Quantum error mitigation (QEM) is crucial for obtaining reliable results on quantum computers by suppressing quantum noise with moderate resources. It is a key factor for successful and practical quantum algorithm implementations in the noisy intermediate scale quantum (NISQ) era. Since quantum-classical hybrid algorithms can be executed with moderate and noisy quantum resources, combining QEM with quantum-classical hybrid schemes is one of the most promising directions toward practical quantum advantages. This work shows how the variational quantum-neural hybrid eigensolver (VQNHE) algorithm, which seamlessly combines the expressive power of a parameterized quantum circuit with a neural network, is inherently noise resilient with a unique QEM capacity, which is absent in vanilla variational quantum eigensolvers (VQE). The study carefully analyzes and elucidates the asymptotic scaling of this unique QEM capacity in VQNHE from both theoretical and experimental perspectives. Finally, a variational basis transformation is proposed for the Hamiltonian to be measured under the VQNHE framework, yielding a powerful tri-optimization setup, dubbed as VQNHE++. VQNHE++ can further enhance the quantum-neural hybrid expressive power and error mitigation capacity.

1. Introduction

Variational quantum algorithms (VQA)[1–3] are under active investigation as they require moderate quantum hardware resources and are promising candidates to deliver practical quantum advantage[4,5] in the NISQ era.[6] VQE is one of the most representative VQAs where the ground state is approximated by variational optimization[7–13] with parameterized quantum circuits (PQC). Quantum error mitigation, as an NISQ alternative for full-fledged quantum error correction, is believed to alleviate the negative effects brought by quantum noise and deliver more reliable results for VQAs. There are already various proposals for QEM techniques[14–27] and specifically some of the proposals are based on the principle of variational optimizations[28–37]. However, the interplay in terms of variational optimization between VQA and QEM remains largely elusive so far. To pave the way toward more practical quantum advantages, it is natural and urgent to investigate the interplay between VQAs and QEM as well as design VQA-native QEM techniques or QEM baked-in VQAs.

Variational quantum-neural hybrid eigensolver (VQNHE) is a powerful VQA approach incorporating the strength of a neural network as a nonunitary post-processing module efficiently.[38] Recently, the idea of adding a non-unitary post-processing module to the variational quantum eigensolver[7–12] has become popular. However, unlike all previous proposals, VQNHE not only enhances the expressive power of the VQAs but also entails just a polynomial scaling of computational resource overhead. For instance, while a previous proposal based on the Jastrow factor[39] could enhance VQE,[40,41] it requires an exponential scaling of resources overhead for the general form of Jastrow factor. In this work, we reveal another important and unique property of VQNHE: intrinsic quantum noise resilience. Through detailed analysis, we demonstrate that the quantum noise resilience is from the introduction of the classical post-processing module and this QEM capacity is absent in the plain VQE. By utilizing the simple idea of adaptive retraining directly on noisy hardware, we obtain much more reliable energy estimations in the presence of quantum noise. In addition, by combining the transformed Hamiltonian approach in the VQNHE++ framework as shown in Figure 1, we further improve the expressive power and the noise resilience of the variational quantum-neural hybrid scheme, resulting in a more efficient and reliable approach for quantum simulation on noisy quantum hardware.

2. VQNHE Setup

We first recapitulate the essence of VQNHE and then elaborate on “adaptive retraining,” a QEM protocol built on top of VQNHE in the following sections. VQNHE is an interesting example of quantum-classical hybrid schemes: it not only requires an outer classical optimizer loop but also features a classical neural network to provide the post-processing enhancement. The aim of VQNHE is the same as VQE, that is to find the ground state of a given Hamiltonian $H$ (without loss of generality, we assume $H$ is a Pauli string below). To approximate such a ground state, we do...
not directly rely on the output state of a parameterized quantum circuit (PQC) $U$ as $|\psi\rangle = U|0\rangle$. Instead, we post-process the output of a PQC with a classical neural network to attain $|\psi_f\rangle = f|\psi\rangle$. Here $f = \sum f(s)|s\rangle\langle s|$ where $f$ is a neural network with trainable weights or any general parameterized function and $s$ is a computational basis in the form of the bitstring. We note that the function $f$ can generally induce a nonunitary transformation on the quantum state. Essentially, via VQNHE, we can apply an arbitrary $2^n \times 2^n$ diagonal matrix on the output quantum state of the PQC. Previously, it was widely believed that the experimental implementation for accurate estimations on the energy $\langle \psi_f | H | \psi_f \rangle / \langle \psi_f | \psi_f \rangle$ requires exponential time. However, as explicated in Ref. [38], this energy estimation can be accurately and efficiently obtained with only a polynomial scaling of hardware resources.

We now describe the experimental protocol for measuring the Hamiltonian expectation with the classical post-processing scheme $f$ on the output of PQC $U$. Without loss of generality, we only show how to measure the expectation for a Pauli string $H$, as the expectation for a general Hamiltonian can be decomposed into a weighted sum of a polynomial number of different Pauli strings in most realistic cases. We define the expectation value for the Pauli string $H$ with classical post-processing $f$ as:

$$\langle \tilde{f} H \rangle_{\psi_f} = \frac{\langle \psi_f | H | \psi_f \rangle}{\langle \psi_f | \psi_f \rangle}$$

(1)

The Pauli string is expressed as $H = \prod_{i \in V} H_i$, where $H_i$ correspond to local Pauli operator I, X, Y, or Z. We denoted the set $i / Z = \{i | H_i = Z\}$, namely, the qubit indices where $H$ hosts Z operator.

If the Pauli string contains no X or Y operator, the energy estimation is straightforward and is given as:

$$\langle \tilde{f} H \rangle_{\psi_f} = \frac{\sum_{s \in U} f(s)^2 \prod_{i \in Z}(1 - 2s_i)}{\sum_{s \in U} f(s)^2}$$

(2)

where $s \in U$ denotes the results collected on the computational basis of circuit $U$, i.e. $s$ is the measurement bitstring results for $U$ circuit.

If the Pauli string contains X or Y operator, we call the first qubit that hosts X or Y operator in the Pauli string as the sign qubit and relabel the qubit as qubit 0 below for notation convenience. We build a measurement circuit block $V$ which is attached after the PQC $U$. The building rule for $V$ is: 1) We apply a control-X gate with control on the sign qubit and target for each qubit $i / X$. We also apply a control-Y gate with control on the sign qubit and target for each qubit $i / Y$. 2) We measure the sign qubit in X or Y direction depending on the operator type on the sign qubit. In other words, we apply an H gate or $e^{-i\pi/4}X$ gate on sign qubit in the circuit $V$. The energy can be estimated in an unbiased and efficient manner as:

$$\langle \tilde{f} H \rangle_{\psi_f} = \frac{\sum_{s \in UV} f(s)^2 \prod_{i \in Z}(1 - 2s_i) f(0_{1:s-1}) f(1_{s-1:s-1})}{\sum_{s \in U} f(s)^2}$$

(3)

where the bitstring $s$ in the denominator is drawn from the PQC $U$ and bitstring $s$ in the numerator is drawn from the PQC with the measurement circuit $V$ appended. $s$ is for bitstring with bitflip on $s$ on qubit indices $i / X$ and $i / Y$. $(|s\rangle \propto H(s)), 1_{s-1:s-1}$ implies that for each bitstring $s$ collected from the experiments, we set the first bit as 1 and flip the following bits if the Pauli string has an X or Y operator on the corresponding position.

In the above, we introduce the scalable protocol on expectation evaluation in VQNHE. To train the model, we need to evaluate the gradient for both neural network and variational circuit parameters if the gradient-based optimizer is adopted. In terms of the circuit parameters, the conventional parameter shift rule still applies since we can regard the process as a plain VQE with the Hamiltonian to be evaluated as $f^+ H f^+ / \langle f^+ f | f | \psi \rangle$. When the measurement results in the form of a collection of bitstrings are fixed, the energy evaluation function as indicated by Equation (3) is a purely classical function with neural parameters, whose gradients can be evaluated via automatic differentiation (back-propagation) method numerically.

In summary, VQNHE jointly optimizes the parameters in the PQC $U$ and the classical post-processing module $f$. As an approach combining the advantages from both VQE and neural variational Monte Carlo (VMC)\textsuperscript{[42-48]} this new setup offers a state-of-the-art approximation on the ground state energy for various quantum spin systems and quantum molecules with a provable bound on the efficiency for the computational complexity.\textsuperscript{[38]}

### 3. Retraining Energy Gain as a Measure for QEM Capacity

We investigate the VQNHE performance on both noisy quantum simulators and real quantum hardware. The quantum noise deteriorates the accuracy of the energy estimation and thus compromises the superior performance that could be attained in an ideal situation, such as a noise-free simulation. Interestingly, we find that VQNHE exhibits inherent noise resilience to a certain extent. Namely, when training the VQNHE in a noisy environment, the neural network can adjust its weights, implicitly mitigating noise-induced disruptions. We term the optimization on
noisy hardware the “adaptive retraining.” The QEM capacity of VQNE can be measured by the difference of energy estimations \( \delta E = E_\theta - E_\phi \), where \( E_\theta \) is the energy estimation with neural weights \( \phi \), and \( E_\phi = \phi(\theta_0) \) is the optimized weights with(out) the presence of quantum noise. Apart from the neural retraining, we can also investigate the energy gain when retraining the PQC or jointly retraining the PQC and the neural network. The energy gains are defined as \( \delta E = E_\theta - E_\phi \) and \( \delta E = E_\theta - E_\phi - \delta_\theta \), respectively, where \( \theta_0 \) is the set of optimized parameters in the PQC trained with(out) noise and both energies are all evaluated in the noisy setting. The more negative energy gain indicates better noise resilience as it reflects the amount of energy that is further lowered via retraining from ideal parameters in noisy devices.

It is worth noting that the so-called retraining can start from any weight initialization, especially in the joint retraining case. The noiseless optimal parameters are only used to define the retraining energy gain theoretically, and are not necessary for the practical QEM. We use this metric instead of the absolute energy to characterize the QEM capacity since we need to separate the contribution of error mitigation from the expressive power enhancement. As we discussed below, the energy gain metric also shows some nice scaling behaviors that can be explained from theoretical understanding.

4. Biased Retraining on the Classical Module

In real experiments, the energy is estimated from a collection of measured bitstrings with finite sampling errors. The number of measurement shots required is often large, especially when near optimum or due to the vanishing gradients. Therefore, it is expensive to run full unbiased retraining. To this end, we propose a very cost-efficient alternative, i.e., biased retraining, which only retrain the neural network part. In the biased retraining, instead of executing the PQC at each training epoch, we fix the bitstring measurement results during the retraining. Since the bitstring results are fixed (with just a finite number of shots), they are biased with measurement uncertainty. As a result, the obtained biased retraining energy gain has two components: the intrinsic QEM and overfitting to the biased measurements. With IBM device-compatible noisy simulations and real IBM hardware experiments, we demonstrate that the average QEM capacity scaling for the biased neural retraining is \( \delta E \propto B + A/M \), where \( M \) is the number of fixed measurement shots and the constant \( B \) stands for intrinsic QEM capacity. The intrinsic QEM part remains when the number of measurement shots is taken to infinity \( M \rightarrow \infty \) where the training bias induced by the finite shot noise vanishes. Figure 2 shows the energy gain results from biased retraining under different noise models including the real hardware experiment, and all results conform to the scaling relation. (See Section S2, Supporting Information, for the data and experiment setup details for the Hamiltonian, circuit ansatz, neural network structures, etc.)

5. QEM Capacity Scaling with the Noise Strength

To investigate the energy gain with a tunable noise strength, we utilize a simple depolarizing error model, where an isotropic depolarization of strength \( p \) is attached after each two-qubit gate.

The 1D five-site transverse field Ising model (TFIM) with an open boundary condition is then utilized as the VQNE target Hamiltonian: \( \hat{H} = \sum_{i=1}^{5} Z_i Z_{i+1} - \sum_{i=1}^{5} X_i \). And the numerical simulation is implemented using TensorCircuit. We study the scaling relation between the energy gain due to the intrinsic QEM and the effective overall noise strength \( p_{\text{eff}} \) of the depolarization. The overall depolarizing probability \( p_{\text{eff}} \) is measured by the energy ratio from the PQC output: \( 1 - p_{\text{eff}} = E_{\text{noise}} / E_{\text{noiseless}} \).

First, retraining solely on the quantum part cannot improve the final energy estimation. With only quantum retraining, the optimal energy estimation in the noisy case is always \((1 - p_{\text{eff}}) E_{\text{noiseless}}\). This fact implies that the adaptive retraining QEM procedure is unique to the pipelines with classical post-processing and plain VQE is not quantum noise resilient in the sense of adaptive retraining.

For depolarizing strength \( p = 0.005, 0.01, 0.015, 0.02 \), the effective overall error strength is correspondingly \( p_{\text{eff}} = 0.017, 0.033, 0.049, 0.065 \) with our circuit ansatz with optimal circuit weights and the intrinsic QEM energy gain from neural retraining are \( \delta E = -0.0031, -0.012, -0.026, -0.044 \), respectively. The scaling relation for the QEM energy gain with neural retraining is thus given by \( \delta E \propto p_{\text{eff}}^2 \).

The intrinsic QEM capacity with neural retraining scales quadratically with \( p_{\text{eff}} \), which is the reason behind the biased neural retraining scaling with the number of measurements we observed before. Note that the effective noise strength \( p_{\text{eff}} \) can only be approximately estimated in experiments from finite measurement shots. We take \( p_{\text{eff}} \) as a random variable and the energy gain is \( \delta E = \langle p_{\text{eff}}^2 \rangle = \langle p_{\text{eff}} \rangle^2 + \Delta p_{\text{eff}} \), where \( \Delta p_{\text{eff}} \propto 1/M \) is the square deviation of the estimation on \( p_{\text{eff}} \) due to finite measurement shots. Therefore, the energy gain after retraining follows a simple scaling form of \( A/M + B \).

In addition, we evaluate the QEM energy gain from joint retraining, with \( \delta E = -0.044, -0.080, -0.12, -0.15 \) for effective error strengths listed above. Therefore, the scaling of QEM capacity with the quantum noise strength from joint retraining is linear instead of quadratic: \( \delta E \propto p_{\text{eff}} \). The different scaling forms are sketched in Figure 3 for both neural retraining and joint retraining. (See more rigorous fitting in log–log scale in Section S4, Supporting Information.)
6. General Picture for The QEM Scaling

To understand the above QEM scaling relations, we first investigate a minimal model involving just one qubit that permits direct analytical analysis. The results are consistent with the observed scaling relation (see Section S1, Supporting Information, for details).

We now discuss the theoretical mechanism behind different QEM capacity scalings. Suppose that the ideal output of a PQC is the exact ground state as $\rho_0 = |\psi_0\rangle\langle \psi_0|$. And the mixed state from the PQC in the presence of depolarizing noise of strength $p$ is $\rho$. The post-processing module is a nonunitary transformation $\hat{f}$ with non-zero elements only appearing in the diagonal. The energy gain with neural retraining is thus $\delta E = E_{\text{QEM}} - E_N$. Here $E_N = (1-p)E_0$ where $E_0 = \text{Tr}(\rho_0H)$ is for the exact ground state. We expand the energy terms as $E_{\text{QEM}} = E_{\text{QEM}}^{(0)} + E_{\text{QEM}}^{(1)}p + E_{\text{QEM}}^{(2)}p^2 + \cdots$ and keep up to the first order of $p$, namely, as long as we have shown that the zeroth and first order of $p$ in the energy gain is zero, the energy gain scaling is at most $p^2$.

Under depolarizing channel $\rho$, we have:

$$E_{\text{QEM}} = \frac{\text{Tr}(\hat{f}\rho_0\hat{H})(1-p) + \text{Tr}(\hat{f}\hat{H})p/2^n}{p + \text{Tr}(\hat{f}\rho_0\hat{H})(1-p)}$$

(4)

The optimized neural module $\hat{f} = I$ when $p = 0$. We assume the optimized $\hat{f} \approx I + pf_j$ to the first order, where $f_j$ is a constant matrix.

The zeroth order of $p$ in the energy gain is trivially zero: $E_{\text{QEM}}^{(0)} = E_0$. Now consider the first order of $p$, and we have $E_{\text{QEM}}^{(1)} = -E_0$ and

$$E_{\text{QEM}}^{(1)} = \text{Tr}(f_1 \rho_0 H) - E_0 \text{Tr}(f_1 \rho_0)$$

$$+ \text{Tr}(\rho_0 f_1 H) - E_0 \text{Tr}(\rho_0 f_1) - E_0$$

(5)

Note that

$$\text{Tr}(f_1 \rho_0 H) - E_0 \text{Tr}(f_1 \rho_0) = \langle \psi_0 | H f_1 | \psi_0 \rangle - E_0 \langle \psi_0 | f_1 | \psi_0 \rangle = 0$$

(6)

thus we have $E_{\text{QEM}}^{(1)} = -E_0$ independent of $f_j$. Therefore, the first order energy gain vanishes $\delta E_{\text{QEM}}^{(1)} = 0$ as well, indicating that the energy gain scales at most quadratically with the noise strength $p$. To summarize, the first order of $p$ in the energy gain is canceled as long as the retraining trajectory can be understood from a simple perturbation, i.e., the optimized post-processing module $f$ is analytically connected to the ideal one $I$ as the noise $p \rightarrow 0$.

To explain why a linear gain emerges in the joint retraining, we note that the perturbative picture fails under the joint-training scenario. As long as we allow joint training, there are infinitely many optimal solutions, constituted by appropriate combinations of PQC and neural-network weights to essentially yield the same output state in the noiseless case. We are no longer restricted to the unique solution as in the neural retraining case, where the PQC generates the exact ground state with an identity neural network (NN). Instead, even when the PQC generates other quantum states than the true ground state, an appropriate post-processing neural module $f$ can still post-process to the ground truth. Therefore, in the ideal case $p = 0$, we have infinitely many combinations of PQC states and neural solutions that would collaboratively lead to the correct groundstate energy. When noise $p > 0$ is introduced, the responses to quantum noise are different and the energy degeneracy (of many possible combinations of PQC and NN setups) is broken. Therefore, the optimal solution in the weak noise case is not connected to the identity one $f = I$ in the joint retraining case. In other words, the optimized $f$ cannot be simply described by $f = I + pf_j$, where the derivation based on the perturbation picture fails and the first order energy gain emerges.

In summary, the energy gains in neural retraining and joint retraining come from different sources and they can be understood using clear and unified physical pictures. The neural retraining perturbatively improves the noisy energy estimation by smoothly shifting the classical module $f$ away from identity $I$. On the contrary, the joint retraining improves energy estimation by breaking the degeneracy of infinitely many possible combinations of PQC and neural setups and selecting the most error-resilient one.

7. VQNHE++: Tri-Optimization with Parameterized Transformed Hamiltonian

VQNHE is a bi-optimization setup, where both parameters $\theta$ in the PQC and parameters $\phi$ in the neural network need to be optimized. The post-processing function $f$ can greatly alter the output states by the PQC. However, $f$ is effectively a diagonal matrix, which certainly cannot represent a universal quantum operation. Therefore, such retraining of the neural post-processing module can only partially mitigate the quantum noise effects. Since a universal nonunitary quantum operation is NP hard to implement in terms of quantum resources, we instead introduce a parameterized gauge Hamiltonian approach to enhance the mitigating power of the post-processing quantum channel $\hat{f}$.

Suppose that $\hat{W}$ is a unitary transformation, then the transformed Hamiltonian $\hat{H}' = \hat{W}^\dagger \hat{H} \hat{W}$ shares the identical spectrum with $\hat{H}$, and thus the ground state energy is the same as that of $\hat{H}$. Therefore, we can utilize VQNHE to simulate the ground state of the transformed Hamiltonian by identifying the Pauli strings in the newly transformed Hamiltonian as observables. To efficiently
implement this idea, we require that the number of Pauli strings in $\hat{H}^\perp$ scales polynomially with the system size $n$. This requirement restricts the possible forms of gauge transformation for $\hat{W}$. For local Hamiltonian such as spin-1 spin models, $\hat{W}$ can be in the form of single-qubit rotation gates $\hat{W} = \prod_i \exp(\text{i} \hat{P}_i \hat{P}_i)$, where $\hat{P}_i$ is the Pauli gate $X$, $Y$, or $Z$. Some special forms or structures of parameterized tensor networks can also play the role of the Hamiltonian transformation with better expressiveness and controllable overhead. (See more details on gauge transformation design and analysis in Section S6, Supporting Information.)

The experimental protocol for VQNHE++ is a straightforward combination of the protocol for VQNHE as explained before and the experimental protocol for the transformed Hamiltonian approach. Namely, we classically track the parameterized transformed Hamiltonian since there are only polynomial Pauli string terms after the transformation that can be efficiently obtained and manipulated on classical computers via simple Pauli matrix commutation algebra. We then can apply VQNHE framework on the transformed Hamiltonian instead of the original Hamiltonian (e.g., Equation (8) for a TFIM transformed Hamiltonian). Similarly, with fixed measurement results, the energy evaluation forward pass is a pure classical function with neural weights and transformation parameters. Therefore, the gradients for both types of parameters can be obtained efficiently by automatic differentiation. From a higher level perspective, we apply the diagonal post-processing on the output quantum state from the PQC using VQNHE protocol as a Schrödinger picture operation and apply the parameterized gauged transformation on the target Hamiltonian as a Heisenberg picture operation. VQNHE++ framework is scalable as it still maintains the polynomial computational complexity.

From a theoretical perspective, we now have the energy estimation as:

$$\langle \hat{H} \rangle = \frac{\text{Tr}\left(\hat{J}_0 \rho_{\hat{J}_0} \hat{W}_c \hat{H} \hat{W}_c^\dagger \right)}{\text{Tr}\left(\hat{J}_0 \rho_{\hat{J}_0} \hat{W}_c \right)}$$

$$= \frac{\text{Tr}\left(\left(\hat{W} \hat{J}_0 \rho_{\hat{J}_0} \hat{W}_c \right) \hat{H} \hat{W}_c^\dagger \right)}{\text{Tr}\left(\hat{J}_0 \rho_{\hat{J}_0} \hat{W}_c \right)}$$

Therefore, the transformed Hamiltonian setup with VQNHE essentially gives a more powerful variational postprocessing channel than the plain diagonal matrix $\hat{f}$. The new effective postprocessing operation is $\hat{W} \hat{f} \hat{W}_c$, which has non-vanishing off-diagonal contributions. The enhanced postprocessing capacity implies better performance on ground state energy optimization and quantum error mitigation as the freedom in the new ansatz is strictly larger than VQNHE. An intuitive limit is by considering $\hat{W}$ as a diagonal transformation for the original Hamiltonian $\hat{H}$, i.e., the transformed Hamiltonian $\hat{H}^\perp$ is a diagonal matrix. In that case, we can train a diagonal $\hat{f}$ to successfully project any PQC output $\rho$ to the exact ground state and thus free from any quantum noise.

In the plain VQE, the transformed Hamiltonian operation can be directly implemented on the circuit instead of tracking the new transformed Hamiltonian virtually (say for local Ry transformation, we directly apply one layer of parameterized Ry gate at the end of the PQC). Nevertheless, there is still a subtle difference between the direct implementation of the transformation on the circuit (Schrödinger picture) and the transformed Hamiltonian tracked classically (Heisenberg picture); the latter is free from quantum noise for the transformation part. In the VQNHE setup, such parameterized transformation cannot be implemented on the circuit as there is an uncommutable neural diagonal module in between. The operation order in VQNHE++ is: PQC $U_\theta$ + neural postprocessing $f_\theta$ + gauge transformation $\hat{W}_c$. If we naively implement the gauge transformation on the circuit, the order is instead PQC $U_\theta$ + gauge transformation $\hat{W}_c$ + neural postprocessing $f_\theta$. The two orders give totally different effective operations even without quantum noise as they are uncommuting with each other, and the latter is more trivial as the transformation can be absorbed into the circuit in the noiseless limit.

We consider the five-qubit TFIM as a specific example to demonstrate the workflow and illustrate the benefits of the transformed Hamiltonian approach. The model is aligned with the one we utilized in VQNHE experiments and is compatible with public IBM hardware devices. We take the gauge transformation $\hat{W} = \prod_i \exp(\text{i} \hat{Y}_i \hat{Y}_i)$, and the corresponding transformed Hamiltonian is

$$\hat{H}_c^\perp = \sum_i (\cos 2\tau \cos 2\tau_i \hat{Z}_i \hat{Z}_{i+1} + \sin 2\tau \sin 2\tau_i \hat{X}_i \hat{X}_{i+1})$$

$$- \sin 2\tau \cos 2\tau_i \hat{X}_i \hat{X}_{i+1} - \cos 2\tau \sin 2\tau_i \hat{Z}_i \hat{Z}_{i+1} - \cos 2\tau \hat{X}_i - \sin 2\tau \hat{Z}_i$$

(8)

which contains a polynomial number of Pauli string terms. We utilize the PQC ansatz of a layered form $[\hat{H}, \hat{Z}\hat{Z}(\theta_1), \hat{X}\hat{X}(\theta_2), \hat{X}\hat{X}(\theta_3)]$ (See SM Section S5 for circuit ansatz representation notation).

With the introduction of the transformed Hamiltonian on top of the VQNHE setup, we are now equipped with more options for noisy adaptive retraining. We run adaptive retraining for all combinations of quantum module, neural module and transformation module. The absolute energy estimated after each kind of retraining is displayed in Figure 4. It is worth noting that the line for the quantum-only retraining also nearly coincides with the line of no retraining and retraining on quantum and transformation parts (not shown), since the quantum part itself cannot be tuned to minimize the depolarizing error as we mentioned before, and the transformation part can be absorbed into the last layer of Ry in the PQC trivially when the neural module is fixed to identity.

The most crucial insights from Figure 4 are that the n+t retraining delivers a similar QEM performance as the joint retraining (q+n), and that the QEM capacity for the n-t retraining is even stronger than the conventional joint retraining (q+n) when the overall noise strength is high. Since the PQC is fixed in the n+t retraining scheme, we can carry out the fast biased retraining based on the same set of measurement results from the PQC, similar to the biased neural retraining (n) case we discussed before. Therefore, the classically tractable biased retraining combining the neural postprocessing and parameterized Hamiltonian transformation can achieve competitive QEM results as joint retraining but avoid issues such as finite sampling errors or quantum gradient vanishing (barren plateau issue). We also report the transformed Hamiltonian approach on the Heisenberg model with various quantum noise models and obtain better error mitigation results. Figure 5 shows the results...
of different retraining strategies for Heisenberg model VQNHE. (See Section S8, Supporting Information, for setup details and other results.) The mitigated energy is at least $E = -9$ in this case since even for the fully mixed state $\rho = 1/2^n$, the transformation, as a quantum channel effectively, can project the system to an averaged energy of $-9$. The consistently promising results for different Hamiltonian systems and under different noise models demonstrate the universal capability of VQNHE++ for ground state problems with built-in error mitigation power.

8. Discussions

In this work, we mainly focus on the noise resilience aspect of VQNHE. The proposed error mitigation method integrated with VQNHE is very promising as it requires fewer hardware resources compared to other well-established QEM schemes. The advantage of resource efficiency is especially prominent for the biased retraining, which only requires the same amount of measurement shots and hardware resources as one round of energy estimation. On the contrary, zero noise extrapolation (ZNE), one of the most common QEM techniques, needs to be conducted on the hardware of different noise strengths. Moreover, virtual distillation method, which prepares multiple copies of the state, and quasi-probability method, which requires tomography on the gates and the exponential scaling with the sampling positions, take much more hardware resources and running times. The neural error mitigation scheme proposed in Ref. [37] essentially provides a good initialization for the neural VMC with expensive resource requirements for the state tomography. And the second stage of the scheme in Ref. [37] is a purely classical VMC training with no input from the PQC. On the contrary, the noisy PQC is always one part of the quantum state generation pipeline in our case.

Several further comments are in order. First, our QEM proposal is strongly correlated with the VQNHE setup and rooted in the energy variational principle. Therefore, the current proposal is not a universal error mitigation method for universal quantum computing tasks. Second, the current QEM scheme can be easily combined with other common error mitigation techniques for further error reduction. Since most QEM schemes focus on error mitigation for the expectation values (from PQC) of some observables, they are compatible with the adaptive retraining for VQNHE. Specifically, we have successfully combined a technique of readout error mitigation with the retraining scheme (see the results in Section S2, Supporting Information). Third, it is also interesting to observe how robust amplitude estimation with parameterized likelihood can also exhibit error mitigation capability. There are some similarities between VQNHE and robust amplitude estimation conceptually. Both methods are amplitude amplification algorithms where the classical neural module acts as the amplifier in VQNHE as compared to the quantum module Grover iteration in robust amplitude estimation (thus the former could be more NISQ friendly). Finally, we recommend two common pipelines suitable for real experiments to mitigate the noise effect based on our work and experiments. If the circuit size under investigation is small and the optimal parameters in the noiseless case can be obtained via numerical simulation, then neural + transformation retraining is strongly recommended since it costs no quantum computational resource overhead and offers sufficiently good performance for the ground state energy prediction. If the scale of the experiment is too large to simulate in silico, then we need to run the joint optimization from scratch for all parameters of different components (PQC, NN and transformation matrix). This pipeline essentially captures both the optimization process and the noisy joint retraining process simultaneously.

Note Added

After the completion of this work, we notice an interesting paper on related topics. This paper shares some similarities with the tri-optimization part in our work. While Ref. utilizes a bi-optimization setup of combining Heisenberg transformed Hamiltonian with variational quantum circuits, the present work employs a tri-optimization setup combining...
Heisenberg transformed Hamiltonian (potentially nonunitary), variational quantum circuit, and additionally neural networks in the middle with the help of VQNHE, which in general has larger expressive power.

9. Conclusion
In this work, we investigate the native QEM scheme for VQNHE and demonstrate that the adaptive retraining manifests excellent error-mitigating effects. We then analyze the QEM capacity and present theoretical explanations for various scaling relations observed in experiments. In addition, we propose an enhancement add-on for VQNHE: the transformed Hamiltonian approach. Equipped with the parameterized gauge Hamiltonian, VQNHE++ shows even better expressive power and QEM capability. An interesting future direction is to apply VQNHE and the baked in error mitigation strategy shown in this work to more applications such as excited state searching problems or combinatorial optimization problems.

Supporting Information
Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements
This work was supported in part by the NSFC under Grant No. 11825404 (S.X.Z., Z.Q.W., and H.Y.), the MOSTC under Grants No.2018YFA0305604 and No. 2021YFA1400100 (H.Y.), and the CAS Strategic Priority Research Program under Grant No. XDB28000000 (H.Y.).

Conflict of Interest
The authors declare no conflict of interest.

Data Availability Statement
The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords
neural network, quantum-classical hybrid scheme, quantum error mitigation, variational quantum eigensolver

Received: May 23, 2023
Revised: July 21, 2023
Published online: August 15, 2023

[1] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, P. J. Coles, Nat. Rev. Phys. 2021, 3, 625.

[2] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, A. Aspuru-Guzik, Rev. Mod. Phys. 2022, 94, 015004.

[3] S. Endo, Z. Cai, S. C. Benjamin, X. Yuan, J. Phys. Soc. Jpn. 2021, 90, 032001.

[4] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtayne, A. Dunssworth, E. Farhi, B. Foxen, A. Fowler, C. Girney, M. Giustina, R. Graff, K. Guerin, S. Habegger, M. P. Harrigan, M. J. Hartmann, A. Ho, M. Hoffmann, T. Huang, et al., Nature 2019, 574, 505.

[5] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-h. Luo, J. Qin, D. Wu, X. Ding, Y. Hu, P. Hu, X.-y. Yang, W.-j. Zhang, H. Li, Y. Li, X. Jiang, L. Gan, C. Yang, L. You, Z. Wang, L. Li, N.-I. Liu, C.-y. Lu, J.-w. Pan, Science 2020, 370, 1460.

[6] J. Preskill, Quantum 2018, 2, 79.

[7] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, J. L. O’Brien, Nat. Commun. 2014, 5, 4213.

[8] F. B. Maciejewski, Z. Zimborás, M. Oszmaniec, Phys. Rev. A 2019, 103, 052315.

[9] S. Endo, S. C. Benjamin, Y. Li, Phys. Rev. X 2017, 7, 021050.

[10] K. Temme, S. Bravyi, J. M. Gambetta, Phys. Rev. Lett. 2017, 119, 180509.

[11] S. Endo, S. C. Benjamin, Y. Li, Phys. Rev. X 2018, 8, 031027.

[12] A. Kandala, K. Temme, A. D. Corcoles, A. Mezzacapo, J. M. Chow, J. M. Gambetta, Nature 2019, 567, 491.

[13] C. Song, J. Cui, A. Fowler, T. C. White, P. V. Coveney, P. J. Love, A. Aspuru-Guzik, J. L. O’Brien, Science 2014, 345, 604.

[14] S. Endo, A. Aspuru-Guzik, J. L. O’Brien, Nat. Commun. 2014, 5, 4213.

[15] S. Endo, S. C. Benjamin, X. Yuan, Rev. Mod. Phys. 2020, 92, 015003.

[16] H. R. Grimsmo, S. E. Economou, E. Barnes, N. J. Mayhall, Nat. Commun. 2019, 10, 3007.

[17] C. Y. Hsieh, Q. Sun, S. Zhang, C. K. Lee, npj Quantum Inf. 2021, 7, 19.

[18] Y. Li, S. C. Benjamin, Phys. Rev. X 2017, 7, 021050.

[19] K. Temme, S. Bravyi, J. M. Gambetta, Phys. Rev. Lett. 2017, 119, 180509.

[20] S. Endo, S. C. Benjamin, Y. Li, Phys. Rev. X 2018, 8, 031027.

[21] A. Kandala, K. Temme, A. D. Corcoles, A. Mezzacapo, J. M. Chow, J. M. Gambetta, Nature 2019, 567, 491.

[22] C. Song, J. Cui, H. Wang, J. Hao, H. Feng, Y. Li, Sci. Adv. 2019, 5, eaaw5686.

[23] S. McCarron, X. Yuan, S. Benjamin, Phys. Rev. Lett. 2019, 122, 180501.

[24] Y. Chen, M. Farahzad, S. Yoo, T.-C. Wei, Phys. Rev. A 2019, 100, 052315.

[25] F. B. Maciejewski, Z. Zimborás, M. Oszmaniec, Quantum 2020, 4, 257.

[26] S. Bravyi, S. Sheldon, A. Kandala, D. C. Mckay, J. M. Gambetta, Phys. Rev. A 2021, 103, 042605.

[27] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, P. J. Coles, Nat. Rev. Phys. 2021, 3, 625.
[35] D. Bultrini, M. H. Gordon, P. Czarnik, A. Arrasmith, P. J. Coles, L. Cincio, 2021, arXiv:2107.13470.
[36] A. A. Zhukov, W. V. Pogosov, 2021, arXiv:2105.07793.
[37] E. R. Bennewitz, F. Hopfmueller, B. Kulchytskyy, J. Carrasquilla, P. Ronagh, Nat. Mach. Intell. 2022, 4, 618.
[38] S.-X. Zhang, Z.-Q. Wan, C.-K. Lee, C.-Y. Hsieh, S. Zhang, H. Yao, Phys. Rev. Lett. 2022, 128, 120502.
[39] R. Jastrow, Phys. Rev. 1955, 98, 1479.
[40] G. Mazzola, P. J., Ollitrault, P. K. Barkoutsos, I. Tavernelli, Phys. Rev. Lett. 2019, 123, 130501.
[41] F. Benfenati, G. Mazzola, C. Capecci, P. K. Barkoutsos, P. J. Ollitrault, I. Tavernelli, L. Guidoni, 2021, arXiv:2101.09316.
[42] G. Carleo, M. Troyer, Science 2017, 355, 602.
[43] D.-L. Deng, X. Li, S. Das Sarma, Phys. Rev. B 2017, 96, 195145.
[44] G. Carleo, Y. Nomura, M. Imada, Nat. Commun. 2018, 9, 5322.
[45] Z. Cai, J. Liu, Phys. Rev. B 2018, 97, 035116.
[46] D. Pfau, J. S. Spencer, A. G. d. G. Matthews, W. M. C. Foulkes, Phys. Rev. Res. 2020, 2, 033429.
[47] J. Hermann, Z. Schätzle, F. Noé, Nat. Chem. 2020, 12, 891.
[48] S.-X. Zhang, Z.-Q. Wan, H. Yao, 2019, arXiv:1911.09117.
[49] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, H. Neven, Nat. Commun. 2018, 9, 4812.
[50] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, P. J. Coles, Nat. Commun. 2021, 12, 6961.
[51] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, P. J. Coles, Nat. Commun. 2021, 12, 1791.
[52] S.-X. Zhang, J. Alcock, Z.-Q. Wan, S. Liu, J. Sun, H. Yu, X.-H. Yang, J. Qiu, Z. Ye, Y.-Q. Chen, C.-K. Lee, Y.-C. Zheng, S.-K. Jian, H. Yao, C.-Y. Hsieh, S. Zhang, 2022, arXiv:2205.10091.
[53] G. Wang, D. E. Koh, P. D. Johnson, Y. Cao, PRX Quantum 2021, 2, 010346.
[54] A. Katabarwa, A. Kunitsa, B. Peropadre, P. Johnson, 2021, arXiv:2110.10664.
[55] Z.-X. Shang, M.-C. Chen, X. Yuan, C.-Y. Lu, J.-W. Pan, 2021, arXiv:2112.07881.