

Experimental quantum computational chemistry with optimized unitary coupled cluster ansatz

In the format provided by the
authors and unedited

CONTENTS

I. Device and calibration	2
A. Device	2
B. System calibration	2
C. Correlated readout error and random state measurements	3
II. Variational quantum state preparation and measurement	5
A. Initial state preparation	5
B. Quantum circuit ansatz design and compilation	6
1. Unitary coupled cluster circuit ansatz	6
2. Point group symmetry constraint	8
3. Selection of excitation operators	8
4. Quantum circuit compilation	9
C. Optimisation by analytic stochastic gradient descent	10
D. Quantum state measurement and classical post-processing	12
III. Error-mitigated observable estimation	12
A. Readout error mitigation	13
B. Gate error learning by Clifford fitting	14
1. Methods	14
2. Experimental procedure	15
C. Symmetry verification	16
D. Connected moments expansion	17
E. Experimental procedure for observable estimation	18
IV. Experimental implementation and results	18
A. Experimental procedure	19
B. Experimental results	19
C. Parameters iteration	22
D. Noise analysis	23
E. Challenges	23
References	25

I. DEVICE AND CALIBRATION

A. Device

The “Zuchongzhi 2.0” quantum processor contains 66 qubits and 110 couplers. In order to control and wire such a large-scale quantum processor, the qubit layer and wiring layer are prepared on separate chips, which are then combined using the flip-chip packaging process. High-purity aluminium thin films are grown on a sapphire substrate by molecular beam epitaxy (MBE) for both chips, which can reduce material loss and improve qubit decoherence performance. The wiring layer employs photolithography and etching processes to fabricate readout circuits and control circuits and utilizes a lift-off process for creating airbridges. These airbridges effectively establish grounding connections, which prevent slot line modes, and simultaneously reduce crosstalk between control circuits. This results in superior qubit isolation and enables high-precision gate operations. The qubits and couplers are fabricated on the qubit layer. In this process, the capacitive elements of the qubits and couplers are defined using photolithography and etching. The Josephson junction is created by defining the junction area using electron beam equipment and then prepared using a dual-angle evaporation process followed by a lift-off process. During the fabrication process, we fine-tuned the junction’s critical current by adjusting the oxidation conditions and the size of the junction region. This allowed us to achieve the optimal frequency adjustment range for both qubits and couplers. Furthermore, a considerable number of indium bumps as superconducting interconnects are employed to establish a solid ground connection between the qubit layer and the wiring layer, effectively reducing parasitic modes.

Following the chip fabrication process, the processor is wire-bonded to a printed circuit board located within a sample box. This sample box is equipped with gold-plated shielding inside and μ -metal shielding outside. Finally, the packaged processor is mounted to the plate of a dilution refrigerator and connected to room-temperature electronics via a series of attenuators, filters, and amplifiers.

B. System calibration

Our experiment is carried out on the “Zuchongzhi 2.0” quantum processor. We selected up to 12 qubits to run the quantum circuits. We turned off the coupling between used and unused qubits. Then, we optimised and calibrated the performance of the selected qubits, especially the quantum gates and readout performance.

The optimisation procedure of our quantum processor can be summarised into the following steps:

1. **Select idle frequencies.** Idle frequencies refer to the qubit operating frequencies when the qubits are in idle operations or single-qubit gates. As the qubit performance varies strongly with the frequency, setting the qubits in good idle frequencies is necessary to achieve high-fidelity gates. To arrange the idle frequencies, we need to consider a combination of factors, including coherence time, residual coupling between qubits, the distance between idle frequencies and CZ interaction frequencies (related to z distortions), two-level systems (TLSs), and XY-crosstalk. XY-crosstalk comes from the interaction between control lines and the distance of the driven frequencies.
2. **Turn off the coupling between qubits.** In the setting of idle frequencies, the frequency distance between nearest-neighbour qubits is about 100MHz. Turning off the coupling to suppress the interaction between qubits is necessary to decrease parallel-gate errors. After setting qubit idle frequencies, we determine the coupler zero-coupling point by measuring the relative two-qubit conditional phase when varying the flux of the coupler. Then, we turn off the qubit-qubit coupling by adjusting the coupler flux bias to the selected zero-coupling point.
3. **Optimise single-qubit gates.** We fine-tune the XY drive pulse amplitudes and apply Derivative Reduction by Adiabatic Gate (DRAG) correction to the pulse envelope with a length of 25 ns.
4. **Optimise the Josephson Parametric Amplifiers (JPAs).** The JPA can significantly improve the readout signal-to-noise ratio (SNR) and fidelity [1, 2]. First, we coarsely tune the JPA direct current (DC) bias, pump signal power and frequency according to our design. Then, these parameters are optimised by the Nelder–Mead algorithm, and the target function is the average SNRs of the qubit readout signals.
5. **Optimise the pulse distortion of each qubit.** Distortion or reflection occurs when the pulse is transmitted from a waveform generator to qubits, and it will cause a shift in the qubit frequency which increases the gate error in the quantum circuit. In our experiment, we first measure the effect of pulse distortion on qubit phase accumulation over a fixed time of 400 ns. Then we calculate the waveform after distortion based on the result of phase accumulation and apply a reverse one to compensate for the pulse. The full protocol is detailed in Ref [3]

6. **Optimise the readout crosstalk.** Readout crosstalk may be generated by factors such as the influence of neighbouring qubits on readout ones, coupling between readout resonators, frequency collision (caused by the overlap of demodulation frequencies, intermodulation of parametric amplifiers, etc.). It can reduce the readout fidelity and make REM unreliable. The readout crosstalk can be well suppressed by optimising the readout parameters, including frequencies and lengths of readout signal and qubits' frequencies for readout.
7. **Calibrate the readout performance.** After optimising the readout crosstalk, we need to confirm whether residual crosstalk is negligible. Here, we first calibrate the standard readout fidelity with qubits preparing in $|0\rangle^{\otimes n}$ and $|1\rangle^{\otimes n}$. Then, we prepare the qubits in a random state set and calibrate the readout fidelity again. If the fidelities are close in both cases, the residual crosstalk is negligible.
8. **Optimise CZ gates.** In the UCC ansatz, the multi-qubit Pauli rotation gate $\hat{V}(\theta)$ is an elementary operator, which is compiled with single-qubit gates and CZ gates. The CZ gate with tunable-coupler structure has been investigated and realised with high fidelity [4, 5]. In our experiment, CZ gates are realised by utilising the resonance coupling between $|11\rangle$ and $|02\rangle$ states and optimised by the following steps:
 - (a) According to the pattern in circuits, The CZ gates and single-qubit gates are divided into different groups. The next steps will be performed individually in these groups.
 - (b) We collect coherence information (T1 and T2) for a large frequency range ($\sim 100\text{MHz}$) of qubits and determine the frequency region that is suitable for CZ interaction accordingly.
 - (c) We first prepare the qubits in $|11\rangle$ states. Then, we measure the conditional phase with varying coupling strength (tuning by coupler) for a fixed 50 ns gate time. Next, we measure the leakage of qubit states to $|02\rangle$ or $|20\rangle$ as qubit interaction frequencies vary. Finally, we coarsely determine the interaction frequencies and coupling strengths by minimising the cost function of conditional phase error and leakage error.
 - (d) Select a CZ gate in the group and finely adjust its parameters by optimising the XEB (cross-entropy benchmarking) fidelity with a fixed cycle of 20 and fixed random circuits of 60 while keeping other parameters of the CZ gate unchanged.
 - (e) We repeat step (d) for the CZ gate with the lowest fidelity in the group until the average fidelity no longer improves.
9. **Calibrate single-qubit gates and CZ gates.** We utilise the SPB and XEB measurements to calibrate single-qubit and CZ gates. For CZ gates, the dynamical phases on both qubits are also determined in the XEB experiment.

All steps can be realised automatically as a directed graph [6], except for step 6, which needs to be checked manually. We refer to Ref. [7] for more details about the performance calibration of quantum devices.

After all the calibrations and optimisations, we summarised the qubit parameters in Fig. S1, and the errors for single-qubit gates, two-qubit gates and readout in Fig. 2a in the main text and Fig. S2. Fig. 2a in the main text shows the accumulated distribution of single-qubit gate, CZ gate and readout errors. Fig. S2 shows the spatial distribution of those gate errors for selected qubits on the “Zuchongzhi 2.0” quantum processor. There is no strong spatial error distribution in our qubit set. We run the automatic calibration steps 3, 7, 8, and 9 every three hours to maintain the system's performance.

C. Correlated readout error and random state measurements

Quantum state measurements determine the information that we can extract from quantum experiments. Measurement errors can significantly impact simulation accuracy since they introduce a bias in the expectation value. Therefore, it is essential to mitigate measurement errors to ensure reliable results.

To mitigate measurement errors, the first step is to learn the measurement noise matrix Λ from readout calibration measurements. Provided an invertible measurement noise matrix Λ , we can mitigate measurement readout errors by applying an inverse matrix Λ^{-1} to the noisy measurement outcomes.

There are two types of measurement calibration matrices: one takes a tensor product form, a tensor product of local single-qubit calibration matrices, while the other one characterises correlated noise. The latter requires more experimental resources, which is not desirable in experiments. When the correlated readout error is negligible, noise can be well described by a measurement noise matrix with a tensor product form.

In general, correlated readout error arises from the residual coupling and the spectral overlap of readout resonators. We make two improvements to ensure that the correlations of measurement errors are weak enough to be neglected.

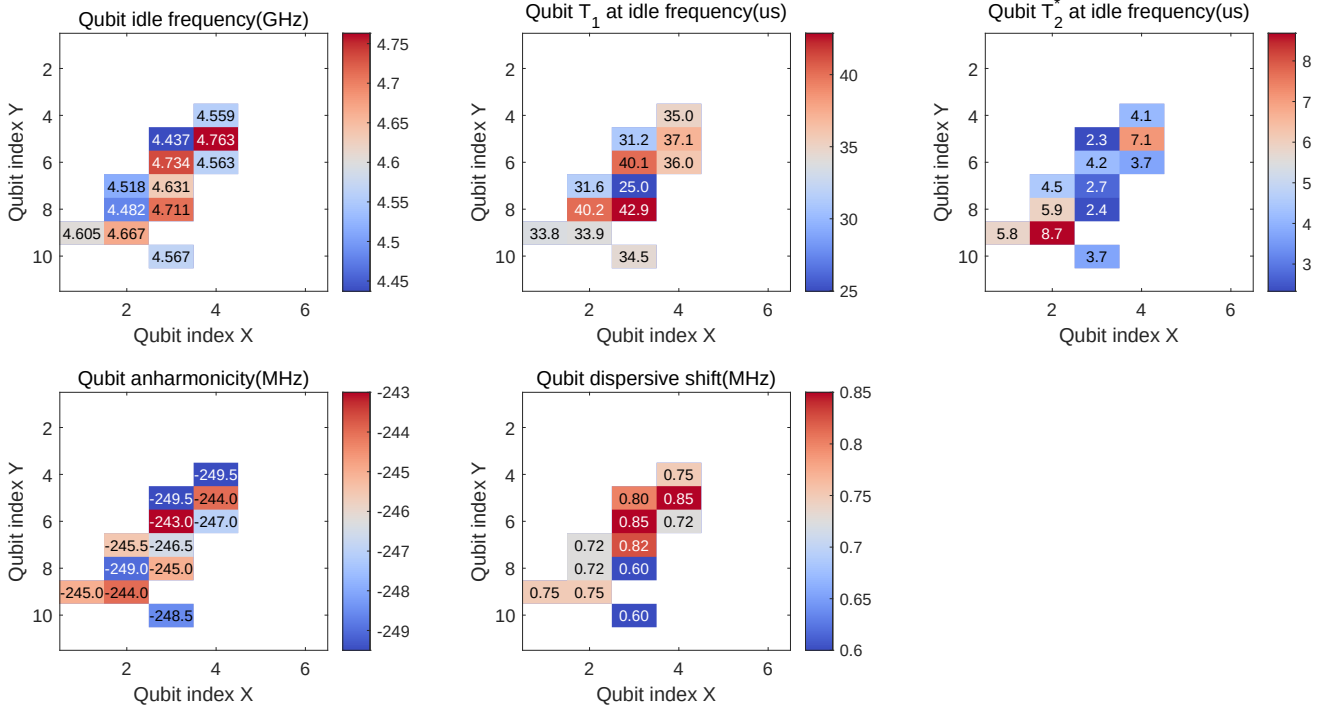


FIG. S1. **The typical distribution of the qubit parameters.** Idle frequency, T_1 , T_2^* , anharmonicity and dispersive shift for the selected 12 qubits.

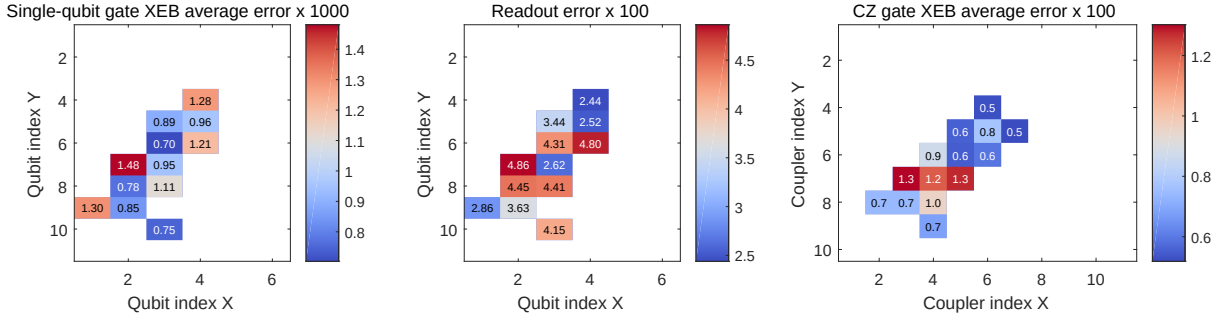


FIG. S2. **The performance of the quantum processor after calibration.** Error for readout, single-qubit, and CZ gates for the selected 12 qubits.

Firstly, we suppressed the residual coupling by adjusting the working point of the coupler delicately. Secondly, we optimised readout parameters simultaneously, including readout pulse frequencies, readout pulse amplitudes, readout pulse length and qubit readout frequencies.

After these improvements, we verified the correlated readout error was suppressed by comparing the results of standard measurement and random state measurement. The verification by in the following steps:

1. Prepare all qubits in $|0\rangle^{\otimes n}$ and $|1\rangle^{\otimes n}$ to acquire F_{00} and F_{11} . Here, F_{00} (F_{11}) means the probability to get a measure result of $|0\rangle$ ($|1\rangle$) when the qubit was prepared in $|0\rangle$ ($|1\rangle$).
2. Random state measurements. Randomly prepare the qubit state and use the same discrimination line of the readout signal demodulated IQ clouds to calculate fidelity. Multiply the fidelities for all qubits, and we refer to the multiplication as the union readout fidelity.
3. With hundreds of repeated experiments, averaged union readout fidelity is computed.

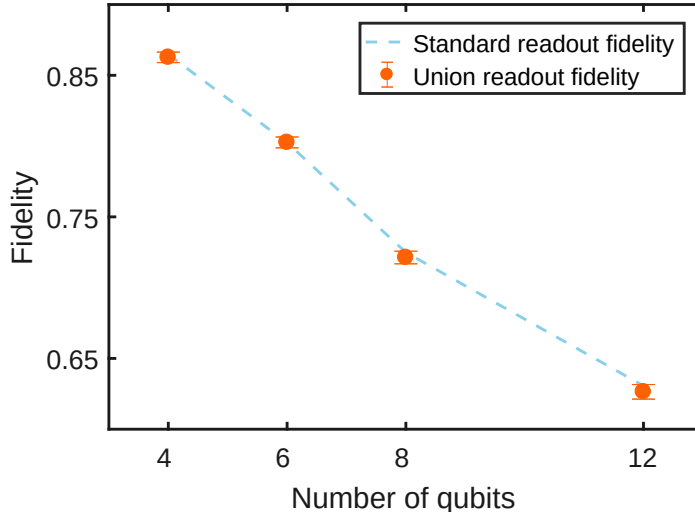


FIG. S3. **Union readout fidelity.** Comparison of average union readout fidelity (orange dots) and standard readout fidelity (blue dashed line) for four qubits, six qubits, eight qubits and twelve qubits.

Fig. S3 compares the average standard and union readout fidelity. These two types of readout fidelity show a good agreement, and the correlated readout error is negligible, which ensures the tensor product form is applicable in this experiment.

II. VARIATIONAL QUANTUM STATE PREPARATION AND MEASUREMENT

This section discusses quantum algorithms for quantum chemistry problems, including variational state preparation, optimisation, and quantum state measurement. This section is organised as follows. In Sec. II A, we first demonstrate multi-reference state preparation on a quantum processor and illustrate the corresponding circuit. In Sec. II B, we introduce quantum circuit design strategies based on unitary coupled cluster ansatz, and propose a circuit reduction strategy to enable the simulation on current hardware. Quantum circuits for the three molecules are shown in Sec. II B 4. In Sec. II C, we introduce an optimisation scheme by analytic stochastic gradient descent and present the analytical form of the gradient estimation. In Sec. II D, we introduce an efficient quantum state measurement strategy for measuring multiple observables.

A. Initial state preparation

As the spin Hamiltonian is constructed according to the $SU(2)$ symmetry, we consider the initial state or the reference state obeying this symmetry by assigning an equal number of electrons to the spin α and β sectors. Also, as only closed-shell molecules are considered, the reference state is restricted to a singlet state. In the first step of generating the molecular Hamiltonians, we adopt the restricted Hartree-Fock (RHF) method, where the ground-state wave function Ψ_0 is expressed by a single Slater determinant of molecular orbitals. The energy $E = \langle \Psi_0 | H | \Psi_0 \rangle$ is then minimised subject to orbital orthogonality. We first consider the initial state to be a single reference state (i.e., a product state in the computational basis). Suppose the total number of electrons is 2η . The single reference state is constructed in a way that η of electrons occupy the lowest energy spin orbitals in the spin α and β sectors, respectively, which is referred to as the Hartree-Fock state $|\Psi_{\text{HF}}\rangle$.

In this work, qubits are arranged in the following way: i) From left to right, each qubit corresponds to a spin-orbital that increases in its molecular-orbital energy; ii) The spin α and β sector are arranged separately; the spin α is assigned first in the left part, then followed by the spin β part. For example, the HF state of the 6-qubit, 2-electron LiH molecule reads $|100100\rangle$, where the first and last 3-qubit state $|100\rangle$ correspond to the spin α and β parts, respectively.

Conventionally, the HF state is a good initial guess when the molecule is close to equilibrium. However, the closer the molecule approaches dissociation, the worse the HF state behaves due to the quasi-degeneration of different configurations. Consequently, the correlation energy, defined as the difference between the ground state energy and HF energy, becomes larger when the bond distance of a bi-atomic molecule stretches. To address this point, we consider the reference state to be a superposition of multi-configurations,

$$|\Psi_{\text{ref}}\rangle = \sum_i c_i |\Psi_i\rangle, \quad (\text{S1})$$

where $\{|\Psi_i\rangle\}$ is the collection of configurations in the reference state with coefficients $\{c_i\}$. The configurations $|\Psi_i\rangle$ are chosen by identifying the fermionic modes that are related to the bond-breaking process and the size $|\{|\Psi_i\rangle\}|$ is set to be a constant as they are only related to a constant number of fermionic modes. Thus, we can optimise efficiently on a classical device to determine the optimal coefficients. This approach is known as a multi-reference method [8] in the quantum chemistry literature. We apply the method for the LiH and F₂ molecule and observe that higher accuracy is achieved compared to the HF state in the dissociation regime. As the two molecules we considered here are bi-atomic, the molecule can be seen as two individual atoms with negligible interaction in the large bond-distance limit. For the LiH molecule, the lowest two molecular orbitals in the active space are established from atomic orbitals in this way: in the dissociation limit, one decouples the molecular orbitals to atomic ones by the equal-superposition state, $\frac{1}{\sqrt{2}}|100100\rangle - \frac{1}{\sqrt{2}}|010010\rangle$. As the molecule is in the intermediate state that interpolates between the equilibrium and the dissociation regimes, the reference state is expected to be a superposition of the above-mentioned two configurations. In the following, we elaborate on the compilation of quantum circuits and optimise the coefficients of the reference state.

In this work, we prepare the reference state as

$$|\Psi_{\text{ref}}\rangle = \frac{1}{\sqrt{1+\beta^2}}(1 - \beta \hat{a}_{\eta+N/2}^\dagger \hat{a}_\eta^\dagger \hat{a}_{\eta+N/2-1} \hat{a}_{\eta-1}) |\Psi_{\text{HF}}\rangle, \quad (\text{S2})$$

which considers the excitations from the highest occupied molecular orbitals of the Hartree-Fock state, and the qubits are numbered from 0. When $\beta = 0$, the initial state reduces to the Hartree-Fock state. The quantum circuit for preparing the multi-reference state in Eq. (S2) is first creating a rotation along the y-axis on the qubit indexed by $(\eta - 1)$ as

$$R_y |0\rangle^{\otimes N} = \frac{1}{\sqrt{1+\beta^2}} |0\rangle^{\otimes(\eta-1)} \otimes (|0\rangle - \beta |1\rangle) \otimes |0\rangle^{\otimes N-\eta}, \quad (\text{S3})$$

followed by 3 CNOT gates and single-qubit X gates, which can be explicitly expressed as

$$\left(\bigotimes_{i=N/2}^{\eta+N/2-1} X_i \bigotimes_{i=0}^{\eta-1} X_i \right) \text{CNOT}_{(N/2+\eta-1) \rightarrow (N/2+\eta)} \text{CNOT}_{\eta \rightarrow (N/2+\eta-1)} \text{CNOT}_{(\eta-1) \rightarrow \eta}. \quad (\text{S4})$$

It is easy to find that compared to the Hartree-Fock state, the state preparation for Eq. (S2) only introduces additional 3 CNOT gates. For instance, we prepare $|\Psi_0\rangle = \frac{1}{\sqrt{1+\beta^2}}(|100100\rangle - \beta |010010\rangle)$ for the LiH molecule. The circuit for initial state preparation is shown in Fig. S4.

The parameter β is determined by minimising the energy on the state, i.e., $\min_\beta \langle \Psi_0 | \hat{H} | \Psi_0 \rangle$. It is easy to check that this minimisation problem can be computed efficiently.

A comparison of the initial state energy associated with the multi-reference state, Hartree-Fock state energy, and the true ground state energy is shown in Fig. S5. The multi-reference state outperforms the Hartree-Fock state at a large bond distance and thus serves as a better initial state in this case.

B. Quantum circuit ansatz design and compilation

1. Unitary coupled cluster circuit ansatz

One crucial step for simulating large molecular systems by variational quantum algorithms is to design an appropriate quantum circuit that can effectively represent the ground state of the target system, and meanwhile, circuit depth

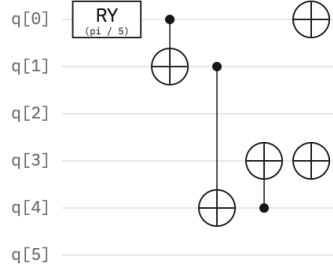


FIG. S4. The initial state preparation for LiH molecule $|\Psi_0\rangle = \frac{1}{\sqrt{1+\beta^2}}(|100100\rangle - \beta|010010\rangle)$. The Pauli rotation along the y-axis prepares the state $R_y|0\rangle = \frac{1}{\sqrt{1+\beta^2}}(|0\rangle - \beta|1\rangle)$.

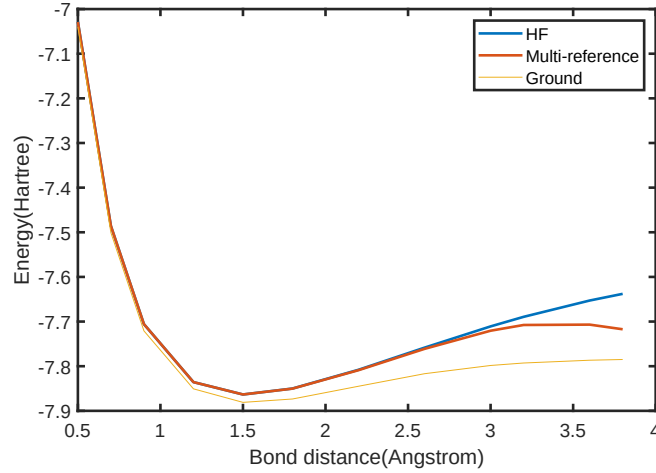


FIG. S5. Comparison of different energies associated with the Hartree-Fock state, the multi-reference state, and the ground state.

is shallow. The unitary coupled cluster (UCC) ansatz has been widely used in the context of quantum computational chemistry [9]. As shown in the main text, the quantum state can be constructed as

$$|\Psi(\theta)\rangle = e^{\hat{T}(\theta) - \hat{T}^\dagger(\theta)} |\Psi_0\rangle, \quad (\text{S5})$$

where $\hat{T}(\theta)$ is the coupled-cluster excitations, and $|\Psi_0\rangle$ is the initial state. The coupled-cluster excitations are usually truncated to single and double excitations, which are also called unitary coupled-cluster singles-and-doubles (UCCSD). The cluster operator truncated at single- and double-excitations has the form

$$\hat{T}(\theta) = \sum_{\substack{p \in \text{vir} \\ i \in \text{occ}}} \theta_{pi} \hat{T}_{pi} + \sum_{\substack{p,q \in \text{vir} \\ i,j \in \text{occ}}} \theta_{pqji} \hat{T}_{pqji}, \quad (\text{S6})$$

where the one- and two-body terms are defined as $\hat{T}_{pi} = \hat{a}_p^\dagger \hat{a}_i$ and $\hat{T}_{pqji} = \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_j \hat{a}_i$, respectively. As the molecular Hamiltonian conserves particles with definite spins, the ansatz is designed to conserve this symmetry as well, whereas the spin degree of freedom is abbreviated for simplicity.

To implement the VQE circuit of UCC ansatz on a quantum device, we apply the first-order Trotter-Suzuki expansion. and the circuit ansatz becomes

$$|\Psi(\theta)\rangle = \prod_{\hat{T}_i \in \mathcal{O}} e^{\hat{T}_i - \hat{T}_i^\dagger} |\Psi_0\rangle, \quad (\text{S7})$$

where we denote the operator pool as $\mathcal{O} := \{\hat{T}_i\}$ with $\hat{T}_i$ representing the excitation operators in Eq. (S6). Here, it is worth remarking that the order of $\hat{T}_i$ will result in a different performance of the circuit.

As can be found from Eq. (S6) and Eq. (S7), the original UCCSD ansatz considering the single- and double-excitations is composed of a large number of operators scaling as $\mathcal{O}((N - \eta)^2 \eta^2)$ with electron number η , and the circuit depth with respect to the two-qubit entangling gates scales as $\mathcal{O}(N(N - \eta)^2 \eta^2)$. As shown in the main text, the CZ gate counts for H_2 , LiH , and F_2 are 56, 280, 2920, respectively. Note that the multi-qubit rotation gate is complied into a ladder of CZ gates and a single-qubit Pauli rotation gate, $\exp(-i\theta\hat{\sigma}/2)$, with $\hat{\sigma} = \{X, Y, Z\}$ being the single-qubit Pauli operator. The high demand for the CZ gates constraints practical implementation of the original UCC ansatz.

2. Point group symmetry constraint

There has been considerable progress in reducing the number of operators in the circuit ansatz, such as using classical pre-calculation to screen out the terms that are small enough [10] or only choosing the important excitation operators [11], for instance, paired double excitation operators in ansatz [12]. In the following, we discuss the circuit reduction strategy using the point group symmetry constraint [13].

To reduce the circuit depth without sacrificing accuracy, we consider using the point group property to reduce the UCC ansatz operators. Point group symmetry (spatial symmetry) is a fundamental property of molecular systems that characterises molecules' geometric and electronic structures. The amplitude associated with the one- and two-electron excitations in both coupled cluster and unitary coupled cluster ansatz is proven to be zero unless the corresponding term preserves the symmetry of the initial state [14]. This indicates that the expansion of the unitary coupled cluster wavefunction will only contain the terms with the same irreducible representation (irrep) as the reference wavefunction [13]. For the simple closed-shell molecules, the excitation operations should preserve the symmetry, and thus the irrep of the ground state should be A_g , which is a totally symmetric irreducible representation, A_g , of which the characters are all 1. This applies to all the cases considered in this work.

In our implementation, we filter out the operators that break the symmetry of the reference Hartree-Fock state when constructing the UCC ansatz. Only the terms excite $|\Psi_0\rangle$ to A_g states remain, and the rest of the terms in the exciting operator are excluded in constructing the cluster operator. As such, we select the operators $\hat{T}_i$ that satisfy the constraint of symmetry, that is,

$$\mathcal{O} := \left\{ \hat{T}_i \mid D \left(e^{\hat{T}_i - \hat{T}_i^\dagger} |\Psi_0\rangle \right) = D(|\Psi_0\rangle) \right\} \quad (\text{S8})$$

where D is the irrep of the corresponding wavefunction.

The point group of LiH is $C_{\infty v}$, which has three irreducible representations D , i.e. $D \in \{A_g, E_{1x}, E_{1y}\}$. All the UCCSD excitation operators satisfy the symmetry constraint in Eq. (S8), and thus there is no circuit reduction for LiH . However, for large molecules, quantum circuits can be reduced without sacrificing accuracy. The point group of F_2 is $D_{\infty h}$. F_2 has six irreducible representations D , $D \in \{A_{1g}, E_{1gx}, E_{1gy}, A_{1u}, E_{1uy}, E_{1ux}\}$. In total, there are ten single excitation operators and 25 double excitation operators. Only nine excitations have the same irreducible representation of A_{1g} and need to be included in the circuit ansatz.

However, we note that the strategy based on point group symmetry constraints may not yield significant reductions in circuit depth for larger molecules that tend to be less symmetric.

3. Selection of excitation operators

The above circuit reduction exploits intrinsic information of the molecules. In Methods, we have provided an overview of how to generate a more compact quantum circuit by selecting the dominant operators. Below, we give some details about the derivations in Methods.

The circuit operator $e^{\theta(\hat{T}_i - \hat{T}_i^\dagger)}$ can be expanded by the Taylor series as

$$\begin{aligned} e^{\theta(\hat{T}_i - \hat{T}_i^\dagger)} &= \sum_{n=0}^{\infty} \frac{\theta^n (\hat{T}_i - \hat{T}_i^\dagger)^n}{n!} \\ &= I + \sum_{n=1}^{\infty} (-1)^n \theta^{2n} \frac{(\hat{T}_i \hat{T}_i^\dagger)^n + (\hat{T}_i^\dagger \hat{T}_i)^n}{(2n)!} + \sum_{n=0}^{\infty} (-1)^n \theta^{2n+1} \frac{-\hat{T}_i^\dagger (\hat{T}_i \hat{T}_i^\dagger)^n + \hat{T}_i (\hat{T}_i^\dagger \hat{T}_i)^n}{(2n+1)!} \\ &= \cos(\theta \sqrt{\hat{T}_i \hat{T}_i^\dagger}) + \cos(\theta \sqrt{\hat{T}_i^\dagger \hat{T}_i}) - \frac{\hat{T}_i^\dagger}{\sqrt{\hat{T}_i \hat{T}_i^\dagger}} \sin(\theta \sqrt{\hat{T}_i \hat{T}_i^\dagger}) + \frac{\hat{T}_i}{\sqrt{\hat{T}_i^\dagger \hat{T}_i}} \sin(\theta \sqrt{\hat{T}_i^\dagger \hat{T}_i}) \end{aligned} \quad (\text{S9})$$

where $\hat{T}_i \in \mathcal{O}$ and $\hat{T}_i^2 = (\hat{T}_i^\dagger)^2 = 0$ is used. From Eq. (S9), the circuit operator comprises four terms only, and each term can be simulated classically. Similar to the discussions in [15], let us take $\cos(\theta\sqrt{\hat{T}_i\hat{T}_i^\dagger})$ for example. Without loss of generality, $\hat{T}_i$ can be expressed as $\hat{T}_i = \hat{a}_3^\dagger\hat{a}_1^\dagger\hat{a}_2\hat{a}_0$, and then

$$\hat{T}_i\hat{T}_i^\dagger = \hat{a}_3^\dagger\hat{a}_1^\dagger\hat{a}_2\hat{a}_0\hat{a}_0^\dagger\hat{a}_2^\dagger\hat{a}_1\hat{a}_3 = \hat{n}_3\hat{n}_1(1-\hat{n}_2)(1-\hat{n}_0). \quad (\text{S10})$$

Since the action of the number operator gives 0 or 1 and $\hat{T}_i\hat{T}_i^\dagger$ is idempotent, the square root of $\hat{T}_i\hat{T}_i^\dagger$ can be defined as

$$\sqrt{\hat{T}_i\hat{T}_i^\dagger} \equiv \hat{n}_3\hat{n}_1(1-\hat{n}_2)(1-\hat{n}_0), \quad (\text{S11})$$

and we have

$$\cos(\theta\sqrt{\hat{T}_i\hat{T}_i^\dagger}) = \cos\theta\sqrt{\hat{T}_i\hat{T}_i^\dagger} = \cos\theta\hat{n}_3\hat{n}_1(1-\hat{n}_2)(1-\hat{n}_0). \quad (\text{S12})$$

The energy expectation $\langle\Psi_0|e^{-\theta(\hat{T}_i-\hat{T}_i^\dagger)}\hat{H}e^{\theta(\hat{T}_i-\hat{T}_i^\dagger)}|\Psi_0\rangle$ contains 16 terms, each of which can be simulated classically. As shown in Methods, the energy calculation with respect to the Hartree-Fock state can be simplified.

Suppose we have selected the operator set $\{\hat{T}_i\}$ in its qubit form. We map $\hat{T}_i$ to its qubit form under the Jordan-Wigner transformation. Since each term is symmetric and the contribution of each term is the same, we only choose one term in the qubit form to further reduce the gate count and hence circuit depth. In the next section, we display the quantum circuits for the three molecules.

4. Quantum circuit compilation

In this section, we present the dominant operators and quantum circuits for the three molecules after circuit compilation. The dominant excitation operators for H_2 are $a_3^\dagger a_1^\dagger a_2 a_0$. The dominant excitation operators for LiH are $a_5^\dagger a_1^\dagger a_3 a_0$, $a_4^\dagger a_2^\dagger a_3 a_0$, and $a_5^\dagger a_2^\dagger a_3 a_0$. The dominant excitation operators for F_2 are $a_{11}^\dagger a_5^\dagger a_6 a_0$, $a_{11}^\dagger a_5^\dagger a_7 a_1$, $a_{11}^\dagger a_5^\dagger a_8 a_2$, $a_{11}^\dagger a_5^\dagger a_9 a_3$, and $a_{11}^\dagger a_5^\dagger a_{10} a_4$.

Next, we optimise the quantum circuits with respect to the circuit depth of CZ gates. We exploit the 2D geometry of the qubit layout of the quantum processor to encode the spin-ups and spin-downs, and we optimise the depth by parallelisation of CZ gates. The parameterised quantum circuits for H_2 , LiH and F_2 are

$$U_{\text{H}_2} = \exp(i\theta X_0 Y_1 X_2 X_3) \exp(-i\theta Y_0 X_1 X_2 X_3), \quad (\text{S13})$$

$$U_{\text{LiH}} = \exp(-i\theta_3 Y_0 Z_1 X_2 X_3 Z_4 X_5) \exp(-i\theta_2 Y_0 Z_1 X_2 X_3 X_4) \exp(-i\theta_1 Y_0 X_1 X_3 Z_4 X_5), \quad (\text{S14})$$

and

$$U_{\text{F}_2} = \exp(-i\theta_5 X_4 X_5 Y_{10} X_{11}) \exp(-i\theta_4 Y_3 Z_4 X_5 X_9 Z_{10} X_{11}) \exp(-i\theta_3 Y_2 Z_3 Z_4 X_5 X_8 Z_9 Z_{10} X_{11}) \\ \exp(-i\theta_2 Y_1 Z_2 Z_3 Z_4 X_5 X_7 Z_8 Z_9 Z_{10} X_{11}) \exp(-i\theta_1 Y_0 Z_1 Z_2 Z_3 Z_4 X_5 X_6 Z_7 Z_8 Z_9 Z_{10} X_{11}), \quad (\text{S15})$$

respectively.

The compiled quantum circuits (excluding the initial state preparation and measurement) for H_2 , LiH , and F_2 are shown in Fig. S6, Fig. S7, and Fig. S8, respectively. In the figures, the parameters are chosen as $\pm\pi/5$ as an example for a clear illustration. The circuit for initial preparation has been discussed in Sec. II A, and the circuit for LiH is shown in Fig. S4.

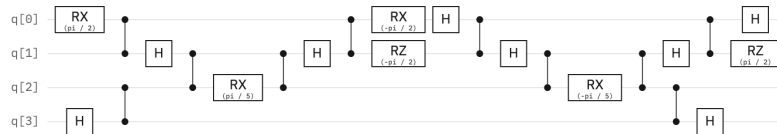


FIG. S6. The compiled quantum circuit ansatz for H_2 .

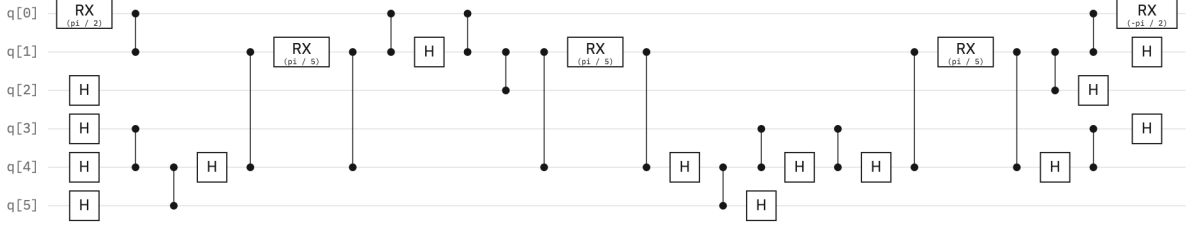
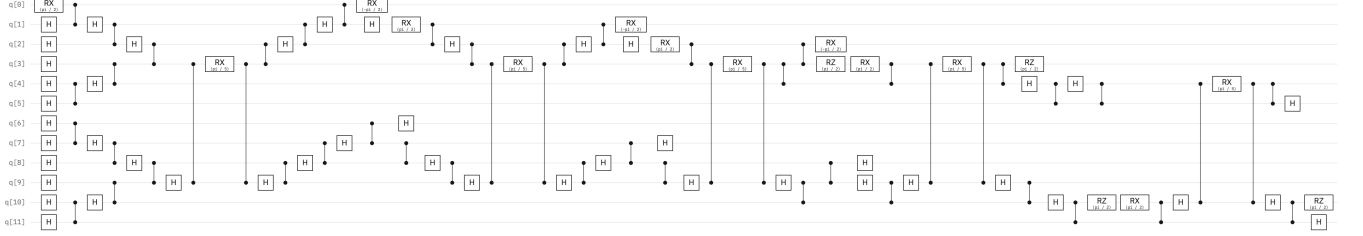


FIG. S7. The compiled quantum circuit ansatz for LiH

FIG. S8. The compiled quantum circuit ansatz for F₂

C. Optimisation by analytic stochastic gradient descent

Now, we show how to analytically estimate the gradient of expectation values of the observable $\hat{O}$. In VQE, the ground state is approximated by solving $\mathcal{L}(\theta) = \min_{\theta} \langle \Psi_0 | U^\dagger(\theta) \hat{O} U(\theta) | \Psi_0 \rangle_{\theta}$ with $\hat{O} = \hat{H}$ using a classical optimiser. The circuit that is constructed using the methods developed in [Sec. II B](#) takes the form of

$$U(\theta) = \prod_j \prod_{l_j}^{L_j} V_{j,l_j}(\theta_{j,l_j}), \quad (\text{S16})$$

where each gate operation V_{j,l_j} is a multi-qubit Pauli rotation $V_{j,l_j} = \exp(-i\theta_{j,l_j} P_{j,l_j}/2)$ with $P_{j,l_j} \in \{I, X, Y, Z\}^{\otimes N}$ being the tensor product of single-qubit Pauli operators over N qubits. Prior works usually consider $L_j = 1$, that is, each parameter is associated with only one term. However, in quantum chemistry experiments, the parameters are often associated with many terms. The objective is to get the first-order derivative with respect to the j th parameter, i.e., $g_j(\theta) = \partial \mathcal{L}(\theta) / \partial \theta_j$. Using the Leibniz rule, the gradient is given by

$$g_j = \sum_{l_j} \langle \Phi_{j,l_j} | \frac{\partial \bar{V}_{j,l_j}(\theta_{j,l_j})}{\partial \theta_{j,l_j}} | \Phi_{j,l_j} \rangle \quad (\text{S17})$$

where we introduce θ_{j,l_j} to represent the parameter associated with the operator V_{j,l_j} with $\theta_{j,l_j} = \theta_j$, and we have

$$|\Phi_{j,l_j}\rangle = \prod_{k < l_j} V_{j,k}(\theta_{j,k}) \prod_{k < j} \prod_{l_k} V_{k,l_k}(\theta_{k,l_k}) |\Psi_0\rangle \quad (\text{S18})$$

and

$$\bar{V}_{j,l_j}(\theta_{j,l_j}) = V_{j,l_j}(\theta_{j,l_j})^\dagger \bar{O} V_{j,l_j}(\theta_{j,l_j}) \quad (\text{S19})$$

with

$$\bar{O} = \left(\prod_{k > j} \prod_{l_k} V_{k,l_k}(\theta_{k,l_k}) \prod_{k > l_j} V_{j,k}(\theta_{j,k}) \right)^\dagger \bar{O} \prod_{k > j} \prod_{l_k} V_{k,l_k}(\theta_{k,l_k}) \prod_{k > l_j} V_{j,k}(\theta_{j,k}). \quad (\text{S20})$$

Since $P_{j,l_j}^2 = I$, each of the gate can be written as

$$V_{j,l_j}(\theta_{j,l_j}) = \cos(\theta_{j,l_j}/2)I - i \sin(\theta_{j,l_j}/2)P_{j,l_j}, \quad (\text{S21})$$

and thus $\bar{V}_{j,l_j}(\boldsymbol{\theta}_{j,l_j})$ can be written as

$$\bar{V}_{j,l_j}(\boldsymbol{\theta}_{j,l_j}) = A + B \cos(\boldsymbol{\theta}_{j,l_j}) + C \sin(\boldsymbol{\theta}_{j,l_j}) \quad (\text{S22})$$

where A , B , and C are operators that rely on V_{j,l_j} and $\bar{O}$ yet independent of the parameters $\boldsymbol{\theta}_{j,l_j}$. From Eq. (S22), we can show that the summand in the gradient is related to a parameter-shifted $\bar{V}_{j,l_j}(\boldsymbol{\theta}_{j,l_j} \pm s)$ with a shift s , which is given by

$$\frac{\partial \bar{V}_{j,l_j}(\boldsymbol{\theta}_{j,l_j})}{\boldsymbol{\theta}_{j,l_j}} = \frac{\bar{V}_{j,l_j}(\boldsymbol{\theta}_{j,l_j} + s) - \bar{V}_{j,l_j}(\boldsymbol{\theta}_{j,l_j} - s)}{2 \sin s}. \quad (\text{S23})$$

The gradient is thus given by

$$\begin{aligned} g_j &= \frac{\sum_{l_j} (\langle \Phi_{j,l_j} | \bar{V}_{j,l_j}(\boldsymbol{\theta} + s \mathbf{e}_{j,l_j}) | \Phi_{j,l_j} \rangle - \langle \Phi_{j,l_j} | \bar{V}_{j,l_j}(\boldsymbol{\theta} - s \mathbf{e}_{j,l_j}) | \Phi_{j,l_j} \rangle)}{2 \sin(s)} \\ &= \frac{\sum_{l_j} \mathcal{L}(\boldsymbol{\theta} + s \mathbf{e}_{j,l_j}) - \mathcal{L}(\boldsymbol{\theta} - s \mathbf{e}_{j,l_j})}{2 \sin(s)} \end{aligned} \quad (\text{S24})$$

with unit vector $\mathbf{e}$. Eq. (S24) indicates that we can get the gradient of the energy expectation analytically by measuring the energy expectation by shifting parameters, termed the parameter-shift rule [16, 17].

In the experiment, the gradient is estimated by measuring the two parameter-shifted quantities. Due to statistical fluctuation, the gradient can only be estimated up to certain precision. The estimator of g_j is given by

$$\hat{g}_j = \frac{\sum_{l_j} \hat{\mathcal{L}}(\boldsymbol{\theta} + s \mathbf{e}_{j,l_j}) - \hat{\mathcal{L}}(\boldsymbol{\theta} - s \mathbf{e}_{j,l_j})}{2 \sin(s)}. \quad (\text{S25})$$

The estimation by the parameter-shift rule has an unbiased estimation of the gradient as

$$\mathbb{E} \hat{g}_j = g_j. \quad (\text{S26})$$

In what follows, we discuss the variance of the estimator. Assume that energy estimation has the same variance as

$$\text{Var}[\hat{\mathcal{L}}] = \text{Var}[\hat{\mathcal{L}}(\boldsymbol{\theta} + s \mathbf{e}_{j,l_j})], \quad \forall j, \quad \forall l_j. \quad (\text{S27})$$

Then the variance of the gradient estimation is given by

$$\text{Var}[\hat{g}_j] = \frac{L_j}{\sin^2(s)} \text{Var}[\hat{\mathcal{L}}] \geq L_j \text{Var}[\hat{\mathcal{L}}]. \quad (\text{S28})$$

It is straightforward that the variance of the estimator is minimised when $s = \pi/2$.

The gradient estimation by the parameter-shift rule shows a clear advantage over the finite difference method where the gradient is estimated as

$$\hat{g}_j^{\text{FD}} = \frac{\sum_{l_j} \hat{\mathcal{L}}(\boldsymbol{\theta} + \delta \mathbf{e}_{j,l_j}) - \hat{\mathcal{L}}(\boldsymbol{\theta} - \delta \mathbf{e}_{j,l_j})}{2\delta}. \quad (\text{S29})$$

The estimation is unbiased when the step size approaches zero, $\delta \rightarrow 0$. To make the estimation close to the true gradient, the step δ should be taken to be a small number. However, the variance of the estimation is

$$\text{Var}[\hat{g}_j^{\text{FD}}] = \frac{L_j}{\delta^2} \text{Var}[\hat{\mathcal{L}}], \quad (\text{S30})$$

which can be much larger than the variance in Eq. (S28).

To make it more efficient, we adopt a stochastic gradient descent strategy. For the k th iteration with parameters $\boldsymbol{\theta}^k$, we obtain the gradient element $g_j^k(\boldsymbol{\theta}^k)$ by the parameter shift rule

$$\hat{g}_j^{\text{SGD},k}(\boldsymbol{\theta}^k) = \frac{\hat{\mathcal{L}}(\boldsymbol{\theta}^+) - \hat{\mathcal{L}}(\boldsymbol{\theta}^-)}{2} \quad (\text{S31})$$

where we define

$$\hat{\mathcal{L}}(\theta^\pm) = \sum_{l_j} \hat{\mathcal{L}}(\theta \pm \pi \Delta_j \mathbf{e}_{j,l_j}/2). \quad (\text{S32})$$

In experiments, the parameters to be optimised are chosen stochastically by $\Delta_j^k = 0$ or 1 and satisfying $|\Delta^k| > N_p/2$ to reduce the sample cost for the gradients. We update the parameters by $\theta^{k+1} = \theta^k + \alpha^k \mathbf{g}^k(\theta^k)$ with an appropriate learning rate α^k . The estimation of the gradient is also unbiased

$$\mathbb{E} \hat{g}_j^{\text{SGD},k}(\theta^k) = g_j^k(\theta^k). \quad (\text{S33})$$

D. Quantum state measurement and classical post-processing

The power of variational quantum algorithms is at the expense of measurements. This becomes a severe problem in molecular systems, in which the number of measurements at each iteration scales up to $\mathcal{O}(N^4)$ with the problem size N . However, as many of the observables have small coefficients and are qubit-wise compatible (hence could be measured simultaneously), we can exploit more efficient measurement schemes to alleviate the measurement cost. In Methods, we have introduced the measurement strategy by exploiting the qubit-wise compatibility of the target observable. In this section, we give some comments on the quantum state measurement in the VQE experiment.

The observables in our experiment include the Hamiltonian $\hat{H}$, the projected Hamiltonian in the symmetry verified space $\Pi \hat{H} \Pi$ with the projector Π , and Hamiltonian moments $\hat{H}^k$ for $k \geq 2$. The target observable $\mathbf{O}$ can be decomposed into the Pauli basis as

$$\mathbf{O} = \sum_l \alpha_l \mathbf{O}_l, \quad (\text{S34})$$

with $\mathbf{O}_l \in \{I, X, Y, Z\}^{\otimes n}$ being the tensor product of single-qubit Pauli operators.

As shown in Methods, an estimator of the target observable $\mathbf{O}$ is expressed as

$$\hat{\mathbf{o}}(\mathbf{P}) = \sum_l \alpha_l f(\mathbf{P}, \mathbf{O}_l, \mathcal{K}) \mu(\mathbf{P}, \text{supp}(\mathbf{O}_l)) \quad (\text{S35})$$

where $\mu(\mathbf{P}, \text{supp}(\mathbf{O}_l)) = \prod_{i \in \text{supp}(\mathbf{O}_l)} \mu(P_i)$ with $\mu(P_i)$ being the single-shot outcome of measurement P_i on the i th qubit, $\text{supp}(\mathbf{Q}) = \{i | Q_i \neq I\}$. Suppose we have chosen the measurement basis set $\{\mathbf{P}_j\}$ and the associated probability distribution $\{\mathcal{K}(\mathbf{P}_j)\}$, the function f is defined as $f(\mathbf{P}, \mathbf{Q}, \mathcal{K}) = \chi(\mathbf{Q})^{-1} \delta_{\mathbf{Q} \triangleright \mathbf{P}}$, where $\chi(\mathbf{Q}) = \sum_{\mathbf{P}: \mathbf{Q} \triangleright \mathbf{P}} \mathcal{K}(\mathbf{P})$ represents the probability that $\mathbf{Q}$ is effectively measured with the measurement basis $\mathbf{P}$.

The variance of the estimator in Eq. (S35) for a single sample could be calculated by the definition as

$$\text{Var}[\hat{\mathbf{o}}] = \mathbb{E}_{\mathbf{P}} \sum_{l,l'} \alpha_l \alpha_{l'} \text{Tr}(\rho \mathbf{O}_l \mathbf{O}_{l'}) f(\mathbf{P}, \mathbf{O}_l, \mathcal{K}) f(\mathbf{P}, \mathbf{O}_{l'}, \mathcal{K}) - \text{Tr}(\rho \mathbf{O})^2, \quad (\text{S36})$$

where we use the equality $\mathbb{E}_{\mu(\mathbf{P})} \mu(\mathbf{P}, \text{supp}(\mathbf{O}_l)) \mu(\mathbf{P}, \text{supp}(\mathbf{O}_{l'})) = \mathbb{E}_{\mu(\mathbf{P})} \mu(\mathbf{P}, \text{supp}(\mathbf{O}_l \mathbf{O}_{l'})) = \text{Tr}(\rho \mathbf{O}_l \mathbf{O}_{l'})$. The detailed proof can be found in Refs. [18, 19]. The strategy to determine the overlapped grouping sets is that we add an observable $\mathbf{O}_l$ that has not been accessed into a new set and all compatible observables into this set. Observables with larger absolute weights are given high priority since they have more contributions to the variance, as can be found from Eq. (S36). The algorithm will generate the measurements $\{\mathbf{P}\}$ with non-optimised probabilities $\{\mathcal{K}\}$. Since our overlapped grouping measurement strategy assumes measurements drawn from the probability distribution, the measurement accuracy may fluctuate. In our experiment, we derandomise the scheme by fixing the measurement basis $\mathbf{P}$. Suppose that we have determined the measurement basis sets $\mathcal{S} = \{\mathbf{e}_1, \dots, \mathbf{e}_s\}$ and the optimised probability distribution $\{(\mathbf{P}, \mathcal{K}(\mathbf{P}))\}$ using the above strategy. For the j th measurement $\mathbf{P}_j$ with sampling probability $\mathcal{K}_j$, we choose the number of measurements for $\mathbf{P}_j$ to be $\lfloor N_s \mathcal{K}_j \rfloor$, and select an additional one $\mathbf{P}_j$ with probability $N_s \mathcal{K}_j - \lfloor N_s \mathcal{K}_j \rfloor$, where the total number of measurements is N_s .

III. ERROR-MITIGATED OBSERVABLE ESTIMATION

The VQE process requires the evaluation of gradient and energy. However, due to gate and measurement errors, these values unavoidably deviate from the ideal ones. It is crucial to mitigate the errors in order to get reliable results. In this section, we show how to obtain error-mitigated expectation values for observables.

A. Readout error mitigation

In experiments, we extract quantum state information by performing projective measurements on a computational basis and analysing the measurement outcomes. The quantum state $|\Psi\rangle$ collapses onto the computational basis $|i\rangle \in \{|0\rangle, |1\rangle\}^{\otimes N}$ with probability $p_i = |\langle i|\Psi\rangle|^2$ by Born's rule, which can be obtained by sampling the classical measurement outcomes. Getting unbiased probability $\{p_i\}$ is central for us to estimate the property of the quantum state. However, due to readout noise, the probability distribution deviates from the ideal one, which results in an error in the estimation. In what follows, we show how to mitigate the measurement readout errors.

The measurement error can be described as a classical random process, where any given ideal probabilistic measured result is transformed into a noisy one. Thus, the error can be fully described by a size $2^n \times 2^n$ transition matrix, also known as the calibration matrix Λ . The noisy measurement outcomes $\tilde{\mathbf{b}}_{\text{noisy}} = \{\tilde{b}_1, \dots, \tilde{b}_n\}$ is related to the ideal measurement outcomes $\mathbf{b}_{\text{ideal}} = \{b_1, \dots, b_n\}$ via the calibration matrix by

$$\tilde{\mathbf{b}}_{\text{noisy}} = \Lambda \mathbf{b}_{\text{ideal}}. \quad (\text{S37})$$

To circumvent the intractability of acquiring the complete calibration matrix, one can model the measurement error by taking tensor products of each local single-qubit calibration matrix of size 2×2 . The underlying assumption is that the measurement crosstalk in the quantum device is small enough that it can be completely omitted.

In our experiment, we optimise qubit-qubit residual coupling and readout parameters and check the correlated readout errors by random state measurement experiments. As shown in Fig. S3, correlated readout errors are negligible. Therefore, it is reasonable to assume that the measurement calibration matrix Λ admits a tensor product form as

$$\Lambda = \bigotimes_{j=1}^N \Lambda_j, \quad (\text{S38})$$

where Λ_j is the calibration matrix for the j th qubit, and it is expressed by

$$\Lambda_j = \begin{pmatrix} p(0|0) & p(0|1) \\ p(1|0) & p(1|1) \end{pmatrix} =: \begin{pmatrix} 1 - \epsilon_j & \gamma_j \\ \epsilon_j & 1 - \gamma_j \end{pmatrix}, \quad (\text{S39})$$

where $p(x|y)$ represents the probability of the measured basis is $|x\rangle$ given the prepared state $|y\rangle$ under computational bases $\{|0\rangle, |1\rangle\}$. With Eq. (S39), $2n$ parameters, $\{\epsilon_j, \gamma_j\}$ for $j \in [1, n]$, can be used to describe the measurement error.

Next, we conduct the calibration process to learn the tensor product model parameters. We first select a set of product states $\mathcal{T}$. For each state $x \in \mathcal{T}$, we prepare the state on the quantum device and collect the measured product state y at each time for N_c times. We denote $m(y, x)$ as the number of times for y is observed when we prepare the state x . Then, as shown in Ref. [20] the estimated two parameters ϵ_j and γ_j for the j th qubit is given by

$$\begin{aligned} \epsilon_j &= \frac{\sum_{x,y} m(y, x) \langle 1|y_j\rangle \langle x_j|0\rangle}{\sum_{x,y} m(y, x) \langle x_j|0\rangle}, \\ \gamma_j &= \frac{\sum_{x,y} m(y, x) \langle 0|y_j\rangle \langle x_j|1\rangle}{\sum_{x,y} m(y, x) \langle x_j|1\rangle}. \end{aligned} \quad (\text{S40})$$

Remark that to effectively learn the parameters, we are supposed to let the selected, prepared product state set $\mathcal{T}$ contain at least once $x_j = 0$ and 1 for each qubit j .

The REM process essentially applies the inverse of Eq. (S38) to the noisy measurement result for estimating the expectation value of an observable. Given an arbitrary Hermitian observable in the form of Eq. (S34), and each $\mathbf{O}_l$ can be expressed in a tensor product form, $\mathbf{O}_l = \bigotimes_{k=1}^N \mathbf{O}_l^k$, where $\mathbf{O}_l^k$ acts on the k th qubit. The estimator of the expectation value of $\mathbf{O}_l$ after REM is given by

$$\hat{\mathbf{o}}_l = \frac{1}{M} \sum_{m=1}^M \prod_{k=1}^N \langle e | \mathbf{O}_l^k \Lambda_k^{-1} | s_k^m \rangle, \quad (\text{S41})$$

where M is the measurement shots assigned for observable $\mathbf{O}_l$, $|e\rangle = |0\rangle + |1\rangle$, and $s_k^m \in \{0, 1\}$ is the obtained measurement outcome for the k th qubit in the m th measurement.

Now, we discuss the estimator of the observable within the grouping measurement framework. Suppose we have determined the measurement basis set $\{\mathbf{P}_j\}$ and the associated measurement times N_j . Similarly to the grouping

method, we denote $\mathcal{S}_j$ containing all $\mathbf{O}_l$ hitted by $\mathbf{P}_j$ (element-wise commute with $\mathbf{P}_j$). The estimator of the expectation value of the observable $\mathbf{O}$ can be expressed by

$$\hat{\mathbf{o}} = \sum_l \alpha_l \hat{\mathbf{o}}_l, \quad (\text{S42})$$

where $\hat{\mathbf{o}}_l$ denotes the estimator of $\mathbf{O}_l$ measured in basis $\mathbf{P}_j$. In our experiment, we estimate the expectation value of the observable $\mathbf{O}_l$ by

$$\hat{\mathbf{o}}_l = \frac{\sum_{j:\mathbf{O}_l \triangleright \mathcal{S}_j} \hat{o}_{l,j}}{s_l}. \quad (\text{S43})$$

Here, the denominator $s_l = \sum_j \delta_{\mathbf{O}_l \triangleright \mathcal{S}_j}$ is the total number of times $\mathbf{O}_l$ being hit by $\mathbf{P}_j$. As such, the variance of $\hat{\mathbf{o}}$ with a single measurement can be estimated as

$$\text{Var}[\hat{\mathbf{o}}] = \sum_l \alpha_l^2 \frac{\sum_{j:\mathbf{O}_l \triangleright \mathcal{S}_j} \text{Var}[\hat{o}_{l,j}]}{s_l^2}. \quad (\text{S44})$$

Without REM, when the observable $\mathbf{O}_l$ is measured by $\mathbf{P}_j$, it yields two potential outcomes: either +1 or -1. Let $t_{l,j}$ represent the count of measurements resulting in an outcome of +1. It is straightforward to see that the measurement result (i.e. estimator) associated with the measurement $\mathbf{P}_j$ for observable $\mathbf{O}_l$ is $\hat{o}_{l,j} = 2t_{l,j}/N_j - 1$.

With REM, $\hat{o}_{l,j}$ is given by Eq. (S41). For later convenience, we give the explicit form of the observable estimation as

$$\hat{\mathbf{o}}_l^{\text{REM}} = \frac{1}{N_j} \sum_{j:\mathbf{O}_l \triangleright \mathcal{S}_j} \sum_{m=1}^{N_j} \prod_{k=1}^N \langle e | \mathbf{O}_l^k \Lambda_k^{-1} | s_k^m \rangle. \quad (\text{S45})$$

It is thus straightforward that the readout error mitigated estimator of the expectation value of observable $\mathbf{O}$ can be explicitly expressed as

$$\hat{\mathbf{o}}^{\text{REM}} = \sum_l \frac{\alpha_l}{s_l} \sum_{j:\mathbf{O}_l \triangleright \mathcal{S}_j} \sum_{m=1}^{N_j} \prod_{k=1}^N \langle e | \mathbf{O}_l^k \Lambda_k^{-1} | s_k^m \rangle. \quad (\text{S46})$$

The readout error mitigated estimation in Eq. (S46) is unbiased if every observable is assigned at least one measurement basis $N_j \geq 1$, otherwise, the estimation has an initial error.

B. Gate error learning by Clifford fitting

1. Methods

We consider Clifford fitting (also called Clifford data regression) [21] to mitigate the dominant CZ gate error. The key idea is to obtain an error-mitigated expectation value by fitting a regression function f^{CF} which maps a noisy expectation value to an error-mitigated one as

$$o^{\text{ideal}} = f^{\text{CF}}(o^{\text{noisy}}).$$

The regression function is obtained in a learning style using data pairs of ideal and noisy results generated from classically simulable (near)-Clifford circuits and noisy quantum circuits. Here, we choose the fitting function to be linear,

$$f^{\text{CF}}(\cdot) = a(\cdot) + b,$$

where a and b are the coefficients to be fitted.

For Clifford fitting in quantum chemistry simulation, which relies on a parameterised quantum circuit, we choose a collection of quantum circuits by randomly replacing the parameterised gates with Clifford gates, and the remaining parameterised gates are assigned with a rotating angle choosing independently, uniformly from $[0, 2\pi]$. Here, we denote n_p as the total number of parameterised gates in a parameterised quantum circuit, K as the number of non-Clifford

gates not being substituted, and $S_u = \{u_i\}$ as the set of the randomly assigned circuit. Hence, there are $n_p - K$ parameterised gates that are replaced with single-qubit Clifford gates.

For a given Pauli observable $\mathbf{O}_j \in P^{\otimes n} \setminus I^{\otimes n}$ with $P \in \{I, X, Y, Z\}$, we aim to establish a training data set, which consists of data pairs of ideal expectation value simulated on a classical computer and noisy value measured from the quantum device with the set of quantum circuits that are constructed above. For efficient classical simulation, it is crucial that they are nearly Clifford in the sense that the stabiliser rank, i.e. the number of non-Clifford gates K in the quantum circuit, of each quantum circuit is a constant. In this case, these quantum circuits can be simulated classically [22]. Compared to other error mitigation methods, such as zero noise extrapolation approaches, the Clifford fitting method does not require stretching the implementation time of the gates. Besides, the effect of noise on the training and corresponding testing (experimental) circuit on the expectation values of observables is expected to remain the same as only single-qubit rotation gates are replaced by Clifford gates through changing the rotation angle and the two-qubit CZ gates are unchanged.

2. Experimental procedure

For our goal to estimate the energy, we run the Clifford fitting process for each observable in the Hamiltonian. The number of quantum circuits could be enormous and impractical for experimental realisation. Here, we discuss how to reduce instances of the circuit to be implemented for Clifford fitting.

We observe that in practical computation, some observables' expectation values remain unchanged or only change in a sufficiently small range for different parameters in different quantum circuits, which are called unchanged observables. It is, therefore, unnecessary to run the algorithm for these unchanged observables on quantum hardware, upon evaluating the energy of the quantum system. We elaborate on the complete protocol in the following.

1. In the first step, we identify and record all the changed and unchanged observables in the Hamiltonian. For each Pauli observable $\mathbf{O}_j$ in the Hamiltonian $\hat{H}$, randomly assign all the parameters uniformly from $[0, 2\pi]$ and obtain a L -size quantum circuit instance set $\mathcal{S}_u^{(Z)} = \{u_i\}_{i=1}^L$, where the upper-script Z stands for the number of non-Clifford gates in the circuit, and u_i denotes the i th quantum circuit instance obtained from the random-parameter assignment. The size L needs to be sufficiently large to create enough quantum circuit instances to determine if the expected value of an observable changes. Compute the expectation value $\{\langle \mathbf{O}_j \rangle_{u_i}\}_{i=1}^L$ for each observable classically for each circuit instance in $\mathcal{S}_u^{(Z)}$. Compute the range value σ_j and the variance μ_j of the expectation values. If both σ_j and μ_j are greater than the thresholds σ_T and μ_T we chose, then $\mathbf{O}_j$ is labelled as a changed observable, otherwise, an unchanged observable. We record the invariant observables, which will not be measured on quantum hardware. Record its value for estimating the energy in the final step.
2. Let M be the total number of parameterised gates. Randomly choose K out of M parameterised gates and assign value independently to each parameter from $[0, 2\pi]$ to construct the quantum circuits set $\mathcal{S}_u^{(K)}$ of size R . Set the left $M - K$ gates to the identity gate. R is chosen to be sufficient for fitting a line for the linear regression, such as 10.
3. Run step 4 for each observable in O_C which is denoted as the set of changed observables. Then go to step 5.
4. For a given observable $\mathbf{O}_j$, calculate the ideal expected value $\{\langle \mathbf{O}_j \rangle_{u_i}\}_{i=1}^R$ with respect to state prepared by each u_i . Here, we want the ideal values to be distributed within $[-1, 1]$ with a broad spectrum. The purpose is that at least these different instances will appear in different intervals so that the values are dispersed enough for better linear regression. To this end, we classify different expected values according to their intervals, such as selecting instances in the intervals of $[-1, 0.9]$, $[-0.9, 0.8]$, $\dots$, $[0.9, 1]$.
5. Set $\{O_S\} = \emptyset$. Update $\{O_S\}$ by adding all the observables that satisfy the dispersion condition in step 4. Repeat step 4 for multiple times for observables in $\{O_C\} \cap \overline{\{O_S\}}$ until $\{O_C\} \cap \overline{\{O_S\}}$ is empty. At each repetition, add qualified observables to $\{O_S\}$, and record the quantum circuit set $\mathcal{S}_u^{(K)}$ as well.
6. For each recorded quantum circuit run on the quantum device to get the noisy expectation value for all its corresponding observables. For each observable O , feed the ideal and corresponding noisy value to fit the linear regression function and optimise the loss function

$$C = \sum_{\phi_i \in \mathcal{S}_\psi^{(K)}} \left(\langle O \rangle_{u_i}^{\text{exact}} - \left(a \langle O \rangle_{u_i}^{\text{noisy}} + b \right) \right)^2. \quad (\text{S47})$$

Therefore we get different linear mapping functions for different observables.

7. Use the fitting function to mitigate the error in expectation value for the corresponding observable.

The estimation of an observable $\mathbf{O}$ can be obtained by Eq. (S42), and is formally expressed as

$$\hat{\mathbf{o}}^{\text{CF}} = \sum_l \alpha_l \hat{\mathbf{o}}_l^{\text{CF}}, \quad (\text{S48})$$

where the estimator for $\mathbf{O}_l$ after REM and CF is given by

$$\hat{\mathbf{o}}_l^{\text{CF}} = f_l^{\text{CF}}(\hat{\mathbf{o}}_l^{\text{REM}}), \quad (\text{S49})$$

and f_l is the fitting function associated with $\mathbf{O}_l$.

C. Symmetry verification

Error mitigation by symmetry verification takes advantage of prior knowledge of the physical subspace where we search for the solution to the problem. There are two types of symmetries in the molecular problems considered here, which are the conservation of particle number and singlet state for the ground state. The particle number, hence the parity, of the ground state should be conserved throughout the simulation process. In addition, it is known in the quantum chemistry community that for closed-shell molecules, the ground state is bound to be a spin singlet state.

Below, we briefly review error mitigation by symmetry verification introduced in [23]. For each symmetry verification operator $\hat{S}_i$, we have,

$$\hat{S}_i|\psi\rangle = s_i|\psi\rangle, \quad (\text{S50})$$

where the $s_i \in \pm 1$. The density matrix, which is projected into the subspace that conserves the symmetry, can be expressed as,

$$\rho_{s_i} = \frac{\Pi_{s_i} \rho \Pi_{s_i}}{\text{Tr}(\Pi_{s_i} \rho)}, \quad (\text{S51})$$

with $\Pi_{s_i} = \frac{I + s_i \hat{S}_i}{2}$. Specifically, as a result of the closed shell molecule that we deal with, the trial wave function we pick is in an even-parity sub-manifold for total particle number and spin. Therefore, the eigenvalues of parity check operators for the correct symmetry are always one, i.e., $s_i = 1$. We project the target molecular Hamiltonian H to the even-parity subspace. For this purpose, we have

$$\text{Tr}(\hat{H} \rho_{s_i}) = \frac{\text{Tr}(\hat{H} \rho) + \text{Tr}(\hat{H} \hat{S}_i \rho)}{1 + \text{Tr}(\hat{S}_i \rho)}, \quad (\text{S52})$$

which effectively conserves the symmetry of the quantum subspace. This is achieved by measuring $\text{Tr}(H\rho)$, $\text{Tr}(H\hat{S}_i\rho)$, and $\text{Tr}(\hat{S}_i\rho)$.

We have the particle number or parity conservation for the molecular problems considered here, i.e., $[\hat{H}, \hat{S}] = 0$ with $\hat{S} = \bigotimes_i \hat{Z}_i$. This symmetry verification expands the to-be-measured observable set, which is composed of the elements in the decomposition of H and $H\hat{S}$. The symmetry verification EM has three ingredients $\langle \hat{H} \rangle$, $\langle \hat{H} \hat{S} \rangle$, and $\langle \hat{S} \rangle$. Each of them is estimated by measuring the predetermined basis, introduced in Sec. II D. The final error mitigated results by SV are computed classically by Eq. (S52) as

$$\hat{\mathbf{o}} = \frac{\langle \hat{H} \rangle^{\text{CF}} + \langle \hat{H} \hat{S} \rangle^{\text{CF}}}{1 + \langle \hat{S} \rangle^{\text{CF}}}. \quad (\text{S53})$$

Here, $\langle \hat{H} \rangle^{\text{CF}}$, $\langle \hat{H} \hat{S} \rangle^{\text{CF}}$, and $\langle \hat{S} \rangle^{\text{CF}}$ are the estimator obtained by sequentially applying REM in Eq. (S45) and CF in Eq. (S48), and classical post-processing by Eq. (S42).

In our experiment, we found that a combination of Clifford fitting and symmetry verification might not perform well. This is because, in our experiment, symmetry verification requires measuring many large-support observables, which demands high-precision measurement for these observables. In Fig. S12, we study the performance of symmetry verification and explain why it results in large fluctuations in the energy estimation, which will be elaborated on below.

D. Connected moments expansion

The connected moments expansion (CMX) method [24] is derived from the imaginary time evolution (ITE) approach that can be expressed in the following formula

$$E(\beta) = \frac{\langle \Phi | H e^{-\beta \hat{H}} | \Phi \rangle}{\langle \Phi | e^{-\beta \hat{H}} | \Phi \rangle} = \sum_{k=0}^{\infty} \frac{(-\beta)^k I_{k+1}}{k!}, \quad (\text{S54})$$

where β is the evolution time of the ITE method, and we expand the formula as the expansion of the connected moments I_k taking advantage of the Taylor expansion. The mechanism of the ITE method is to project the state $|\Psi\rangle$ towards the ground state of H , and Eq. (S54) gives the exact ground state energy as β goes to infinity as long as the state $|\Psi\rangle$ is non-orthogonal to the ground state. The evolution time β in practice is chosen to be a finite number and is essentially related to the overlap between the state $|\Psi\rangle$ and the ground state.

The ITE operator $e^{-\beta \hat{H}}$ is nonunitary and thus generally hard to implement on a quantum device. The CMX method provides a way to circumvent this problem through measuring the higher moments of $\hat{H}$ in Eq. (S54). Specifically, the connected moment is defined recursively by

$$I_k = \langle \Phi | \hat{H}^k | \Phi \rangle - \sum_{i=0}^{k-2} \binom{k-1}{i} I_{i+1} \langle \Phi | \hat{H}^{k-i-1} | \Phi \rangle. \quad (\text{S55})$$

The analytical form for exact energy as β goes to infinity can be represented by

$$E_{\text{ground}} = I_1 - \frac{S_{2,1}^2}{S_{3,1}} \left(1 + \frac{S_{2,2}^2}{S_{2,1}^2 S_{3,2}} \cdots \left(1 + \frac{S_{2,m}^2}{S_{2,m-1}^2 S_{3,m}} \right) \cdots \right), \quad (\text{S56})$$

where $S_{k,1} = I_k$, $S_{k,i+1} = S_{k,1} S_{k+2,i} - S_{k+1,i}^2$ and it is straightforward to have $S_{2,1} = I_2 = \langle \hat{H}^2 \rangle - \langle \hat{H} \rangle^2$ and $S_{3,1} = \langle \hat{H}^3 \rangle - 3 \langle \hat{H} \rangle \langle \hat{H}^2 \rangle + 2 \langle \hat{H} \rangle^3$. For practical concerns, one truncates the expansion up some order k . In our experiment, we measure the higher-order moment $\langle \Phi | \hat{H}^k | \Phi \rangle$ for $k \leq 3$ using the grouping measurement schemes introduced in Sec. IID. It is reasonable for our choice of $k \leq 3$ in that we expect the application of various quantum error mitigation (QEM) methods to be able to reproduce most parts of the ground state energy, but in some cases, it may not be able to reach the chemical-accuracy requirement. As such, the state we prepared in the energy estimation step corresponds to an effective state which has a fairly good overlap with the ground state, but is not exactly the same. Assuming to perform the ITE approach in this case, it follows that the evolution time β of the ITE method can be chosen to a relatively small one. Thus one can equivalently truncate the expansion Eq. (S55) up to some constant order as the expansion is monotonically decreasing. Truncating the expansion to a low order also makes the method affordable experimentally. Note that CMX is not a variational method, and thus the energies obtained by CMX may be lower than the true ground state energy.

In the experiment, after the VQE procedure, we would obtain the final circuit parameters. We then execute the circuit under different measurement bases to measure the expectation values of different observables ($\hat{H}$, $\hat{H}^2$ and $\hat{H}^3$). Finally, the ground state energy is estimated by Eq. (S56).

The variance of the energy estimation using the connected moments' expansion up to the second order is calculated by

$$\text{Var}[E_{\text{ground}}] = \left(\frac{\partial E_{\text{ground}}}{\partial a} \right)^2 \text{Var}[a] + \left(\frac{\partial E_{\text{ground}}}{\partial b} \right)^2 \text{Var}[b] + \left(\frac{\partial E_{\text{ground}}}{\partial c} \right)^2 \text{Var}[c], \quad (\text{S57})$$

and it takes the form of

$$\text{Var}[E_{\text{ground}}] = \left(1 + \frac{4aS_{2,1}}{S_{3,1}} - \frac{S_{2,1}^2(-3b+6a^2)}{S_{3,1}^2} \right)^2 \text{Var}[a] + \left(\frac{2S_{2,1}}{S_{3,1}} + \frac{3S_{2,1}^2 a}{S_{3,1}^2} \right)^2 \text{Var}[b] + \frac{S_{2,1}^4}{S_{3,1}^4} \text{Var}[c] \quad (\text{S58})$$

where we have defined $a := \langle \hat{H} \rangle$, $b := \langle \hat{H}^2 \rangle$ and $c := \langle \hat{H}^3 \rangle$. It is worth noting that this calculation for the variance is based on the variance propagation, and the actual variance could be smaller. The error bar for the energy estimation using CMX is calculated according to Eq. (S58).

E. Experimental procedure for observable estimation

In the experiment, according to the target observable set, $\{\mathbf{O}_l\}$, including the Hamiltonian $\hat{H}$, the symmetry verified Hamiltonian $\hat{H}\hat{S}$, parity $\hat{S}$, and the Hamiltonian power $\hat{H}^k$, the measurement basis $\{\mathbf{P}\}$ and the associated measurement shots are generated using the method introduced in Sec. II D. The experimental procedure for obtaining error-mitigated observable estimation in order is summarised as follows.

1. Experimental benchmark for measurement noise. Test whether the measurement errors admit a tensor product type by random state measurement. Obtain the calibration matrix Λ by measurement error calibration.
2. Generate a series of random quantum circuits for Clifford noise learning. Measure the state prepared by Clifford circuits in the predetermined basis $\{\mathbf{P}\}$ and obtain the fitting function f^{CF} for each observable.
3. During the VQE process, measure the observable on the state in basis $\{\mathbf{P}\}$. For each $\mathbf{O}_l$, apply readout error mitigation by Eq. (S45). Get the observable estimation after CF in Eq. (S48) or Eq. (S49).
4. Get symmetry verified results using Eq. (S52). In our experiment, we apply symmetry verification to correct the energy for H_2 and LiH .
5. The final ground state energy is estimated using the CMX in Eq. (S56).

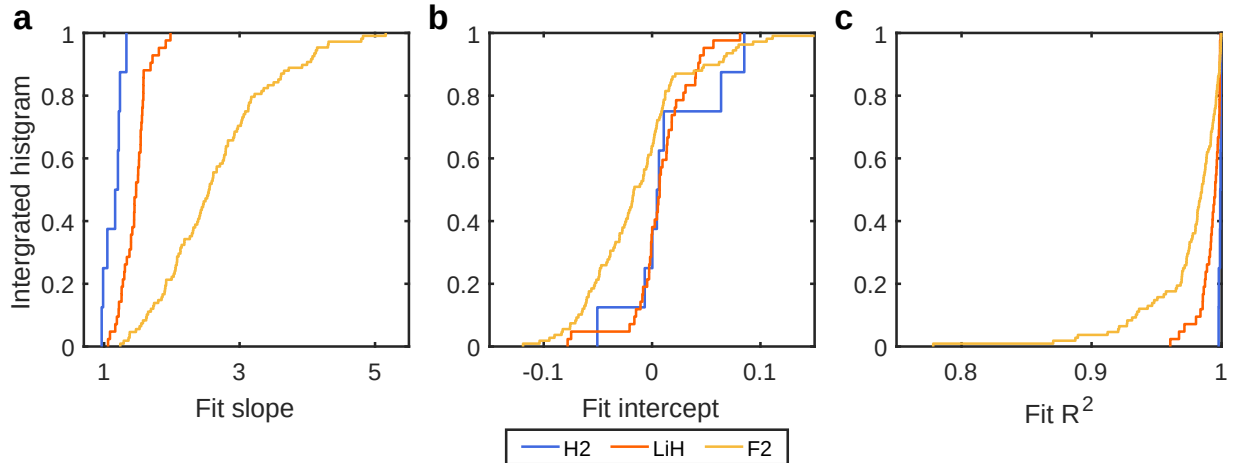


FIG. S9. **Distribution of Clifford fit parameters.** The accumulated distribution of slope, intercept and goodness of fit for H_2 ($R=2.6$), LiH ($R=1.5$) and F_2 ($R=2.0$) are shown in the diagram. With the system size and circuit depth increasing, we can observe that (a) the slopes of Clifford fit become larger due to the impact of decoherence. (b) the variance of intercept becomes larger, and the mean value is off zero, which may be caused by gate correlated errors. (c) the goodness of fit becomes worse. Two main issues may cause it. One is gate-correlated errors that make our model nonlinear. The other is insufficient fit data, which will increase our experiment resources.

IV. EXPERIMENTAL IMPLEMENTATION AND RESULTS

In Sec. I, we show how to optimise the Zuchongzhi processor to obtain high-fidelity gates and readout. In Sec. II, we discuss quantum algorithms for the ground state search and quantum state measurement with fewer quantum resources. Subsequently, in Sec. III, we show how to obtain error-mitigated ground state energy estimation. With these experimental and theoretical advances, we study systematically the ground state energy estimation for three molecules, H_2 , LiH , F_2 . Finally, we point out the challenges in experimental quantum computational chemistry.

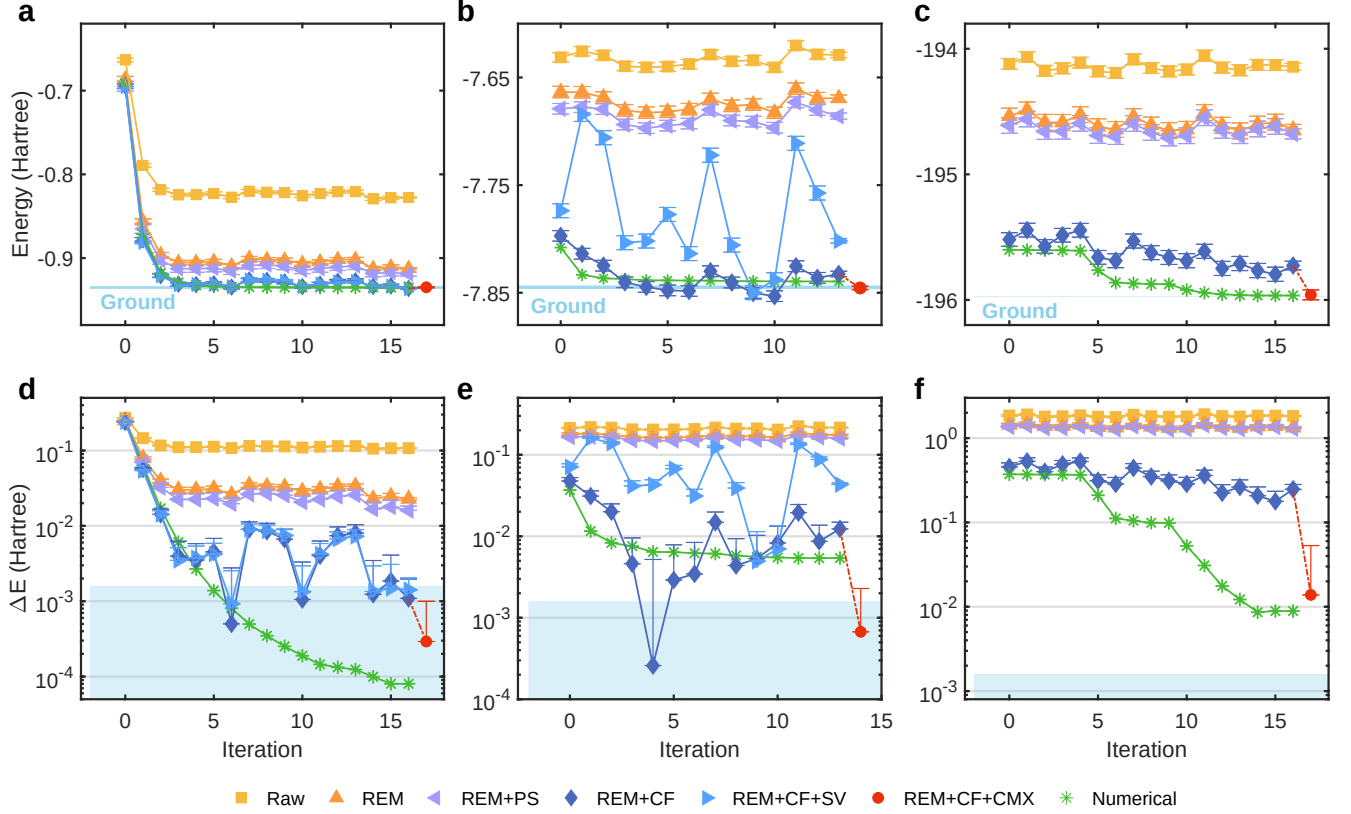


FIG. S10. **Optimisation procedure.** The optimisation procedure traces of energy and abs energy error for H_2 ($R=2.6$), LiH ($R=2.2$) and F_2 ($R=2.6$) are shown in the picture.

Here in (a)(b)(c), we add energy with REM + PS (purple leftward triangles), while the others are the same as in Figure 2b.

Figure (d)(e)(f) show the traces of abs energy error during the optimisation procedure.

A. Experimental procedure

The benchmarking and the overall experimental procedure have been discussed in Methods. The experimental procedures for error mitigation have been shown in Sec. III E. The VQE optimisation is summarised as follows:

1. Gradient estimation and optimisation. Determine θ associated with the single-qubit Pauli-rotation gate, in which we select $N/2$ parameters to update. For each selected parameter, the gradient is estimated using Eq. (S31), which requires measuring the Hamiltonian on the parameter-shifted state. Get the gradient for all selected parameters and update the parameters. Here, we obtain the error-mitigated measurement outcome $\mathcal{L}(\theta^\pm)$ defined in Eq. II C by REM and CF error mitigation schemes.
2. For each iteration, we measure the energy with the updated parameters and hence record the optimisation trace.
3. Run the 1 and 2 steps to update the parameters until convergence. For all the molecules, the energy converges within 15 iterations.

Finally, we measure the state with the optimised parameters. We correct the ground state energy using connected moments expansion (CMX), which requires measuring $\langle \hat{H}^2 \rangle$ and $\langle \hat{H}^3 \rangle$, as described in Sec. III D. The expectation value of $\langle \hat{H}^k \rangle$ for $k \leq 3$ will be measured and then error mitigated in a similar fashion to that of $\langle \hat{H} \rangle$. With the error mitigated results, we calculate the CMX energy using equations Eq. (S55) and Eq. (S56).

B. Experimental results

This section shows experimental results for the ground state search and ground state energy estimation.

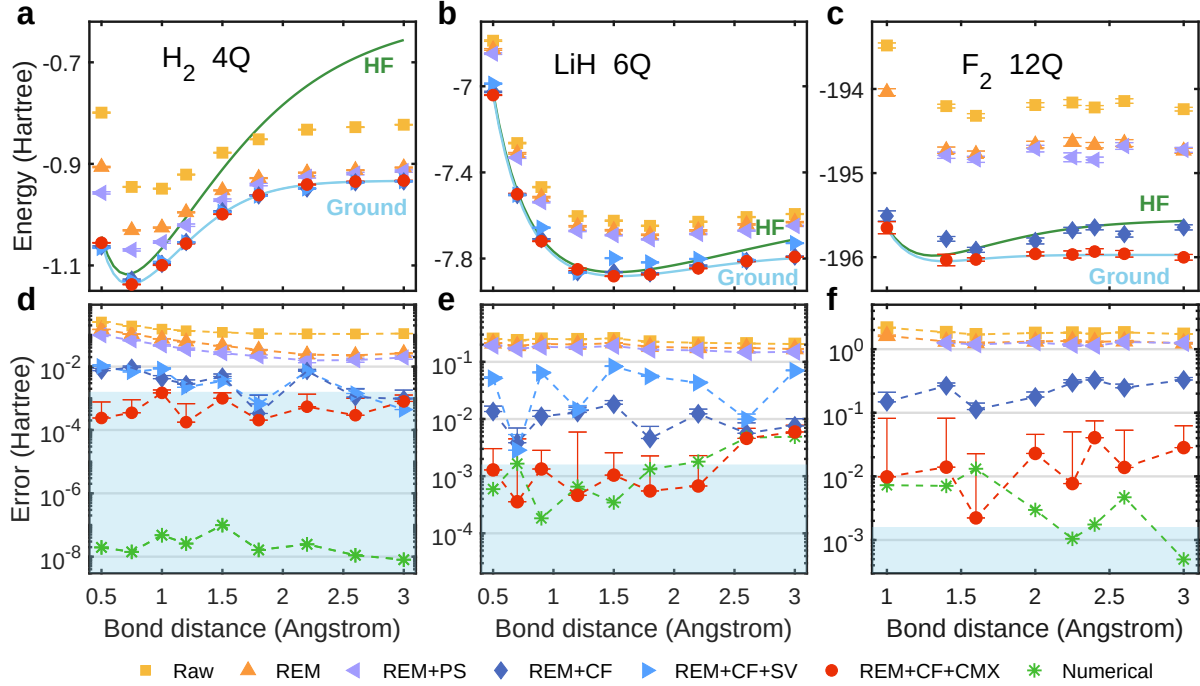


FIG. S11. **The VQE simulations for potential energy curves for different molecules.** Similar to Figure 3, more results with different error mitigation strategies are shown here. (a-c) Potential energy curves as functions of the bond distance for H_2 (4 qubits), LiH (6 qubits) and F_2 (12 qubits) molecules with various error mitigation strategies. (d-f) Absolute errors are compared to the FCI results. We compare the raw data (yellow squares) with the application of REM (orange triangles), REM and PS (purple triangles), REM and CF (deep blue diamonds), SV (blue triangles) and CMX (red circles). The results marked by green asterisks are energies calculated by classical simulation with the parameters searched in the experiment. The ground state energy with chemical accuracy (blue regime) is calculated theoretically as a reference.

We select at most 12 qubits on the Zuchongzhi 2.0 quantum processor. Several patterns of 12 qubits are available in our processor. Different pattern choices could lead to significantly different results in our experiment. Since learning the gate error is crucial in our experiment, we run the Clifford fitting experiment with different qubit patterns to characterise the performance and finally select the pattern based on the goodness of fit and the fit slope.

We first select three bond distances for different molecules, including H_2 ($R=2.6$), LiH ($R=2.2$) and F_2 ($R=2.6$), to study the performance of the Clifford fitting. Fig. S9 shows the performance of the Clifford fitting for the three molecules at the selected bond distances by the accumulated distribution of the slope, intercept, and the R^2 value. Through the comparison of a, b, and c, we find that the fitting becomes worse with the system size increases, which may be attributed to the correlated gate errors and the fluctuations of the data.

Then, we show the ground state searching with the variational algorithms for the three molecules. We show the optimisation processes for H_2 ($R=2.6$), LiH ($R=2.2$) and F_2 ($R=2.6$) in Fig. S10. The figure presents error-mitigated energy estimations with different error mitigation schemes, including REM, PS, CF, SV, CMX, and their combinations. Fig. S10 shows that although measured energies (yellow square) have fluctuations, energies after applying the combination of REM and CF clearly improve the energy accuracy along the optimisation. Error-mitigated energies tend to decrease consistently in successive iterations.

Next, we show the calculation of potential energy curves in Fig. S11. Fig. S11a, b, c show the energy iteration during the VQE optimisation process for H_2 ($R=2.6$), LiH ($R=1.5$) and F_2 ($R=2.0$), respectively. These results provide a much more detailed comparison of the error mitigation schemes used in the experiment. In the following, we discuss the performance of these error mitigation schemes.

We first discuss error mitigation with symmetry verification. As can be found in Fig. S10 and Fig. S11, SV results in a large fluctuation in the energy estimation. In particular, as illustrated in Fig. S11b, the energy with REM, CF and SV (grey right triangles) have a greater fluctuation than REM + CF during the optimisation procedure. The energy with REM, CF and SV show a larger error than REM and CF for all bond distances, as can be found in Fig. S12a. This can be understood as follows. The fitting coefficient for the symmetry verified Hamiltonian $\hat{H}\hat{S}$ is much larger

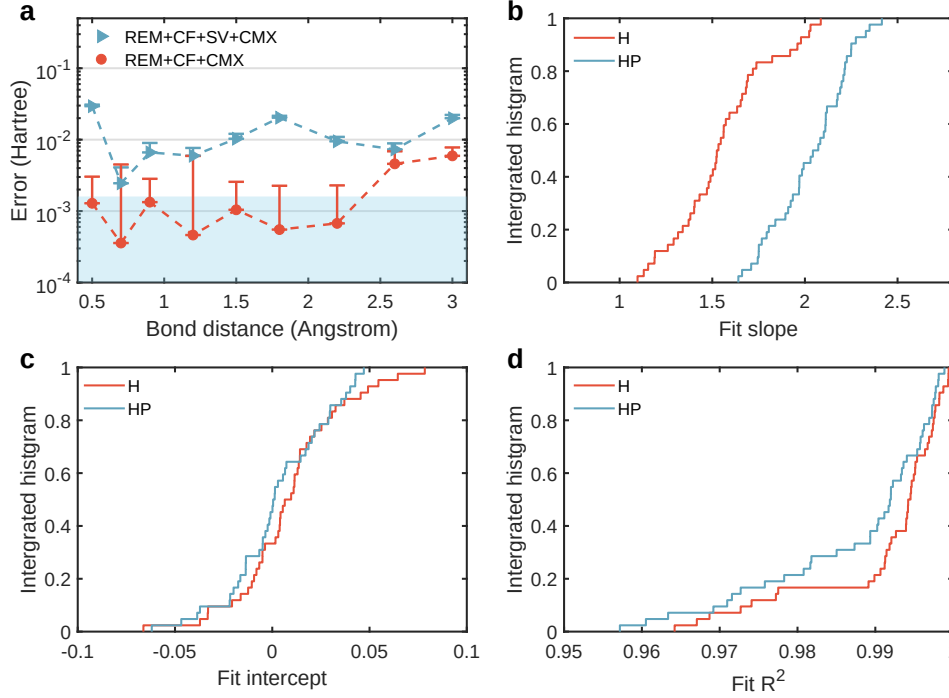


FIG. S12. **Symmetry Verification.** (a) The comparison of the abs energy error with different error mitigation setups for LiH. The red dot is the energy with REM + Clifford fitting + CMX, and the grey-blue triangle is the energy with REM + Clifford fitting + CMX + SV. (b)(c)(d) are the slope, intercept and R^2 of the LiH($R = 2.2$), respectively. “H” refers to the original Hamiltonian $\hat{H}$ and “HP” refers to the symmetry verified Hamiltonian $\hat{H}\hat{S}$.

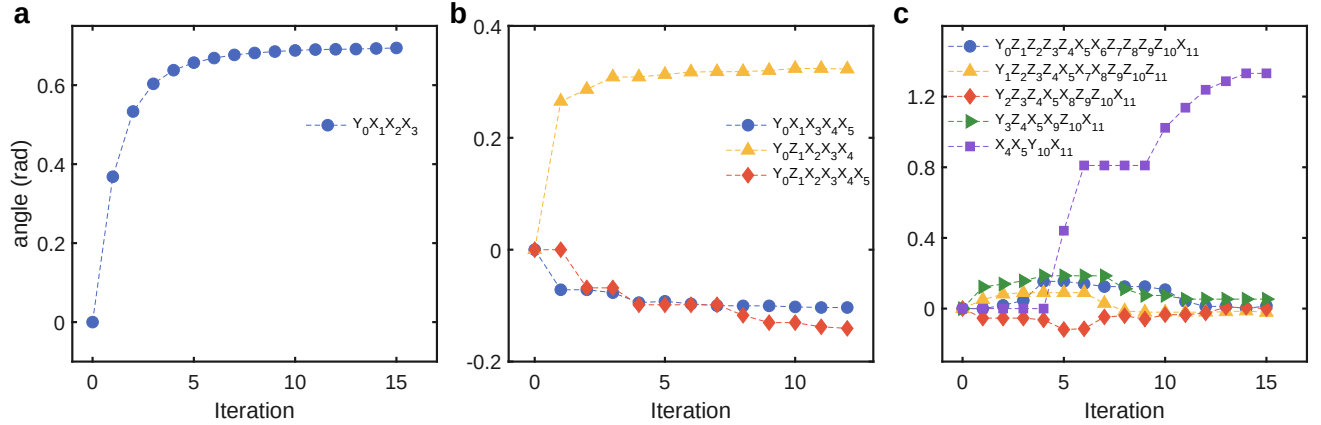


FIG. S13. **Parameter optimisation processes.** The evolution of variational parameters in optimisation processes for H_2 (a), LiH (b) and F_2 (c), respectively. For each iteration, we select $N/2$ parameters to update, and the parameter converges within 15 iterations for all the molecules.

than $\hat{H}$ and thus results in large fluctuation after applying the combination of CF and SV. This is verified by the slope in Fig. S12b and the fitting performance in Fig. S12d.

These results indicate the combination of SV and CF could bring a large fluctuation. Therefore, we mainly focused on the readout error mitigation, Clifford fitting and ground state energy correction by using connected moment expansion in the main text, while we did not show the results energy of LiH and F_2 using symmetry verification.

In addition, we add the energy with REM and post-selection (PS). In this scheme, the measurement result of $Z^{\otimes n}$ Pauli basis is post-selected based on the conservation of particle numbers, and other Pauli bases are still error-mitigated with REM. The REM and post-selection (PS) energy perform better than single REM yet worse than

REM and CF. With an increasing system size, a combination of REM and PS shows diminishing energy accuracy improvements, implying PS is not a suitable error mitigation strategy for scalable quantum computational chemistry.

Finally, we show the performance of VQE after the optimisation process. Fig. S14 shows the relative error compared to the initial energy for the three molecules at different bond distances. The relative error compared to the initial state energy is calculated by

$$\frac{E_i - E_{\text{ground}}}{E_0^{\text{expm}} - E_{\text{ground}}}, \quad (\text{S59})$$

where E_i is the error-mitigated energy or the energy, which is computed using the parameters that are found until convergence, E_0^{expm} is the energy measured on the initial state, E_{ground} is the ideal ground state energy. This relative error characterises the improvement with respect to the initial state energy by VQE.

We can see from Fig. S14 that compared to the initial state energy, the relative average energy errors with only REM for the three molecules are 28.98%, 78.86% and 75.42%.

The results are improved by 0.278%, 0.778% (0.326% except R2.6 and R3.0), and 1.00% for molecules with REM+CF+CMX EM schemes, and by 1.56×10^{-7} , 0.783% (0.392% except R2.6 and R3.0), and 0.268% for numerical result. The average absolute error associated with different methods, including REM, REM + PS, REM+CF+SV, REM+CF, REM+CF+CMX, and numerical, are displayed in Table I.

Molecules	Raw	REM	REM + PS	REM+CF+SV	REM+CF	REM+CF+CMX	Numerical
H ₂ (milli-Hartree)	143.661 $\pm$ 0.807	60.580 $\pm$ 1.015	38.650 $\pm$ 2.367	4.552 $\pm$ 0.740	4.254 $\pm$ 0.997	0.565 $\pm$ 0.542	2.96×10^{-5}
LiH (milli-Hartree)	233.564 $\pm$ 2.704	191.358 $\pm$ 3.313	169.987 $\pm$ 3.110	43.888 $\pm$ 2.383	9.965 $\pm$ 2.923	1.802 $\pm$ 2.423	1.794
F ₂ (Hartree)	1.8315 $\pm$ 0.0264	1.3419 $\pm$ 0.0387	1.2231 $\pm$ 0.0382	–	0.2378 $\pm$ 0.0354	0.0174 $\pm$ 0.0417	0.0048

TABLE I. Comparison of average absolute energy error for H₂, LiH and F₂ using different error mitigation schemes. The first two rows use milli-Hartree.

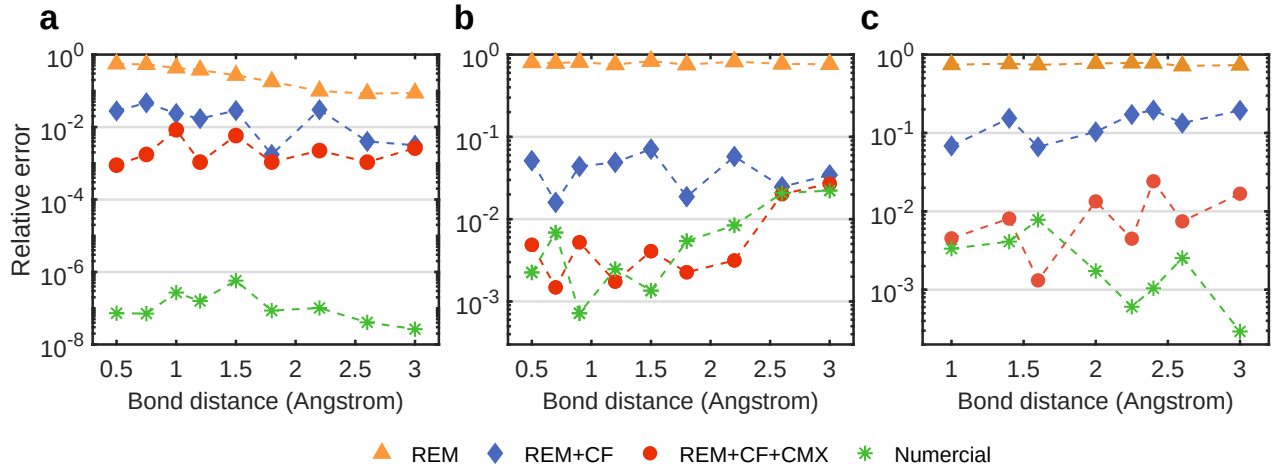


FIG. S14. **Relative error compared to the initial state energy.** Relative errors for H₂, LiH and F₂ with different error mitigation schemes, only REM (yellow triangles), REM and Clifford fitting schemes (blue diamonds), REM and Clifford fitting and CMX (red circles), the numerical result (green stars).

C. Parameters iteration

In Fig. S13, we present the parameters iteration for H₂ (R=2.6), LiH (R=1.5) and F₂ (R=2.0). For H₂, LiH and F₂, there are 1, 3 and 5 parameters in total to iterate respectively. All parameters are optimised from zero. 1, 2 and 3 parameters are randomly selected to optimise during each iteration for H₂, LiH and F₂, respectively. All optimisation runs for 5 to 15 iterations.

D. Noise analysis

In this section, we analyse the noise effect in our chemistry simulation. We assume noise in the circuit can be described by a depolarising channel

$$\mathcal{E}(\rho) = p\rho + (1-p)\frac{I}{2^N}, \quad (\text{S60})$$

where $1-p$ characterises the error rate, and p characterises the fidelity of the quantum state. Under depolarising noise, the energy is given by

$$E^{\text{noisy}} = \text{Tr}(\rho^{\text{ideal}} \hat{H}) = p\text{Tr}(\rho^{\text{ideal}}) = pE^{\text{ideal}}. \quad (\text{S61})$$

Here, the identity operator in the Hamiltonian has been excluded so that $\text{Tr}(\hat{H}) = 0$. We can get the error rate p by fitting the measured energies E^{noisy} and the corresponding ideal energies E^{ideal} .

We show the ideal energies and measured noisy energies in Fig. S15 along the optimisation iterations for H_2 (a), LiH (b) and F_2 (c) at all the bond distances. The measured noisy energies represent the measurement energies only with Readout Error Mitigation (REM). We observe that the ideal energies and measured energies have a well-linear dependence. The slope of the fitting curves gives us the depolarising fidelity p , which decreases with a deeper circuit.

To further our understanding of noise with an increasing circuit size, we compare the circuit fidelity estimated using different methods. As a reference, we calculate the circuit fidelity by multiplying the fidelity of each individual gate, which is obtained by individual XEB. Specifically, we first calculate the fidelity of each layer using the Pauli fidelity, and calculate the whole circuit fidelity using the depolarising fidelity of different layers. Since dynamical decoupling is used in our experiment to avoid dephasing during the idle time, we also take account of the idle time decay. The idle time fidelity is calculated by the fidelity of a single-qubit gate with an equal time length.

Fig. S15d shows the comparison of the estimation using individual XEB results with (red) and without idle time decay (blue), and the fitted depolarising fidelity by fitting the ideal energies and measured energies (orange). Remarkably, we can find that the actual circuit error rate is much lower than the error estimated with individual XEB results when the system size increases. This result implies the potential for the simulation of larger molecules. A detailed investigation will be an interesting future work.

E. Challenges

As we have discussed here and in the main text, there are many challenges in VQE experiments. Compared to dynamics simulation or random circuit sampling, VQE experiments are more challenging. The former examples only require running one instance on a quantum processor. In contrast, in the VQE experiment, the final results are highly related to the results that are obtained in the prior incidents. In spite of these challenges, our experiments for the first time achieved chemical accuracy for the six-qubit VQE simulation of molecular systems. Good control of hardware and efficient quantum algorithms enable such high-precision simulation.

For future large-scale quantum computational chemistry experiments, it is crucial to suppress gates and measurement readout errors in order to achieve chimerical accuracy with quantum hardware. We point out these challenges as follows.

Stability during experiments. For large-scale problems, VQE experiments could take several hours or days to obtain the ground state. Therefore, the performance of VQE highly depends on the stability of the experimental system. In our experiment, the readout error mitigation and Clifford fitting schemes are applied. These techniques require maintaining the system feature during the REM matrix extraction experiment, Clifford fitting experiment and VQE experiment. Only when the system float is suppressed can we make the error mitigation schemes work. The fluctuation mainly comes from the following:

1. **Electronic outcome fluctuation.** The electronic outcome determines the system conditions and manipulation accuracy, such as the qubit frequencies, coupler flux bias and the control pulse amplitude. The fluctuation of electronics directly affects gate fidelities as well as the results of the quantum circuit. The outcome accuracy of the electronic itself could be suppressed by increasing the samples, but the average outcome amplitude is greatly influenced by ambient temperature.
2. **Unstable Two Level System (TLSs).** When the unstable TLSs get close to the qubit operating frequency, the coherence of the qubits would decrease abruptly. The actual noisy may be changed, causing the Clifford fitting to fail.

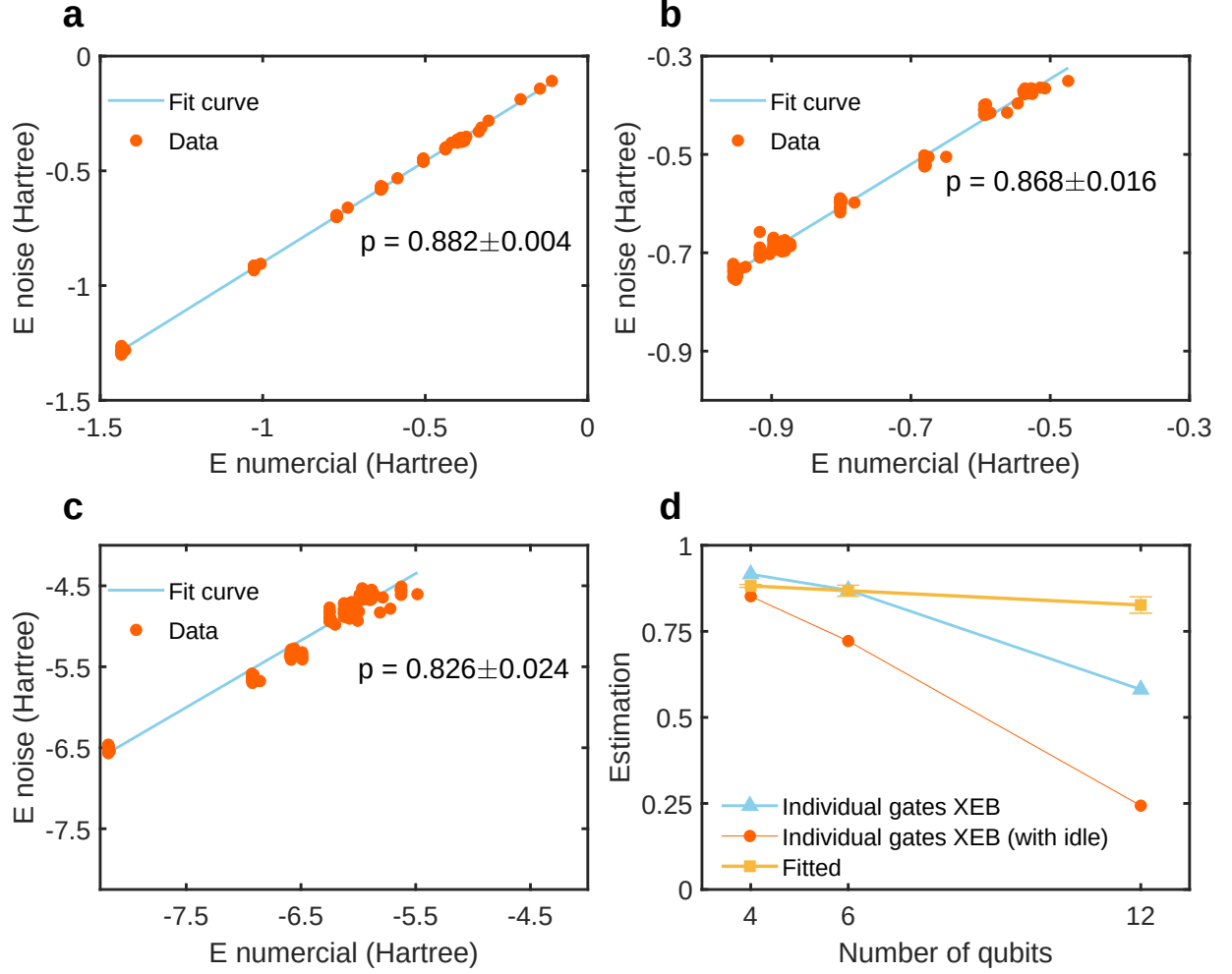


FIG. S15. **Fitted results using a depolarising noise model.** The figures show the ideal energies and measured energies along the optimisation iterations for H_2 (a), LiH (b) and F_2 (c) at all the bond distances. The noisy energies represent the measurement energies only with Readout Error Mitigation (REM).

The ideal energies and measured energies have a linear dependence. The slope of the fitting curves characterises the depolarising fidelity p . (d) The circuit fidelity estimation by individual XEB results with and without idle time decay, and the fitted depolarising fidelity obtained by fitting the data points in (a)(b)(c).

Errors and samples. In our experiment, as we pointed out, we require the expectation values of observables for the optimisation and energy evaluation, instead of single-shot measurements. To suppress the statistical errors less than ϵ , the number of samples scales as $1/\epsilon^2$. With the increase of qubit number, coherent errors become a main error source, which leads to a large slope for the Clifford fitting function f^{CF} , as shown in Fig. S9, and hence a large variance. As a result of fitting, more samples are required for a larger system to suppress the fluctuation. Nevertheless, this can be overcome by further improving hardware performance. This is because a smaller error rate will lower the inherent expectation value variance, and hence fewer samples are required.

In our experiment, the average number of shots for a single bond length VQE procedure of H_2 , LiH , and F_2 is 1.3×10^7 , 4.1×10^7 , and 7.3×10^7 , respectively. However, the number of observables will increase with the system size, and thus more samples are needed in a larger VQE experiment. Moreover, as extensively discussed, the CMX and Clifford fitting methods will introduce a fairly large measurement overhead, which may become problematic for larger systems. The requirement for more measurements will need a long time to run experiments, but it could be hard to

stabilise the system for a long time. Better measurement schemes would be highly desirable from the algorithm side.

-
- [1] Z. R. Lin, K. Inomata, W. D. Oliver, K. Koshino, Y. Nakamura, J. S. Tsai, and T. Yamamoto. Single-shot readout of a superconducting flux qubit with a flux-driven josephson parametric amplifier. *Applied Physics Letters*, 103(13):132602, 2013.
 - [2] J. Y. Mutus, T. C. White, R. Barends, Yu Chen, Z. Chen, B. Chiaro, A. Dunsworth, E. Jeffrey, J. Kelly, A. Megrant, C. Neill, P. J. J. O'Malley, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, K. M. Sundqvist, A. N. Cleland, and John M. Martinis. Strong environmental coupling in a josephson parametric amplifier. *Applied Physics Letters*, 104(26):263513, 2014.
 - [3] Zhiguang Yan, Yu-Ran Zhang, Ming Gong, Yulin Wu, Yarui Zheng, Shaowei Li, Can Wang, Futian Liang, Jin Lin, Yu Xu, Cheng Guo, Lihua Sun, Cheng-Zhi Peng, Keyu Xia, Hui Deng, Hao Rong, J. Q. You, Franco Nori, Heng Fan, Xiaobo Zhu, and Jian-Wei Pan. Strongly correlated quantum walks with a 12-qubit superconducting processor. *Science*, 364(6442):753–756, 2019.
 - [4] Yangsen Ye, Sirui Cao, Yulin Wu, Xiawei Chen, Qingling Zhu, Shaowei Li, Fusheng Chen, Ming Gong, Chen Zha, He-Liang Huang, Youwei Zhao, Shiyu Wang, Shaojun Guo, Haoran Qian, Futian Liang, Jin Lin, Yu Xu, Cheng Guo, Lihua Sun, Na Li, Hui Deng, Xiaobo Zhu, and Jian-Wei Pan. Realization of high-fidelity controlled-phase gates in extensible superconducting qubits design with a tunable coupler. *Chinese Physics Letters*, 38(10):100301, nov 2021.
 - [5] Youngkyu Sung, Leon Ding, Jochen Braumüller, Antti Vepsäläinen, Bharath Kannan, Morten Kjaergaard, Ami Greene, Gabriel O. Samach, Chris McNally, David Kim, Alexander Melville, Bethany M. Niedzielski, Mollie E. Schwartz, Jonilyn L. Yoder, Terry P. Orlando, Simon Gustavsson, and William D. Oliver. Realization of high-fidelity cz and zz-free iswap gates with a tunable coupler. *Phys. Rev. X*, 11:021058, Jun 2021.
 - [6] Julian Kelly, Peter O'Malley, Matthew Neeley, Hartmut Neven, and John M Martinis. Physical qubit calibration on a directed acyclic graph. *arXiv preprint arXiv:1803.03226*, 2018.
 - [7] Yulin Wu, Wan-Su Bao, Sirui Cao, Fusheng Chen, Ming-Cheng Chen, Xiawei Chen, Tung-Hsun Chung, Hui Deng, Yajie Du, Daojin Fan, et al. Strong quantum computational advantage using a superconducting quantum processor. *Physical review letters*, 127(18):180501, 2021.
 - [8] Sam McArdle, Suguru Endo, Alán Aspuru-Guzik, Simon C. Benjamin, and Xiao Yuan. Quantum computational chemistry. *Rev. Mod. Phys.*, 92:015003, Mar 2020.
 - [9] Rodney J. Bartlett, Stanislaw A. Kucharski, and Jozef Noga. Alternative coupled-cluster ansätze II. the unitary coupled-cluster method. *Chem. Phys. Lett.*, 155(1):133–140, 1989.
 - [10] Jonathan Romero, Ryan Babbush, Jarrod R McClean, Cornelius Hempel, Peter J Love, and Alán Aspuru-Guzik. Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz. *Quantum Science and Technology*, 4(1):014008, 2018.
 - [11] Dave Wecker, Matthew B. Hastings, and Matthias Troyer. Progress towards practical quantum variational algorithms. *Phys. Rev. A*, 92:042303, Oct 2015.
 - [12] Joonho Lee, William J Huggins, Martin Head-Gordon, and K Birgitta Whaley. Generalized unitary coupled cluster wave functions for quantum computation. *Journal of chemical theory and computation*, 15(1):311–324, 2018.
 - [13] Changsu Cao, Jiaqi Hu, Wengang Zhang, Xusheng Xu, Dechin Chen, Fan Yu, Jun Li, Han-Shi Hu, Dingshun Lv, and Man-Hong Yung. Progress toward larger molecular simulation on a quantum computer: Simulating a system with up to 28 qubits accelerated by point-group symmetry. *Phys. Rev. A*, 105:062452, Jun 2022.
 - [14] John F Stanton, Jürgen Gauss, John D Watts, and Rodney J Bartlett. A direct product decomposition approach for symmetry exploitation in many-body methods. i. energy calculations. *The Journal of Chemical Physics*, 94(6):4334–4345, 1991.
 - [15] Nicholas C. Rubin, Klaas Gunst, Alec White, Leon Freitag, Kyle Throssell, Garnet Kin-Lic Chan, Ryan Babbush, and Toru Shiozaki. The fermionic quantum emulator. *Quantum*, 5:568, Oct 2021.
 - [16] Kosuke Mitarai, Makoto Negoro, Masahiro Kitagawa, and Keisuke Fujii. Quantum circuit learning. *Physical Review A*, 98(3):032309, 2018.
 - [17] Andrea Mari, Thomas R Bromley, and Nathan Killoran. Estimating the gradient and higher-order derivatives on quantum hardware. *Physical Review A*, 103(1):012405, 2021.
 - [18] Bujiao Wu, Jinzhao Sun, Qi Huang, and Xiao Yuan. Overlapped grouping measurement: A unified framework for measuring quantum states. *Quantum*, 7:896, 2023.
 - [19] Charles Hadfield, Sergey Bravyi, Rudy Raymond, and Antonio Mezzacapo. Measurements of quantum hamiltonians with locally-biased classical shadows. *Communications in Mathematical Physics*, 391(3):951–967, 2022.
 - [20] Sergey Bravyi, Sarah Sheldon, Abhinav Kandala, David C McKay, and Jay M Gambetta. Mitigating measurement errors in multiqubit experiments. *Phys. Rev. A*, 103(4):042605, 2021.
 - [21] Piotr Czarnik, Andrew Arrasmith, Patrick J Coles, and Lukasz Cincio. Error mitigation with clifford quantum-circuit data. *Quantum*, 5:592, 2021.
 - [22] Sergey Bravyi, Dan Browne, Pádraic Calpin, Earl Campbell, David Gosset, and Mark Howard. Simulation of quantum circuits by low-rank stabilizer decompositions. *Quantum*, 3:181, 2019.
 - [23] X. Bonet-Monroig, R. Sagastizabal, M. Singh, and T. E. O'Brien. Low-cost error mitigation by symmetry verification.

Phys. Rev. A, 98:062339, Dec 2018.

- [24] Karol Kowalski and Bo Peng. Quantum simulations employing connected moments expansions. The Journal of Chemical Physics, 153(20):201102, 2020.