attributed to a lower correlation between the noisy expectation values $m_{\mathbf{z}}^{(\text{noisy})}$ and the ground-truth values $m_{\mathbf{z}}$. In fact, the corresponding Pearson correlation coefficient for quantum circuits in configuration A with, e.g., $N = 16$ and $P = 20$, is $\rho = 0.945$, while for quantum circuits of the same size in configuration B it is only $\rho = 0.664$. Hence, it is natural to ask if and how much the predictions of the CNN which processes also $\mathbf{z}^{(\text{noisy})}$, beyond the classical descriptors, improve when the quantum hardware is less affected by noise. We analyse this effect by reducing the amount of errors associated to the CNOT gates, as discussed in Sect. 2 (see Appendix A for further details). The prediction accuracy is shown in Fig. 6, as a function of the noise level p_{noise} . We reiterate that the errors beyond those associated to the CNOT gates are not tuned compared to the default *FakeGuadalupe* model. Interestingly, we find that even small improvements in the noisy quantum data lead to a substantial boost in the accuracy of $\text{CNN}(\theta, \mathbf{q}, \mathbf{z}^{(\text{noisy})})$. Chiefly, this model systematically outperforms the network with only classical inputs $\text{CNN}(\theta, \mathbf{q})$, as well as the ones corresponding to the noisy expectation values $\mathbf{z}^{(\text{noisy})}$, even when these are corrected via ZNE. It is worth mentioning that at $p_{\text{noise}} = 0$ ZNE does not affect the result. This is because certain types of noise, like readout errors, cannot be addressed using this error mitigation technique. To further visualize the comparison between the CNN predictions and the noisy outputs of the simulated quantum chip, in Fig. 7 we show scatter plots of the average magnetizations per qubit for a representative testing circuit size. For configuration A, one notices an appreciable correlation between noisy expectation values and ground-truth results. The correlation is less pronounced for configuration B. Furthermore, in the latter case the noisy expectation values are rather concentrated, and this contributes to the difficulty of the discriminative learning task. The behaviour of the hybrid network $\text{CNN}(\theta^{(N)/(P)}, \mathbf{q}, \mathbf{z}^{(\text{noisy})})$ can be further characterized by making comparison against a linear model. Specifically, we define a



simple error mitigation strategy based on the linear fit $m_{\mathbf{z}} = c_1 m_{\mathbf{z}}^{(\text{noisy})} + c_2$, where c_1 and c_2 are fitting parameters. Its performance is analyzed in Fig. 8, where the same test cases of Fig. 7 are considered, but the results are shown as a function of $m_{\mathbf{z}}^{(\text{noisy})}$. Notably, in the tests of panels (a) and (c) the linear model outperforms ZNE, indicating that training even simple models against exact expectation values leads to effective error mitigation schemes. However, the hybrid CNN always outperforms also the linear model, reaching the scores $R^2 \simeq 0.98$, $R^2 \simeq 0.91$, and $R^2 \simeq 0.98$ for the tests in panels (a), (b), and (c), respectively, while the corresponding scores of the linear model are: $R^2 \simeq 0.88$, $R^2 \simeq 0.54$, and $R^2 \simeq 0.94$. Indeed, Fig. 8 allows one to appreciate that the variations around the linear scaling are reproduced by the hybrid CNN with good accuracy. To facilitate replication of our findings and further investigations on the synergy between noisy quantum computers and classical deep learning, the descriptors and target values of the exemplary tests of Figs. 7 and 8 are made publicly available at the repository of Ref. [39]. The codes used to simulate the quantum circuits and to implement the neural networks are accessible through the same repository.

The last test we discuss is the extrapolation on the circuit depth. Specifically, we train the CNNs on relatively shallow circuits featuring $P \leq 12$ layers, and test them on circuits

Figure 9 Prediction error $1 - R^2$ as a function of the number of layers of gates P of the quantum circuits in the test set. The CNNs are trained on $K_{\text{train}} \simeq 6 \times 10^5$ quantum circuits in configuration B with $P \leq 12$ layers of gates (see vertical dotted line). The number of qubits of the quantum circuits is $N = 10$. The different datasets are defined as in Fig. 3



with equal and larger depths. In this test, $N = 10$ qubits in configuration B are considered. The results are shown in Fig. 9. One notices that the hybrid CNN is able to predict the output of deeper circuits, but the accuracy gradually diminishes as a function of P . This effect can be attributed to the increased effect of hardware errors for deeper circuits, which causes the noisy expectation values to become less informative. Anyway, it is worth pointing out that, while useful, the scalability with the circuit depth is not strictly necessary. In principle, the CNNs can be trained on computationally feasible circuits featuring fewer qubits, exploiting the (more stable) extrapolation on the qubit number to address computationally challenging circuits.

The above findings underscore the promising synergy between classical deep learning and quantum circuit outputs. Noisy expectation values offer valuable insights to the neural networks, enabling them to predict expectation values significantly more accurately, even in setups where supervised learning with only classical descriptors drastically fails. Meanwhile, employing CNNs to mitigate noisy expectation value errors yields superior accuracies compared to those achieved with simulated noisy quantum computers, even when using a prominent error mitigation technique such as ZNE.

It is useful to discuss our approach vis-à-vis the machine-learning technique for quantum error mitigation discussed in Ref. [12]. The significant distinction lies in the training method and in the scope of the network. In Ref. [12], the size of the training circuits is equal to the size of the test circuits and zero-noise extrapolated expectation values obtained from a quantum computer are used as training targets. In fact, the main goal of Ref. [12] is not outperforming the accuracy of ZNE, but rather reproducing equivalent results with a reduced sampling overhead. In contrast, our scalable architecture eliminates the need to train the neural network directly on large quantum circuits and, consequently, it can be trained with exact target values associated to small-scale circuits. Due to the different training method, our model can be used as a way of reducing the sampling overhead but also as a way of improving the estimation accuracy compared to standard error mitigation. Indeed, in our numerical simulations, we observe a better accuracy compared with ZNE, despite paying the same sampling cost of direct estimation.

4 Conclusions

In this work, we spotlighted the effectiveness of combining scalable classical neural networks with noisy quantum computers. We applied our approach to predict the output expectation values of quantum circuits describing the Trotter-decomposed dynamics of quantum Ising models, similarly to recent investigations on quantum utility experiments [4]. We considered the connectivity allowed by the Guadalupe IBM chip, account-

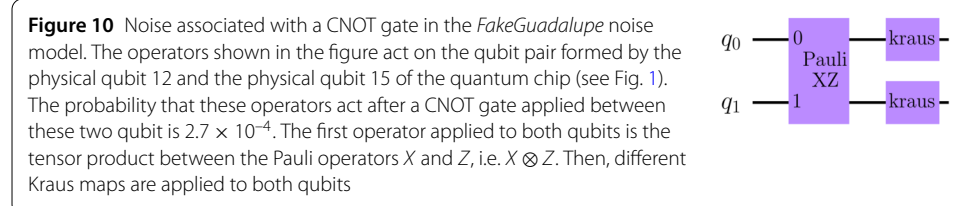
ing for hardware errors via the *FakeGuadalupe* noise model implemented in the Qiskit library.

In detail, the inputs of our CNNs include single-qubit noisy output expectation values, beyond the classical circuit descriptors – in this study, rotation angles and qubit indices – which were already considered in previous supervised learning studies. Training and testing circuits are implemented across various regions of the physical chip. This strategic arrangement enables the CNN to visualize and learn from all potential connections between physical qubits during the training process. Two circuit configurations were addressed, featuring either intra-layer or inter-layer random variations of the single-qubit rotation angles. The former angle configuration was already shown to be amenable to supervised learning [26]. Yet, here we found that the inclusion of noisy expectation values leads to systematically superior performances. In the second configuration the boost is extreme. Indeed, while supervised learning with only classical descriptors drastically fails, the combination with noisy quantum circuit outputs leads to accurate predictions. A modified error model was implemented to allow us to tune the noise, and we quantified how the synergetic predictions improve when the quantum expectation values become more precise.

Notably, the CNNs trained (also) on noisy expectation values produce results more efficiently and with greater accuracy than a prominent error mitigation method, namely, ZNE. Moreover, our approach is a viable alternative to the one presented in Ref. [12], which relies on noisy expectation values mitigated via ZNE as training target values. Transfer-learning from small-scale to large scale circuits is a key feature of our network, allowing the prediction of expectation values for larger circuits than those in the training set, without the requirement for target values at these larger sizes. Extrapolation on the circuit depth is also possible, but the prediction accuracy gradually diminishes, arguably due to the increased role of hardware errors in deeper circuits, which makes the noisy expectation values less informative. In general, our strategy enables the integration of the strengths of classical deep learning and of noisy quantum computers, potentially outperforming existing quantum error mitigation methods.

Appendix A: Manipulation of the noise strength associated with the CNOT gates

In the noise model of *FakeGuadalupe*, the CNOT noise is represented by a set of operators applied after each CNOT gate with varying probabilities. An example is depicted in Fig. 10. The specific operators and their associated probabilities depend on the considered pair of qubits. To tune the noise strength and the corresponding error, we introduce a circuit containing one identity gate per qubit into each circuit set, assigning it a probability



of $1 - p_{\text{noise}}$ to take effect. Therefore, the remaining circuits, representing CNOT errors, collectively carry a probability of p_{noise} to occur after the CNOT application.

Appendix B: Zero-noise extrapolation

To apply ZNE to the noisy expectation values, we utilize the Mitiq python library [40]. Specifically, we employ Richardson extrapolation with noise scale factors $\lambda = 1, 2, 3$. To manipulate the noise level, the unitary folding map $G \rightarrow GG^\dagger G$ is applied to all the gates of the investigated quantum circuits for $\lambda = 3$, and to a half of the gates (randomly selected) for $\lambda = 2$.

Acknowledgements

Not applicable.

Author contributions

All authors contributed to the conceptualization of the project. S.C. implemented the software and performed the calculations. All authors contributed to the data analysis and manuscript writing.

Funding

This work was supported by the PNRR MUR Project No. PE0000023-NQSTI and by the Italian Ministry of University and Research under the PRIN2022 project “Hybrid algorithms for quantum simulators” No. 2022H77XB7. S.P. acknowledges support from the CINECA award IsCb2_NEMCASRA and the CINECA-INFN agreement, for the availability of high-performance computing resources and support. S.P. also acknowledges the EuroHPC Joint Undertaking for awarding access to the EuroHPC supercomputer LUMI, hosted by CSC (Finland) and the LUMI consortium through a EuroHPC Benchmark Access call.

Data Availability

Some benchmark data and the codes used to simulate the quantum circuits and to implement the neural networks are made available through a Zenodo repository. All other data are available from the authors upon reasonable request.

Declarations

Competing interests

The authors declare no competing interests.

Received: 16 April 2024 Accepted: 3 July 2024 Published online: 15 July 2024

References

1. Shor PW. Algorithms for quantum computation: discrete logarithms and factoring. In: Proceedings 35th annual symposium on foundations of computer science. 1994. p. 124–34. <https://doi.org/10.1109/SFCS.1994.365700>.
2. Daley AJ, Bloch I, Kokail C, Flannigan S, Pearson N, Troyer M, Zoller P. Practical quantum advantage in quantum simulation. *Nature*. 2022;607(7920):667–76. <https://doi.org/10.1038/s41586-022-04940-6>.
3. Cai Z, Babbush R, Benjamin SC, Endo S, Huggins WJ, Li Y, McClean JR, O'Brien TE. Quantum error mitigation. *Rev Mod Phys*. 2023;95:045005. <https://doi.org/10.1103/RevModPhys.95.045005>.
4. Kim Y, Eddins A, Anand S, Wei KX, Berg E, Rosenblatt S, Nayfeh H, Wu Y, Zaletel M, Temme K, Kandala A. Evidence for the utility of quantum computing before fault tolerance. *Nature*. 2023;618(7965):500–5. <https://doi.org/10.1038/s41586-023-06096-3>.
5. Kim Y, Wood CJ, Yoder TJ, Merkel ST, Gambetta JM, Temme K, Kandala A. Scalable error mitigation for noisy quantum circuits produces competitive expectation values. *Nat Phys*. 2023;19(5):752–9. <https://doi.org/10.1038/s41567-022-01914-3>.
6. Temme K, Bravyi S, Gambetta JM. Error mitigation for short-depth quantum circuits. *Phys Rev Lett*. 2017;119:180509. <https://doi.org/10.1103/PhysRevLett.119.180509>.
7. Berg E, Mineev ZK, Kandala A, Temme K. Probabilistic error cancellation with sparse Pauli–Lindblad models on noisy quantum processors. *Nat Phys*. 2023;19(8):1116–21. <https://doi.org/10.1038/s41567-023-02042-2>.
8. Strikis A, Qin D, Chen Y, Benjamin SC, Li Y. Learning-based quantum error mitigation. *PRX Quantum*. 2021;2:040330. <https://doi.org/10.1103/PRXQuantum.2.040330>.
9. Takagi R, Tajima H, Gu M. Universal sampling lower bounds for quantum error mitigation. *Phys Rev Lett*. 2023;131:210602. <https://doi.org/10.1103/PhysRevLett.131.210602>.
10. Quek Y, França DS, Khatri S, Meyer JJ, Eisert J. Exponentially tighter bounds on limitations of quantum error mitigation. 2024. [arXiv:2210.11505](https://arxiv.org/abs/2210.11505) [quant-ph].
11. Tsubouchi K, Sagawa T, Yoshioka N. Universal cost bound of quantum error mitigation based on quantum estimation theory. *Phys Rev Lett*. 2023;131:210601. <https://doi.org/10.1103/PhysRevLett.131.210601>.
12. Liao H, Wang DS, Sitdikov I, Salcedo C, Seif A, Mineev ZK. Machine learning for practical quantum error mitigation. 2023. [arXiv:2309.17368](https://arxiv.org/abs/2309.17368) [quant-ph].
13. Sack SH, Egger DJ. Large-scale quantum approximate optimization on nonplanar graphs with machine learning noise mitigation. *Phys Rev Res*. 2024;6:013223. <https://doi.org/10.1103/PhysRevResearch.6.013223>.

14. Huang H-Y, Kueng R, Torlai G, Albert VV, Preskill J. Provably efficient machine learning for quantum many-body problems. *Science*. 2022;377:6613. <https://doi.org/10.1126/science.abk3333>.
15. Carleo G, Cirac I, Cranmer K, Daudet L, Schuld M, Tishby N, Vogt-Maranto L, Zdeborová L. Machine learning and the physical sciences. *Rev Mod Phys*. 2019;91:045002. <https://doi.org/10.1103/RevModPhys.91.045002>.
16. Schütt KT, Chmiela S, Von Lilienfeld OA, Tkatchenko A, Tsuda K, Müller K-R. Machine learning meets quantum physics. *Lect. Notes Phys*. 2020. <https://doi.org/10.1007/978-3-030-40245-7>.
17. Kulik HJ, Hammerschmidt T, Schmidt J, Botti S, Marques MAL, Boley M, Scheffler M, Todorović M, Rinke P, Oses C, Smolyanyuk A, Curtarolo S, Tkatchenko A, Bartók AP, Manzhos S, Ihara M, Carrington T, Behler J, Isayev O, Veit M, Grisafi A, Nigam J, Ceriotti M, Schütt KT, Westermayr J, Gastegger M, Maurer RJ, Kalita B, Burke K, Nagai R, Akashi R, Sugino O, Hermann J, Noé F, Pilati S, Draxl C, Kuban M, Rigamonti S, Scheidgen M, Esters M, Hicks D, Toher C, Balachandran PV, Tamblyn I, Whitelam S, Bellinger C, Ghiringhelli LM. Roadmap on machine learning in electronic structure. *Electron Struct*. 2022;4(2):023004. <https://doi.org/10.1088/2516-1075/ac572f>.
18. Carrasquilla J, Torlai G. How to use neural networks to investigate quantum many-body physics. *PRX Quantum*. 2021;2:040201. <https://doi.org/10.1103/PRXQuantum.2.040201>.
19. Baireuther P, Caio MD, Criger B, Beenakker CWJ, O'Brien TE. Neural network decoder for topological color codes with circuit level noise. *New J Phys*. 2019;21(1):013003. <https://doi.org/10.1088/1367-2630/aaf29e>.
20. Baireuther P, O'Brien TE, Tarasinski B, Beenakker CWJ. Machine-learning-assisted correction of correlated qubit errors in a topological code. *Quantum*. 2018;2:48. <https://doi.org/10.22331/q-2018-01-29-48>.
21. Chamberland C, Ronagh P. Deep neural decoders for near term fault-tolerant experiments. *Quantum Sci Technol*. 2018;3(4):044002. <https://doi.org/10.1088/2058-9565/aad1f7>.
22. Torlai G, Mazzola G, Carrasquilla J, Troyer M, Melko R, Carleo G. Neural-network quantum state tomography. *Nat Phys*. 2018;14(5):447–50. <https://doi.org/10.1038/s41567-018-0048-5>.
23. Zlokapa A, Gheorghiu A. A deep learning model for noise prediction on near-term quantum devices. 2020. [arXiv:2005.10811](https://arxiv.org/abs/2005.10811).
24. Cantori S, Vitali D, Pilati S. Supervised learning of random quantum circuits via scalable neural networks. *Quantum Sci Technol*. 2023;8(2):025022. <https://doi.org/10.1088/2058-9565/acc4e2>.
25. Mohseni N, Shi J, Byrnes T, Hartmann M. Deep learning of many-body observables and quantum information scrambling. 2023. [arXiv:2302.04621](https://arxiv.org/abs/2302.04621) [quant-ph].
26. Cantori S, Pilati S. Challenges and opportunities in the supervised learning of quantum circuit outputs. 2024. [arXiv:2402.04992](https://arxiv.org/abs/2402.04992) [cond-mat.dis-nn].
27. Melko RG, Carrasquilla J. Language models for quantum simulation. *Nat Comput Sci*. 2024;4:11–8. <https://doi.org/10.1038/s43588-023-00578-0>.
28. Mills K, Ryczko K, Luchak I, Domurad A, Beeler C, Tamblyn I. Extensive deep neural networks for transferring small scale learning to large scale systems. *Chem Sci*. 2019;10:4129–40. <https://doi.org/10.1039/C8SC04578J>.
29. Saraceni N, Cantori S, Pilati S. Scalable neural networks for the efficient learning of disordered quantum systems. *Phys Rev E*. 2020;102:033301. <https://doi.org/10.1103/PhysRevE.102.033301>.
30. Jung H, Stocker S, Kunkel C, Oberhofer H, Han B, Reuter K, Margraf JT. Size-extensive molecular machine learning with global representations. *ChemSystemsChem*. 2020;2(4):1900052. <https://doi.org/10.1002/syst.201900052>.
31. Mujal P, Miguel AM, Polls A, Juliá-Díaz B, Pilati S. Supervised learning of few dirty bosons with variable particle number. *SciPost Phys*. 2021;10:073. <https://doi.org/10.21468/SciPostPhys.10.3.073>.
32. Mohseni N, Navarrete-Benlloch C, Byrnes T, Marquardt F. Deep recurrent networks predicting the gap evolution in adiabatic quantum computing. *Quantum*. 2023;7:1039. <https://doi.org/10.22331/q-2023-06-12-1039>.
33. Narasimhan P, Humeniuk S, Roy A, Drouin-Touchette V. Simulating the transverse field ising model on the kagome lattice using a programmable quantum annealer. 2023. [arXiv:2310.06698](https://arxiv.org/abs/2310.06698) [cond-mat.stat-mech].
34. Zhang S-X, Hsieh C-Y, Zhang S, Yao H. Neural predictor based quantum architecture search. *Mach Learn: Sci Technol*. 2021;2(4):045027. <https://doi.org/10.1088/2632-2153/ac28dd>.
35. Kandala A, Temme K, Córcoles AD, Mezzacapo A, Chow JM, Gambetta JM. Error mitigation extends the computational reach of a noisy quantum processor. *Nature*. 2019;567(7749):491–5. <https://doi.org/10.1038/s41586-019-1040-7>.
36. Li Y, Benjamin SC. Efficient variational quantum simulator incorporating active error minimization. *Phys Rev X*. 2017;7:021050. <https://doi.org/10.1103/PhysRevX.7.021050>.
37. Qiskit contributors. Qiskit: an open-source framework for quantum computing. 2023. <https://doi.org/10.5281/zenodo.2573505>.
38. Kingma DP, Ba J. Adam: a method for stochastic optimization. 2017. [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) [cs.LG].
39. Cantori S, Mari A, Vitali D, Pilati S. Synergy between noisy quantum computers and scalable classical deep learning. 2024. <https://doi.org/10.5281/zenodo.12527150>.
40. LaRose R, Mari A, Kaiser S, Karalekas PJ, Alves AA, Czarnik P, Mandouh ME, Gordon MH, Hindy Y, Robertson A, Thakre P, Wahl M, Samuel D, Mistri R, Tremblay M, Gardner N, Stemen NT, Shammah N, Zeng WJ. Mitiq: a software package for error mitigation on noisy quantum computers. *Quantum*. 2022;6:774. <https://doi.org/10.22331/q-2022-08-11-774>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.