Article

Noise-agnostic quantum error mitigation with data augmented neural models

Check for updates

Manwen Liao[1,4], Yan Zhu[1,4], Giulio Chiribella ![ORCID][1,2,3] & Yuxiang Yang ![ORCID][1] ✉

Quantum error mitigation, a data processing technique for recovering the statistics of target processes from their noisy version, is a crucial task for near-term quantum technologies. Most existing methods require prior knowledge of the noise model or the noise parameters. Deep neural networks have the potential to lift this requirement, but current models require training data produced by ideal processes in the absence of noise. Here we build a neural model that achieves quantum error mitigation without any prior knowledge of the noise and without training on noise-free data. To achieve this feature, we introduce a quantum augmentation technique for error mitigation. Our approach applies to quantum circuits and to the dynamics of many-body and continuous-variable quantum systems, accommodating various types of noise models. We demonstrate its effectiveness by testing it both on simulated noisy circuits and on real quantum hardware.

Quantum technologies have potentially disruptive applications in the fields of computing, communication, and sensing. In the near term, however, the demonstration of practical quantum advantages remains challenging due to the presence of noise. A promising technique to restore quantum advantages in realistic noisy devices is quantum error mitigation[1], including zero-noise extrapolation (ZNE)[2–6], Clifford data regression (CDR)[7–10], probablistic error cancellation[2,11,12], and virtual purification[13–15].

A limitation of the existing error mitigation methods is that they generally require prior knowledge about the noise model, leading to an overhead in terms of noise characterization operations[16–18]. A promising approach to circumvent this issue is to exploit deep neural networks, which have been successfully applied to other quantum tasks such as quantum state characterization[19–23], quantum property estimation[24,25], quantum verification[26,27], and quantum simulations[28,29]. Previous research[30–32] has also explored the use of machine learning models to estimate the output fidelity of quantum circuits. Although these approaches do not eliminate noise in the circuit output, they hold promise in guiding the generation of quantum circuits with fewer errors. Recently, a series of works explored the application of deep neural networks directly to quantum error mitigation[33–36]. Nevertheless, training these networks generally requires access to noise-free data, which can be hard to obtain from experiments or from classical simulations.

Here we propose a neural model that achieves quantum error mitigation without prior knowledge of the noise and without any access to noise-free data. To achieve this feature, we introduce a technique, called quantum data augmentation, to expand the original data set by generating new data from a fiducial set of noisy processes. Our technique provides a quantum

version of classical data augmentation techniques[37,38], which have proven valuable in scenarios where the available training data is limited[39,40].

Our model exhibits four major features. (1) No need of noise-free statistics from the target quantum process. Thanks to this feature, our model is applicable to relevant real-world scenarios where the ideal target process is hard to simulate. (2) Noise-agnostic error mitigation. The model does not require prior knowledge about the noise model, nor about the values of the noise parameter. As a result, it avoids overheads due to noise characterization, and works both for Markovian and non-Markovian types of noise. (3) Versatility. The model works in a broad range of applications and enables error mitigation for quantum algorithms, dynamics of many-body systems, and continuous-variable quantum information processing. In addition, it accepts the input data in a variety of different forms, including expectation values of quantum observables, statistics of measurement outcomes, and estimates of the Wigner function. (4) Transferability. The trained model exhibits the capability to mitigate errors for circuits sharing the same circuit skeleton as those considered in the training, all without the need for retraining. This feature makes it possible to apply the model to a wide range of quantum circuits, enhancing its practical utility and scalability. To demonstrate these features, we test our model on a series of paradigmatic quantum algorithms, such as variational quantum eigensolvers[41] and quantum approximate optimization[42], and quantum dynamics, such as the many-body dynamics of the Ising model and the Kerr Hamiltonian[43] for continuous variable systems. Furthermore, we tested our model on real quantum hardware. The results demonstrate its superior performance compared to previous methods, including ZNE and CDR.

[1]QICI Quantum Information and Computation Initiative, Department of Computer Science, The University of Hong Kong, Pok Fu Lam, Hong Kong. [2]Department of Computer Science, Oxford, UK. [3]Perimeter Institute for Theoretical Physics, Waterloo, ON, Canada. [4]These authors contributed equally: Manwen Liao, Yan Zhu. ✉e-mail: yuxiang@cs.hku.hk

Results

The DAEM model

Let us start by specifying our error mitigation framework. We consider a target quantum process $\mathcal{E}$ corresponding to a quantum circuit composed of a specific sequence of single-qubit and CNOT gates (which is general as single-qubit gates and CNOT gates can form a universal gate set), with the restriction that only Pauli measurements are performed. The circuit represents the action of an ideal quantum device in the absence of noise. Note that our method also applies to some reversible processes without explicit circuit representations (see "Error mitigation for continuous-variable processes"). In the real world, however, one has only access to noisy versions of the process $\mathcal{E}$. Such noisy versions will be denoted by $\mathcal{N}_\lambda(\mathcal{E})$, where $\mathcal{N}_\lambda$ represents the noise model and λ indicates the noise parameter. The input state of the process $\mathcal{N}_\lambda(\mathcal{E})$ is randomly selected from an ensemble $\mathcal{S} = \{\rho_s\}_{s=1}^n$, which can generally contain multiple quantum states. For every input state $\rho_s \in \mathcal{S}$, the goal of error mitigation is to estimate the statistics of measurements performed on the ideal output state $\mathcal{E}(\rho_s)$ given access to data from its noisy version $\mathcal{N}_\lambda(\mathcal{E})(\rho_s)$.

Here we focus on a set of Pauli measurements $\mathcal{M}$ of interest, such as a set of Pauli measurements performed on a subset of the output qubits in a quantum computation. Each measurement $\boldsymbol{M}_i = (M_{ij})_j$ in $\mathcal{M}$ is a positive operator-valued measure consisting of m positive operators that satisfy the normalization condition $\sum_{j=1}^m M_{ij} = \mathbb{1}$. This general setup covers tomography (when $\mathcal{M}$ is informationally complete), as well as quantum algorithms, where a single measurement is used for read-out.

To achieve error mitigation, we now introduce a neural network-based model. Our model, called data augmentation-empowered error mitigation model (DAEM), is illustrated in Fig. 1. Its high-level structure consists of two phases. In the first phase, the Noise-Awareness phase, we train a neural network to remove the action of the noise $\mathcal{N}_\lambda$ from the measurement statistics. The training is boosted by a technique called *quantum data augmentation*. The key idea is to train the network on data generated by fiducial processes, a set of quantum circuits derived from the target quantum process.

The fiducial process $\mathcal{F}$ is expected to have two desired features: First, in the absence of noise, a classical computer should be capable of efficiently generating adequate measurement data corresponding to various input states and Pauli measurements. In the presence of noise, the fiducial process $\mathcal{N}_\lambda(\mathcal{F})$ should be implementable using the same quantum computing hardware that executes the target process. Second, when implemented on the quantum hardware, the noise pattern of the fiducial process should be as close to that of the target process as possible. This ensures that error mitigation techniques learned on the fiducial process can be effectively transferred to the target process.

Following this spirit, we construct the fiducial process by making changes to the execution of the original implementation according to the following recipe: (1) For every single-qubit gate R, we instead ask the quantum computer to execute $\sqrt{R^\dagger}\sqrt{R}$, which equals an identity gate in the ideal case. The motivation for this is to make the noise pattern of the fiducial process emulates that of the target process. For example, in trapped ion systems, the replaced gates can be implemented with the same execution time by adjusting the duration of the interaction. This leads to similar noise patterns assuming the dissipative part of the qubit dynamics to be fixed. Note that the implemented fiducial process will not be an identity process in general, since the implementation is not perfect. (2) All CNOT gates are executed according to the original circuit.

With this recipe, since $\mathcal{F}$ consists only of CNOT gates, the measurement statistics for the output state $\mathcal{F}(\sigma)$ with respect to Pauli measurements can be efficiently computed for any product state σ: Note that $\mathcal{F}$ is Clifford since it consists of CNOTs only. Taking the Heisenberg picture, the evolution of any Pauli observable P under $\mathcal{F}^\dagger$, which results in an N-qubit Pauli observable $\mathcal{F}^\dagger(P)$, can be efficiently simulated. Since σ is a product state, the desired expectation $\text{tr}(\mathcal{F}^\dagger(P)\sigma)$ can be computed classically in $O(N)$ time.

To generate the training data, the experimenter collects measurement statistics by executing the noisy fiducial processes $\mathcal{N}_\lambda(\mathcal{F})$ on a set of product states $\{\sigma_s\}$, using the same hardware as the target quantum process. The acquired measurement statistics will be denoted by $\{\boldsymbol{p}'^{(1)}_{i,s}\}$, while the ideal
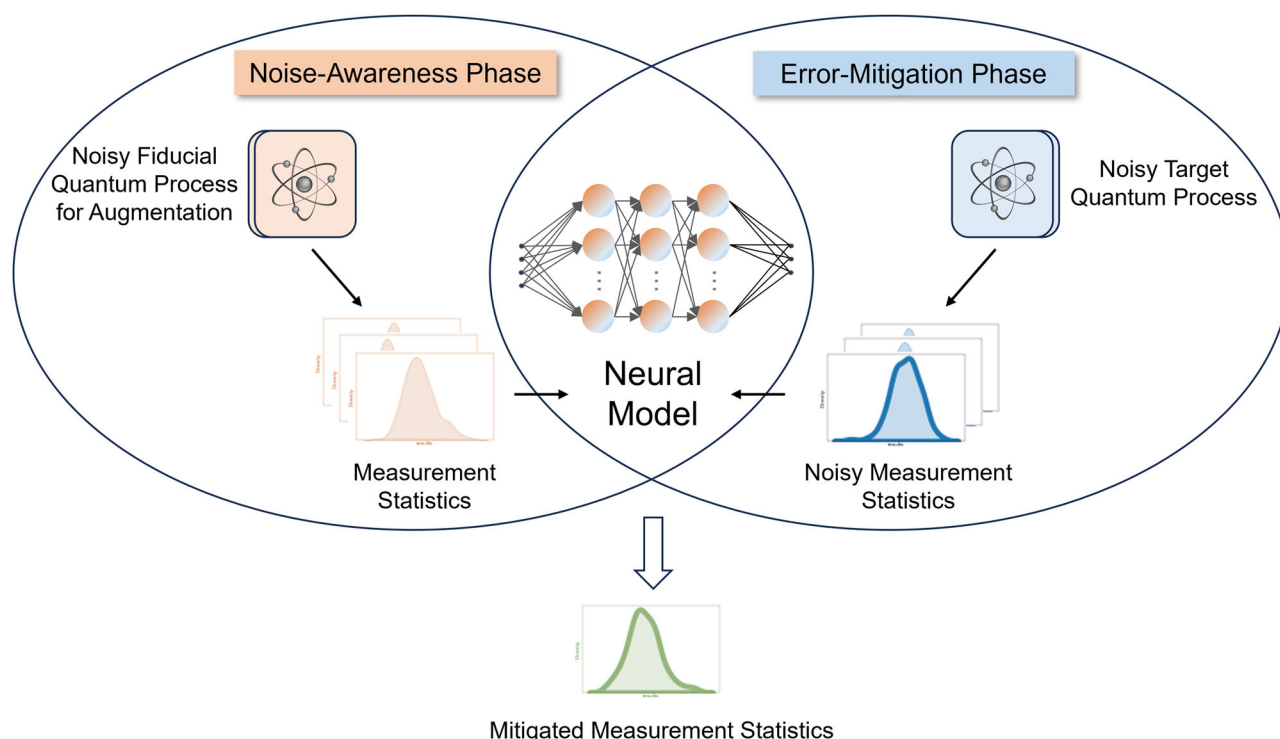

Fig. 1 | Framework of DAEM model. The entire procedure is divided into two phases. In the first phase, known as Noise-Awareness phase, we train the neural model for error mitigation with the assistance of a fiducial process responsible for data augmentation. In the second phase, known as Error-Mitigation phase, we apply the trained neural model to mitigate the errors in noisy measurement statistics collected from the noisy version of the target quantum process.

measurement statistics will be denoted by $\boldsymbol{p}'^{(0)}_{i,s} := (\mathrm{tr}(\mathcal{F}(\sigma_s)M_{ij}))_j$. In scenarios where varying the noise is possible, statistics can also be collected with various noise parameters $\{\lambda_k\}^K_{k=1}$. In such cases, the acquired measurement statistics are denoted by $\{\boldsymbol{p}'^{(k)}_{i,s}\}^K_{k=1}$. Note that we do not assume knowledge of the exact values of $\{\lambda_k\}^K_{k=1}$ and, consequently, do not require any extra noise estimation procedure. We then train the neural model by providing tuples of the form $(\boldsymbol{p}'^{(k)}_{i,s})^K_{k=1}$ corresponding to a given input state σ_s and a given measurement $\boldsymbol{M}_i$. Thanks to the aforementioned features of the fiducial processes, the ideal statistics $\boldsymbol{p}'^{(0)}_{i,s}$ can be computed efficiently from the input product state σ_s. In the training, we optimize the parameters of the model with respect to a loss function $\mathcal{L}$ that quantifies the deviation between the predicted statistics and the noise-free one (see "Methods" for details).

After the training is concluded, the model can be used for error mitigation on the target process $\mathcal{E}$. The experimenter collects measurement statistics by performing the noisy process $\mathcal{N}_\lambda(\mathcal{E})$ on an arbitrary input state $\rho \in \mathcal{S}$, with different noise parameters $\{\lambda_k\}^K_{k=1}$. The corresponding statistics will be denoted by $\boldsymbol{p}^{(k)}_i = (\mathrm{tr}(\mathcal{N}_{\lambda_k}(\mathcal{E})(\rho)M_{ij}))_j$. The neural model then outputs the inferred ideal statistics $\boldsymbol{p}^{(0)}_i := (\mathrm{tr}(\mathcal{E}(\rho)M_{ij}))_j$ pertaining the target process $\mathcal{E}$. A detailed description of the implementation of the neural model in various examples is provided in the "Methods" section.

An important feature of the DAEM model is that it can be applied to ensembles containing multiple input states. In addition, the states appearing in the Error Mitigation phase do not need to be the same states used in the training. Furthermore, it is worth stressing that the model does not require any ideal measurement data (neither experimentally generated nor classically simulated) for the target process $\mathcal{E}$. As a consequence, it has the potential to be applied to large-scale systems where the classical simulations are not feasible, and realistic experiments are affected by non-negligible amounts of noise.

Additionally, our model can be trained for multiple target processes that share the same circuit skeleton but have different parameters. This is achieved using a set of fiducial processes. The underlying intuition is that circuits with the same skeleton are likely affected by similar noise patterns. Consequently, the knowledge gained from mitigating errors in one such circuit can be transferred to others within the same structural framework. This transferability enhances the efficiency and applicability of our model, reducing the need for extensive retraining for each new set of parameters.

Error mitigation for quantum algorithms

The domain most suitable for testing our error mitigation model is quantum circuits, which are widely employed in various quantum algorithms. Our framework applies generally to quantum algorithms, where the goal is to obtain noise-free statistics from noisy quantum circuits. In this section, we test the performance of DAEM on prototypical NISQ algorithms, including the Variational Quantum Eigensolvers (VQEs)[41], the swap test[44], and the Quantum Approximate Optimization Algorithm (QAOA)[42].

Variational quantum eigensolvers. VQEs, widely utilized in the realms of quantum chemistry and quantum computation, leverage parameterized quantum circuits to approximate the ground states of specified Hamiltonians. However, in practical scenarios, these circuits inevitably grapple with noise, leading to deviations in the ground state energy from the ideal scenario. In this context, we consider a scenario where an experimenter possesses the optimal parameters of a well-trained VQE circuit and intends to employ it on a real noisy quantum device. The experimenter's goal is to derive the ideal measurement statistics of the ground state based on the gathered noisy measurement data.

In the following, we consider the VQEs for the transverse Ising chain with Hamiltonian

$$H_{\mathrm{Ising}} = -g\sum_{i=1}^{N} X_i - J\sum_{i=1}^{N-1} Z_i Z_{i+1}, \tag{1}$$

where X, Z are Pauli operators, and N is the number of qubits. The variational ansatz used to prepare the ground state is a hardware-efficient ansatz, composed of single-qubit Euler rotation gates and CNOT gates, as illustrated in Fig. 2a. We choose 16 circuits, varying the parameter g within the range of [0.4, 2.0] with a stride of 0.1. Additionally, we set the values of J and N to be 1 and 4 respectively for all experiments. For the set of measurements $\mathcal{M}$, we choose all two-qubit Pauli measurements on nearest-neighbor qubits. In the Noise-Awareness phase, we construct the fiducial circuit by replacing each single-qubit rotation gate with two single-qubit rotation gates, while keeping the CNOT gates unchanged. These rotation gates are parameterized to mutually cancel each other out. For instance, an $R_x(\phi)$ gate is replaced by an identity gate, which is specifically constructed as $R_x(-\phi/2)R_x(\phi/2)$. We let $\mathcal{S}$ be all of the 4-qubit mixed states and we randomly select $n = 100$ states $\{\sigma_{s_1}\}^n_{s_1=1}$ in the Noise-Awareness phase of all our experiments for each g.

During the Error-Mitigation phase, we evaluate our mitigation model by using the prepared initial state $\rho_0 = |0\rangle\langle0|^{\otimes N}$.

First, we evaluate our model's performance under two Markovian noise models: amplitude damping and phase damping. In all of the experiments, noise is applied after each gate in Fig. 2a. The amplitude damping noise channel and the phase damping noise channel are mathematically defined by Equation (3) and Equation (6), respectively. Throughout all of our experiments, we consider a set of noise parameters denoted as $\{\lambda_k\}^K_{k=1} \in [0.05, 0.29]$, with stride 0.02.

$$\rho \rightarrow V_0 \rho V_0^\dagger + V_1 \rho V_1^\dagger, \tag{2}$$

with $V_0 = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-\lambda} \end{bmatrix}$ and $V_1 = \begin{bmatrix} 0 & \sqrt{\lambda} \\ 0 & 0 \end{bmatrix}$. Figure 2b illustrates the mitigation results obtained using various error mitigation techniques for VQE circuits affected by phase damping noise, while Fig. 2c presents the mitigation results for VQE circuits affected by amplitude damping noise. The results clearly demonstrate that DAEM consistently outperforms other mitigation methods for each VQE circuit, regardless of the specific value of g.

Furthermore, we tested our trained model on circuits for preparing ground states of the Ising model with parameters not included in the training set, using the same variational ansatz. Specifically, we varied the parameter g within the range of [0.45, 1.95] with a stride of 0.1. We present the experimental results for mitigating phase damping noise and amplitude damping noise in Fig. 2d, e. The results demonstrate that our model can efficiently transfer error mitigation knowledge to circuits sharing the same ansatz but with different parameters, without requiring further training.

In addition to the Markovian noise model, we also investigate the impact of Non-Markovian noise, which, despite its relevance in real-world quantum experiments[45–47], has received limited attention in previous error mitigation studies. Specifically, we consider the multi-qubit spin-boson model[48] for phase damping to exemplify this scenario, in which a quantum system interacts with the environment, namely, a heat bath, and evolves jointly. This is a potential noise happening in superconducting quantum circuits[49]. In this setup, depicted in Fig. 3a, the system Hamiltonian H_S corresponds to the VQE circuit, while the heat bath is modeled as a bosonic system. We assume each gate in the circuit interacts independently with a bath attached locally. The bath Hamiltonian is $H_B = \sum_k \omega_k b_k^\dagger b_k$. Here b_k is the annihilation operator for mode k, and ω_k is the corresponding energy. The interaction between the system and the bath is captured by the Hamiltonian $H_{SB} = \sum_k \sigma_z \otimes [\lambda_k b_k + \lambda_k^* b_k^\dagger]$, where σ_z is Pauli-Z operator, and $\lambda_k \propto 1/\sqrt{\omega_k}$. We initiate the system and bath as a product state state $\rho_0 \otimes \rho_B$, where ρ_0 is the initial state of the system, namely $|0\rangle\langle0|$. The bath
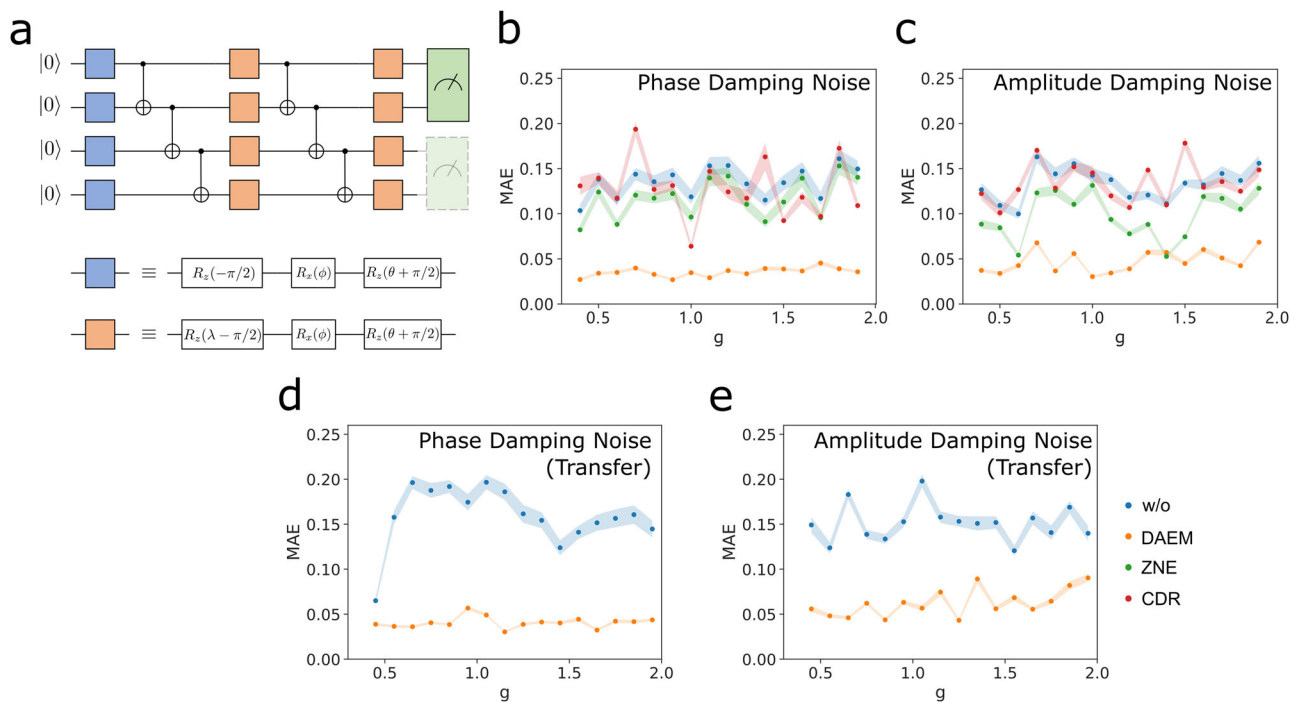
Fig. 2 | Error mitigation for variational quantum eigensolvers. a The variational ansatz for preparing the ground states of 4-qubit transverse Ising models. **b** Mean Absolute Errors (MAE) between the mitigated measurement expectation values for phase damping noise model and ideal expectation values. **c** Mean Absolute Errors (MAE) between the mitigated measurement expectation values for amplitude damping noise model and ideal expectation values. **d** Mean Absolute Errors (MAE) between the mitigated measurement expectation values for phase damping noise model and ideal expectation values for the circuits not included in the training. **e** Mean Absolute Errors (MAE) between the mitigated measurement expectation values for amplitude damping noise model and ideal expectation values for the circuits not included in the training. It is noteworthy that ZNE requires knowledge of noise parameters associated with statistics while our proposed DAEM not. Despite being under an unfair comparison, DAEM still demonstrates superior performance.
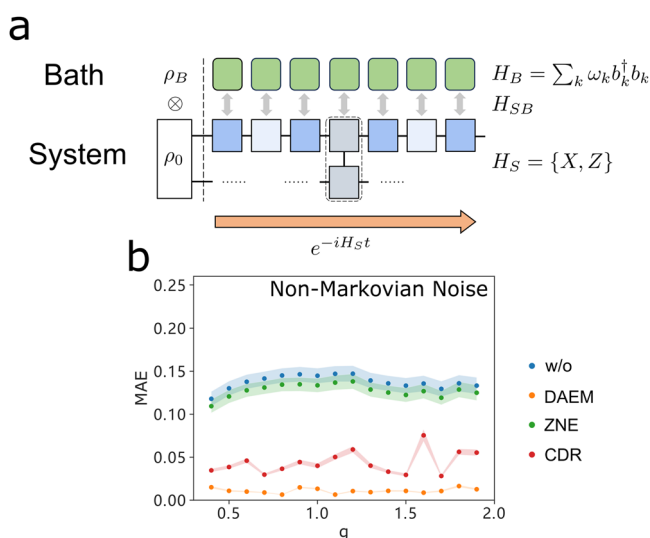

Fig. 3 | Error mitigation for variational quantum eigensolvers affected by Non-Markovian Noise. a Schematic diagram of quantum circuits affected by Non-Markovian noise. **b** Mean Absolute Errors (MAE) between the mitigated measurement expectation values for considered Non-Markovian noise model and ideal expectation values.

$\rho_B = e^{-\beta H_B}/Z$ is a Gibbs state, with $\beta = 1/(k_B T)$ and Z being a normalization factor. The evolution of the system under noisy conditions is represented as $\rho_S(t) = \mathrm{Tr}_B(U(t)(\rho_0 \otimes \rho_B)U^\dagger(t))$. Here, $U(t) = \exp[-i \int_0^t H(\tau)d\tau]$ is the unitary describing the joint evolution of the whole system, with $H = H_S + H_B + H_{SB}$. Importantly, it should be noted that this noise is gate-dependent, as

gate parameters rely on the evolution time of the Hamiltonians, leading to varying noise effects.

To modulate this noise, we vary the Hamiltonian evolution time within the range of $[0.05, 0.3]$, while maintaining the computational impact constant. The numerical results, as depicted in Fig. 3b, consistently demonstrate the remarkable effectiveness of our model in handling Non-Markovian noise scenarios. This performance is particularly significant because Non-Markovian noise, a common occurrence in practical quantum experiments[45-47], poses a substantial challenge for error mitigation techniques. The robustness of our model in such conditions significantly enhances its practical utility and reliability in the realm of quantum computing.

To further illustrate the effectiveness of our model in practical scenarios, we tested its performance on the real quantum computing hardware provided by OriginQ Cloud[50]. Based on the types of gates available on this hardware, we adopted the variational ansatz shown in Fig. 4a. The circuit is composed of three layers of U_3 gates and two layers of controlled-Z gates. We selected 10 circuits, varying the parameter g within the range of $[1.0, 2.0)$ with a stride of 0.1. The choices of J, N, $\mathcal{M}$, and $\mathcal{S}$ were kept consistent with those used in the simulation experiments above. During the Noise-Awareness phase, we construct the fiducial circuit by replacing each U_3 gate with two U_3 gates that cancel each other out under noiseless conditions, while keeping the CZ gates unchanged, as shown in Fig. 4b.

Given the difficulty of adjusting noise strength in real experiments, we consider only one noise level ($K = 1$) for our model. As the results shown in Fig. 4c, our DAEM model demonstrates superior performance in most cases compared to CDR. Note that ZNE cannot be applied in this scenario directly as it requires measurement data from circuits with varying noise strengths. For comparison, we used the method of unitary folding[51] to generate data
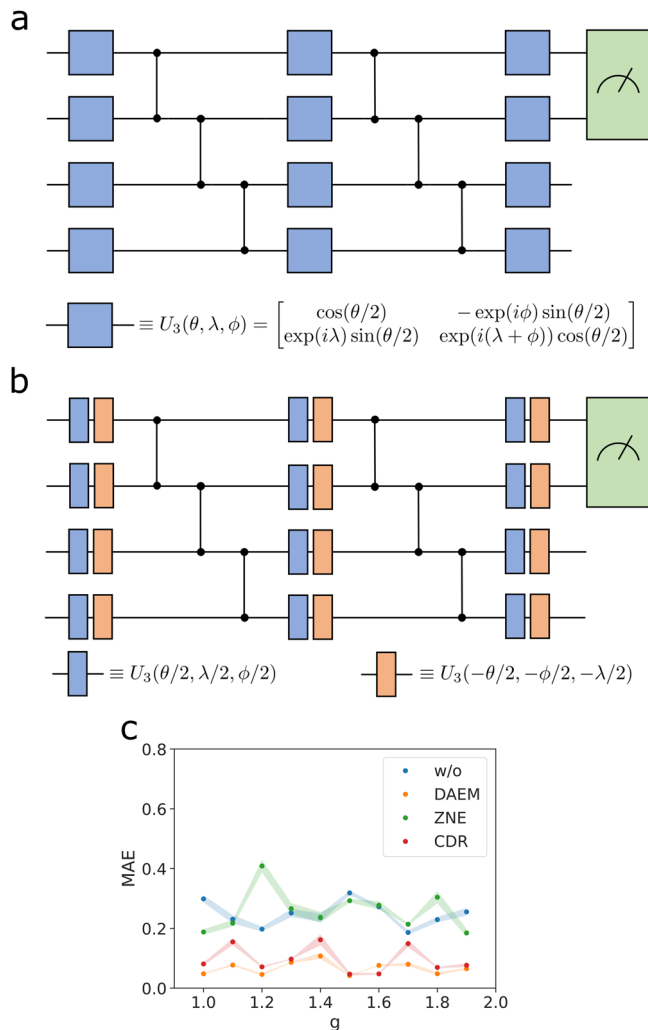
Fig. 4 | **Error mitigation for variational quantum eigensolvers on the OriginQ Cloud quantum hardware. a** The variational ansatz for preparing the ground states of 4-qubit transverse Ising models. **b** The structure of fiducial circuits. **c** Mean Absolute Errors (MAE) between the mitigated measurement expectation values and ideal expectation values. The average MAE of w/o, DAEM, ZNE, CDR are 0.247, 0.067, 0.259, and 0.095 respectively.
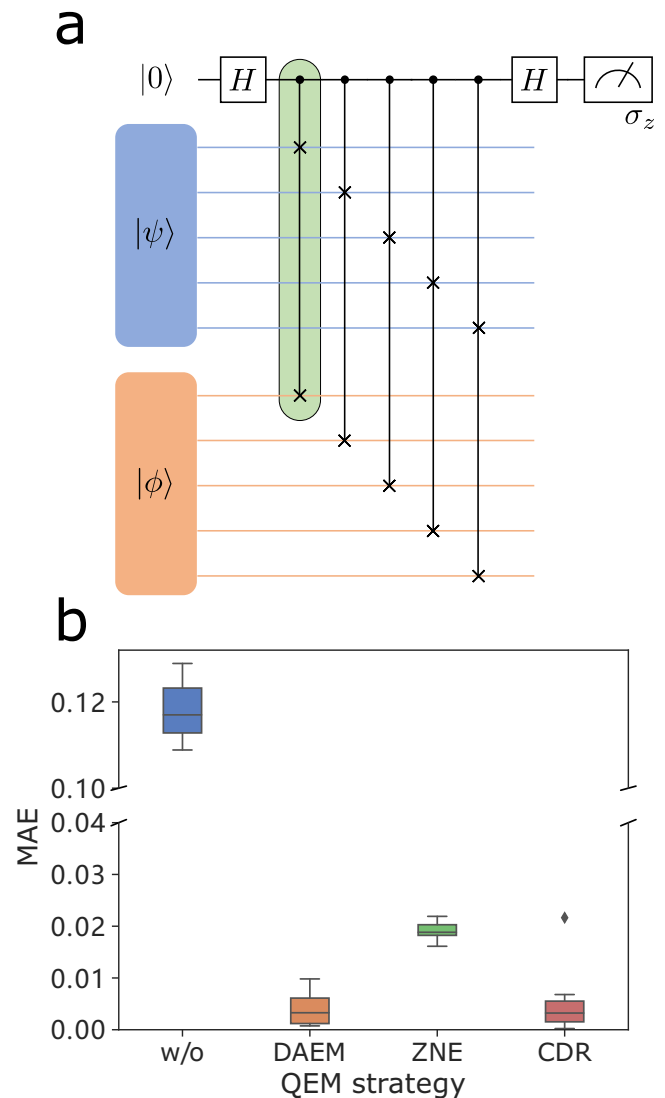

Fig. 5 | **Error mitigation for the swap test. a** The swap test circuit for comparing two 5-qubit states. The gate within the green box is the controlled-SWAP gate. **b** Mean Absolute Errors (MAE) between the mitigated fidelity values and the ground truth values.

with four different noise levels for ZNE. Although our model uses data from fewer noise levels than ZNE, it still achieves significant advantages over ZNE.

Swap test. The swap test is a technique used to measure the dissimilarity between two quantum states. In Fig. 5a, we illustrate the circuit designed for comparing two 5-qubit states. When executed on the quantum device, the CSWAP gate is implemented by decomposing it into three Toffoli gates, which are further decomposed using Hadamard, S, T, and CNOT gates. The details are provided in "Methods" section. The circuit takes two input quantum states, $|\psi\rangle$ and $|\phi\rangle$, for comparison. It initializes the first control qubit as $|0\rangle$ and produces expectation value that equals the fidelity between two pure states, i.e., $|\langle\psi|\phi\rangle|^2$, by performing a Pauli Z measurement on the first control qubit. Here, we assume that noise takes place before each controlled-SWAP gate. Specifically, we examine phase damping channel, which can be characterized by the following equation:

$$\rho \to V_0 \rho V_0^\dagger + V_1 \rho V_1^\dagger, \qquad (3)$$

with $V_0 = \begin{bmatrix} 1 & 0 \\ 0 & e^{-2\lambda} \end{bmatrix}$ and $V_1 = \begin{bmatrix} 0 & 0 \\ 0 & \sqrt{1-e^{-4\lambda}} \end{bmatrix}$. λ represents the scale of the noise and P_i are three Pauli gates. We use a set of noise parameters denoted as $\{\lambda_k\}_{k=1}^{K} = \{0.05, 0.08, 0.12, 0.15\}$ in this experiment.

In the Noise-Awareness phase, the controlled-SWAP gates in Fig. 5 are first decomposed into single-qubit and CNOT gates. $\mathcal{N}_\lambda(\mathcal{F})$ is constructed by replacing all single-qubit gates by two gates that cancel each other. Specifically, for a quantum gate G, we replace it by $\sqrt{G^\dagger}\sqrt{G}$. We randomly select n input states $\{\sigma_{s_1}\}_{s_1=1}^{n}$, where each $\sigma_{s_1} = |\psi\rangle\langle\psi| \otimes \sigma_{s_1}^1 \otimes \sigma_{s_1}^2$, with $|\psi\rangle$ being a random 1-qubit pure state, and $\sigma_{s_1}^1$ and $\sigma_{s_1}^2$ representing two random 5-qubit product states.

In the Error-Mitigation phase, we evaluate our model using 20 pairs of input states ρ_1, ρ_2 in the swap test circuit. We collect statistics by conducting Pauli Z measurements on the first qubit within the noisy swap test circuit. These measurements are subsequently used to compute the overlap between ρ_1 and ρ_2. The noisy expectation values obtained from these measurements are then input into the trained neural model, which produces mitigated values as output. In Fig. 5b, we present Mean Absolute Errors (MAE)
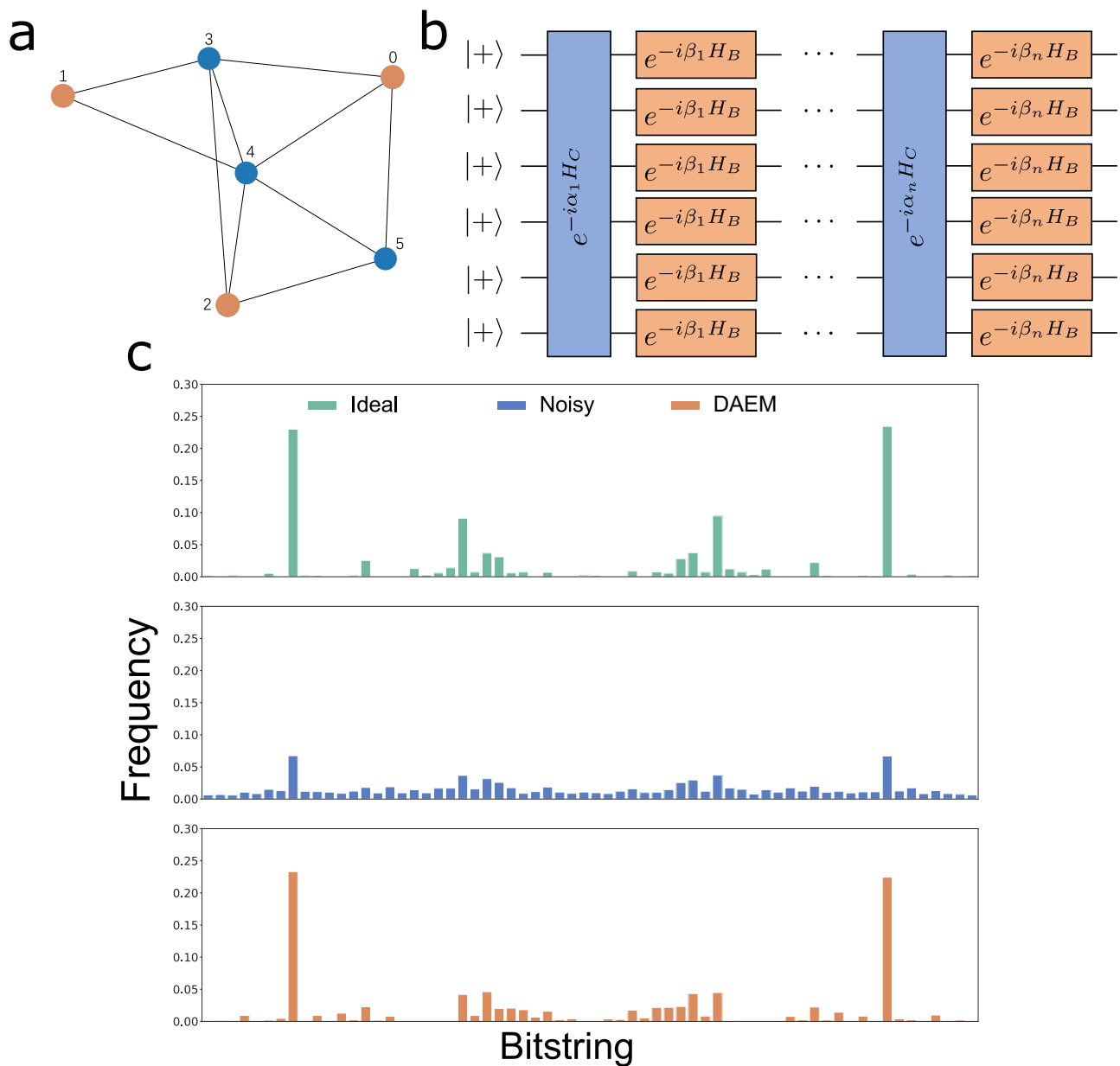
Fig. 6 | Error mitigation for quantum approximate optimization. a An instance of a graph for the Max-cut problem. **b** The variational ansatz for implementing QAOA algorithm. **c** Ideal, Noisy, and Mitigated frequency of measurement results.

between the mitigated values and the ground truth values, providing a comparative analysis with two other quantum error mitigation techniques, ZNE and CDR. The performance of DAEM stands out, showcasing significant improvements over ZNE and demonstrating comparable performance with CDR.

Quantum approximate optimization algorithms. QAOA[52] is a quantum algorithm specifically designed for solving combinatorial optimization problems. The core of this algorithm involves encoding the objective function of the target optimization problem into a Hamiltonian, and trains an elaborately designed parameterized circuit to approximate the ground state. The final solution is derived by sampling bitstrings from the circuit's output in the computational basis. However, when running a well-trained QAOA circuit on noisy quantum computers, the resulting output distribution deviates from the ideal scenario, which results in less accurate solutions. Hence, our goal is to mitigate this noise-induced bias in the output distribution, thereby providing experimenters with more precise solutions.

In this specific application, we focus on implementing QAOA for the maximum cut (Max-cut) problem[53]. The goal is to find a bi-partition of graph G, namely subsets A and B, in which the partition contains the maximum number of edges. This can be defined as an optimization with objective

$$\max_{z} L(z) = \frac{1}{2} \sum_{(i,j) \in E} (1 - z_i z_j), \tag{4}$$

where i, j denote the indices of vertices, (i, j) represents the edge connecting vertex i and vertex j, and E is the set containing all edges of the graph. If vertex i belongs to subset A, then $z_j = 1$, otherwise $z_j = 0$. We provide an instance of G with six vertices in Fig. 6a. The corresponding Hamiltonian of this problem in QAOA can be described by the following:

$$H_C = \frac{1}{2} \sum_{(i,j) \in E} (I - Z_i Z_j), \tag{5}$$

Fig. 7 | Error mitigation for quantum spin dynamics. a Mean Absolute Errors (MAE) between the mitigated measurement expectation values for phase damping noise model and ideal expectation values. **b** Mean Absolute Errors (MAE) between the mitigated measurement expectation values for amplitude damping noise model and ideal expectation values.
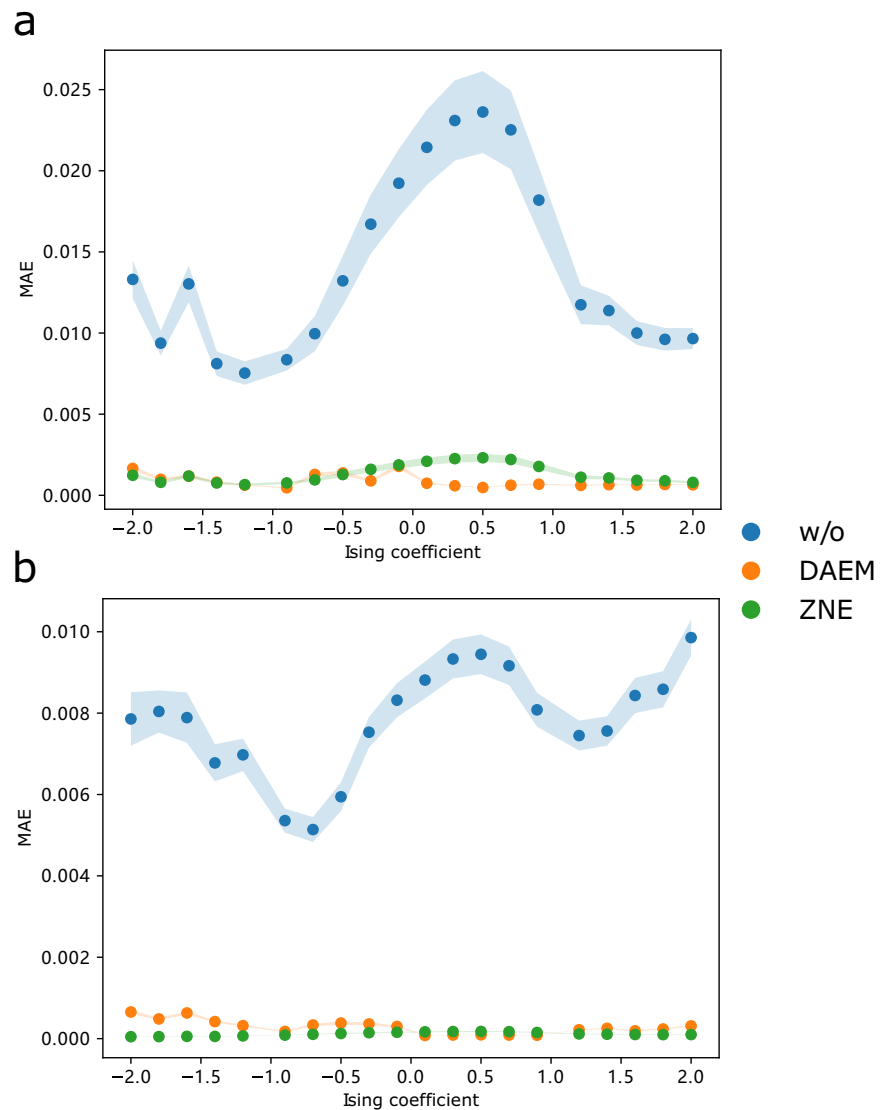
where Z is Pauli-Z operator. The circuit for QAOA, shown in Fig. 6b, typically comprises two sets of parameterized quantum gates, alternating between a mixing operator $H_B = \sum_{n=1}^{N} \sigma_n^x$ and the problem-specific cost operator H_C. In the Noise-awareness phase, to generate fiducial process, after replacing all single-qubit gates with identity gates, the CNOT gates automatically cancel each other. In this case, the fiducial process is trivially identity ideally, i.e., $\mathcal{N}_0(\mathcal{F}) = I$. Again, we execute the fiducial process on noisy quantum devices with different noise parameters to acquire noisy bitstring distributions. Besides, we sample input states $|\psi\rangle$ in computational basis to obtain labels for training. In Fig. 6c, we present the mitigated results concerning the output state of a trained QAOA circuit applied to the graph depicted in Fig. 6a. Here, we consider the depolarizing noise model and it can be described by

$$\rho \rightarrow (1 - \lambda)\rho + \frac{\lambda}{4^N - 1} \sum_i P_i \rho P_i, \tag{6}$$

where P_i are the $4^N - 1$ Pauli gates excluding the identity gate. It's evident that the mitigated frequency of measurement results closely approximates the ideal scenario, signifying that we can obtain more reliable solutions to the original Max-Cut problem through our DAEM model. It's worth noting that both ZNE and CDR are designed specifically for mitigating errors in expectation values, and they cannot be applied directly to the probability distribution of measurement results, in contrast to our proposed DAEM.

Error mitigation for many-body dynamics

Our model works for quantum processes beyond the circuit model. It applies to, for example, the dynamics of physical systems. In this section, we delve into the challenge of error mitigation within the domain of many-body dynamics, which is fundamental to various applications in quantum physics and materials science.

Here, our focus is on the dynamics of a 50-qubit quantum system with an Ising Hamiltonian H_{Ising}, described in Eq. (1). We consider the whole system's evolution for time t, given as $U = \exp(-iH_{\text{Ising}}t)$. This specific process is characterized by the following parameters: $J = 1$, $g = 2$, and a time duration of $t = 5$. For the initial states involved in this Ising Hamiltonian evolution, we have selected the ground states of the Ising model, varying J within the range $[-2, 2]$, while keeping g constant at 1. To simulate these processes, we employ a combination of two powerful techniques: the density matrix renormalization group (DMRG)[54] and time-evolving block decimation (TEBD)[55,56]. In this setup, noise is introduced after the completion of the unitary quantum process. Specifically, we evaluate our model's performance under two distinct noise models: phase damping and amplitude damping channels. For the set of measurements $\mathcal{M}$, we also consider all two-qubit Pauli measurements on nearest-neighbor qubits. The results, corresponding to different values of g in the input states, are presented in Fig. 7. We can observe that both our model and ZNE have achieved nearly perfect mitigation, as evidenced by the MAE between the mitigated

expectation values and the ground-truth values, which are close to zero. We conjecture the reason for the nearly perfect performance of ZNE in this experiment is that no SPAM errors have been introduced. This makes the actual measurement expectation values decay quadratically with respect to the noise parameters, perfectly fitting the ansatz of ZNE. It is important to highlight that ZNE relies on precise knowledge of the noise parameters corresponding to the noisy measurement data, whereas our DAEM model does not have this requirement. CDR is a technique tailored for quantum circuits, and therefore, it cannot be employed to mitigate errors in spin-system dynamics.

Error mitigation for continuous-variable processes

Continuous-variable quantum systems have demonstrated their potential in diverse applications including quantum cryptography[57] and quantum computing[58]. However, despite the growing significance of this type of system, no prior research has discussed the issue of error mitigation within continuous-variable quantum systems as far as we know. In this section, we applied our proposed DAEM model to address this long-unexplored challenge first.

We assess the effectiveness of our method on the dynamics induced by Kerr's nonlinear interaction[43], which is important for continuous variable quantum computing[58]. Consider a quantum system initially prepared in a coherent state $|\alpha\rangle$ and subjected to the Kerr Hamiltonian $H_{\text{kerr}} = \pi \hat{a}^{\dagger 2} \hat{a}^2$, where $\hat{a}$ and $\hat{a}^\dagger$ represent the annihilation and creation operators. In this scenario, we model the noisy process by a lossy open system, whose dynamics are described by the Lindblad master equation:

$$\dot{\rho} = -\mathrm{i}[H_{\text{kerr}}, \rho] + \lambda \mathcal{D}(\hat{a})(\rho). \tag{7}$$

Here, $\mathcal{D}(\hat{a})(\rho) := \hat{a}\rho\hat{a}^\dagger - \frac{1}{2}(\rho\hat{a}^\dagger\hat{a} + \hat{a}^\dagger\hat{a}\rho)$, and λ represents the loss rate. Our objective is to mitigate the errors in this process, making it closely resemble the ideal closed-system dynamics governed by Schrödinger's equation with Hamiltonian H_{kerr}. In this setting, we consider the measurement results associated with the point-wise Wigner function[59,60]. In our numerical experiments, we initialize the state as coherent state $|\alpha\rangle$ with $\alpha = 1.5$ and dimension $N = 15$. We vary the evolution time over the interval $t \in [0, 1]$, considering different loss rates $\lambda \in \{0.6, 0.65, 0.7, 0.75, 0.8\}$. To train the neural model within DAEM, we construct the fiducial process $\mathcal{N}_\lambda(\mathcal{F})$ by implementing the evolution of two inverse Hamiltonians, ensuring that the overall effect on the input state is an identity operation under noiseless conditions. Specifically, assuming a total evolution time of t_0, the state evolves with the Hamiltonian H_{kerr} for $t \leq t_0/2$, and with $-H_{\text{kerr}}$ for $t_0/2 \leq t \leq t_0$. We assess the effectiveness of our DAEM by computing the fidelities between the mitigated states and their noiseless counterparts, employing the values of the point-wise Wigner function. As shown in the numerical results presented in Fig. 8a, the fidelity between the state affected by the noise and the ideal state decreases rapidly as time increases. In contrast, our method excels in mitigating this effect, resulting in a dramatic improvement in fidelity. We also present snapshots of the state at different time points in Fig. 8b, and our mitigated point-wise measurement results are notably closer to the ideal ones, particularly for longer evolution.

Discussion

The workhorse of our model is the quantum data augmentation method, which generates the training data by letting the unknown noise act on a set of ideal fiducial processes. This technique is not only applicable to quantum error mitigation, but also to other tasks in quantum information processing, including in particular the task of enhancing parameter estimation in quantum metrology[61]. By combining quantum data augmentation with the representational capability of deep neural networks, our model becomes able to effectively handle complex noise scenarios. For example, it deals effectively with Non-Markovian noise (viz. Fig. 2e), for which other neural models, like ZNE, tend to perform poorly due to reliance on a predefined extrapolation algorithm.

Our model also offers appealing features compared to conventional error mitigation methods. In CDR, an error mitigation model is better trained with classically simulated quantum circuits that resemble the target circuit[7]. Therefore, the effectiveness of CDR can depend heavily on how closely the training circuits match the target circuit, and achieving such a close match might not always be feasible in practical experiments. In contrast, our proposed DAEM model conducts training directly on the data collected from the hardware, targeting the specific noise to be mitigated. This approach ensures more accurate and effective error mitigation tailored to the actual noise characteristics of the hardware. It is also worth observing that CDR and ZNE are effective at mitigating expectation values, but generally less effective at mitigating the whole probability distribution of the measurement outcomes, a task that is necessary in quantum algorithms like QAOA. For noise preservation, our approach shares a similar spirit with ref. 36, which targets mitigating errors in QAOA circuits. Reference [36] works by considering a modified version of the original circuit, where all single-qubit R_Z gates in the cost gates are ignored. This corresponds to a modified problem Hamiltonian whose ZZ coupling strength is zero, and thus the modified circuit can be simulated efficiently with a classical computer. The modified circuit is then executed on a (noisy) quantum computer. As the (pairwise) CNOT gates are left unchanged in the modified circuits, the outputs of the real-device execution can be compared with the simulation to learn the pattern of noise propagation. Likewise in our framework, we elaborately design our data augmentation strategy that preserves the skeletons of the circuits. Our approach can also be compared to probabilistic error cancellation[2,11,12], which estimates noise-free expectation values by representing them as linear combinations of expectation values from a set of noisy quantum circuits. To work out the appropriate decomposition, this method requires a tomography of the noise, thus resulting in an overhead in sample complexity. A benefit of our approach is that it removes the need for tomography and replaces it with the quantum data augmentation procedure, which is generally less demanding in terms of number of measurement settings.

In terms of scalability, our model can potentially be scaled to larger systems. We conduct experiments on mitigating errors of circuits with different number of qubits and different circuit depth, with a fixed number of training data. Results show that our model can keep stable performance with respect to different circuit configurations. Further information can be found in Supplementary Note 4.

Finally, our model has the potential for extension to mitigate a broader range of realistic quantum errors, including crosstalk errors[62,63], which are common in quantum computing systems. Crosstalk errors result from hardware imperfections that violate the assumption of locality and independence of quantum operations, and are therefore challenging to model[64]. Despite these challenges, error mitigation for crosstalk errors could become approachable in our framework, which does not require prior error modeling.

Methods

Neural model in DAEM

Our error mitigation is model-agnostic thus the structure of the neural model can be flexibly chosen, ranging from simple linear models[65], multi-layer perceptrons (MLP)[66], to deep neural networks like convolutional neural networks[67] and Transformers[68]. In practice, we adopt a problem-aware strategy to design the specific construction of the model. In general, we prefer non-linear models for they have stronger ability to capture the intrinsic characteristics of various noise models.

For error mitigation in quantum algorithms and many-body dynamics, we use MLP as the architecture of the neural model. The neural network is composed of multiple layers of fully connected neurons. Each neuron involves one linear transform followed by one non-linear activation. The stack of neurons allows for complex non-linear function fitting, which is powerful for estimating the expectation values and probability distributions in our error mitigation settings. The model's inputs are parameters indicating the target circuit for mitigation, the observable to be measured, and the measurement statistics. The model's output is either a real number or a
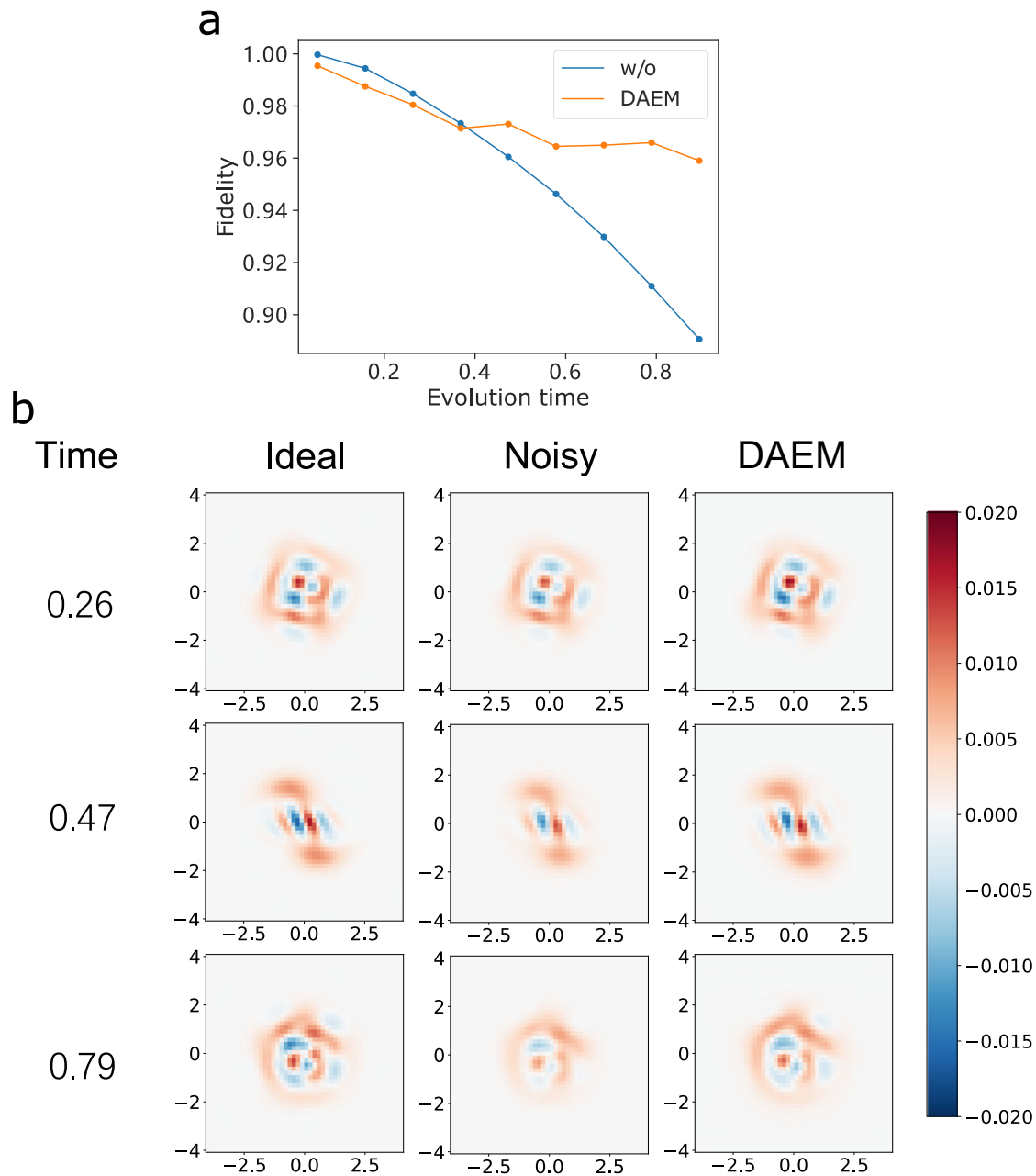
Fig. 8 | Error mitigation for the Kerr gate in a continuous variable system. a Fidelity values between the noisy/mitigated state and the ideal output state. **b** Snapshots of the point-wise measurement results of the state at different time points.

probability distribution obtained by passing through a softmax function, depending on the measurement statistics to be mitigated. The cost function for mitigating expectation values is L_2 loss, namely

$$\mathcal{L}(\boldsymbol{y}, \boldsymbol{y}_{\text{fid}}) = \frac{1}{n} \sum_{i=1}^{n} (y^{(i)} - y_{\text{fid}}^{(i)})^2, \tag{8}$$

where $\boldsymbol{y}$ denotes the output of the neural model and $\boldsymbol{y}_{\text{fid}}$ are the observable expectation values generated from data augmentation using the fiducial channel. The cost function for mitigating probability distribution is the average relative entropy[69], defined as

$$\mathcal{L}(\boldsymbol{p}, \boldsymbol{p}_{\text{fid}}) = \frac{1}{n} \sum_{i=1}^{n} \sum_{x} p^{(i)}(x) \log \left(\frac{p^{(i)}(x)}{p_{\text{fid}}^{(i)}(x)} \right), \tag{9}$$

where $\boldsymbol{p}$ are probability distribution predicted by our model, and $\boldsymbol{p}_{\text{fid}}$ are distribution obtained from fiducial channel by sampling 10000 shots.

To mitigate errors in continuous-variable processes, we adopt U-Net[70], a convolutional neural network originally designed for image segmentation[71], to be the neural model denoising the 2-dimensional Wigner function. U-Net possesses the strong ability to extract spatial features and construct 2-dimensional distributions. In this sense, it helps learn the distribution of the point-wise Wigner function. The inputs to the model are evolution time and Wigner functions corresponding to different photon loss rates. The output is a single 2-dimensional feature map, which represents the denoised Wigner function. To train the neural model, we use L_1 loss as cost function, defined as

$$\mathcal{L}(\boldsymbol{y}, \boldsymbol{y}_{\text{fid}}) = \frac{1}{n} \sum_{i=1}^{n} |y^{(i)} - y_{\text{fid}}^{(i)}|. \tag{10}$$

Fig. 9 | CSWAP gate decomposition. a CSWAP gate is decomposed into 3 Toffoli gates. **b** Further decomposition of Toffoli gate into single-qubit and CNOT gates. **c** The corresponding fiducial circuit for training.
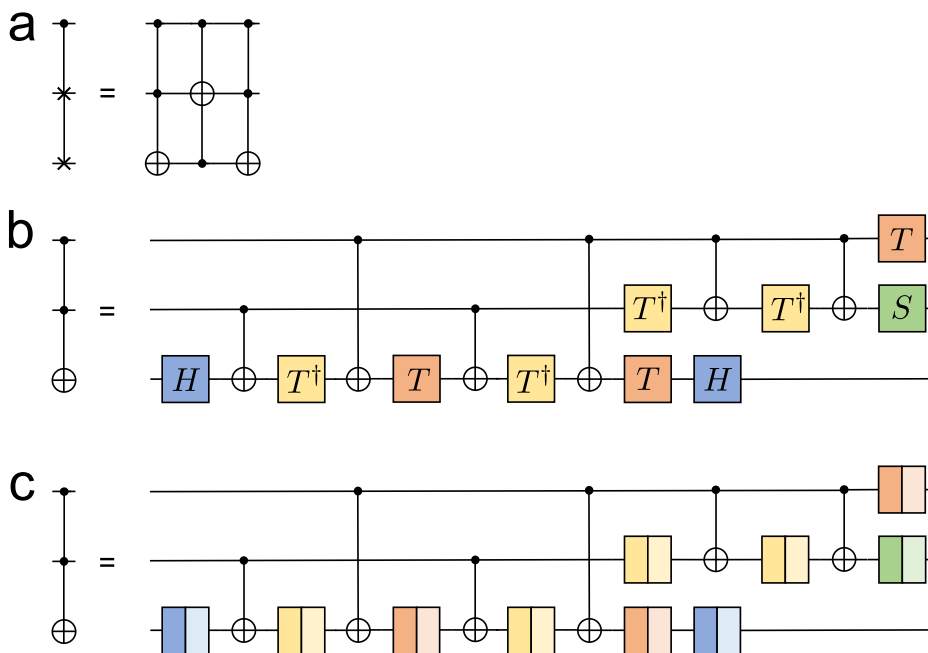
It encourages sparse output distribution, which conforms to the Wigner quasiprobability distribution of our target states. More details can be found in Supplementary Note 1.

Data augmentation strategy

In practice, the noise-free labels are not available unless we know the exact noiseless output states of the circuits. However, if the input and output states are the same, or under known transformation in the noise-free scenario, we can directly measure the input states to generate labels. Here, we introduce a fiducial process, i.e., $\mathcal{N}_\lambda(\mathcal{F})$, to achieve this goal. The process is trivially identity or contains only CNOT gates that can be absorbed into observables in the noise-free conditions, but share similar noise pattern as the target process in a noisy quantum device. To generate the training set for Noise-Awareness phase, we send input states through the fiducial process, measure the noisy outputs as data and measure the original input states as labels. Note that the input states can be either noisy or noiseless.

We detail the choice of input states and specification of $\mathcal{N}_\lambda(\mathcal{F})$ for different applications as follows.

For swap test circuits, the states to be compared are pure. We decompose the CSWAP gates into three Toffoli gates. The Toffoli gates are further transformed into CNOT and single qubit gates, as shown in Fig. 9.

In Noise-Awarenes phase, we sample 170 random pairs of pure states from the single-qubit Haar measure, in which 100 are used for training, 50 for validation, and 20 for Error-Mitigation phase. For ancilla, we choose random single-qubit mixed states. We construct the fiducial process by replacing every single-qubit gate G with identity gate, which is implemented by $\sqrt{G^\dagger}\sqrt{G}$, leaving the CNOT gates unchanged. This results in a channel $\mathcal{N}_0(\mathcal{F}) = U$, in which U describes the effects of all CNOT gates in the original circuit. Then we execute the fiducial process in noisy environment with varying noise parameters, and measure the noisy outputs using observable $M = Z_1$, which denotes the Pauli-Z observable on the ancilla qubit, and calculate the expectation values. Meanwhile, we measure the input states with observable $\tilde{M} = U^\dagger Z_1 U$. The measurement expectation values are the corresponding labels.

In VQEs, the augmentation strategy is generally the same as in Swap test. One difference is that, rather than pure states, we randomly sample 100

states as inputs in the Noise-Awareness phase. Whereas in the Error-Mitigation phase, the input states are chosen to be ground state $|0\rangle$.

For QAOA circuits, note that the distributions of output bitstrings possess symmetry, e.g., if 00011 is one solution, 11100 should also be a solution. To boost the performance of the neural model, we want to make the output distributions of the dataset in Noise-Awareness phase more aligned with those in the Error-Mitigation phase, i.e., the output distributions in the training set also possess this symmetry. It can be mathematically described as $X^{\otimes n}|\psi\rangle = |\psi\rangle$, where X is Pauli-X operator. This shows that the input states $|\psi\rangle$ are the eigenvectors of $X^{\otimes n}$ with eigenvalue 1. In our implementation, we sample 100 vectors from the eigenspace of $X^{\otimes n}$ with eigenvalue 1 as the input states. In the Noise-awareness phase, to generate a fiducial process, after replacing all single-qubit gates with identity gates, the CNOT gates automatically cancel each other. In this case, the fiducial process is trivially identity ideally, i.e., $\mathcal{N}_0(\mathcal{F}) = I$. Again, we execute the fiducial process on noisy quantum devices with different noise parameters to acquire noisy bitstring distributions. Besides, we sample input states $|\psi\rangle$ in computational basis to obtain labels for training.

For spin systems, the input states in both Noise-Awareness phase and Error-Mitigation phase are sampled from the same distribution. We have 100 different states for training in the first phase and 20 for testing in the second phase. The fiducial process is constructed by simply setting $H_{\text{Ising}} = I$.

For continuous-variable processes, to generate initial states, we first record intermediate states during noisy evolution of H_{kerr} in the time interval $t \in [0, 1]$, each with a timestep 0.05. With this procedure, we obtain 20 noisy states. Next, for each state, we evolve it with fiducial process $\mathcal{N}_0(\mathcal{F}) = I$ under different loss rates. The fiducial process is generated by H_{kerr} followed by $-H_{\text{kerr}}$, which can be simulated on hardware by ref. 72. The evolution times of the states are uniformly chosen in the range [0, 1] with a gap of 0.1.

After the completion of this work, another work[73] appeared where a data generation technique is employed to mitigate errors in the transverse field Ising model. This approach can be viewed as a specific means of obtaining the fiducial process.

Data availability

Datasets utilized during this study are available from the corresponding author upon reasonable request.

Code availability

The code supporting the findings of this study is available from the corresponding author upon reasonable request.

References

1. Cai, Z. et al. Quantum error mitigation. *Rev. Mod. Phys.* **95**, 045005 (2023).
2. Temme, K., Bravyi, S. & Gambetta, J. M. Error mitigation for short-depth quantum circuits. *Phys. Rev. Lett.* **119**, 180509 (2017).
3. Li, Y. & Benjamin, S. C. Efficient variational quantum simulator incorporating active error minimization. *Phys. Rev. X* **7**, 021050 (2017).
4. Kandala, A. et al. Error mitigation extends the computational reach of a noisy quantum processor. *Nature* **567**, 491–495 (2019).
5. Kim, Y. et al. Scalable error mitigation for noisy quantum circuits produces competitive expectation values. *Nat. Phys.* **19**, 752–759 (2023).
6. Kim, Y. et al. Evidence for the utility of quantum computing before fault tolerance. *Nature* **618**, 500–505 (2023).
7. Czarnik, P., Arrasmith, A., Coles, P. J. & Cincio, L. Error mitigation with Clifford quantum-circuit data. *Quantum* **5**, 592 (2021).
8. Czarnik, P., McKerns, M., Sornborger, A. T. & Cincio, L. Improving the efficiency of learning-based error mitigation, http://arxiv.org/abs/2204.07109 [quant-ph] (2022).
9. Strikis, A., Qin, D., Chen, Y., Benjamin, S. C. & Li, Y. Learning-based quantum error mitigation. *PRX Quantum* **2**, 040330 (2021).
10. Lowe, A. et al. Unified approach to data-driven quantum error mitigation. *Phys. Rev. Res.* **3**, 033098 (2021).
11. Endo, S., Benjamin, S. C. & Li, Y. Practical quantum error mitigation for near-future applications. *Phys. Rev. X* **8**, 031027 (2018).
12. Van Den Berg, E., Minev, Z. K., Kandala, A. & Temme, K. Probabilistic error cancellation with sparse Pauli–Lindblad models on noisy quantum processors. *Nat. Phys.* **19**, 1116–1121 (2023).
13. Huggins, W. J. et al. Virtual distillation for quantum error mitigation. *Phys. Rev. X* **11**, 041036 (2021).
14. Koczor, B. álint Exponential error suppression for near-term quantum devices. *Phys. Rev. X* **11**, 031057 (2021).
15. Seif, A., Cian, Ze-Pei, Zhou, S., Chen, S. & Jiang, L. Shadow distillation: quantum error mitigation with classical shadows for near-term quantum processors. *PRX Quantum* **4**, 010303 (2023).
16. Harper, R., Yu, W. & Flammia, S. T. Fast estimation of sparse quantum noise. *PRX Quantum* **2**, 010322 (2021).
17. Harper, R., Flammia, S. T. & Wallman, J. J. Efficient learning of quantum noise. *Nat. Phys.* **16**, 1184–1188 (2020).
18. Figueroa-Romero, P., Modi, K., Harris, R. J., Stace, T. M. & Hsieh, Min-Hsiu Randomized benchmarking for non-markovian noise. *PRX Quantum* **2**, 040351 (2021).
19. Torlai, G. et al. Neural-network quantum state tomography. *Nat. Phys.* **14**, 447–450 (2018).
20. Zhu, Y. et al. Flexible learning of quantum states with generative query neural networks. *Nat. Commun.* **13**, 6222 (2022).
21. Ahmed, S., Mu noz, CarlosS. ánchez, Nori, F. & Kockum, AntonFrisk Quantum state tomography with conditional generative adversarial networks. *Phys. Rev. Lett.* **127**, 140502 (2021).
22. Carrasquilla, J., Torlai, G., Melko, R. G. & Aolita, L. Reconstructing quantum states with generative models. *Nat. Mach. Intell.* **1**, 155–161 (2019).
23. Du, Y. et al. Shadownet for data-centric quantum system learning, *arXiv preprint arXiv:2308.11290* (2023).
24. Wu, YD. et al. Learning quantum properties from short-range correlations using multi-task networks. *Nat. Commun.* **15**, 8796 (2024).
25. Tang, Y., Xiong, H., Yang, N., Xiao, T. & Yan, J. Q-tape: A task-agnostic pre-trained approach for quantum properties estimation. In *Proc. Twelfth International Conference on Learning Representations* (2023).
26. Wu, Ya-Dong, Zhu, Y., Bai, G., Wang, Y. & Chiribella, G. Quantum similarity testing with convolutional neural networks. *Phys. Rev. Lett.* **130**, 210601 (2023).
27. Qian, Y., Du, Y., He, Z., Hsieh, M. & Tao, D. Multimodal deep representation learning for quantum cross-platform verification. *Phys. Rev. Lett.* **133**, 130601 (2024).
28. Sharir, O., Levine, Y., Wies, N., Carleo, G. & Shashua, A. Deep autoregressive models for the efficient variational simulation of many-body quantum systems. *Phys. Rev. Lett.* **124**, 020503 (2020).
29. Bennewitz, E. R., Hopfmueller, F., Kulchytskyy, B., Carrasquilla, J. & Ronagh, P. Neural error mitigation of near-term quantum simulations. *Nat. Mach. Intell.* **4**, 618–624 (2022).
30. Zhang, X. et al. Direct fidelity estimation of quantum states using machine learning. *Phys. Rev. Lett.* **127**, 130503 (2021).
31. Qin, H. et al. Experimental direct quantum fidelity learning via a data-driven approach. *Phys. Rev. Lett.* **132**, 190801 (2024).
32. Vadali, A., Kshirsagar, R., Shyamsundar, P. & Perdue, G. N. Quantum circuit fidelity estimation using machine learning. *Quantum Mach. Intell.* **6**, 1 (2024).
33. Kim, C., Park, KyungdeockDaniel & Rhee, June-Koo Quantum error mitigation with artificial neural network. *IEEE Access* **8**, 188853–188860 (2020).
34. Liao, H. et al. Machine learning for practical quantum error mitigation. *Nat. Mach. Intell.* **6**, 1478–1486 (2024).
35. Zhukov, A. & Pogosov, W. Quantum error reduction with deep neural network applied at the post-processing stage. *Quantum Inf. Process.* **21**, 93 (2022).
36. Sack, S. H. & Egger, D. J. Large-scale quantum approximate optimization on nonplanar graphs with machine learning noise mitigation. *Phys. Rev. Res.* **6**, 013223 (2024).
37. Shorten, C. & Khoshgoftaar, T. M. A survey on image data augmentation for deep learning. *J. Big Data* **6**, 1–48 (2019).
38. Taylor, L. & Nitschke, G. Improving deep learning with generic data augmentation. In *Proc. IEEE Symposium Series on Computational Intelligence (SSCI)* 1542–1547 (IEEE, 2018).
39. Wang, Y., Yao, Q., Kwok, J. T. & Ni, L. M. Generalizing from a few examples: a survey on few-shot learning. *ACM Comput. Surv. (csur)* **53**, 1–34 (2020).
40. Wang, W., Zheng, V. W., Yu, H. & Miao, C. A survey of zero-shot learning: settings, methods, and applications. *ACM Trans. Intell. Syst. Technol. (TIST)* **10**, 1–37 (2019).
41. Peruzzo, A. et al. A variational eigenvalue solver on a photonic quantum processor. *Nat. Commun.* **5**, 4213 (2014).
42. Farhi, E., Goldstone, J. and Gutmann, S. A quantum approximate optimization algorithm, *arXiv preprint arXiv:1411.4028* (2014).
43. Dykman, M. *Fluctuating Nonlinear Oscillators: from Nanomechanics to Quantum Superconducting Circuits* (Oxford University Press, 2012).
44. Buhrman, H., Cleve, R., Watrous, J. & De Wolf, R. Quantum fingerprinting. *Phys. Rev. Lett.* **87**, 167902 (2001).
45. White, G. A. L., Hill, C. D., Pollock, F. A., Hollenberg, L. C. L. & Modi, K. Demonstration of non-Markovian process characterisation and control on a quantum processor. *Nat. Commun.* **11**, 6301 (2020).
46. Agarwal, A., Lindoy, L. P., Lall, D., Jamet, F. & Rungger, I. Modelling non-markovian noise in driven superconducting qubits. *Quantum Sci. Technol.* **9**, 035017 (2024).
47. Groszkowski, P., Seif, A., Koch, J. & Clerk, A. A. Simple master equations for describing driven systems subject to classical non-markovian noise. *Quantum* **7**, 972 (2023).
48. Breuer, H.-P. & Petruccione, F. https://doi.org/10.1093/acprof:oso/9780199213900.001.0001 *The Theory of Open Quantum Systems* (Oxford University Press, 2007).

49. Magazzù, L. et al. Probing the strongly driven spin-boson model in a superconducting quantum circuit. *Nat. Commun.* **9**, 1403 (2018).

50. https://originqc.com.cn/en/ Origin quantum.

51. Giurgica-Tiron, T., Hindy, Y., LaRose, R., Mari, A. & Zeng, W. J. Digital zero noise extrapolation for quantum error mitigation. In *Proc. IEEE International Conference on Quantum Computing and Engineering (QCE)* 306–316 (IEEE, 2020).

52. Zhou, L., Wang, Sheng-Tao, Choi, S., Pichler, H. & Lukin, M. D. Quantum approximate optimization algorithm: Performance, mechanism, and implementation on near-term devices. *Phys. Rev. X* **10**, 021067 (2020).

53. Guerreschi, GianGiacomo & Matsuura, A. Y. Qaoa for max-cut requires hundreds of qubits for quantum speed-up. *Sci. Rep.* **9**, 6903 (2019).

54. White, S. R. Density matrix formulation for quantum renormalization groups. *Phys. Rev. Lett.* **69**, 2863 (1992).

55. White, S. R. & Feiguin, A. E. Real-time evolution using the density matrix renormalization group. *Phys. Rev. Lett.* **93**, 076401 (2004).

56. Daley, AndrewJohn, Kollath, C., Schollwöck, U. & Vidal, Guifré Time-dependent density-matrix renormalization-group using adaptive effective Hilbert spaces. *J. Stat. Mech. Theory Exp.* **2004**, P04005 (2004).

57. Jouguet, P., Kunz-Jacques, S. ébastien, Leverrier, A., Grangier, P. & Diamanti, E. Experimental demonstration of long-distance continuous-variable quantum key distribution. *Nat. Photonics* **7**, 378–381 (2013).

58. Lloyd, S. & Braunstein, S. L. Quantum computation over continuous variables. *Phys. Rev. Lett.* **82**, 1784 (1999).

59. Lutterbach, L. G. & Davidovich, L. Method for direct measurement of the Wigner function in cavity QED and ion traps. *Phys. Rev. Lett.* **78**, 2547–2550 (1997).

60. Bertet, P. et al. Direct measurement of the Wigner function of a one-photon fock state in a cavity. *Phys. Rev. Lett.* **89**, 200402 (2002).

61. Giovannetti, V., Lloyd, S. & Maccone, L. Quantum metrology. *Phys. Rev. Lett.* **96**, 010401 (2006).

62. Ash-Saki, A., Alam, M. & Ghosh, S. Experimental characterization, modeling, and analysis of crosstalk in a quantum computer. *IEEE Trans. Quantum Eng.* **1**, 1–6 (2020).

63. Sarovar, M. et al. Detecting crosstalk errors in quantum information processors. *Quantum* **4**, 321 (2020).

64. Murali, P., Mckay, D. C., Martonosi, M. & Javadi-Abhari, A. Software mitigation of crosstalk on noisy intermediate-scale quantum computers. In *Proc. Twenty-Fifth International Conference on Architectural Support for Programming Languages and Operating Systems*, ASPLOS '20 1001–1016 (Association for Computing Machinery, 2020). https://doi.org/10.1145/3373376.3378477.

65. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning*, Springer Series in Statistics (Springer New York Inc., 2001).

66. Haykin, S. *Neural Networks: a Comprehensive Foundation* (Prentice Hall PTR, 1994).

67. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition, http://arxiv.org/abs/1512.03385 [cs.CV](2015).

68. Vaswani, A. et al. Attention is all you need. In *Proc. 31st International Conference on Neural Information Processing Systems*, NIPS'17. 6000–6010 (Curran Associates Inc., 2017).

69. Cover, T. M. & Thomas, J. A. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)* (Wiley-Interscience, 2006).

70. Ronneberger, O., Fischer, P. & Brox, T. U-net: convolutional networks for biomedical image segmentation, http://arxiv.org/abs/1505.04597 [cs.CV] (2015).

71. Minaee, S. et al. Image segmentation using deep learning: a survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**, 3523–3542 (2022).

72. Rossini, M., Maile, D., Ankerhold, J. & Donvil, BrechtI. C. Single-qubit error mitigation by simulating non-Markovian dynamics. *Phys. Rev. Lett.* **131**, 110603 (2023).

73. Babukhin, D. V. Echo-evolution data generation for quantum error mitigation via neural networks. *Quantum Inf. Process.* **23**, 1–25 (2024).

Acknowledgements

Author contributions

M.L., Y.Z., and Y.Y. established the key idea in the paper. M.L. developed the neural network model and did the numerical experiments. Y.Z. wrote the draft paper. All the authors contributed to the preparation of the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41534-025-00960-y.

Correspondence and requests for materials should be addressed to Yuxiang Yang.

Reprints and permissions information is available at http://www.nature.com/reprints

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