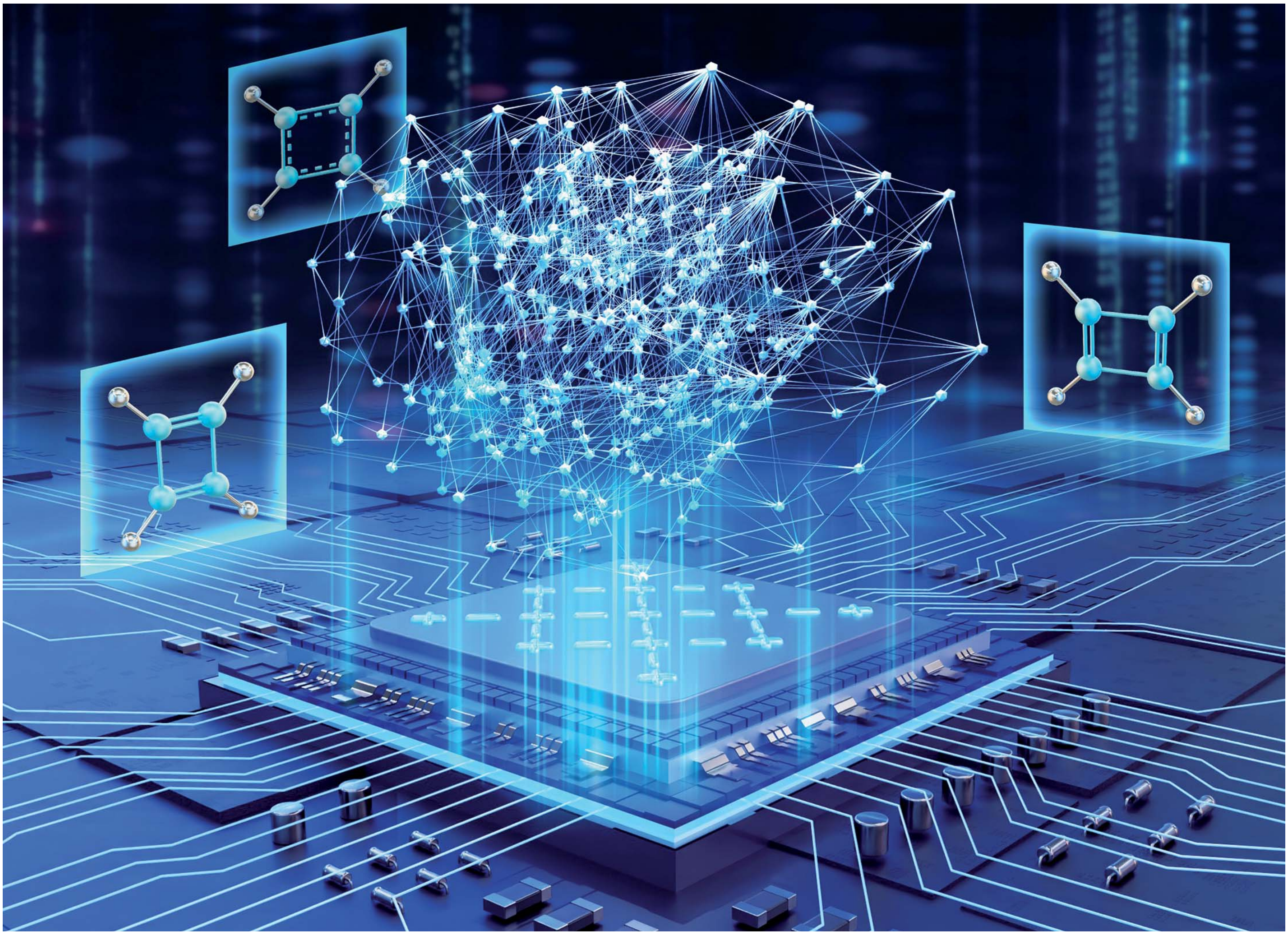

Showcasing research from Professor Weitang Li's laboratory, School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen, China.

Quantum machine learning of molecular energies with hybrid quantum-neural wavefunction

Quantum computation and machine learning are revolutionizing computational chemistry because of their superior ability to represent and manipulate molecular wavefunctions. This work introduces a hybrid framework that learns molecular wavefunction using a combination of an efficient quantum circuit and a deep neural network. The results show not only high accuracy but also significant resilience to noise, highlighting the robustness of our method in practical quantum applications.

Image reproduced by permission of Weitang Li from *Digital Discovery*, 2025, **4**, 2697.

As featured in:

See Weitang Li *et al.*, *Digital Discovery*, 2025, **4**, 2697.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/digitaldiscovery

Registered charity number: 207890

Digital Discovery

PAPER

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Digital Discovery*, 2025, **4**, 2697

Quantum machine learning of molecular energies with hybrid quantum-neural wavefunction†

Weitang Li,*[a] Shi-Xin Zhang,[b] Zirui Sheng,[a] Cunxi Gong,[a] Jianpeng Chen[a] and Zhigang Shuai[ac]

Received 23rd May 2025
Accepted 14th July 2025

DOI: 10.1039/d5dd00222b

rsc.li/digitaldiscovery

Quantum computational chemistry holds great promise for simulating molecular systems more efficiently than classical methods by leveraging quantum bits to represent molecular wavefunctions. However, current implementations face significant limitations in accuracy due to hardware noise and algorithmic constraints. To overcome these challenges, we introduce a hybrid framework that learns molecular wavefunction using a combination of an efficient quantum circuit and a neural network. Numerical benchmarking on molecular systems shows that our hybrid quantum-neural wavefunction approach achieves near-chemical accuracy, comparable to advanced quantum and classical techniques. Based on the isomerization reaction of cyclobutadiene, a challenging multi-reference model, our approach is further validated on a superconducting quantum computer with high accuracy and significant resilience to noise.

I Introduction

Quantum computers leverage quantum effects to store and manipulate data, making them particularly suitable for the simulation of microscopic quantum systems.[1–3] The Variational Quantum Eigensolver (VQE) algorithm is the most widely adopted framework for quantum computational chemistry.[4–9] The key component of the VQE algorithm is the parameterized quantum circuit, which learns the quantum state of the system under study variationally.[10] The challenge of VQE lies in striking a delicate balance between circuit depth and accuracy.[11–14] While deeper circuits tend to improve accuracy, they also make the algorithm more sensitive to noise and can suffer from barren plateaus.[15] In contrast, shallow circuits may not capture the system's complexity adequately. Parallel to the evolution of VQE, Neural Networks (NNs) have shown remarkable success in representing quantum wavefunctions of chemical systems.[16] Based on variational Monte Carlo, these NNs are trained to minimize the energy expectation, similar to the VQE approach. Efforts along this line include DeepWF,[17] FermiNet,[18] PauliNet,[19] QiankunNet,[20] and so on.[21–24] Thanks to the expressive power of NNs, these methods demonstrate accuracy comparable to Coupled Cluster with Single and Double excitations (CCSD) but with significantly lower computational scaling, typically $\mathcal{O}(N^4)$.

The success in these new wavefunction representations has inspired the development of hybrid quantum-neural wavefunctions, where quantum circuits and neural networks are jointly trained to represent the wavefunction of quantum systems.[25] In this hybrid approach, quantum circuits are responsible for learning the quantum phase structure of the target state, which is a difficult task for neural networks alone,[26] and the neural network correctly describes the amplitude. The combination of quantum computation and variational Monte Carlo has also demonstrated considerable potential in simulating quantum systems,[27] and the inclusion of neural networks significantly enhances the expressiveness of trial wave functions, thereby leading to more accurate and scalable simulations. The intersection between quantum computing and machine learning, known as quantum machine learning, is developing at a rapid pace.[28–31] Chemistry applications include the construction of shallow depth ansatz, energy eigenstate filtration, material phase prediction, neural network pertaining, and so on.[32–38]

In this work, we propose a quantum machine learning framework for efficient representation of molecular wavefunction and accurate computation of molecular energies. The method employs the linear-depth paired Unitary Coupled-Cluster (UCC) with double excitations (pUCCD) circuit to learn molecular wavefunction in the seniority-zero subspace,[39–43] and a neural network to correctly account for the contributions from unpaired configurations. We propose an efficient algorithm to compute the expectations of physical observables for the hybrid quantum-neural wavefunction, which avoids calculating the overlap between the quantum circuit state and classical state, or the costly process of quantum state tomography. This represents an enhancement of scalability over the previously

[a] School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen, Guangdong 518172, China. E-mail: liwt31@gmail.com

[b] Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China

[c] Department of Chemistry, MOE Key Laboratory for Organic OptoElectronics and Molecular Engineering, Tsinghua University, Beijing 100084, China

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d5dd00222b

proposed quantum-classical hybrid quantum Monte-Carlo method.[44] We name our method as pUNN, which stands for paired unitary coupled-cluster with neural networks. pUNN retains the low qubit count (N qubits) and shallow circuit depth of pUCCD, while achieving accuracy comparable to the most precise quantum and classical computational chemistry methods, such as UCCSD (UCC with single and double excitations) and CCSD(T) (CCSD with perturbative triple excitations). We demonstrate the efficacy of pUNN through numerical simulations of various diatomic and polyatomic molecular systems, such as N_2 and CH_4. To test pUNN in a real quantum computing scenario, we compute the reaction barrier for the isomerization of cyclobutadiene on a programmable superconducting quantum computer. The results demonstrate that the pUNN is highly accurate and noise resilient for a real quantum computing task.

II Theory and methodology

In this section, we present our pUNN algorithm and focus on our contribution. General backgrounds, such as the electronic structure problem and the UCC types of ansatz for quantum computational chemistry are briefly overviewed in the Appendix A. We start by employing the pUCCD ansatz to represent molecular wavefunction, which is encoded in the parameterized quantum circuit $\hat{U}(\vec{\theta})$. In the computational basis, the pUCCD circuit state can be expressed as

$$|\psi\rangle = \sum_k a_k |k\rangle, \tag{1}$$

where $|k\rangle$ represents the occupation of a pair of electrons in the original N-qubit Hilbert space. For ground state problems, the coefficients a_k can be assumed to be real numbers. To correctly describe the configurations outside of the seniority-zero subspace, we add N ancilla qubits to the circuit and expand the Hilbert space from N qubits to $2N$ qubits. In the expanded $2N$-qubit space, the equivalent state is

$$|\Phi\rangle = \sum_k a_k |k\rangle \otimes |k\rangle, \tag{2}$$

with the two $|k\rangle$ terms now representing the occupation of the alpha and beta spin sectors, respectively. We note that these N ancilla qubits can be treated classically, which will be explained later.

In the context of quantum circuits, the expanded state $|\Phi\rangle$ is constructed from $|\psi\rangle$ using the ancilla qubits and an entanglement circuit $\hat{E}$:

$$|\Phi\rangle = \hat{E}(|\psi\rangle \otimes |0\rangle). \tag{3}$$

The entanglement circuit $\hat{E}$ creates the necessary correlations between the original qubits and the ancilla qubits. $\hat{E}$ can be decomposed into N parallel CNOT gates:

$$\hat{E} = \prod_i^N \mathrm{CNOT}_{i,i+N}, \tag{4}$$

where each CNOT gate entangles the i-th original qubit with the corresponding i-th ancilla qubit.

Although $|\Phi\rangle$ has $2N$ qubits while $|\psi\rangle$ has N qubits, from a quantum chemistry perspective, they represent the same state in the seniority-zero space and therefore have the same energy. We then apply the neural network, acting as an quantum operator $\hat{N}$, on the quantum state. $\hat{N}$ is a non-unitary post-processing operator[25] defined in the expanded Hilbert space. After applying $\hat{N}$, the overall state becomes $\hat{N}\hat{E}(|\psi\rangle \otimes |0\rangle)$. The method is inspired by variational quantum-neural hybrid eigensolver (VQNHE) and it provides exponential acceleration for nonunitary postprocessing in VQE than naive transformed Hamiltonian approach.[45–47] The neural network operator $\hat{N}$ modulates the state $|\Phi\rangle$ as follows:

$$\hat{N} = \sum_{kj} b_{kj} |k\rangle|j\rangle\langle j|\langle k|, \tag{5}$$

where b_{kj} is a real tensor represented by a continuous neural network $\mathscr{B}(k,j)$, such that $b_{kj} = \mathscr{B}(k,j)$. To drive $\hat{N}\hat{E}(|\psi\rangle \otimes |0\rangle)$ out of the seniority-zero subspace, we apply a perturbation circuit $\hat{P}$ to the ancilla qubits at the beginning, diverting the state of the ancilla qubits $|\phi\rangle = \hat{P}|0\rangle$ from $|0\rangle$

$$|\phi\rangle = \hat{P}|0\rangle = \frac{|0\rangle + \sum_{j\neq 0} \varepsilon_j |j\rangle}{1 + \sum_{j\neq 0} {\varepsilon_j}^2} \tag{6}$$

where $\in_j$ are small coefficients satisfying $\sum_{j\neq 0} {\varepsilon_j}^2 \ll 1$. As a result, our algorithm is expected to be resilient to noise,[48] making it well-suited for implementation on real quantum devices. The conservation of the particle number is enforced by the neural network introduced in the following. The values of $\in_j$ and the exact form of $\hat{P}$ are flexible. The only key requirement for $\hat{P}$ is that it should have a low circuit depth, which allows efficient simulation of $\hat{P}|0\rangle$ on classical computers. To this end, we adopt a perturbation circuit with single qubit rotation gates R_y for each qubit and the rotation angle is set to 0.2. $\hat{P}$ produces real coefficients, a desired property for the ground state of the molecular Hamiltonian.

After describing the quantum circuit part, we turn to the neural network structure used for $\mathscr{B}(k,j)$. $\mathscr{B}(k,j)$ accepts the two bitstring k and j as input and outputs the coefficients b_{kj}. The first component of the neural network is embedding the bitstring $|k\rangle \otimes |j\rangle$ into a vector. We employ a binary representation, where $|k\rangle \otimes |j\rangle$ is converted to a vector of size $2N$, with each element being either -1 or 1. The vector $\mathbf{x}_0(k, j)$ is then passed through a neural network consisting of L dense layers and ReLU activation functions

$$\mathbf{x}_{i+1}(k,j) = \mathrm{ReLU}[\mathbf{W}_i \mathbf{x}_i(k,j) + \mathbf{c}_i]. \tag{7}$$

In the hidden layers, the number of neurons is set to $2KN$ where K is a tunable integer that controls the size of the neural network. In this work we set $K = 2$ unless otherwise specified. The number of layers L is set to $N - 3$, proportional to the size of the molecule. The number of parameters in the neural network scales as K^2N^3 considering both the width and depth of the neural network. The computational complexity is also $\mathcal{O}(K^2N^3)$ for each input bitstring.

The final dense layer outputs the desired coefficient b_{kj}, before multiplying with the particle number conservation mask $m(k,j)$

$$b_{kj} = m(k,j)[\mathbf{W}_L\mathbf{x}_L(k,j) + \mathbf{c}_L]. \tag{8}$$

The mask $m(k,j)$ is defined as

$$m(k,j) = \begin{cases} 1 & \text{if} \quad \sum_i k_i = N_\alpha \text{ and } \sum_i j_i = N_\beta, \\ 0 & \qquad \text{otherwise,} \end{cases} \tag{9}$$

where $N_{\alpha/\beta}$ is the number of spin up/down electrons. The mask eliminates configurations $|k\rangle \otimes |j\rangle$ that do not conserve the number of spin up and down electrons.

To summarize, the overall wavefunction is given by

$$|\Psi\rangle = \hat{N}\hat{E}\left(\hat{U}\left(\overrightarrow{\theta}\right)|0\rangle \otimes \hat{P}|0\rangle\right), \tag{10}$$

which consists of four components: the pUCCD circuit $\hat{U}(\overrightarrow{\theta})$, the perturbation circuit $\hat{P}$, the entanglement circuit $\hat{E}$ and the neural network $\hat{N}$. The next challenge is to measure the expectation value of the physical observables such as the energy based on eqn (10), which is highly nontrivial without resorting to quantum tomography or incurring exponential measurement overhead. Without an efficient measurement protocol, the pUNN approach could be rendered impractical. Besides, in quantum computational chemistry, the number of measurements required to estimate expectation values is a key indicator of efficiency for variational algorithms like pUNN. In fact, the ansatz represented by eqn (10) is carefully designed in such a way that an efficient algorithm for computing expectation values is possible.

Since $|\Psi\rangle$ is not normalized, the energy expectation is

$$\langle E\rangle = \frac{\langle\Psi|\hat{H}|\Psi\rangle}{\langle\Psi|\Psi\rangle}. \tag{11}$$

Here we outline the key points of the measurement protocol that enables the computation of both $\langle\Psi|\hat{H}|\Psi\rangle$ and $\langle\Psi|\Psi\rangle$ using the measurement outcome of the quantum circuit $\hat{U}(\overrightarrow{\theta})|0\rangle$ and the output from the neural network. The full measurement protocol is provided in the Appendix B. For brevity, we assume there is only a single Pauli string in $\hat{H}$, and the summation over many Pauli strings can be handled straightforwardly. We also note that the estimation of the norm $\langle\Psi|\Psi\rangle$ can be considered as a special case when $\hat{H} = \hat{I}$.

To perform the measurement, we transform the Hamiltonian $\hat{H}$ and the neural network $\hat{N}$ with $\hat{E}$

$$\langle\Psi|\hat{H}|\Psi\rangle = \langle\psi \otimes \phi|(\hat{E}^\dagger\hat{N}^\dagger\hat{E})(\hat{E}^\dagger\hat{H}\hat{E})(\hat{E}^\dagger\hat{N}\hat{E})|\psi \otimes \phi\rangle. \tag{12}$$

Since $\hat{E}$ is a Clifford circuit, $\hat{H}' = \hat{E}^\dagger\hat{H}\hat{E}$ is also a Pauli string. Additionally, since $\hat{E}$ is composed of CNOT gates, it reversibly maps one bitstring to another, rather than a linear combination of bitstrings. Specifically,

$$\hat{E}(|k\rangle \otimes |j\rangle) = |k\rangle \otimes |k \oplus j\rangle. \tag{13}$$

The transformed neural network $\hat{N}' = \hat{E}^\dagger\hat{N}\hat{E}$ is

$$\hat{N}^{'} = \sum_{kj} b_{kj}|k\rangle|k\oplus j\rangle\langle k\oplus j|\langle k| = \sum_{kj} b_{k,k\oplus j}|k\rangle|j\rangle\langle j|\langle k|. \tag{14}$$

$\hat{N}'$ is thus formally the same as $\hat{N}$ but with a permuted index for the coefficient b.

After the transformation, the entanglement circuit $\hat{E}$ is removed from eqn (12)

$$\langle\Psi|\hat{H}|\Psi\rangle = \langle\Psi \otimes \phi|\hat{N}'^\dagger\hat{H}'\hat{N}'|\Psi \otimes \phi\rangle. \tag{15}$$

eqn (15) corresponds to the measurement of $\hat{N}'^\dagger\hat{H}'\hat{N}'$ on two unentangled circuits $|\psi\rangle$ and $|\phi\rangle$. If $\hat{N}'$ is absent or if $\hat{N}' = \hat{I}$, the measurement of $\hat{H}$ can be performed efficiently by measuring the two separate circuits $|\psi\rangle$ and $|\phi\rangle$. In Appendix B, we show that, by carefully designing the measurement circuit, $\hat{N}'\hat{H}'\hat{N}'$ can also be measured by separate measurement of $|\psi\rangle$ and $|\phi\rangle$, with a constant overhead. Therefore, the evaluation of $\langle\Psi|\hat{H}|\Psi\rangle$ is cast into the separate measurement of $|\psi\rangle$ and $|\phi\rangle$. Since $|\phi\rangle$ is designed to be a shallow circuit that can be efficiently simulated classically, the only circuit that needs to be executed on real quantum devices is the pUCCD circuit $|\psi\rangle$. Nonetheless, the number of terms to measure in the Hamiltonian increases from N^2 in the pUCCD circuit to N^4 for more general electronic structure problems. Thus, in terms of measurement shots, the pUNN method is as efficient as other quantum computational methods such as UCCSD, but with significantly reduced circuit depth and higher accuracy. Compared with entanglement forging,[49] which utilizes classical sampling to recover the entanglement between two sub-systems, our method encodes the entanglement between two sub-systems into $\hat{H}$ and $\hat{N}$ and avoids excessive sampling.

A schematic diagram of the pUNN framework is depicted in Fig. 1. In the whole algorithm, only the pUCCD circuit within the grey box in dashed lines is executed on quantum computers, which allows pUNN to maintain the N-qubit requirement for the computation instead of $2N$. The perturbation circuit and entanglement circuit can be efficiently processed on classical computers. The measured bitstring of the composite circuit is fed into the neural network for $\mathscr{B}(k,j)$, which is then used to adjust the measurement outcome. The entire ansatz is then set up in a VQE workflow, where both the parameters in the quantum circuit and the neural network are trained to minimize the molecular energy. This process ultimately yields the ground state through the variational principle.

In the noiseless simulation described in Sec. III A, we use the L-BFGS-B algorithm to optimize the parameters in the quantum circuit. For circuit optimization on real quantum hardware, we employ the SOAP method.[50] For both the noiseless simulations and the experiments on quantum computers, the neural network is trained using the AdaMax optimizer,[51] a variant of the widely adopted Adam optimizer. The optimizer begins with a learning rate schedule of $\alpha = 0.01$, $b_1 = 0.8$ and $b_2 = 0.99$. The learning rate decays linearly to $\alpha = 0.001$ between the 8000th and 32000th steps. This learning rate schedule helps ensure stable convergence by gradually decreasing the learning rate as the training progresses. For noiseless simulation, the maximum

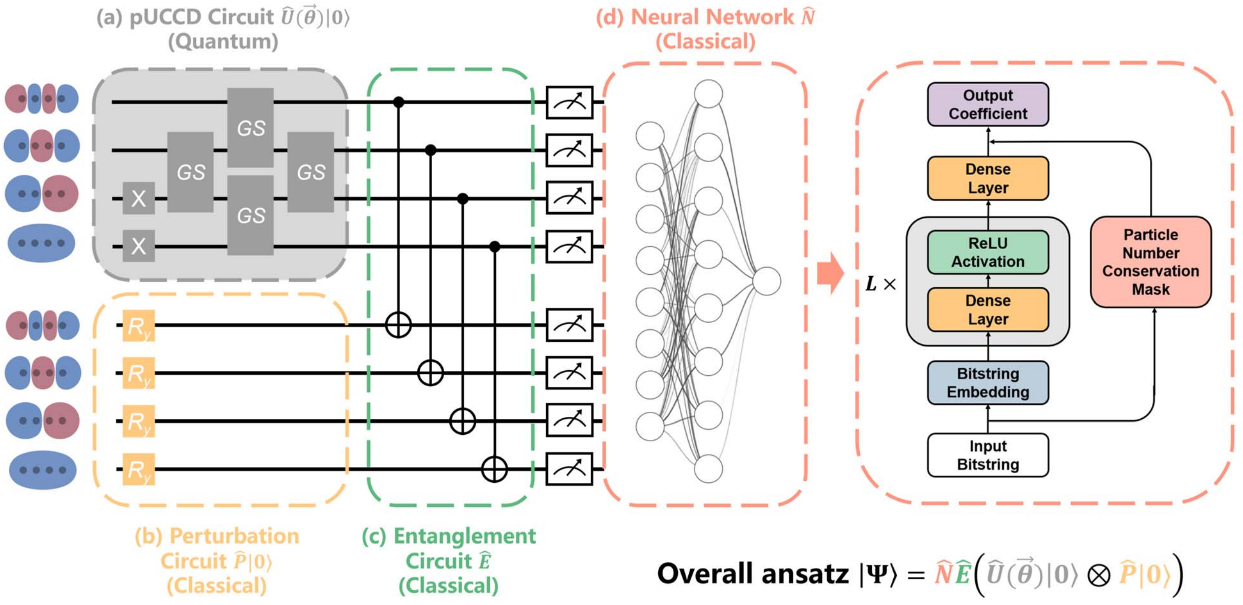
(a) pUCCD Circuit $\hat{U}(\vec{\theta})|0\rangle$
(Quantum)
(d) Neural Network $\hat{N}$
(Classical)
GS
X
Output Coefficient
Dense Layer
ReLU Activation
Dense Layer
Bitstring Embedding
Input Bitstring
Particle Number Conservation Mask
$L\times$
R_y
(b) Perturbation Circuit $\hat{P}|0\rangle$
(Classical)
(c) Entanglement Circuit $\hat{E}$
(Classical)
Overall ansatz $|\Psi\rangle = \hat{N}\hat{E}\left(\hat{U}(\vec{\theta})|0\rangle \otimes \hat{P}|0\rangle\right)$

Fig. 1 A schematic diagram for the pUNN framework. The (a) pUCCD circuit in the grey box is the only component executed on a real quantum computer. "GS" denotes Givens–Swap gate. Meanwhile, the (b) perturbation circuit and the (c) entanglement circuit are processed classically. Together, the quantum circuit and the (d) neural network serves as an ansatz and are trained jointly to represent the molecular wavefunction.

number of steps is set to 64 000. A summary table for the hyperparameters can be found in the ESI.† For the noiseless simulation, we initialize the neural network with five different random seeds, and the lowest energy found across these seeds is reported. For quantum circuit manipulation, including both noiseless and noisy emulation as well as interfacing with real quantum hardware, we use the TensorCircuit framework.[52] General quantum computational chemistry tasks, including Hamiltonian construction, reference value calculation, and parameter optimization are handled by TenCirChem,[53] a specialized package built on top of TensorCircuit designed for quantum computational chemistry. TenCirChem also relies on PySCF for evaluating the integrals and performing calculations based on classical computational chemistry.[54]

III Results

A Accuracy and scalability

We first compare the accuracy of pUNN with other quantum computational methods in Fig. 2. For this comparison, we perform noiseless numerical calculations on molecular systems corresponding to 8 spatial orbitals and 16 qubits. The basis set employed is STO-3G and the 1s orbitals are frozen. The exact geometries of the molecules are reported in the ESI.† The full configuration interaction (FCI) energy for these molecules is computed as the reference energy. As shown in Fig. 2, the standard pUCCD approach improves over the HF method but consistently shows the highest error across all molecules. The results indicate that the neglect of configurations outside of the seniority-zero subspace limits the accuracy. The orbital optimization pUCCD (oo-pUCCD) method[42,55,56] reduces the error to a modest extent for most molecules, except for N_2 and CO, where the errors of pUCCD and oo-pUCCD are comparable. This demonstrates the limitation of oo-pUCCD, as it still assumes electron paring. The UCCSD method, known for its high accuracy, performs well across the board. However, UCCSD requires $2N$ qubits for N molecular orbitals and has a very deep circuit, which is computationally expensive. The typical circuit depth of UCCSD for these molecules is approximately 1000. Our

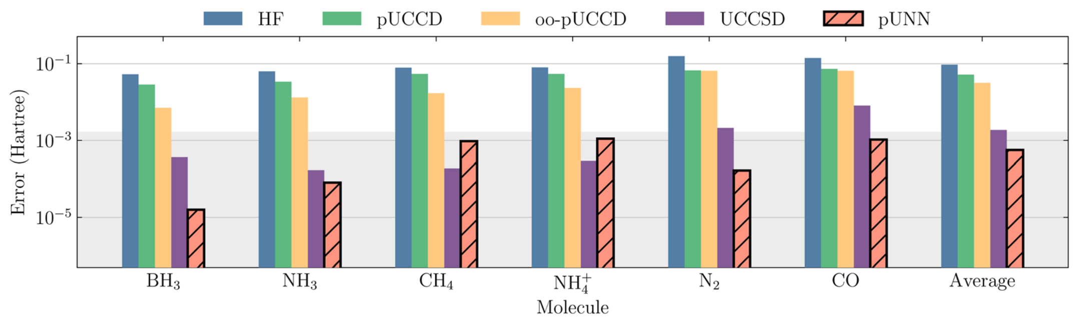

Fig. 2 Compare the accuracy of pUNN with other quantum computational chemistry methods. The 1s orbitals are frozen and the reference energy is FCI. The shaded area indicates the chemical accuracy.

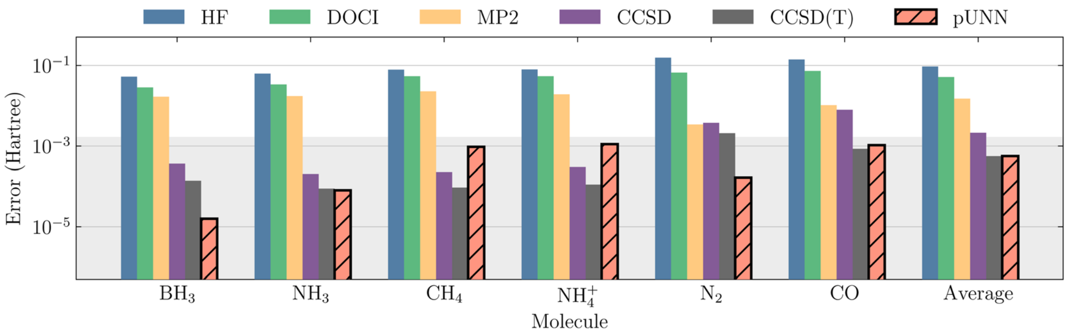

Fig. 3 Compare the accuracy of pUNN with other classical computational chemistry methods. The 1s orbitals are frozen and the reference energy is FCI. The shaded area indicates the chemical accuracy.

proposed pUNN method stands out as the most accurate approach for the majority of the molecules studied. In the meantime, pUNN uses only N qubits for N molecular orbitals, and its circuit depth is the same as the circuit depth of pUCCD. In contrast to UCCSD, the circuit depth of pUNN is approximately 20. pUNN frequently achieves or approaches the chemical accuracy threshold of 1.6 m hartree, as indicated by the shaded area on the graph. By comparing pUNN and pUCCD, we find that, the mean absolute error (MAE) decreases from 51.9 m hartree for pUCCD to 0.6 m hartree for pUNN. This corresponds to a reduction in error by two orders of magnitude. The MAE of pUNN is comparable to the MAE of UCCSD, which is 1.9 m hartree.

In Fig. 3 we compare the error of pUNN with several classical computational methods. The doubly occupied configuration interaction (DOCI) method is the classical counterpart of the pUCCD method since it also assumes electron pairing. Based on the results in Fig. 2 we can expect DOCI will perform poorly, which is confirmed by the data in Fig. 3. The second order Møller–Plesset perturbation theory (MP2) improves over DOCI, particularly for diatomic molecules. This suggests that including the configurations with singly occupied orbitals is crucial for accurately describing the molecular wavefunction. The coupled-cluster methods, CCSD and its perturbative extension CCSD(T), are considered some of the most accurate techniques in quantum chemistry. Both CCSD and CCSD(T) demonstrate high accuracy, with CCSD(T) achieving chemical accuracy for most of the molecules studied. When comparing pUNN to these classical methods, we find that pUNN achieves accuracy comparable to that of CCSD(T), indicating that pUNN is a high-accuracy method for quantum chemistry calculations.

We next investigate the factors that determine the accuracy of the pUNN method. Fig. 4(a) compares the accuracy of pUCCD and pUNN methods against the size of hydrogen chain molecules (H_5^+, H_6, H_7^+, and H_8) for two different bond lengths (d = 1.0 Å and d = 2.5 Å). The results clearly demonstrate that pUNN consistently outperforms standard pUCCD, achieving lower error across all molecule sizes and bond lengths. Notably, pUNN maintains high accuracy even as the molecule size increases especially for the longer bond length of 2.5 Å. When d = 1.0 Å, the errors of pUNN seem to fluctuate when the system size varies. However, the magnitude of the fluctuation, in the order of 10^{-4} hartree, is well below the chemical accuracy threshold and thus insignificant. Fig. 4(b) showcases the impact of neural network size on the error of pUNN for various molecules. The x-axis represents the neural network size K, and the number of hidden neurons is $2KN$ where N is the number of molecular orbitals. The atomic distance in H_8 is d = 1.0 Å. As the network size increases from 2 to 8, there's a clear trend of a logarithmic decreasing error for all molecules. Most molecules achieve chemical accuracy (indicated by the shaded area) with larger neural networks, with NH_3 and BH_3 showing

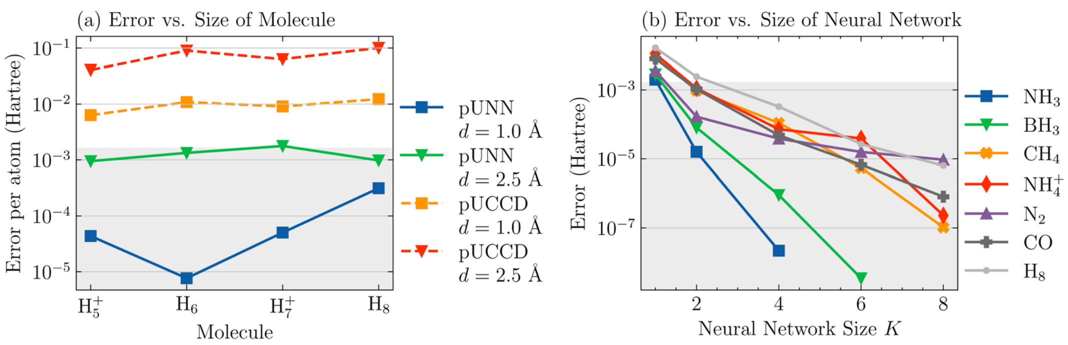

Fig. 4 Factors for the accuracy of the pUNN method. (a) The error of pUNN *versus* the size of the molecule under study. (b) The error of pUNN *versus* the size of the neural network K. The number of hidden neurons in the neural network is $2KN$ where N is the number of molecular orbitals. The shaded area indicates the chemical accuracy.

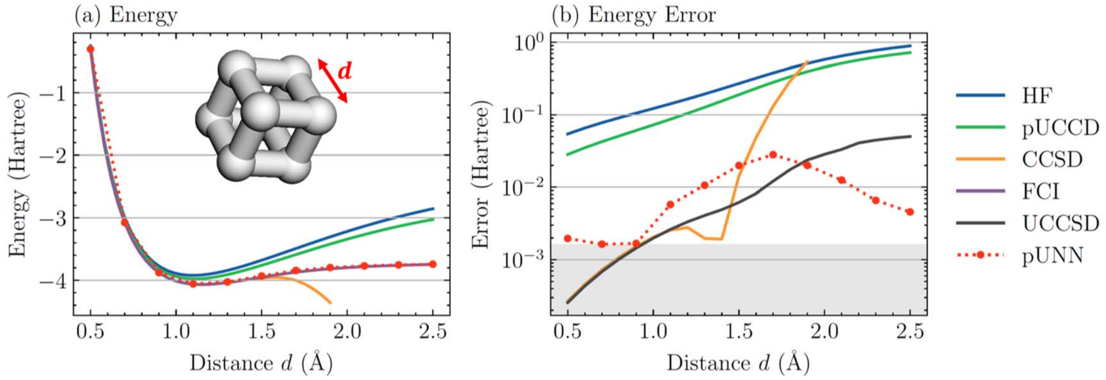

Fig. 5 Benchmarking pUNN based on the potential energy profile of cubic H_8. (a) The potential energy profile of cubic H_8 by different computational methods. (b) The error compared with the exact solution *versus* the H–H distance in the H_8 cube.

particularly significant improvements in accuracy as the network size grows. Although molecules studied here are still much smaller than those encountered in practical chemistry problems, the promising scaling shown in Fig. 4(a) suggests that pUNN has the potential to accurately learn the wavefunction of complex chemical systems.

We finally test the accuracy of pUNN based on cubic H_8 molecule at different H–H distance d. The system is particularly challenging due to the strong correlation as d increases. In Fig. 5(a) we show the potential energy profile computed by both pUNN. As expected, pUNN shows much higher accuracy than other methods. From $d = 0.5$ Å to 2.5 Å pUNN coincides well with the FCI solution. For reference, we also include the CCSD method, which shows high accuracy at intermediate d. However, due to its single-reference and non-variational nature, the error of CCSD quickly increases as d becomes larger than 1.5 Å and it fails to reach convergence for larger d. CCSD(T) is not expected to improve CCSD when it fails because CCSD(T) relies on good CCSD wavefunction to account for perturbative triple excitation. Thus, although the 16 qubit system represents a relatively small variational space compared to challenging strongly correlated systems,[57,58] it is sufficient to reveal the limitations of methods like CCSD, which fail in strongly correlated regimes, while pUNN maintains relatively high accuracy. In Fig. 5(b) we depicted the error of the methods in logarithmic scale. All methods except pUNN show an increase in error as d increases. The maximum error for pUNN appears at $d = 1.7$ Å and the magnitude of the error is 10^{-2} hartree. The relatively large error highlights the complexity of the cubic H_8 molecule. We anticipate that integrating alternative quantum circuits into our pUNN framework, such as those based on valence bond theory,[59] could enhance accuracy in strong correlation. The UCCSD method is also included in Fig. 5(b). While UCCSD shows high accuracy at smaller d, it suffers from significant error at the large d limit, similar to CCSD. We perform additional benchmarks for strongly correlated systems based on the potential energy profile of N_2 and CH_4 and the trend is similar to Fig. 5(b). The results are included in the ESI.†

In Table 1, we present a breakdown of parameters for hydrogen systems studied in Fig. 4 and 5, comparing pUNN with FCI. For pUNN, the pUCCD circuit has $\mathcal{O}(N^2)$ parameters, while the NN has $\mathcal{O}(K^2N^2L)$ parameters, with $K = 2$ and $L = N - 3$. From Table 1, pUNN's total parameters grow polynomially with N, while FCI's determinant space grows exponentially. For H_8, pUNN uses fewer parameters than FCI, and achieves high accuracy across both weak and strong correlation, as shown in Fig. 4(a) and 5. As our main contribution is the novel and unique quantum-neural hybrid framework, our choice of a dense MLP for the neural network is a proof-of-concept. More efficient architectures, such as restricted Boltzmann machines or graph neural networks, could further optimize pUNN,[60,61].

3.2 Experiments on a superconducting quantum computer

To evaluate the performance of pUNN in a real quantum computing scenario, we conduct experiments on a superconducting quantum computer. We choose the isomerization reaction of cyclobutadiene as our model system, as shown in Fig. 6(a). The transition state of this system is particularly challenging due to strong correlations arising from degeneracy.[19,62] In this reaction, the reactant and product are identical molecules, with a 90-degree rotation between them. The

Table 1 Parameter counts for pUNN and FCI for hydrogen systems

System	FCI determinants	NN parameters	pUCCD parameters
H_5^+	100	661	6
H_6	400	1537	9
H_7^+	1225	2885	12
H_8	4900	4801	16
H_{2n}	$[(2n)!/(n!)^2]^2$	$128n^3 - 208n^2 - 16n + 1$	n^2

electronic structures of the reactant and the product are considerably simpler than that of the transition state. Therefore, in the following analysis, we focus on the transition state, and calculate the reaction barrier by subtracting the exact energy of the reactant and product from the energy of the transition state.

We employ the cc-pVDZ basis set[63] for HF calculation and select the four frontier orbitals as the active space. Using the paired ansatz, the active space is represented by a 4-qubit quantum circuit, with four parameters corresponding to four double excitations. The superconducting quantum chip used in this work consists of 13 qubits. Since the Givens–Swap gate is not a native gate on this chip, we carefully select 4 qubits from the 13-qubit system, which follows a ring topology, as shown in Fig. 6(d). This selection allows us to implement all four excitation operators using only Givens rotation gates, eliminating the need for the more expensive swap gates, which would otherwise require 3 CNOT gates. The Givens rotation gates should be further compiled into 4 native CNOT gates, along with several single-qubit gates. To reduce circuit depth, we introduce an approximation that breaks the symmetry and removes the control qubit of the controlled R_y gate.[64] The resulting circuit does not conserve the total particle number anymore but the overall error could be smaller than the gate error by 8 additional CNOT gates, especially when some of the rotation gates have small rotation angles. Each Givens rotation gate is thus compiled into 2 CNOT gates, resulting in a total of 8 CNOT gates in the circuit. Standard readout error mitigation based on a direct product calibration matrix is applied to enhance the precision.

We obtain the circuit parameters by optimizing the pUCCD Hamiltonian on this chip using the SOAP optimizer,[50] which is an efficient optimizer tailored for parameter optimization on quantum circuits. Next, we train a neural network based on the sampling output from the optimized quantum circuit. In Fig. 6(b), we report the energy estimates during the optimization process. Sampling from the quantum circuit occurs every 30 steps, with the macro iteration performed 15 times, for a total of 450 iterations. The number of iterations is determined by trial classical simulation, which ensures convergence. For each quantum circuit, we perform 1024 shots of measurement for each Pauli string. The optimization is repeated with three different neural network initializations and the lowest energy is employed for reaction barrier calculation.

As shown in Fig. 6(c), the reaction barrier predicted by pUNN on the quantum circuit is approximately 16 kcal mol^{-1}. While this value is still higher than the experimentally reported range of 2 ∼ 10 kcal mol^{-1},[65] it represents a notable improvement over the HF and MP2 energies, and is comparable to the noiseless UCCSD prediction. When using a noiseless pUNN model, obtained *via* a statevector simulator, the predicted reaction barrier is around 9 kcal mol^{-1}, which aligns well with the FCI results and experimental observations. This highlights the importance of addressing errors introduced by quantum circuit gates and measurement uncertainties. In particular, the neural network parameters with quantum computers are different from the neural network parameters with noiseless simulation. We

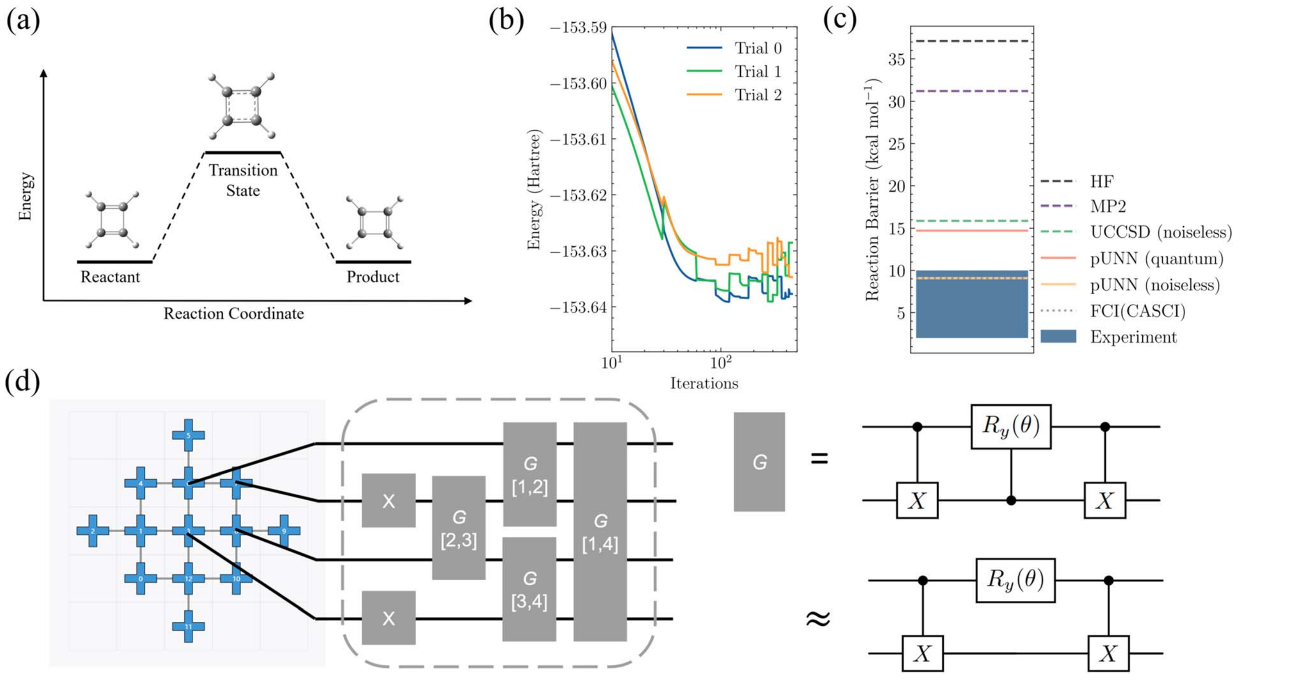

Fig. 6 Experiments on a superconducting quantum computer. (a) The isomerization reaction of cyclobutadiene, with the transition state energy calculated using pUNN on a superconducting quantum computer. (b) The estimated energy during the optimization process. Results by three independent random initializations of the neural network are shown. (c) The computed reaction barrier from pUNN, compared with results from several other computational methods. "Experiment" means the reaction barrier calculated by experimentally observed chemical reaction rate. (d) The 13-qubit superconducting quantum chip and the quantum circuit used for the calculation.

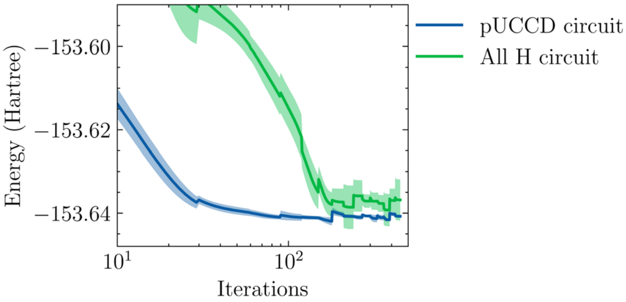

Fig. 7 Energy estimates during neural network training with different quantum circuits. This figure illustrates the effect of quantum circuits on energy estimation within the pUNN framework: (1) the pUCCD circuit, which is the circuit used throughout this paper, and (2) the Hadamard superposition circuit, where Hadamard gates are applied to all qubits, creating a superposition of all possible states. The standard deviation across five different neural network initializations is shown as the shaded area.

conjecture that pUNN(quantum) predicts a higher energy because the neural network parameters are stuck in a local minimum. To improve the performance of pUNN in the presence of these errors, advanced optimizers, such as KFAC,[18,66] could be considered. Yet adaption of the KFAC optimizer will likely be necessary due to the unique algorithmic structure of pUNN.

Next, we investigate the advantage of incorporating quantum computing into the pUNN framework. Since neural networks are widely known for their effectiveness across a variety of tasks including representing molecular wavefunction, it is important to assess whether a quantum circuit is truly necessary for this framework. To explore this, we replace the pUCCD circuit in pUNN with a Hadamard superposition circuit, where Hadamard gates are applied to all qubits. The Hadamard superposition circuit can be easily emulated on classical computers and can be considered as a "dummy" sample generator when used to compute the energy with the neural network. To isolate the impact of quantum gate noise, we perform the comparison using a shot-based classical emulator which is free of gate noise. We use the transition state of the cyclobutadiene isomerization reaction as our model system. As shown in Fig. 7, replacing the pUCCD circuit with a Hadamard superposition leads to a noticeable decrease in accuracy, along with a significant increase in energy variance. In fact, for large molecules, a Hadamard superposition circuit greatly reduces the probability of sampling the dominant configuration, making it less effective for energy estimation. In contrast, the pUCCD circuit provides a suitable starting point for further refinement through neural network training, demonstrating the advantage of quantum computing in this context.

IV Conclusion and outlook

The pUNN framework combines an efficient quantum circuit with the expressive power of a neural network to accurately and robustly compute molecular energies. Through a carefully designed algorithmic structure—including the pUCCD circuit, entanglement circuit, perturbation circuit, and neural network augmentation—the method achieves high accuracy with low quantum resource requirements, utilizing only N qubits instead of the $2N$ qubits typically required by comparable methods. The incorporation of a neural network allows the framework to mitigate errors effectively, making it robust to gate noise and capable of delivering consistent accuracy on noisy quantum hardware. The design also ensures manageable measurement overhead for the interaction between the quantum circuit and the neural network.

Extensive numerical benchmarks demonstrate that pUNN achieves accuracy comparable to advanced methods like UCCSD, while being more resource-efficient and scalable to larger molecular systems. Experimental validation on a superconducting quantum computer demonstrates the practicality of this approach. With a 4-qubit quantum circuit, pUNN successfully computes the transition state energy of cyclobutadiene isomerization, yielding energy estimates with accuracy comparable to noiseless UCCSD. Based on this model reaction, we also demonstrate that the quantum circuit plays an indispensable role in the hybrid framework, as replacing it with a neural network alone leads to a higher error and, crucially, a significantly larger variance in energy estimation. This observation serves as an evidence for the advantage of this hybrid design than pure classical neural networks, where the quantum circuit reduces the representational burden on the neural network. Thus, we expect that pUNN is able to demonstrate quantum advantage as we tackle larger systems where classical simulation of the pUCCD circuit becomes intractable.

While this work focuses on closed–shell systems, the pUNN framework can be directly extended to open-shell systems by modifying the particle number conservation mask in the neural network. However, since the pUCCD quantum circuit may not accurately approximate open-shell wavefunctions, further adaptations will likely be necessary to maintain accuracy for open–shell systems. Future work could enhance the neural network architecture by incorporating more sophisticated neural layers with physical insights. Additionally, pretraining the neural network on a diverse set of molecules offers a possible avenue for creating a generalizable model that can be fine-tuned for specific systems.

Code availability

The code for this study is available from the TenCirChem-NG package hosted on GitHub **https://github.com/tensorcircuit/TenCirChem-NG** and from Zenodo.[67]

Data availability

The source data for the figures in this study is available from Zenodo.[68] DOI: **https://www.doi.org/10.5281/zenodo.15859709**.

Author contributions

Weitang Li: conceptualization, data curation, investigation, methodology, validation, visualization, writing – original draft, supervision, funding acquisition; Shi-Xin Zhang: conceptualization, formal analysis, software, writing – review & editing;

Zirui Sheng: investigation, writing – review & editing; Cunxi Gong: investigation, visualization, writing – review & editing; Jianpeng Chen: investigation, writing – review & editing; Zhigang Shuai: conceptualization, investigation, resources, writing – review & editing, supervision, project administration, funding acquisition.

Conflicts of interest

The authors declare no competing interests.

Appendices

A. The electronic structure problem and the pUCCD ansatz

In this work, we are interested in the second-quantized *ab initio* electronic structure Hamiltonian

$$\hat{H} = \sum_{pq} h_{pq} \hat{a}_p^\dagger \hat{a}_q + \frac{1}{2} \sum_{pqrs} h_{pqrs} \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_r \hat{a}_s + E_{\text{nuc}}, \tag{16}$$

where h_{pq} and $h_{pqrs} = [ps|qr]$ are one-electron and two-electron integrals, and $\hat{a}_p^\dagger$,$\hat{a}_p$ are fermionic creation and annihilation operators, respectively, acting on the p-th spin-orbital.

In order to compute the expectation of eqn (16) on a programmable quantum computer, the symmetry of the creation and annihilation operators has to be taken care of. Creation and annihilation operators for fermions obey the anticommutation relations

$$\begin{aligned} \left\{\hat{a}_i, \hat{a}_j^\dagger\right\} &= \delta_{ij} \\ \left\{\hat{a}_i^\dagger, \hat{a}_j^\dagger\right\} &= \left\{\hat{a}_i, \hat{a}_j\right\} = 0 \end{aligned} \tag{17}$$

On the other hand, the qubit creation operator $\hat{c}^\dagger = \frac{1}{2}(X - iY)$ and annihilation operator $\hat{c} = \frac{1}{2}(X + iY)$ obey the commutation relations

$$\{\hat{c}_i,\hat{c}_j^\dagger\} = \delta_{ij}, [\hat{c}_i^\dagger,\hat{c}_j^\dagger] = [\hat{c}_i,\hat{c}_j] = 0. \tag{18}$$

In this work, when necessary, we employ the Jordan-Wigner transformation to map fermionic ladder operators into qubit operators.

In general, UCC types of ansatz can be written as

$$|\psi(\theta)\rangle = \Pi e^{\theta_k \hat{G}_k} |\phi\rangle. \tag{19}$$

Here, $|\phi\rangle$ is the Hartree–Fock state. For the UCCSD method, $\hat{G}_k$ has the form

$$\hat{G}_k = \begin{cases} \hat{a}_{\text{p}}^\dagger \hat{a}_{\text{q}} - \text{h.c.}, \\ \hat{a}_{\text{p}}^\dagger \hat{a}_{\text{q}}^\dagger \hat{a}_{\text{r}} \hat{a}_{\text{s}} - \text{h.c.} \end{cases} \tag{20}$$

pUCCD is an efficient ansatz requiring only $\mathcal{O}(N)$ circuit depth and half as many qubits as other UCC ansatze.[39,40] pUCCD allows only paired double excitations, which enables one qubit to represent one spatial orbital instead of one spin orbital, and removes the need to perform the fermion–qubit mapping. The subspace in which all states have paired configuration is called the seniority-zero subspace. In this subspace, there are $\mathcal{O}(N^2)$ double excitations, which can be executed on a quantum computer efficiently using a compact circuit. The circuit is composed of a linear depth of Givens–Swap gates, assuming linear qubit connectivity.[40] In the seniority-zero subspace, the Hamiltonian also takes a simpler form, with only N^2 terms:

$$\hat{H} = \sum_p h_p \hat{c}_p^\dagger \hat{c}_p + \sum_{pq} v_{pq} \hat{c}_p^\dagger \hat{c}_q + \sum_{p \neq q} w_{pq} \hat{c}_p^\dagger \hat{c}_p \hat{c}_q^\dagger \hat{c}_q + E_{\text{nuc}}, \tag{21}$$

where $h_p = 2h_{pp}$, $v_{pq} = (pq|pq)$ and $\omega_{pq} = 2(pp|qq) - (pq|pq)$. Here p and q are indices for spatial orbitals. If we use $\hat{n}_{\text{p}} = \hat{c}_{\text{p}}^\dagger \hat{c}_{\text{p}} = \frac{1 - Z}{2}$ to denote occupation number operator, eqn (21) can be converted to a sum of Pauli string where the maximum length of Pauli string is 2. Meanwhile, the first and the third term in eqn (21) have only Z terms and the second term will contribute to XX and YY terms. Thus, the expectation of eqn (21) can be measured in 3 different bases, regardless the number of qubit involved.

B. The measurement protocol for pUNN

To begin with, we describe the measurement protocol when a single quantum circuit is integrated with a neural network, following the ref. 25. Then, we move on to our measurement method that enables efficient measurement of two separate circuits in the pUNN algorithm, defined in eqn (15). In the following, for clarity, we omit the prime symbol for both $\hat{H}'$ and $\hat{N}'$, since $\hat{H}'$ is Pauli string similar to $\hat{H}$, and $\hat{N}'$ follows the definition of $\hat{N}$ in eqn (5).

1. A single quantum circuit. For a single circuit $|\psi\rangle = \sum_k a_k |k\rangle$, where $|k\rangle$ is the computational basis, $\hat{N}$ is written as

$$\hat{N} = \sum_k b_k |k\rangle\langle k|. \tag{22}$$

We then focus on deriving an appropriate form of $\hat{N}\hat{H}\hat{N}$. We assume that both $|\psi\rangle$ and $\hat{N}$ are real-valued. We first derive the measurement protocol for the norm of $|\Psi\rangle = \hat{N}|\psi\rangle$, given by

$$\langle \Psi | \Psi \rangle = \langle \psi | \hat{N}^\dagger \hat{N} | \psi \rangle, \tag{23}$$

where

$$\hat{N}^\dagger \hat{N} = \sum_k {b_k}^2 |k\rangle\langle k|. \tag{24}$$

Clearly, the eigenvectors of $\hat{N}^\dagger \hat{N}$ are $|k\rangle$ and their eigenvalues are $\text{b}_k{}^2$. To compute the norm, we sample bitstrings from $|\psi\rangle$ and multiply the probability of k by ${b_k}^2$. For efficient sampling, b_k should not be too large or small. In other words, a_k must provide a good first-order approximation to the ground state. The same is also true for our measurement protocol for 2 circuits and it highlights the role of the quantum computer in this framework.

Next, we consider the measurement of a Pauli string $\hat{H}$. The main focus is to derive $\hat{N}^\dagger \hat{H} \hat{N}$. In general, a Pauli string $\hat{H}$ can be written as

$$\hat{H} = \sum_{k} S_{\tilde{k}} \left|\tilde{k}\right\rangle \langle k|, \tag{25}$$

where the summation is over k rather than k and $\tilde{k}$. In other words, applying the Pauli string $\hat{H}$ on $|k\rangle$ will produce only one bitstring $|\tilde{k}\rangle$ up to a phase $S_{\tilde{k}}$

$$\hat{H}|k\rangle = S_{\tilde{k}}|\tilde{k}\rangle. \tag{26}$$

Since $\hat{H}^2 = I$, we also have $\hat{H}|\tilde{k}\rangle = S_k|k\rangle$ and $S_{\tilde{k}}S_k = 1$.

Let's first consider the case where $\hat{H}$ has only Z operators, *i.e.* $|k\rangle = |\tilde{k}\rangle$. In this case, the term to measure is

$$\hat{N}^{\dagger}\hat{H}\hat{N} = \sum_{k} {b_k}^2 S_k |k\rangle\langle k|. \tag{27}$$

eqn (27) is similar to the expression for $\hat{N}^{\dagger}\hat{N}$ in eqn (24). As a result, the measurement protocol when $\hat{H}$ only involves Z operators is very similar to the procedure for measuring the norm of the state.

Now, consider the general case where $\hat{H}$ includes at least one X or Y operator, where it is ensured that $|k\rangle \neq |\tilde{k}\rangle$. In this case, we can rewrite $\hat{H}$ as

$$\hat{H} = \sum_{k\in\Omega} \left(S_k|k\rangle \left\langle\tilde{k}\right| + S_{\tilde{k}}\left|\tilde{k}\right\rangle\langle k| \right), \tag{28}$$

where $\Omega = \{k|\text{bin}(k) < \text{bin}(\tilde{k})\}$ and bin(k) refers to the corresbonding binary integer of k.

The Hamiltonian transformed by $\hat{N}$ is given by

$$\hat{N}^{\dagger}\hat{H}\hat{N} = \sum_{k\in\Omega} b_k b_{\tilde{k}} \hat{H}_k, \tag{29}$$

where

$$\hat{H}_k = S_k|k\rangle\langle\tilde{k}| + S_{\tilde{k}}|\tilde{k}\rangle\langle k|. \tag{30}$$

To measure $\hat{N}^{\dagger}\hat{H}\hat{N}$, we need to derive the eigenvectors of $\hat{H}_k$. $\hat{H}_k$ is defined by two basis $|k\rangle$ and $|\tilde{k}\rangle$ and therefore $\hat{H}_k$ has two eigenvectors with eigenvalues +1 and −1. Denote the two eigenvectors as $|k^+\rangle$ and $|k^-\rangle$, we can then write $\hat{H}_k$ as

$$\hat{H}_k = |k^+\rangle\langle k^+| - |k^-\rangle\langle k^-|. \tag{31}$$

In the computational basis, $|k^+\rangle$ and $|k^-\rangle$ are written as

$$\begin{aligned}\sqrt{2}|k^+\rangle &= S_{\tilde{k}}\left|\tilde{k}\right\rangle + |k\rangle = \left(\hat{H}_k + 1\right)|k\rangle,\\ \sqrt{2}|k^-\rangle &= S_{\tilde{k}}\left|\tilde{k}\right\rangle - |k\rangle = \left(\hat{H}_k - 1\right)|k\rangle.\end{aligned} \tag{32}$$

These eigenvectors have eigenvalues +1 and −1, respectively. The neural network transformed Hamiltonian is then

$$\hat{N}^{\dagger}\hat{H}\hat{N} = \sum_{k\in\Omega} b_k b_{\tilde{k}} \left(|k^+\rangle\langle k^+| - |k^-\rangle\langle k^-|\right). \tag{33}$$

To perform the measurement in the $|k^+\rangle$ and $|k^-\rangle$ bases, we append a unitary measurement circuit V to the original quantum circuit $|\psi\rangle$. V satisfies

$$\begin{aligned}V^{\dagger}|k\rangle &= \left|k^+\right\rangle,\\ V^{\dagger}\left|\tilde{k}\right\rangle &= |k\ \rangle,\end{aligned} \tag{34}$$

for any $k \in \Omega$. The unitary property can be proven by considering $\langle k'|VV^{\dagger}|k\rangle$ or by noting that V is a permutation between two sets of orthonormal basis states. The construction of the transformation circuit $\hat{V}$ is a standard procedure in quantum computation, because $\hat{V}$ is a circuit that diagonalizes the Pauli string $\hat{H}$. If the number of X and Y operators in $\hat{H}$ is m, then the number of two-qubit gates in $\hat{V}$ is $m - 1$.

To summarize, the quantum circuit used for the measurement is $\hat{V}|\psi\rangle$, and the term to measure is

$$\hat{V}\hat{N}^{\dagger}\hat{H}\hat{N}\hat{V}^{\dagger} = \sum_{k\in\Omega} b_k b_{\tilde{k}} \left(|k\rangle\langle k| - \left|\tilde{k}\right\rangle\left\langle\tilde{k}\right|\right). \tag{35}$$

The expectation value of this term is readily accessible from the quantum circuit $\hat{V}|\psi\rangle$ by performing a projection measurement in the computational basis.

2. Two separate quantum circuits. If we take the two separate quantum circuit $|\psi \otimes \phi\rangle$ as a whole, the measurement protocol developed in Appendix B 1 can be applied to measure the expectation when a neural network is integrated with $|\psi \otimes \phi\rangle$. However, in this case, the unitary transformation for measurement V will generally entangle the two originally unentangled quantum circuits. This results in a quantum circuit of $2N$ qubits. If we wish to avoid this entanglement and measure the expectation using two separate quantum circuits, a special measurement procedure is needed. This procedure will be described in the following.

The total wavefunction is expressed as:

$$|\Psi\rangle = \hat{N}|\psi \otimes \phi\rangle. \tag{36}$$

In the pUNN framework, $|\psi\rangle$ is the pUCCD quantum circuit, and $|\phi\rangle$ is the perturbation circuit to be simulated classically. However, the procedure outlined below is general and can be readily applied to other cases involving uncorrelated circuits.

Consider Hamiltonian in the form:

$$\hat{H} = \hat{H}_{\psi} \otimes \hat{H}_{\phi}, \tag{37}$$

where $\hat{H}_{\psi}$ and $\hat{H}_{\phi}$ are Pauli strings for the two separate circuits. If either of $\hat{H}_{\psi}$ and $\hat{H}\phi$ does not contain X or Y, the measurement procedure simplifies to the standard approach described in Appendix B 1. Therefore, we will focus on the general case where both $\hat{H}_{\psi}$ and $\hat{H}_{\phi}$ contain X or Y. Similar to eqn (26), $\hat{H}_{\psi}$ and $\hat{H}_{\phi}$ satisfy the following relations:

$$\begin{aligned}\hat{H}_{\psi}|k\rangle &= S_{\tilde{k}}\left|\tilde{k}\right\rangle,\\ \hat{H}_{\phi}|j\rangle &= S_{\tilde{j}}|\tilde{j}\rangle.\end{aligned} \tag{38}$$

Here, $\hat{H}_{\psi}$ and $\hat{H}_{\phi}$ act independently on the circuit $|\psi\rangle$ and $|\phi\rangle$, transforming the states $|k\rangle$ and $|j\rangle$ into $|\tilde{k}\rangle$ and $|\tilde{j}\rangle$, with corresponding signs $S_{\tilde{k}}$ and $S_{\tilde{j}}$.

The eigenvectors of $\hat{H}$ are given by

$$2|k^{\pm}\rangle|j^{\pm}\rangle = (S_{\tilde{k}}|\tilde{k}\rangle \pm |k\rangle)(S_{\tilde{j}}|\tilde{j}\rangle \pm |j\rangle), \tag{39}$$

where we again require $k \in \Omega_{\psi}$ and $j \in \Omega_{\phi}$ to avoid double-counting. In the following, we use $k, j \in \Omega$ as a short-hand notation for the condition.

The Hamiltonian in the computational basis is

$$\begin{aligned}\hat{H} &= \sum_{k,j\in\Omega}\left(S_k|k\rangle\left\langle\tilde{k}\right| + S_{\tilde{k}}\left|\tilde{k}\right\rangle\langle k|\right)\otimes\left(S_j|j\rangle\langle\tilde{j}| + S_{\tilde{j}}|\tilde{j}\rangle\langle j|\right)\\ &= \sum_{k,j\in\Omega}\left(S_kS_j|k,j\rangle\left\langle\tilde{k},\tilde{j}\right| + S_{\tilde{k}}S_{\tilde{j}}\left|\tilde{k},\tilde{j}\right\rangle\langle k,j|\right)\\ &\quad+\sum_{k,j\in\Omega}\left(S_kS_{\tilde{j}}\left|k,\tilde{j}\right\rangle\left\langle\tilde{k},j\right| + S_{\tilde{k}}S_j\left|\tilde{k},j\right\rangle\left\langle k,\tilde{j}\right|\right).\end{aligned} \tag{40}$$

After applying the transformation $\hat{N}$, the transformed Hamiltonian becomes

$$\begin{aligned}N^{\dagger}\hat{\hat{H}}\hat{N} = &\sum_{k,j\in\Omega}b_{kj}b_{\tilde{k}\tilde{j}}\left(S_kS_j|k,j\rangle\left\langle\tilde{k},\tilde{j}\right| + S_{\tilde{k}}S_{\tilde{j}}\left|\tilde{k},\tilde{j}\right\rangle\langle k,j|\right)\\ &+\sum_{k,j\in\Omega}b_{k\tilde{j}}b_{\tilde{k}j}\left(S_kS_{\tilde{j}}\left|k,\tilde{j}\right\rangle\left\langle\tilde{k},j\right| + S_{\tilde{k}}S_j\left|\tilde{k},j\right\rangle\left\langle k,\tilde{j}\right|\right).\end{aligned} \tag{41}$$

The structure of $\hat{N}^{\dagger}\hat{H}\hat{N}$ remains similar to $\hat{H}$, but the terms are now weighted by the neural network coefficients b_{kj}. Eqn (41) is more complex than eqn (29) since each term can not be readily factored into the direct product of operators acting on the two separate circuits. Consequently, finding a measurement circuit that diagonalizes eqn (41) without introducing entanglement between the two circuits is not straightforward.

To proceed, it is instructive to consider a 2-qubit system and with the Hamiltonian $\hat{H} = \mathrm{XX}$ as an example. In this case, we can express the neural network transformed Hamiltonian as:

$$\begin{aligned}\hat{N}^{\dagger}\hat{H}\hat{N} &= b_{00}b_{11}(|00\rangle\langle 11| + |11\rangle\langle 00|) + b_{01}b_{10}(|01\rangle\langle 10| + |10\rangle\langle 01|)\\ &= \frac{1}{2}b_{00}b_{11}(\mathrm{XX}-\mathrm{YY}) + \frac{1}{2}b_{01}b_{10}(\mathrm{XX}+\mathrm{YY})\\ &= \frac{1}{2}(b_{00}b_{11}+b_{01}b_{10})\mathrm{XX} + \frac{1}{2}(-b_{00}b_{11}+b_{01}b_{10})\mathrm{YY}.\end{aligned} \tag{42}$$

Thus, to measure the expectation value of XX in the presence of a NN, one needs to measure both XX and YY to avoid measurement circuit that entangles the two separate circuits.

More generally, consider a Hamiltonian $\hat{J} = \hat{J}_{\psi} \otimes \hat{J}_{\phi}$ such that (as in eqn (38))

$$\begin{aligned}\hat{J}_{\psi}|k\rangle &= iS_{\tilde{k}}\left|\tilde{k}\right\rangle,\\ \hat{J}_{\phi}|j\rangle &= iS_{\tilde{j}}|\tilde{j}\rangle.\end{aligned} \tag{43}$$

$\hat{J}$ can be constructed by replacing an X operator with $-Y$ or a Y operator with X in $\hat{H}_{\psi}$ and $\hat{H}_{\phi}$. The eigenvectors of $\hat{J}$ are

$$2|k^{i\pm}\rangle|j^{i\pm}\rangle = (iS_{\tilde{k}}|\tilde{k}\rangle \pm |k\rangle)(iS_{\tilde{j}}|\tilde{j}\rangle \pm |j\rangle). \tag{44}$$

We define short-hand notation for the projectors

$$\hat{h_k}^{\pm} = |k^{\pm}\rangle\langle k^{\pm}| \tag{45}$$

which form the diagonal bases for $\hat{H}$ and $\hat{J}$. The first term of $\hat{N}^{\dagger}\hat{H}\hat{N}$ from eqn (41) is then transformed to:

$$\begin{aligned}S_kS_j|k,j\rangle\left\langle\tilde{k},\tilde{j}\right| + S_{\tilde{k}}S_{\tilde{j}}\left|\tilde{k},\tilde{j}\right\rangle\langle k,j| = \frac{1}{2}\left(\hat{h}_k{}^{+} - \hat{h}_k{}^{-}\right)\otimes\\ \left(\hat{h}_j{}^{+} - \hat{h}_j{}^{-}\right) - \frac{1}{2}\left(\hat{h}_k{}^{i+} - \hat{h}_k{}^{i-}\right)\otimes\left(\hat{h}_j{}^{i+} - \hat{h}_j{}^{i-}\right).\end{aligned} \tag{46}$$

Similarly, the second term becomes

$$\begin{aligned}S_kS_{\tilde{j}}\left|k,\tilde{j}\right\rangle\left\langle\tilde{k},j\right| + S_{\tilde{k}}S_j\left|\tilde{k},j\right\rangle\left\langle k,\tilde{j}\right| = \frac{1}{2}\left(\hat{h}_k{}^{+} - \hat{h}_k{}^{-}\right)\otimes\\ \left(\hat{h}_j{}^{+} - \hat{h}_j{}^{-}\right) + \frac{1}{2}\left(\hat{h}_k{}^{i+} - \hat{h}_k{}^{i-}\right)\otimes\left(\hat{h}_j{}^{i+} - \hat{h}_j{}^{i-}\right).\end{aligned} \tag{47}$$

The overall expression for $\hat{N}^{\dagger}\hat{H}\hat{N}$ is

$$\begin{aligned}\hat{N}^{\dagger}\hat{H}\hat{N} = &\sum_{k,j\in\Omega}\frac{1}{2}\left(b_{kj}b_{\tilde{k}\tilde{j}} + b_{k\tilde{j}}b_{\tilde{k}j}\right)\left(\hat{h}_k{}^{+} - \hat{h}_k{}^{-}\right)\otimes\left(\hat{h}_j{}^{+} - \hat{h}_j{}^{-}\right)\\ &+\sum_{k,j\in\Omega}\frac{1}{2}\left(-b_{kj}b_{\tilde{k}\tilde{j}} + b_{k\tilde{j}}b_{\tilde{k}j}\right)\left(\hat{h}_k{}^{i+} - \hat{h}_k{}^{i-}\right)\otimes\left(\hat{h}_j{}^{i+} - \hat{h}_j{}^{i-}\right)\end{aligned} \tag{48}$$

One may verify the equation by setting $b = 1$ and $\hat{N}$ becomes $\hat{I}$. In this case, the second term vanishes and the first term reduces to the original Hamiltonian $\hat{H}$. In eqn (48), the operators are factored into the direct product of operators acting on the two separate circuits. Consequently, they can be diagonalized to the computational basis separately following the approach discussed in Appendix B 1.

Thus, in order to measure the expectation in eqn (15), one has to sample bitstrings from both $\hat{H}$ and $\hat{J}$ and calculate the expectation following eqn (48) accordingly. In the framework of pUNN, we first sample bitstrings that correspond to $\hat{h}_k^{\pm}$ and $\hat{h}_k^{i\pm}$ on quantum computers, and then sample bitstrings that correspond to $\hat{h}_j^{\pm}$ and $\hat{h}_j^{i\pm}$ on classical simulators. Then we query the neural network $\mathscr{B}$ for $b_{k,j}$, and finally calculate the expectation based on the sampling statistics and the output from the neural network.

Acknowledgements

We acknowledge Tong Jiang for helpful discussions. This work is supported by National Natural Science Foundation of China (T2350009, 22433007), the Young Elite Scientists Sponsorship Program by CAST (2023QNRC001), University Development Fund (UDF01003789), and the Shenzhen Science and Technology Program (No. KQTD20240729102028011). Shi-Xin Zhang acknowledges the support from a start-up grant at IOP-CAS.

References

1 A. J. Daley, I. Bloch, C. Kokail, S. Flannigan, N. Pearson, M. Troyer, *et al.*, Practical quantum advantage in quantum simulation, *Nature*, 2022, **607**(7920), 667–676.

2 A. D. King, A. Nocera, M. M. Rams, J. Dziarmaga, R. Wiersema, W. Bernoudy, *et al.*, Computational supremacy in quantum simulation, *arXiv*, 2024, preprint, arXiv: 240300910, DOI: 10.48550/arXiv.240300910.

3 G. K. L. Chan, Quantum chemistry, classical heuristics, and quantum advantage, *Faraday Discuss.*, 2024, **254**, 11–52.

4 A. Peruzzo, J. McClean, P. Shadbolt, M. H. Yung, X. Q. Zhou, P. J. Love, *et al.*, A variational eigenvalue solver on a photonic quantum processor, *Nat. Commun.*, 2014, **5**, 4213.

5 Google AI Quantum and Collaborators, F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, *et al.*, Hartree-Fock on a superconducting qubit quantum computer, *Science*, 2020, **369**(6507), 1084–1089.

6 M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, *et al.*, Variational quantum algorithms, *Nat. Rev. Phys.*, 2021, **3**(9), 625–644.

7 J. Tilly, H. Chen, S. Cao, D. Picozzi, K. Setia, Y. Li, *et al.*, The variational quantum eigensolver: a review of methods and best practices, *Phys. Rep.*, 2022, **986**, 1–128.

8 K. Huang, X. Cai, H. Li, Z. Y. Ge, R. Hou, H. Li, *et al.*, Variational quantum computation of molecular linear response properties on a superconducting quantum processor, *J. Phys. Chem. Lett.*, 2022, **13**(39), 9114–9121.

9 S. Guo, J. Sun, H. Qian, M. Gong, Y. Zhang, F. Chen, *et al.*, Experimental quantum computational chemistry with optimized unitary coupled cluster ansatz, *Nat. Phys.*, 2024, **20**, 1240–1246.

10 X. Yuan, S. Endo, Q. Zhao, Y. Li and S. C. Benjamin, Theory of variational quantum simulation, *Quantum*, 2019, **3**, 191.

11 J. Lee, W. J. Huggins, M. Head-Gordon and K. B. Whaley, Generalized unitary coupled cluster wave functions for quantum computation, *J. Chem. Theory Comput.*, 2018, **15**(1), 311–324.

12 H. R. Grimsley, S. E. Economou, E. Barnes and N. J. Mayhall, An adaptive variational algorithm for exact molecular simulations on a quantum computer, *Nat. Commun.*, 2019, **10**, 3007.

13 J. Sun, L. Cheng and W. Li, Toward Chemical Accuracy with Shallow Quantum Circuits: A Clifford-Based Hamiltonian Engineering Approach, *J. Chem. Theory Comput.*, 2024, **20**(2), 695–707.

14 X. Xiao, H. Zhao, J. Ren, W. H. Fang and Z. Li, Physics-constrained hardware-efficient ansatz on quantum computers that is universal, systematically improvable, and size-consistent, *J. Chem. Theory Comput.*, 2024, **20**(5), 1912–1922.

15 J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush and H. Neven, Barren plateaus in quantum neural network training landscapes, *Nat. Commun.*, 2018, **9**(1), 4812.

16 J. Hermann, J. Spencer, K. Choo, A. Mezzacapo, W. M. C. Foulkes, D. Pfau, *et al.*, Ab initio quantum chemistry with neural-network wavefunctions, *Nat. Rev. Chem.*, 2023, **7**(10), 692–709.

17 J. Han, L. Zhang and E. Weinan, Solving many-electron Schrödinger equation using deep neural networks, *J. Comput. Phys.*, 2019, **399**, 108929.

18 D. Pfau, J. S. Spencer, A. G. Matthews and W. M. C. Foulkes, Ab initio solution of the many-electron Schrödinger equation with deep neural networks, *Phys. Rev. Res.*, 2020, **2**(3), 033429.

19 J. Hermann, Z. Schätzle and F. Noé, Deep-neural-network solution of the electronic Schrödinger equation, *Nat. Chem.*, 2020, **12**(10), 891–897.

20 H. Shang, C. Guo, Y. Wu, Z. Li, and J. Yang, Solving Schrödinger equation with a language model, *arXiv*, 2023, preprint, arXiv: 230709343, DOI: **10.48550/arXiv.230709343**.

21 X. Li, Z. Li and J. Chen, Ab initio calculation of real solids via neural network ansatz, *Nat. Commun.*, 2022, **13**(1), 7895.

22 M. Scherbela, L. Gerard and P. Grohs, Towards a transferable fermionic neural wavefunction for molecules, *Nat. Commun.*, 2024, **15**(1), 120.

23 X. Li, J. C. Huang, G. Z. Zhang, H. E. Li, Z. P. Shen, C. Zhao, *et al.*, Improved optimization for the neural-network quantum states and tests on the chromium dimer, *J. Chem. Phys.*, 2024, **160**(23), 234102.

24 J. Nys, G. Pescia, and G. Carleo, Ab-initio variational wave functions for the time-dependent many-electron Schrödinger equation, *arXiv*, 2024, preprint arXiv: 240307447, DOI: **10.48550/arXiv.240307447**.

25 S. X. Zhang, Z. Q. Wan, C. K. Lee, C. Y. Hsieh, S. Zhang and H. Yao, Variational quantum-neural hybrid eigensolver, *Phys. Rev. Lett.*, 2022, **128**(12), 120502.

26 T. Westerhout, N. Astrakhantsev, K. S. Tikhonov, M. I. Katsnelson and A. A. Bagrov, Generalization properties of neural network approximations to frustrated magnet ground states, *Nat. Commun.*, 2020, **11**(1), 1593.

27 T. Jiang, J. Zhang, M. K. Baumgarten, M. F. Chen, H. Q. Dinh, A. Ganeshram, *et al.*, Walking through Hilbert space with quantum computers, *arXiv*, 2024, preprint, arXiv: 240711672, DOI: **10.48550/arXiv.240711672**.

28 J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe and S. Lloyd, Quantum machine learning, *Nature*, 2017, **549**(7671), 195–202.

29 M. Cerezo, G. Verdon, H. Y. Huang, L. Cincio and P. J. Coles, Challenges and opportunities in quantum machine learning, *Nat. Comput. Sci.*, 2022, **2**(9), 567–576.

30 W. Ren, W. Li, S. Xu, K. Wang, W. Jiang, F. Jin, *et al.*, Experimental quantum adversarial learning with programmable superconducting qubits, *Nat. Comput. Sci.*, 2022, **2**(11), 711–717.

31 X. K. Li, J. X. Ma, X. Y. Li, J. J. Hu, C. Y. Ding, F. K. Han, *et al.*, High-efficiency reinforcement learning with hybrid architecture photonic integrated circuit, *Nat. Commun.*, 2024, **15**(1), 1044.

32 J. Li and S. Kais, Quantum cluster algorithm for data classification, *Mat. Theory*, 2021, **5**(6), 1–14.

33 M. Sajjan, S. H. Sureshbabu and S. Kais, Quantum machine-learning for eigenstate filtration in two-dimensional materials, *J. Am. Chem. Soc.*, 2021, **143**(44), 18426–18445.

34 M. Sajjan, J. Li, R. Selvarajan, S. H. Sureshbabu, S. S. Kale, R. Gupta, *et al.*, Quantum machine learning for chemistry and physics, *Chem. Soc. Rev.*, 2022, **51**(15), 6475–6573.

35 X. Zeng, Y. Fan, J. Liu, Z. Li and J. Yang, Quantum neural network inspired hardware adaptable ansatz for efficient

quantum simulation of chemical systems, *J. Chem. Theory Comput.*, 2023, **19**(23), 8587–8597.

36 S. Halder, C. Patra, D. Mondal and R. Maitra, Machine learning aided dimensionality reduction toward a resource efficient projective quantum eigensolver: Formal development and pilot applications, *J. Chem. Phys.*, 2023, **158**(24), 244101.

37 S. Halder, A. Dey, C. Shrikhande and R. Maitra, Machine learning assisted construction of a shallow depth dynamic ansatz for noisy quantum hardware, *Chem. Sci.*, 2024, **15**(9), 3279–3289.

38 H. Shang, X. Zeng, M. Gong, Y. Wu, S. Guo, H. Qian, *et al.*, Rapidly Achieving Chemical Accuracy with Quantum Computing Enforced Language Model, *arXiv*, 2024, preprint, arXiv: 240509164, DOI: **10.48550/arXiv.240509164**.

39 T. M. Henderson, I. W. Bulik and G. E. Scuseria, Pair extended coupled cluster doubles, *J. Chem. Phys.*, 2015, **142**(21), 214116.

40 V. E. Elfving, M. Millaruelo, J. A. Gámez and C. Gogolin, Simulating quantum chemistry in the seniority-zero space on qubit-based quantum computers, *Phys. Rev. A*, 2021, **103**(3), 032605.

41 T. E. O'Brien, G. Anselmetti, F. Gkritsis, V. E. Elfving, S. Polla, W. J. Huggins, *et al.*, Purification-based quantum error mitigation of pair-correlated electron simulations, *Nat. Phys.*, 2023, **19**(12), 1787–1792, DOI: **10.1038/s41567-023-02240-y**.

42 L. Zhao, J. Goings, K. Shin, W. Kyoung, J. I. Fuks, J. K. Kevin Rhee, *et al.*, Orbital-optimized pair-correlated electron simulations on trapped-ion quantum computers, *Quantum Inf.*, 2023, **9**(1), 60.

43 I. Khan, M. Tudorovskaya, J. Kirsopp, D. Muñoz Ramo, P. Warrier, D. Papanastasiou, *et al.*, Chemically aware unitary coupled cluster with ab initio calculations on an ion trap quantum computer: A refrigerant chemicals' application, *J. Chem. Phys.*, 2023, **158**(21), 214114.

44 W. J. Huggins, B. A. O'Gorman, N. C. Rubin, D. R. Reichman, R. Babbush and J. Lee, Unbiasing fermionic quantum Monte Carlo with a quantum computer, *Nature*, 2022, **603**(7901), 416–420.

45 G. Mazzola, P. J. Ollitrault, P. K. Barkoutsos and I. Tavernelli, Nonunitary operations for ground-state calculations in near-term quantum computers, *Phys. Rev. Lett.*, 2019, **123**(13), 130501, DOI: **10.1103/PhysRevLett.123.130501**.

46 F. Benfenati, G. Mazzola, C. Capecci, P. K. Barkoutsos, P. J. Ollitrault, I. Tavernelli, *et al.*, Improved accuracy on noisy devices by nonunitary variational quantum eigensolver for chemistry applications, *J. Chem. Theory Comput.*, 2021, **17**(7), 3946–3954, DOI: **10.1021/acs.jctc.1c00091**.

47 Z. X. Shang, M. C. Chen, X. Yuan, C. Y. Lu and J. W. Pan, Schrödinger-Heisenberg variational quantum algorithms, *Phys. Rev. Lett.*, 2023, **131**(6), 060406. Available from: **https://link.aps.org/doi/10.1103/PhysRevLett.131.060406**.

48 S. Zhang, Z. Q. Wan, C. Y. Hsieh, H. Yao and S. Zhang, Variational Quantum-Neural Hybrid Error Mitigation, *Adv. Quantum Technol.*, 2023, **6**(10), 202300147, DOI: **10.1002/qute.202300147**.

49 A. Eddins, M. Motta, T. P. Gujarati, S. Bravyi, A. Mezzacapo, C. Hadfield, *et al.*, Doubling the size of quantum simulators by entanglement forging, *PRX Quantum*, 2022, **3**(1), 010309.

50 W. Li, Y. Ge, S. X. Zhang, Y. Q. Chen and S. Zhang, Efficient and Robust Parameter Optimization of the Unitary Coupled-Cluster Ansatz, *J. Chem. Theory Comput.*, 2024, **20**(9), 3683–3696.

51 D. P. Kingma, Adam: A method for stochastic optimization, *arXiv*, 2014, preprint, arXiv: 14126980, DOI: **10.48550/arXiv.14126980**.

52 S. X. Zhang, J. Allcock, Z. Q. Wan, S. Liu, J. Sun, H. Yu, *et al.*, Tensorcircuit: a quantum software framework for the NISQ era, *Quantum*, 2023, **7**, 912.

53 W. Li, J. Allcock, L. Cheng, S. X. Zhang, Y. Q. Chen, J. P. Mailoa, *et al.*, TenCirChem: An Efficient Quantum Computational Chemistry Package for the NISQ Era, *J. Chem. Theory Comput.*, 2023, **19**(13), 3966–3981.

54 Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt, *et al.*, Recent developments in the PySCF program package, *J. Chem. Phys.*, 2020, **153**(2), 024109.

55 I. O. Sokolov, P. K. Barkoutsos, P. J. Ollitrault, D. Greenberg, J. Rice, M. Pistoia, *et al.*, Quantum orbital-optimized unitary coupled cluster methods in the strongly correlated regime: Can quantum algorithms outperform their classical equivalents?, *J. Chem. Phys.*, 2020, **152**(12), 124107.

56 W. Mizukami, K. Mitarai, Y. O. Nakagawa, T. Yamamoto, T. Yan and Y. Y. Ohnishi, Orbital optimized unitary coupled cluster theory for quantum computer, *Phys. Rev. Res.*, 2020, **2**(3), 033421.

57 Z. Li, S. Guo, Q. Sun and G. K. L. Chan, Electronic landscape of the P-cluster of nitrogenase as revealed through many-electron quantum wavefunction simulations, *Nat. Chem.*, 2019, **11**(11), 1026–1033.

58 H. R. Larsson, H. Zhai, C. J. Umrigar and G. K. L. Chan, The chromium dimer: closing a chapter of quantum chemistry, *J. Am. Chem. Soc.*, 2022, **144**(35), 15932–15937.

59 J. Hu, Q. Wang and S. Li, Unitary Block-Correlated Coupled Cluster Ansatz Based on the Generalized Valence Bond Wave Function for Quantum Simulation, *J. Chem. Theory Comput.*, 2025, **21**(9), 4579–4590.

60 G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, *Science*, 2017, **355**(6325), 602–606.

61 G. Pescia, J. Nys, J. Kim, A. Lovato and G. Carleo, Message-passing neural quantum states for the homogeneous electron gas, *Phys. Rev. B*, 2024, **110**(3), 035108.

62 D. I. Lyakh, M. Musiał, V. F. Lotrich and R. J. Bartlett, Multireference nature of chemistry: The coupled-cluster view, *Chem. Rev.*, 2012, **112**(1), 182–243.

63 T. H. Dunning, Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen, *J. Chem. Phys.*, 1989, **90**(2), 1007–1023, DOI: **10.1063/1.456153**.

64 I. Magoulas and F. A. Evangelista, Linear-scaling quantum circuits for computational chemistry, *J. Chem. Theory Comput.*, 2023, **19**(15), 4815–4821.

65 D. W. Whitman and B. K. Carpenter, Limits on the activation parameters for automerization of cyclobutadiene-1, 2-*d*2, *J. Am. Chem. Soc.*, 1982, **104**(23), 6473–6474.

66 J. Martens and R. Grosse, Optimizing neural networks with Kronecker-factored approximate curvature, in *International Conference on Machine Learning*, PMLR, 2015, pp. 2408–2417.

67 TenCirChem-NG, available from DOI: **10.5281/zenodo.15860167**.

68 Data to Reproduce Figures, available from DOI: **10.5281/zenodo.15859709**.