

Chemistry Beyond Exact Solutions on a Quantum-Centric Supercomputer

Javier Robledo-Moreno,^{1,*} Mario Motta,^{1,†} Holger Haas,¹ Ali Javadi-Abhari,¹ Petar Jurcevic,¹ William Kirby,² Simon Martiel,³ Kunal Sharma,¹ Sandeep Sharma,⁴ Tomonori Shirakawa,^{5,6,7} Iskandar Sitedikov,¹ Rong-Yang Sun,^{5,6,7} Kevin J. Sung,¹ Maika Takita,¹ Minh C. Tran,² Seiji Yunoki,^{5,6,7,8} and Antonio Mezzacapo^{1,‡}

¹IBM Quantum, IBM T.J. Watson Research Center, Yorktown Heights, NY 10598

²IBM Quantum, IBM Research Cambridge, Cambridge, MA 02142, USA

³IBM Quantum, IBM France Lab, Orsay, France

⁴Department of Chemistry, University of Colorado, Boulder, CO 80302, USA

⁵Computational Materials Science Research Team,

RIKEN Center for Computational Science (R-CCS), Kobe, Hyogo, 650-0047, Japan

⁶Quantum Computational Science Research Team,

RIKEN Center for Quantum Computing (RQC), Wako, Saitama, 351-0198, Japan

⁷RIKEN Interdisciplinary Theoretical and Mathematical Sciences Program (iTHEMS), Wako, Saitama 351-0198, Japan

⁸RIKEN Center for Emergent Matter Science (CEMS), Wako, Saitama 351-0198, Japan

A universal quantum computer can be used as a simulator capable of predicting properties of diverse quantum systems [1]. Electronic structure problems in chemistry offer practical use cases around the hundred-qubit mark [2]. This appears promising since current quantum processors have reached these sizes. However, mapping these use cases onto quantum computers yields deep circuits, and for pre-fault-tolerant quantum processors, the large number of measurements to estimate molecular energies leads to prohibitive runtimes [3]. As a result, realistic chemistry is out of reach of current quantum computers in isolation. A natural question is whether classical distributed computation can relieve quantum processors from parsing all but a core, intrinsically quantum component of a chemistry workflow. Here, we incorporate quantum computations of chemistry in a quantum-centric supercomputing architecture, using up to 6400 nodes of the supercomputer Fugaku to assist a Heron superconducting quantum processor. We simulate the N₂ triple bond breaking in a correlation-consistent cc-pVDZ basis set, and the active-space electronic structure of [2Fe–2S] and [4Fe–4S] clusters [4], using 58, 45 and 77 qubits respectively, with quantum circuits of up to 10570 (3590 2-qubit) quantum gates. We obtain our results using a class of quantum circuits that approximates molecular eigenstates, and a hybrid estimator. The estimator processes quantum samples, produces upper bounds to the ground-state energy and wavefunctions supported on a polynomial number of states. This guarantees an unconditional quality metric for quantum advantage, certifiable by classical computers at polynomial cost. For cur-

rent error rates, our results show that classical distributed computing coupled to quantum processors can produce good approximate solutions for practical problems beyond sizes amenable to exact diagonalization.

The most common task in theoretical quantum chemistry is the computation of ground-state energies by solving the Schrödinger equation $H|\Psi\rangle = E|\Psi\rangle$ in the Born-Oppenheimer approximation. Exact numerical solutions in a finite basis set have a cost growing combinatorially in the number of electrons and orbitals. This limits exact diagonalization in the full configuration interaction (FCI) to system sizes close to 22 electrons in 22 orbitals (22e,22o) [5] and (26e,23o) [6]. For system sizes beyond the reach of FCI, one must rely on approximate methods, e.g., diagrammatic techniques, wavefunction ansatzes, and Monte Carlo integration [7, 8].

Progress in quantum computing has triggered a flurry of theoretical proposals for computational chemistry over the last decade (e.g., [9–11]). At the same time, attempts have been made at implementations on pre-fault-tolerant quantum processors [12–18], but these have so far been limited to small systems, for two main reasons. First, despite numerous efforts to improve on the measurement problem (e.g., [19–21]), runtime for energy expectation value estimation on interesting systems remains out of any reasonable timescale. Second, the depths of chemically-motivated quantum circuits for computations of chemistry are very high. For unitary coupled cluster [22] and a single step of time evolution, these quantities scale as M^4 [23] on a system with M spin-orbitals. While this scaling can be improved with various techniques [24], on pre-fault-tolerant devices the signal emerging from circuits of such size is weakened by the accumulation of gate errors and qubit decoherence.

I. METHODS

We set up the discussion of our results by considering the quantum-centric supercomputing architecture [25]

* j.robledomoreno@ibm.com

† mario.motta@ibm.com

‡ mezzacapo@ibm.com

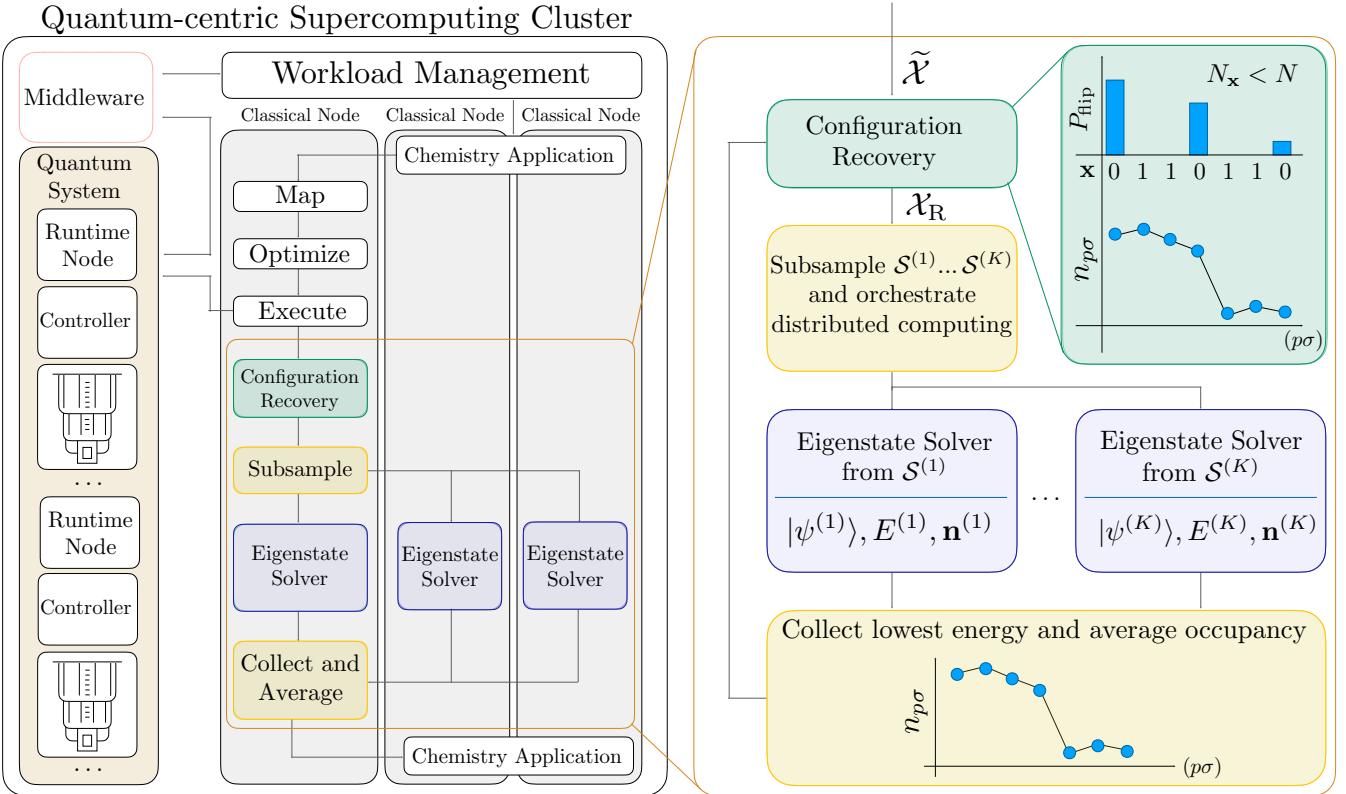


Figure 1. Quantum-centric supercomputing architecture and workflow diagram. **Left:** We illustrate a simplified architecture used to execute our workflow. The architecture has a cluster with a quantum system alongside classical runtime nodes within an isolated environment. A workload management system controls hybrid quantum-classical jobs through middleware. Our workflow is distributed on a set of classical nodes. It includes standard quantum chemistry application routines such as computing electronic integrals, mapping to qubits, and preparing circuits to be executed. **Right:** Details of the classical post-processing step. The input is a set of noisy samples $\tilde{\mathcal{X}}$ from the quantum execution that are processed with our configuration recovery step, using information from a vector \mathbf{n} of reference orbital occupancies. The green inset shows an example where a configuration with $N_x < N$ is corrected. The set of recovered configurations \mathcal{X}_R is sub-sampled and distributed for projection and diagonalization on parallel classical nodes. A new average reference occupancy vector \mathbf{n} is computed from the results, and the configuration recovery loop is repeated self-consistently until convergence.

schematized in Fig. 1 a . The architecture enables scaling of computational capacity by leveraging quantum processors for their natural task: executing a limited number of large quantum circuits. We follow the workflow in Fig. 1 to summarize our methods.

Our main goal is to find the ground state of chemistry Hamiltonians

$$\hat{H} = \sum_{pr} h_{pr} \hat{a}_{p\sigma}^\dagger \hat{a}_{r\sigma} + \sum_{\sigma\tau} \frac{(pr|qs)}{2} \hat{a}_{p\sigma}^\dagger \hat{a}_{q\tau}^\dagger \hat{a}_{s\tau} \hat{a}_{r\sigma}, \quad (1)$$

expanded over a discrete basis set. Here we have defined the fermionic creation/annihilation operator $\hat{a}_{p\sigma}^\dagger/\hat{a}_{p\sigma}$ associated to the p -th basis set element and the spin σ , while h_{pr} and $(pr|qs)$ are the one- and two-body electronic integrals, obtained from standard chemistry software [26]. Throughout this manuscript we use molecular orbitals as basis set elements. We map the degrees of freedom of Eq. (1) to qubits with a Jordan-Wigner (JW) transformation [27]. We then construct a quantum

circuit to be executed on quantum hardware, preparing a state $|\Psi\rangle$ on M qubits, which represents a molecular wavefunction on M molecular spin-orbitals. In the JW mapping, the single-qubit basis states $|0\rangle/|1\rangle$ represent empty/occupied spin-orbitals. These mapping and optimization steps are performed on classical nodes, see Fig. 1. We execute the circuit on a quantum computer and measure $|\Psi\rangle$ in the computational basis. Repeating this produces a set of measurement outcomes

$$\tilde{\mathcal{X}} = \left\{ \mathbf{x} \mid \mathbf{x} \sim \tilde{P}_\Psi(\mathbf{x}) \right\} \quad (2)$$

in the form of bitstrings $\mathbf{x} \in \{0, 1\}^M$ distributed according to some \tilde{P}_Ψ ; the bitstrings represent electronic configurations (Slater determinants).

On a pre-fault-tolerant quantum computer, the action of noise alters the distribution from its ideal form $P_\Psi = |\langle \mathbf{x} | \Psi \rangle|^2$ to some other \tilde{P}_Ψ , which generates the noisy set of configurations $\tilde{\mathcal{X}}$, accessible to us via quan-

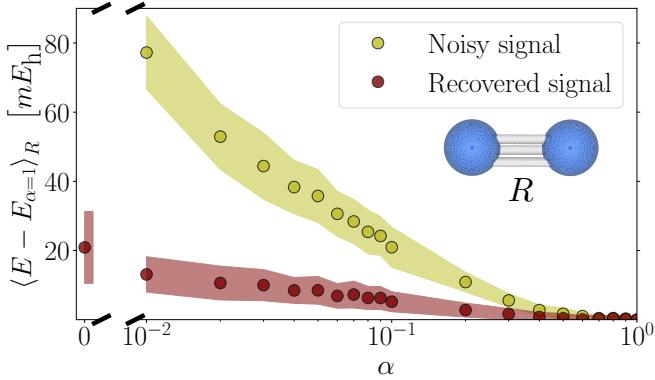


Figure 2. Self-consistent configuration recovery. Energy error (mean and standard deviation) in the dissociation of N₂ (6-31g), averaged over 24 bond lengths equally spaced between $R = 0.7 \text{ \AA}$ and $R = 3.0 \text{ \AA}$, as a function of quantum signal to noise ratio, parametrized by α . The point at $\alpha = 0$ corresponds to sampling from the uniform distribution (no signal). The energies are obtained via projection and diagonalization of $d = 10^6$ raw noisy samples against a subspace of the same size obtained by the configuration recovery routine.

tum measurement. Noise in the quantum system broadens the distribution P_Ψ over configurations that do not contribute to low-energy states, so-called *deadwood* [28]. As a result, only a fraction of $\tilde{\mathcal{X}}$ contains a meaningful quantum signal. To improve this scenario, we introduce a probabilistic *self-consistent configuration recovery* technique, which allows a probabilistic partial recovery of noiseless configuration samples from $\tilde{\mathcal{X}}$.

The configuration recovery scheme is inspired by the structure of chemistry problems. The Hamiltonian in Eq. (1) conserves the number of particles separately for each spin species. The recovery routine targets configurations \mathbf{x} that have the wrong particle number $N_{\mathbf{x}} \neq N$ due to the accumulation of errors in the execution of the quantum circuit. Repeated rounds of recovery can be carried out self-consistently. The first step of each recovery round is to iterate through the set $\tilde{\mathcal{X}}$ and find configurations \mathbf{x} with $N_{\mathbf{x}} \neq N$ particles. If $N_{\mathbf{x}} > N$ (or $N_{\mathbf{x}} < N$), $|N_{\mathbf{x}} - N|$ bits are sampled to be flipped from the set of occupied (or empty) spin-orbitals, according to a distribution proportional to a monotonically-increasing function [29] of $|x_{p\sigma} - n_{p\sigma}|$, the distance from the current value of the bit to the average occupancy of the spin-orbital $p\sigma$, obtained from the previous recovery round. This generates a new set of recovered configurations \mathcal{X}_R .

Following the next step of Fig. 1, we build K batches of d configurations $\mathcal{S}^{(1)}, \dots, \mathcal{S}^{(K)}$ using samples from the set \mathcal{X}_R , according to a distribution proportional to the empirical frequencies of each \mathbf{x} in \mathcal{X}_R . We project and diagonalize the Hamiltonian over each $\mathcal{S}^{(k)} : k = 1, \dots, K$, as proposed recently in the quantum selected configuration interaction method [30, 31], which draws inspiration from the classical selected configuration interaction

framework [32–38].

Each batch of sampled configurations spans a subspace $\mathcal{S}^{(k)}$ in which the many-body Hamiltonian is projected:

$$\hat{H}_{\mathcal{S}^{(k)}} = \hat{P}_{\mathcal{S}^{(k)}} \hat{H} \hat{P}_{\mathcal{S}^{(k)}}, \text{ with } \hat{P}_{\mathcal{S}^{(k)}} = \sum_{\mathbf{x} \in \mathcal{S}^{(k)}} |\mathbf{x}\rangle\langle\mathbf{x}|. \quad (3)$$

The ground states and energies of $\hat{H}_{\mathcal{S}^{(k)}}$, which we label $|\psi^{(k)}\rangle$ and $E^{(k)}$, are then computed using the iterative Davidson method on multiple classical nodes. The computational cost – both quantum and classical – to produce $|\psi^{(k)}\rangle$ is polynomial in d , the dimension of the subspace.

The ground states are then used to obtain new occupancies

$$n_{p\sigma} = \frac{1}{K} \sum_{1 \leq k \leq K} \langle \psi^{(k)} | \hat{n}_{p\sigma} | \psi^{(k)} \rangle, \quad (4)$$

for each spin-orbital tuple $(p\sigma)$, averaged on the K batches. These occupancies are sent back to the configuration recovery step, and this entire self-consistent iteration is repeated until convergence. The initial guess for \mathbf{n} used for the first round of recovery comes from the raw quantum samples in the correct particle sector. The configuration recovery routine can be seen effectively as a problem-informed clustering of a noisy signal around the occupations \mathbf{n} .

Let us now discuss the advantages and limitations of the method. First, on a noiseless signal \mathcal{X} , it is guaranteed to succeed efficiently if the ground state has a support \mathcal{X}_G of polynomial size, and if the wavefunction $|\Psi\rangle$ prepared on the quantum processor has a support similar to that of the ground state. In general, independently of the wavefunction support size, one can unambiguously benchmark by comparing with other methods that can only produce upper bounds to the ground state energy. In this regard, our method shares the same working assumptions as classical CI methods. Within the CI framework, quantum computing will allow us to test the performance of unitary circuits that are hard to simulate classically [39].

To establish some insight on the configuration recovery, we analytically compute a lower bound on the probability of recovering electronic configurations that have been corrupted by noise, as a function of their Hamming distance from an integer rounding of \mathbf{n} , using a global depolarizing noise model $\tilde{P}_\Psi(\mathbf{x}) = \alpha P_\Psi(\mathbf{x}) + (1 - \alpha) \frac{1}{2^M}$, where $\alpha \in [0, 1]$ is the parameter that controls the amount of noiseless quantum signal [29]. We find that the likelihood of recovery decays exponentially with the system size and also with the Hamming distance from the configuration that best approximates \mathbf{n} . Since classical CI methods also require exponential time to reach configurations with high Hamming distances from a reference state, it would be possible for the quantum-centric architecture presented here to perform a more efficient simulation at some finite system size and error rates.

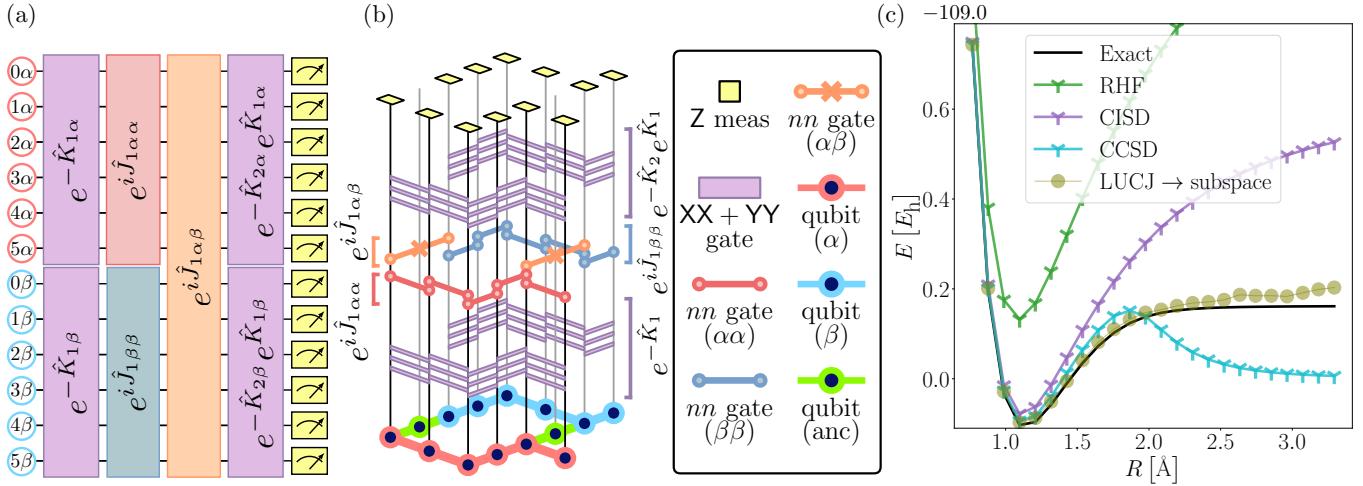


Figure 3. Quantum circuits for chemistry: local unitary cluster Jastrow. (a) Schematic representation of the truncated LUCJ circuit used to generate the set of samples. The circuit is comprised of orbital rotation unitaries $\exp(-K_{\mu\sigma})$, same-spin cluster operators $\exp(i(J_{\mu\alpha\alpha} + J_{\mu\beta\beta}))$, and opposite-spin cluster operators $\exp(iJ_{\mu\alpha\beta})$, with α/β denoting spin-up/down. (b) A single unit of the heavy-hex lattice with a compilation of the LUCJ circuit into single- and two-qubit gates, with the nn gate defined as $U_{nn}(\varphi) = \exp(i\varphi(1 - Z_0)(1 - Z_1)/4)$. (c) Comparison of the potential energy surface of N₂ (with a 6-31G basis) obtained from a noiseless subspace simulation of the LUCJ circuit in panel (a) with $L = 1$, and optimized parameters, against restricted coupled cluster singles and doubles (CCSD). Restricted Hartree-Fock (RHF), configuration interaction singles and doubles (CISD), and exact energies are shown for reference.

To further test noise robustness, we perform numerical simulations that confirm the improvements of applying the configuration recovery routine to the dissociation of N₂ (6-31G basis set). In this test we sample from the exact ground state $P_{\Psi_G}(\mathbf{x}) = |\langle \mathbf{x} | \psi_G \rangle|^2$ and we set a subspace dimension of $d = 10^6$.

Fig. 2 shows the error in the ground-state energy relative to the noiseless case ($\alpha = 1$), as a function of the amount of signal α , for the estimator both with and without configuration recovery. On the N₂ model, errors below $10mE_h$ can be obtained from $\sim 20\%$ signal using the raw noisy samples. However, by using configuration recovery we can tolerate a $\sim 2\%$ signal to reach the same error. This numerical experiment hints that the use of configuration recovery will be crucial for large-scale experiments.

Before presenting our experimental results, we discuss the circuits $|\Psi\rangle$ used to produce the candidate ground states. We employ a truncated version of the local unitary cluster Jastrow (LUCJ) ansatz [40], shown in Fig. 3 (a),

$$|\Psi\rangle = \prod_{\mu=1}^L e^{\hat{K}_\mu} e^{i\hat{J}_\mu} e^{-\hat{K}_\mu} |\mathbf{x}_{\text{RHF}}\rangle. \quad (5)$$

Here $\hat{K}_\mu = \sum_{pr,\sigma} K_{pr}^\mu \hat{a}_{p\sigma}^\dagger \hat{a}_{r\sigma}$ are generic one-body operators, $\hat{J}_\mu = \sum_{pr,\sigma\tau} J_{pr,r\tau}^\mu \hat{n}_{p\sigma} \hat{n}_{r\tau}$ are density-density operators restricted to spin-orbitals that are mapped onto adjacent qubits [40], and \mathbf{x}_{RHF} is the bitstring representing the restricted Hartree-Fock (RHF) state in the JW mapping. Through this local approximation, the

LUCJ ansatz allows for moderate circuit depths. Its accuracy derives from the connection with unitary coupled cluster theory and adiabatic state preparation [39–41]. The moderate depths of LUCJ are due to the use of exponentials of one-body operators, implementable in linear depth and a quadratic number of 2-qubit gates, and density-density operators, implementable in constant depth and a linear number of ZZ rotations (due to the locality approximation) [40]. The LUCJ circuit, compiled into one- and two-qubit gates, is shown in Fig. 3 (b).

In Fig. 3 (c) we show a numerical experiment comparing the potential energy curve of N₂ (6-31G basis) obtained by restricted CCSD to one obtained from the LUCJ ansatz, numerically optimized using the subspace energy as the objective function [29]. The dimension of the diagonalization subspace for different bond lengths ranging from $d = 209764$ to $d = 1340964$, with a median of $d = 563250$. Due to the presence of strong static correlation, CCSD fails in the description of the dissociation curve, while the optimized LUCJ ansatz produces a qualitatively correct dissociation curve. We simulated the LUCJ ansatz using ffsim [42].

II. RESULTS

In the following, we present the experimental results obtained using the methods discussed so far, on Heron quantum processors and the Fugaku supercomputer. The largest experiment is run on a subset of 77 qubits of a 133-qubit Heron quantum processor. The median fidelites for this subset are 99.77% for two-qubits gates, 99.97% for

single-qubit, and readout fidelity of 98.37%, with median coherence times $T_1 = 180\mu\text{s}$ and $T_2 = 150\mu\text{s}$. The quantum circuits are executed a maximum of three million times with a ~ 45 minute QPU runtime per experiment. We employ a reset-mitigation scheme by adding an additional measurement instruction before the circuit execution and post-selecting outcomes based on this first measurement returning the initial state $|000\dots000\rangle$. This post-selection results in $\sim 1/3$ retention rate of all the executions.

The classical projection and diagonalizations are obtained with the Davidson method implemented in the library PySCF [26] on a single node, or DICE [35, 38] for distributed computing on multiple nodes. Convergence to the most accurate solution can be obtained in two ways: increasing the accuracy per diagonalization with the subspace size d , and increasing the number of batches K , which will reduce statistical errors in the analysis. For our largest experiment on the [4Fe-4S] cluster, we use up to $d = 100\text{M}$, distributing a single projection and diagonalization to 64 nodes of Fugaku, and $K = 100$ batches, for a total of 6400 nodes. We analyze runtime performance as a function of d and K versus the number of nodes used [29]. At 64 nodes per diagonalization on the largest experiments, classical runtimes are about 1.5 hours. The largest HCI calculation that we performed on 16 nodes at $d = 2.3\text{M}$ took about 16 minutes.

We perform two classes of experiments: the breaking of the triple bond of N_2 (cc-pVDZ basis), in the top panel in Fig. 4 (a), and the ground states of [2Fe-2S] and [4Fe-4S] clusters (active spaces of the TZP-DKH basis), shown in the top panels in Fig. 4 (b) and (c). We study the ground-state properties of these molecular systems in the $S_z = 0$ and $S^2 = 0$ subspace, where S_z is the total \hat{z} component of the spin and $S^2 = S_x^2 + S_y^2 + S_z^2$.

Throughout our experiments we use the truncated LUCJ circuit $|\Psi\rangle = e^{-\hat{K}_2}e^{\hat{K}_1}e^{i\hat{J}_1}e^{-\hat{K}_1}|\mathbf{x}_{\text{RHF}}\rangle$, which is the result of considering the $L = 2$ circuit and removing the last orbital rotation and Jastrow operations. We parameterize the LUCJ circuits converting the CCSD wavefunction in Jastrow form and imposing a locality approximation to the resulting J^μ tensors, i.e., zeroing out the components $J_{p\sigma,r\tau}^\mu$ not corresponding to adjacent qubits [40]. For systems where the $J_{p\sigma,r\tau}^\mu$ parameters obtained from CCSD have small amplitude, the $e^{-\hat{K}_1}$ and $e^{\hat{K}_1}$ terms in the ansatz approximately cancel: in such situation, without $e^{\hat{K}_2}$, the resulting wavefunction is over-concentrated around the Hartree-Fock configuration [29].

The breaking of the N_2 bond is a well-known test of the accuracy of electronic structure methods in the presence of static electronic correlation [43, 44]. Restricted CCSD theory, a dominant paradigm for the accurate description of weakly correlated systems in quantum chemistry, fails in the description of N_2 dissociation due to static correlation effects: as correlations become stronger, RHF becomes unstable toward a symmetry-broken unrestricted

Hartree-Fock (UHF) state. CCSD built from RHF predicts an artificial barrier to binding and over-correlates at dissociation, whereas CCSD built from a UHF reference dissociates correctly at the cost of spin contamination, a manifestation of Löwdin's symmetry dilemma. We use a correlation-consistent cc-pVDZ basis set, to place emphasis on a theory's ability to treat both dynamic and static correlation in an accurate and balanced manner. We map the N_2 molecule onto a Heron processor as shown in the middle panel of Fig. 4 (a). We project and diagonalize a Hamiltonian using $d = 16 \cdot 10^6$ configurations. We consider $K = 10$ batches of configurations and each point in the dissociation curve has $|\tilde{\mathcal{X}}| = 100 \cdot 10^3$ measurement outcomes and 10 iterations of recovery. The experimental data is reported in the bottom panel of Fig. 4 (a), showing the potential energy surface of N_2 compared to classical approximate methods. The data from our experiments are consistent with other classical methods except for CCSD, which fails in the description of the dissociation as seen for the smaller basis set considered in Fig. 4(b). Amongst the classical SCI methods, *Heat-Bath Configuration Interaction* (HCI) [35] obtains the best results for N_2 and will be our reference classical method in all the other experiments. The difference between our method and HCI energies is everywhere within tens of mE_h . We further analyze the accuracy of our experiments as a function of d , and the effect of orbital optimizations in the accuracy of the predictions [29]. This first test demonstrates that we are capable of addressing multi-reference ground states and builds confidence for the next set of experiments, which will focus on assessing the ability of the quantum-classical architecture to do precision many-body physics.

Iron–sulfur (FeS) clusters are molecular ensembles of sulfide-linked 1- to 8-iron centers in variable oxidation states. They are important cofactors in biological processes ranging from nitrogen fixation to photosynthesis and respiration [46]. Their electronic structure, with multiple low-lying states of differing electronic and magnetic character, is responsible for their rich chemistry. At the same time, they pose considerable challenges for experimental studies and numerical tools. For our experiments we consider the synthetic $[\text{Fe}_2\text{S}_2(\text{SCH}_3)_4]^{-2}$ cluster [4], abbreviated [2Fe-2S] and used in numerical studies to mimic the oxidized dimers prominently found in ferredoxins [47].

The qubit mapping of the LUCJ circuit on the Heron processor for [2Fe-2S] is shown in the middle panel of Fig. 4 (b). We consider $K = 10$ batches of configurations and $|\tilde{\mathcal{X}}| = 2.4576 \cdot 10^6$ measurement outcomes. For the [2Fe-2S] cluster, we perform an energy-variance analysis of the low-energy spectrum of the molecule. The energy-variance analysis is a tool routinely used in classical computational electronic structure to capture the convergence of the approximate eigenstate energy for different levels of accuracy of a computational method [48]. Here we use energy-variance analysis to assess the convergence as a function of d , which is directly related to

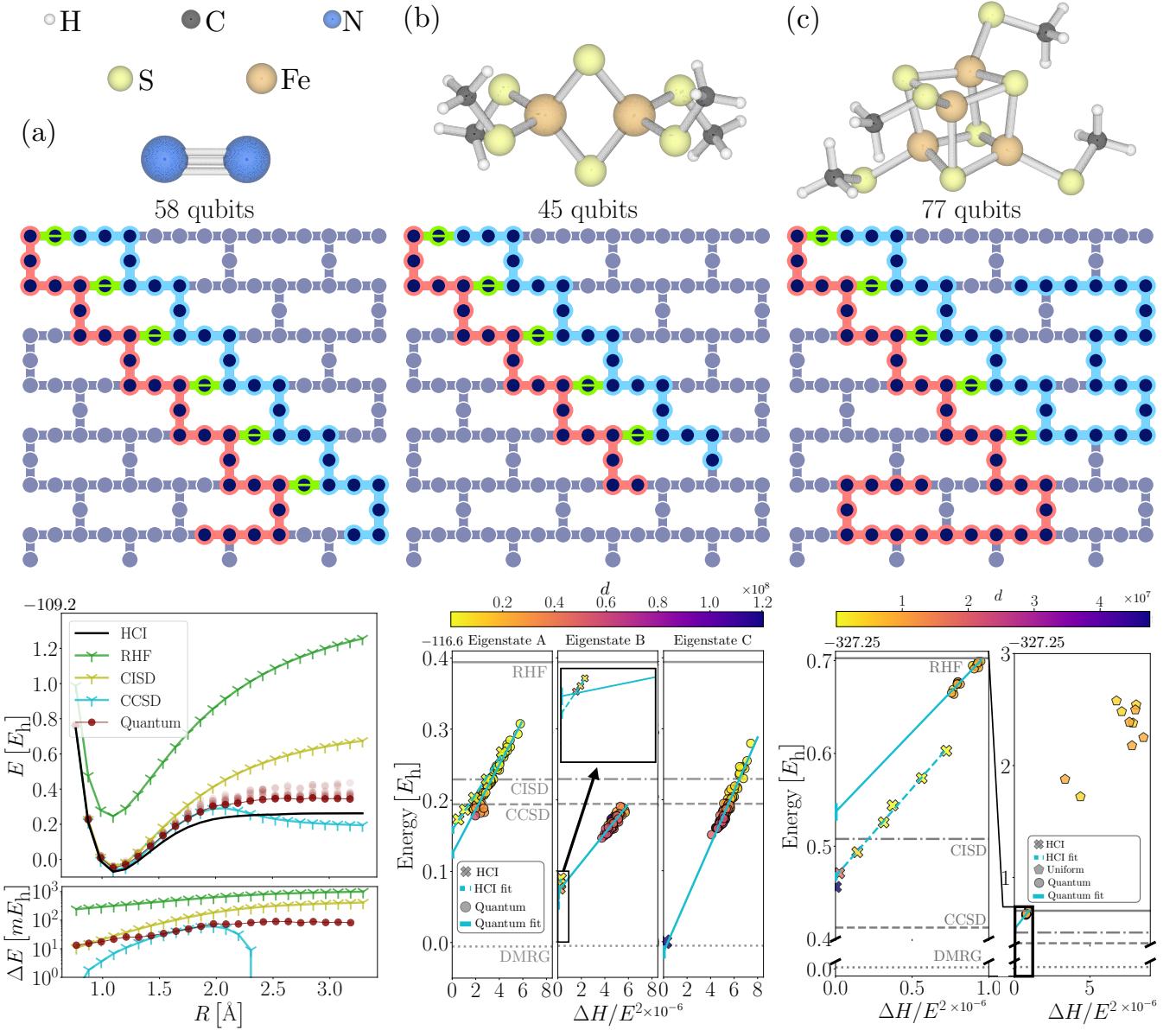


Figure 4. Experiments: Chemistry on large basis sets. (a) 58 qubits are used to model the N₂ dissociation (cc-pVDZ basis set). (b) 45 qubits are used for the [2Fe-2S] cluster (TZP-DKH basis set) and (c) 77 qubits for the [4Fe-4S] cluster (TZP-DKH basis set). The top panels show a 3-dimensional representation of the geometry of each molecule. The middle panels show the qubits selected on a Heron quantum processor layout, following the same color convention as panel (b) in Fig. 3. The **bottom panel in (a)** shows the potential energy surface comparison, as well as the energy difference ΔE between the Heat-Bath Configuration Interaction (HCI) energy and the energies obtained from different methods, including the quantum estimator. The brown scatter plot shows the value of $E^{(k)}$ for all batches of configurations and the connected dots show $\min_k(E^{(k)})$. The **bottom panel in (b)** shows the energy-variance analysis for three different eigenstates that both HCI and our method find upon increasing the value of d , as labeled by the color bar. For each approximate eigenstate $|\psi^{(k)}\rangle$, the horizontal axis $\Delta H = \langle\psi^{(k)}|\hat{H}^2|\psi^{(k)}\rangle - \langle\psi^{(k)}|\hat{H}|\psi^{(k)}\rangle^2$. The **bottom panel in (c)** shows a comparison of the energy-variance analysis applied to quantum measurement outcomes and bitstrings (with the correct particle number) sampled from the uniform distribution. The DMRG energy in panels (b) and (c) is from Ref. [45].

quantum and classical accuracy, runtimes, and costs. If one can statistically sample from a good approximation of an eigenstate, points at finite number of samples will be distributed linearly in the energy-variance plane [48].

This gives us a tool to detect eigenstates for both quantum and classical methods.

The bottom panel in Fig. 4(b) shows an energy-variance comparison of HCI and our estimator. The ac-

curacy of both estimators is quantified by the number of electronic configurations d that span the subspace for diagonalization. For $d > 250 \cdot 10^3$, we warm start with the \mathbf{n} obtained from the method at $d = 250 \cdot 10^3$, and only two recovery iterations are considered after. As we increase d , three eigenstates appear, which we label A, B, and C. At lower values of d , the data follows the expected linear relation for eigenstate A. Beyond a critical value of d , both the energy and variance change abruptly, discovering another set of points following a linear relation, which correspond to a second eigenstate B, of lower energy. The resulting mean occupations \mathbf{n} of eigenstates A and B are substantially different [29], confirming that we have found two distinct eigenstates. The third cluster of points in the energy-variance plane (Eigenstate C) was found by warm starting the procedure, for all values of d considered, with the \mathbf{n} obtained from the second cluster (Eigenstate B) at $d = 25 \cdot 10^6$. We then extrapolate the quantum data following the linear energy-variance relations, finding agreement between the quantum extrapolated data and the best HCl calculation performed and the Density Matrix Renormalization Group (DMRG) calculation from Ref. [45], showcasing the ability of our methods to perform precision many-body physics.

The circuits considered for the N₂ experiments and [2Fe-2S] reached sizes of approximately 1-1.5k two-qubit gates. We now test the quality of the signal in noisy circuits that test the limits of Heron devices, using up to 6400 nodes of Fugaku for the classical processing. We consider the synthetic [Fe₄S₄(SCH₃)₄]⁻² cluster [4], abbreviated [4Fe-4S], a representative of nature's cubanes, whose ground state deduction from experimental measurements was an early success of inorganic spectroscopy [49]. The LUCJ circuit employed for this molecular species contains approximately 3.5k two-qubit gates. The qubit mapping on the Heron processor for [4Fe-4S] is shown in the middle panel of Fig. 4 (c). As in the previous experiment, we consider $K = 10$ batches of configurations and $|\tilde{\mathcal{X}}| = 2.4576 \cdot 10^6$ measurement outcomes. For $d > 250 \cdot 10^3$, configuration recovery is warm started with the \mathbf{n} obtained from the method at $d = 250 \cdot 10^3$, and only two iterations are then performed.

This last set of experiments sheds light on the quality of the quantum signal that is passed to the configuration recovery at these large circuit sizes. The bottom right panel in Fig. 4 (c) shows a comparison of the energy-variance analysis from measurement outcomes obtained from the Heron device and configurations sampled from the uniform distribution. We see that, even if the quantum solutions produced are worse than other classical methods, the energy and variance obtained from quantum data are significantly lower than those obtained from uniformly distributed configurations (i.e. pure noise), on subspaces of the same size. This confirms that there is a valuable signal at circuit sizes of approximately 3.5k two-qubit gates.

III. DISCUSSION

Current quantum computers in isolation can be used as tools to explore calculations beyond brute-force [50, 51] on problems that match the interactions of the quantum devices. The assistance of classical distributed computations enables addressing problems that go beyond this constraint, with workflows such as the one used here for chemistry.

We now discuss the implications of our results in the search for quantum advantage. Any computation is ranked against three parameters: runtime, energy or cost, and accuracy. While the first two are often easily measured, ranking by accuracy is not always straightforward. However, the class of computational methods that strictly produce upper bounds to the ground-state energy defines an unconditional accuracy metric, which can be used to assess quantum advantage. Within this class of methods, a lower energy is ranked as higher quality, with no extra information required. This, for example, allowed us to benchmark our results against classical selected configuration interaction and uniform configuration sampling on an experiment on 77 qubits, without access to exact solutions. Additionally, the approximate solutions found here can be stored in classical memory, which permits a classical prover to certify them, and allows their manipulation by further classical processing.

The estimator can be applied to simulation tasks other than quantum chemistry, if the target ground state energy can be captured by a sparse wavefunction. This motivates looking into useful quantum circuits with supports in the computational basis scaling polynomially with problem size. Although in principle other estimators, such as the standard ones used in variational quantum eigensolvers, have bounded variance for any wavefunction, the dire scaling of number of measurements to estimate energies makes them impractical for the molecules targeted in this work [3].

Reconstructing the exact ground state probability distribution on a quantum computer is not required to get accurate energy approximations, as long as one is sampling from the ground state support. It will be compelling to use these methods on future quantum devices with lower error rates and correspondingly reduced classical processing requirements.

To counter wavefunction broadening on current quantum hardware, we have used a self-consistent configuration recovery method, exploiting a problem-inspired clustering that leverages the average occupation numbers of the molecular orbitals. We foresee generalizations of our configuration recovery technique to problems other than quantum chemistry that are not informed by the physics of the problem. Conversely, for specific applications one could use even more information about the problem.

We have used a LUCJ class of quantum circuits that can reproduce a low-rank decomposition and sparsification of the quantum unitary CCSD (qUCCSD), which allowed us to keep circuit depths manageable [40]. Bet-

ter quantum noise rates would allow to access deeper quantum circuits, improving their approximation of the qUCCSD theory and eventually surpassing qUCCSD itself if enough parameters are given. The inclusion of more density-density interactions, takes LUCJ further away from non-interacting fermionic computations, increasing the classical simulation overhead [52]. We have performed optimization-free experiments, exploiting the connection between LUCJ and classical coupled cluster theory, yet closing a quantum-classical optimization could further improve the quality of the solutions.

It is important to evaluate the methods presented here in the context of variational classical heuristics to highlight their different domains of applicability [53]. To improve on other classical CI methods, one would need to sample important configurations with a significant number of excitations over the HF state. Variational quantum Monte Carlo methods [54, 55], of paramount importance in many-body physics, are sensitive to the structures of the probability distributions they are modeling, not just to their supports. Methods based on tensor networks [56, 57] are successfully used to tackle strongly correlated problems in chemistry, granted the ability to converge their energies with bond dimensions. Convergence can be challenging in some cases [58, 59] because of its computational cost and its sensitivity to the nature and ordering of the basis-set orbitals. Characterizing the domain of applicability of different classical and quantum heuristics facilitates combining them in quantum-centric supercomputing environments, shifting the search

for quantum advantage towards practical problems.

AUTHOR CONTRIBUTIONS

Design of the workflow and experiments: J. R.-M., M. M., W. K., K. S., M. T., A. M. Implementation of the workflow: J. R.-M., M. M., A. J.-A., S. M., S. S., T. S., I. S., R.-Y. S., K. J. S. Tuning and calibration of the Heron processor: H.H. Execution of the quantum part of the workflow on Heron: M. M., P. J., M. T. Execution of the classical part of the workflow on Fugaku: T. S., R.-Y. S., S. Y. Numerical benchmarks: J. R.-M., M.M., T. S., K. J. S. Analytical derivations: J. R.-M., M.M., W. K., K. S., M. T., A. M. All authors contributed to the manuscript writing and data analysis.

ACKNOWLEDGMENTS

We thank the IBM Quantum Service and Data team for help with the workflow execution. The authors acknowledge feedback and insightful conversations with Alberto Baiardi, Giuseppe Carleo, Garnet Kin-Lic Chan, Antonio Córcoles, Oliver Dial, Jay Gambetta, Artur Izmaylov, Yukio Kawashima, David Kremer, Isaac Lauer, Peter Love, Guglielmo Mazzola, Takahito Nakajima, Paul Nation, Hanhee Paik, Emily Pritchett, Max Rossmanek, Paul Schweigert, James E.T. Smith, Miles Stoudenmire, Andrew Wack, Christa Zoufal.

-
- [1] R. P. Feynman, Simulating physics with computers, *Int. J. Theor. Phys.* **21**, 467 (1982).
 - [2] A. Aspuru-Guzik, A. D. Dutoi, P. J. Love, and M. Head-Gordon, Simulated quantum computation of molecular energies, *Science* **309**, 1704 (2005).
 - [3] D. Wecker, M. B. Hastings, and M. Troyer, Progress towards practical quantum variational algorithms, *Phys. Rev. A* **92**, 042303 (2015).
 - [4] S. Sharma, K. Sivalingam, F. Neese, and G. K.-L. Chan, Low-energy spectrum of iron-sulfur clusters directly from many-particle quantum mechanics, *Nat. Chem.* **6**, 927 (2014).
 - [5] K. D. Vogiatzis, D. Ma, J. Olsen, L. Gagliardi, and W. A. De Jong, Pushing configuration-interaction to the limit: towards massively parallel MCSCF calculations, *J. Chem. Phys.* **147**, 184111 (2017).
 - [6] H. Gao, S. Imamura, A. Kasagi, and E. Yoshida, Distributed implementation of full configuration interaction for one trillion determinants, *J. Chem. Theory Comput.* (2024).
 - [7] J. P. LeBlanc, A. E. Antipov, F. Becca, I. W. Bulik, G. K.-L. Chan, C.-M. Chung, Y. Deng, M. Ferrero, T. M. Henderson, C. A. Jiménez-Hoyos, *et al.*, Solutions of the two-dimensional Hubbard model: benchmarks and results from a wide range of numerical algorithms, *Phys. Rev. X* **5**, 041041 (2015).
 - [8] M. Motta, D. M. Ceperley, G. K.-L. Chan, J. A. Gomez, E. Gull, S. Guo, C. A. Jiménez-Hoyos, T. N. Lan, J. Li, F. Ma, *et al.*, Towards the solution of the many-electron problem in real materials: equation of state of the hydrogen chain with state-of-the-art many-body methods, *Phys. Rev. X* **7**, 031059 (2017).
 - [9] Y. Cao, J. Romero, J. P. Olson, M. Degroote, P. D. Johnson, M. Kieferová, I. D. Kivlichan, T. Menke, B. Peropadre, N. P. Sawaya, *et al.*, Quantum chemistry in the age of quantum computing, *Chem. Rev.* **119**, 10856 (2019).
 - [10] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, Quantum computational chemistry, *Rev. Mod. Phys.* **92**, 015003 (2020).
 - [11] B. Bauer, S. Bravyi, M. Motta, and G. Kin-Lic Chan, Quantum algorithms for quantum chemistry and quantum materials science, *Chem. Rev.* **120**, 12685 (2020).
 - [12] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, *Nature* **549**, 242 (2017).
 - [13] Google AI Quantum and Collaborators, Hartree-Fock on a superconducting qubit quantum computer, *Science* **369**, 1084 (2020).
 - [14] 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, *Na-*

- ture **603**, 416 (2022).
- [15] M. Motta, G. O. Jones, J. E. Rice, T. P. Gujarati, R. Sakuma, I. Liepuoniute, J. M. Garcia, and Y.-y. Ohnishi, Quantum chemistry simulation of ground- and excited-state properties of the sulfonium cation on a superconducting quantum processor, *Chem. Sci* **14**, 2915 (2023).
- [16] L. Zhao, J. Goings, K. Shin, W. Kyoung, J. I. Fuks, J.-K. Kevin Rhee, Y. M. Rhee, K. Wright, J. Nguyen, J. Kim, *et al.*, Orbital-optimized pair-correlated electron simulations on trapped-ion quantum computers, *npj Quantum Inf* **9**, 60 (2023).
- [17] Google AI Quantum and Collaborators, Purification-based quantum error mitigation of pair-correlated electron simulations, *Nat. Phys* **19**, 1787 (2023).
- [18] T. Weaving, A. Ralli, P. J. Love, S. Succi, and P. V. Coveney, Contextual subspace variational quantum eigensolver calculation of the dissociation curve of molecular nitrogen on a superconducting quantum computer (2023), [arXiv:2312.04392 \[quant-ph\]](#).
- [19] V. Verteletskyi, T.-C. Yen, and A. F. Izmaylov, Measurement optimization in the variational quantum eigensolver using a minimum clique cover, *J. Chem. Phys* **152**, 124114 (2020).
- [20] W. J. Huggins, J. R. McClean, N. C. Rubin, Z. Jiang, N. Wiebe, K. B. Whaley, and R. Babbush, Efficient and noise resilient measurements for quantum chemistry on near-term quantum computers, *npj Quantum Inf* **7**, 1 (2021).
- [21] A. Dutt, W. Kirby, R. Raymond, C. Hadfield, S. Sheldon, I. L. Chuang, and A. Mezzacapo, Practical benchmarking of randomized measurement methods for quantum chemistry Hamiltonians, [arXiv:2312.07497](#) (2023).
- [22] A. Anand, P. Schleich, S. Alperin-Lea, P. W. Jensen, S. Sim, M. Díaz-Tinoco, J. S. Kottmann, M. Degroote, A. F. Izmaylov, and A. Aspuru-Guzik, A quantum computing view on unitary coupled cluster theory, *Chem. Soc. Rev* **51**, 1659 (2022).
- [23] M. B. Hastings, D. Wecker, B. Bauer, and M. Troyer, Improving quantum algorithms for quantum chemistry, *Quant. Info. Comput* **15**, 1–21 (2015).
- [24] J. Lee, D. W. Berry, C. Gidney, W. J. Huggins, J. R. McClean, N. Wiebe, and R. Babbush, Even more efficient quantum computations of chemistry through tensor hypercontraction, *Phys. Rev. X Quantum* **2**, 030305 (2021).
- [25] Y. Alexeev, M. Amsler, P. Baity, M. A. Barroca, S. Bassini, T. Battelle, D. Camps, D. Casanova, F. T. Chong, C. Chung, *et al.*, Quantum-centric supercomputing for materials science: A perspective on challenges and future directions, arXiv preprint [arXiv:2312.09733](#) (2023).
- [26] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, *et al.*, PySCF: the python-based simulations of chemistry framework, *WIREs Comput. Mol. Sci* **8**, e1340 (2018).
- [27] P. Jordan and E. Wigner, Über das paulische äquivalenzverbot, *Zeit. Phys* **47**, 631 (1928).
- [28] J. Ivanic and K. Ruedenberg, Identification of deadwood in configuration spaces through general direct configuration interaction, *Theor. Chem. Acc* **106**, 339 (2001).
- [29] See supplementary information.
- [30] K. Kanno, M. Kohda, R. Imai, S. Koh, K. Mitarai, W. Mizukami, and Y. O. Nakagawa, Quantum-selected configuration interaction: classical diagonalization of Hamiltonians in subspaces selected by quantum computers (2023), [arXiv:2302.11320 \[quant-ph\]](#).
- [31] Y. O. Nakagawa, M. Kamoshita, W. Mizukami, S. Sudo, and Y. ya Ohnishi, ADAPT-QSCI: Adaptive construction of input state for quantum-selected configuration interaction (2023), [arXiv:2311.01105 \[quant-ph\]](#).
- [32] F. A. Evangelista, Adaptive multiconfigurational wave functions, *J. Chem. Phys* **140**, 124114 (2014).
- [33] J. B. Schriber and F. A. Evangelista, Communication: An adaptive configuration interaction approach for strongly correlated electrons with tunable accuracy, *J. Chem. Phys* **144**, 161106 (2016).
- [34] A. A. Holmes, H. J. Changlani, and C. Umrigar, Efficient heat-bath sampling in Fock space, *J. Chem. Theory Comput* **12**, 1561 (2016).
- [35] A. A. Holmes, N. M. Tubman, and C. Umrigar, Heat-bath configuration interaction: An efficient selected configuration interaction algorithm inspired by heat-bath sampling, *J. Chem. Theory Comput* **12**, 3674 (2016).
- [36] N. M. Tubman, J. Lee, T. Y. Takeshita, M. Head-Gordon, and K. B. Whaley, A deterministic alternative to the full configuration interaction quantum Monte Carlo method, *J. Chem. Phys* **145**, 044112 (2016).
- [37] J. B. Schriber and F. A. Evangelista, Adaptive configuration interaction for computing challenging electronic excited states with tunable accuracy, *J. Chem. Theory Comput* **13**, 5354 (2017).
- [38] S. Sharma, A. A. Holmes, G. Jeanmairet, A. Alavi, and C. J. Umrigar, Semistochastic heat-bath configuration interaction method: Selected configuration interaction with semistochastic perturbation theory, *J. Chem. Theory Comput* **13**, 1595 (2017).
- [39] F. A. Evangelista, G. K.-L. Chan, and G. E. Scuseria, Exact parameterization of fermionic wave functions via unitary coupled cluster theory, *J. Chem. Phys* **151**, 244112 (2019).
- [40] M. Motta, K. J. Sung, K. B. Whaley, M. Head-Gordon, and J. Shee, Bridging physical intuition and hardware efficiency for correlated electronic states: the local unitary cluster Jastrow ansatz for electronic structure, *Chem. Sci* **14**, 11213 (2023).
- [41] Y. Matsuzawa and Y. Kurashige, Jastrow-type decomposition in quantum chemistry for low-depth quantum circuits, *J. Chem. Theory Comput* **16**, 944 (2020).
- [42] The ffsim developers, *ffsim: Faster simulations of fermionic quantum circuits* (2024).
- [43] P.-D. Fan and P. Piecuch, The usefulness of exponential wave function expansions employing one-and two-body cluster operators in electronic structure theory: The extended and generalized coupled-cluster methods, *Adv. Quantum Chem* **51**, 1 (2006).
- [44] I. W. Bulik, T. M. Henderson, and G. E. Scuseria, Can single-reference coupled cluster theory describe static correlation?, *J. Chem. Theory Comput* **11**, 3171 (2015).
- [45] Z. Li and G. K.-L. Chan, Spin-projected matrix product states: versatile tool for strongly correlated systems, *J. Chem. Theory Comput* **13**, 2681 (2017).
- [46] H. Beinert, R. H. Holm, and E. Munck, Iron-sulfur clusters: nature's modular, multipurpose structures, *Science* **277**, 653 (1997).
- [47] P. Venkateswara Rao and R. Holm, Synthetic analogues

- of the active sites of iron-sulfur proteins, *Chem. Rev.* **104**, 527 (2004).
- [48] T. Kashima and M. Imada, Path-integral renormalization group method for numerical study on ground states of strongly correlated electronic systems, *J. Phys. Soc. Jpn.* **70**, 2287 (2001).
- [49] V. Papaefthymiou, M. M. Millar, and E. Muenck, Moessbauer and EPR studies of a synthetic analog for the iron-sulfur Fe_4S_4 core of oxidized and reduced high-potential iron proteins, *Inorg. Chem.* **25**, 3010 (1986).
- [50] Y. Kim, A. Eddins, S. Anand, K. X. Wei, E. Van Den Berg, S. Rosenblatt, H. Nayfeh, Y. Wu, M. Zalatel, K. Temme, *et al.*, Evidence for the utility of quantum computing before fault tolerance, *Nature* **618**, 500 (2023).
- [51] K. Shinjo, K. Seki, T. Shirakawa, R.-Y. Sun, and S. Yunoki, Unveiling clean two-dimensional discrete time quasicrystals on a digital quantum computer, arXiv preprint arXiv:2403.16718 (2024).
- [52] O. Reardon-Smith, M. Oszmaniec, and K. Kozekwa, Improved simulation of quantum circuits dominated by free fermionic operations, arXiv preprint arXiv:2307.12702 (2023).
- [53] S. Lee, J. Lee, H. Zhai, Y. Tong, A. M. Dalzell, A. Kumar, P. Helms, J. Gray, Z.-H. Cui, W. Liu, *et al.*, Evaluating the evidence for exponential quantum advantage in ground-state quantum chemistry, *Nat. Commun.* **14**, 1952 (2023).
- [54] F. Becca and S. Sorella, *Quantum Monte Carlo approaches for correlated systems* (Cambridge University Press, 2017).
- [55] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, *Science* **355**, 602 (2017).
- [56] S. R. White and R. L. Martin, Ab initio quantum chemistry using the density matrix renormalization group, *J. Chem. Phys.* **110**, 4127 (1999).
- [57] G. K.-L. Chan and S. Sharma, The density matrix renormalization group in quantum chemistry, *Annu. Rev. Phys. Chem.* **62**, 465 (2011).
- [58] C. Hubig, J. Haegeman, and U. Schollwöck, Error estimates for extrapolations with matrix-product states, *Phys. Rev. B* **97**, 045125 (2018).
- [59] Z. Li, J. Li, N. S. Dattani, C. Umrigar, and G. K. Chan, The electronic complexity of the ground-state of the FeMo cofactor of nitrogenase as relevant to quantum simulations, *J. Chem. Phys.* **150**, 024302 (2019).
- [60] W. J. Hehre, R. F. Stewart, and J. A. Pople, Self-consistent molecular-orbital methods. I. use of Gaussian expansions of Slater-type atomic orbitals, *J. Chem. Phys.* **51**, 2657 (1969).
- [61] W. J. Hehre, R. Ditchfield, and J. A. Pople, Self-consistent molecular orbital methods. XII. Further extensions of Gaussian-type basis sets for use in molecular orbital studies of organic molecules, *J. Chem. Phys.* **56**, 2257 (1972).
- [62] T. H. Dunning, Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen, *J. Chem. Phys.* **90**, 1007 (1989).
- [63] Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt, N. A. Bogdanov, G. H. Booth, J. Chen, Z.-H. Cui, *et al.*, Recent developments in the PySCF program package, *J. Chem. Phys.* **153**, 024109 (2020).
- [64] A. D. Becke, Density-functional exchange-energy approximation with correct asymptotic behavior, *Phys. Rev. A* **38**, 3098 (1988).
- [65] J. P. Perdew, Density-functional approximation for the correlation energy of the inhomogeneous electron gas, *Phys. Rev. B* **33**, 8822 (1986).
- [66] F. Jorge, A. Canal Neto, G. Camiletti, and S. Machado, Contracted Gaussian basis sets for Douglas-Kroll-Hess calculations: estimating scalar relativistic effects of some atomic and molecular properties, *J. Chem. Phys.* **130** (2009).
- [67] W. Liu, Ideas of relativistic quantum chemistry, *Mol. Phys.* **108**, 1679 (2010).
- [68] Z. Li, Y. Xiao, and W. Liu, On the spin separation of algebraic two-component relativistic Hamiltonians, *J. Chem. Phys.* **137**, 154114 (2012).
- [69] H. L. Tang, V. Shkolnikov, G. S. Barron, H. R. Grimsley, N. J. Mayhall, E. Barnes, and S. E. Economou, Qubit-ADAPT-VQE: an adaptive algorithm for constructing hardware-efficient ansaatzes on a quantum processor, *PRX Quantum* **2**, 020310 (2021).
- [70] M. J. D. Powell, A direct search optimization method that models the objective and constraint functions by linear interpolation, in *Advances in Optimization and Numerical Analysis*, edited by S. Gomez and J.-P. Hennart (Springer Netherlands, Dordrecht, 1994) pp. 51–67.
- [71] B. O. Roos, P. R. Taylor, and P. E. Sigbahn, A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach, *Chem. Phys.* **48**, 157 (1980).
- [72] M. Head-Gordon and J. A. Pople, Optimization of wave function and geometry in the finite basis Hartree-Fock method, *J. Phys. Chem.* **92**, 3063 (1988).
- [73] H.-J. Werner and P. J. Knowles, A second order multiconfiguration SCF procedure with optimum convergence, *J. Chem. Phys.* **82**, 5053 (1985).
- [74] J. Olsen, The CASSCF method: A perspective and commentary, *Int. J. Quantum Chem.* **111**, 3267 (2011).
- [75] D. Zgid and M. Nooijen, The density matrix renormalization group self-consistent field method: orbital optimization with the density matrix renormalization group method in the active space, *J. Chem. Phys.* **128**, 144116 (2008).
- [76] A.-K. Wu, M. T. Fishman, J. Pixley, and E. Stoudenmire, Disentangling interacting systems with fermionic gaussian circuits: application to the single impurity Anderson model, arXiv:2212.09798 (2022).
- [77] S. Wouters and D. Van Neck, The density matrix renormalization group for ab initio quantum chemistry, *Eur. Phys. J. D* **68**, 272 (2014).
- [78] J. R. Moreno, J. Cohn, D. Sels, and M. Motta, Enhancing the expressivity of variational neural, and hardware-efficient quantum states through orbital rotations (2023), arXiv:2302.11588 [quant-ph].
- [79] I. O. Sokolov, P. K. Barkoutsos, P. J. Ollitrault, D. Greenberg, J. Rice, M. Pistoia, and I. Tavernelli, Quantum orbital-optimized unitary coupled cluster methods in the strongly correlated regime: can quantum algorithms outperform their classical equivalents?, *J. Chem. Phys.* **152**, 124107 (2020).
- [80] 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. Research* **2**, 033421 (2020).
- [81] J. Bierman, Y. Li, and J. Lu, Improving the accuracy

- of variational quantum eigensolvers with fewer qubits using orbital optimization, *J. Chem. Theory Comput.* **19**, 790 (2023).
- [82] J. E. Smith, B. Mussard, A. A. Holmes, and S. Sharma, Cheap and near exact casscf with large active spaces, *Journal of chemical theory and computation* **13**, 5468 (2017).
- [83] D. Thouless, Stability conditions and nuclear rotations in the Hartree-Fock theory, *Nuc. Phys.* **21**, 225 (1960).
- [84] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, *JAX: composable transformations of Python+NumPy programs* (2018).
- [85] D. Chowdhury, A. Georges, O. Parcollet, and S. Sachdev, Sachdev-Ye-Kitaev models and beyond: window into non-Fermi liquids, *Rev. Mod. Phys.* **94**, 035004 (2022).
- [86] B. Friedman and G. Levine, Configuration-interaction approach to the two-dimensional Hubbard model near half-filling, *Phys. Rev. B* **55**, 9558 (1997).
- [87] E. Louis, F. Guinea, M. P. López Sancho, and J. A. Vergés, Configuration-interaction approach to hole pairing in the two-dimensional Hubbard model, *Phys. Rev. B* **59**, 14005 (1999).
- [88] L. R. Schwarz, G. H. Booth, and A. Alavi, Insights into the structure of many-electron wave functions of mott-insulating antiferromagnets: The three-band Hubbard model in full configuration interaction quantum Monte Carlo, *Phys. Rev. B* **91**, 045139 (2015).
- [89] W. Dobratz, H. Luo, and A. Alavi, Compact numerical solutions to the two-dimensional repulsive Hubbard model obtained via nonunitary similarity transformations, *Phys. Rev. B* **99**, 075119 (2019).
- [90] M. J. Rozenberg, G. Kotliar, and X. Y. Zhang, Mott-Hubbard transition in infinite dimensions. II, *Phys. Rev. B* **49**, 10181 (1994).
- [91] A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions, *Rev. Mod. Phys.* **68**, 13 (1996).
- [92] S.-S. B. Lee, J. von Delft, and A. Weichselbaum, Doublon-holon origin of the subpeaks at the Hubbard band edges, *Phys. Rev. Lett.* **119**, 236402 (2017).
- [93] C. Gauvin-Ndiaye, J. Tindall, J. R. Moreno, and A. Georges, Mott transition and volume law entanglement with neural quantum states (2023), arXiv:2311.05749 [cond-mat.str-el].
- [94] Y. Kwon, D. M. Ceperley, and R. M. Martin, Effects of backflow correlation in the three-dimensional electron gas: Quantum Monte Carlo study, *Phys. Rev. B* **58**, 6800 (1998).
- [95] M. Imada and T. Kashima, Path-integral renormalization group method for numerical study of strongly correlated electron systems, *J. Phys. Soc. Jpn.* **69**, 2723 (2000).
- [96] Y. Nomura, A. S. Darmawan, Y. Yamaji, and M. Imada, Restricted Boltzmann machine learning for solving strongly correlated quantum systems, *Phys. Rev. B* **96**, 205152 (2017).
- [97] S. Sorella, Generalized Lanczos algorithm for variational quantum Monte Carlo, *Phys. Rev. B* **64**, 024512 (2001).
- [98] T. Mizusaki and M. Imada, Extrapolation method for shell model calculations, *Phys. Rev. C* **65**, 064319 (2002).
- [99] M. Motta, E. Ye, J. R. McClean, Z. Li, A. J. Minnich, R. Babbush, and G. K.-L. Chan, Low rank representations for quantum simulation of electronic structure, *npj Quantum Inf.* **7**, 1 (2021).
- [100] M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani, Experimental realization of any discrete unitary operator, *Phys. Rev. Lett.* **73**, 58 (1994).
- [101] W. R. Clements, P. C. Humphreys, B. J. Metcalf, W. S. Kolthammer, and I. A. Walmsley, Optimal design for universal multiport interferometers, *Optica* **3**, 1460 (2016).
- [102] Z. Jiang, K. J. Sung, K. Kechedzhi, V. N. Smelyanskiy, and S. Boixo, Quantum algorithms to simulate many-body physics of correlated fermions, *Phys. Rev. Appl.* **9**, 044036 (2018).
- [103] P. D. Nation, H. Kang, N. Sundaresan, and J. M. Gambetta, Scalable mitigation of measurement errors on quantum computers, *Phys. Rev. X Quantum* **2**, 040326 (2021).
- [104] L. Viola and S. Lloyd, Dynamical suppression of decoherence in two-state quantum systems, *Phys. Rev. A* **58**, 2733 (1998).
- [105] A. Kofman and G. Kurizki, Universal dynamical control of quantum mechanical decay: modulation of the coupling to the continuum, *Phys. Rev. Lett.* **87**, 270405 (2001).
- [106] M. J. Biercuk, H. Uys, A. P. VanDevender, N. Shiga, W. M. Itano, and J. J. Bollinger, Optimized dynamical decoupling in a model quantum memory, *Nature* **458**, 996 (2009).
- [107] S. Niu and A. Todri-Sanial, Effects of dynamical decoupling and pulse-level optimizations on IBM quantum computers, *IEEE Trans. Quantum Eng.* **3**, 1 (2022).
- [108] Qiskit contributors, *Qiskit: An open-source framework for quantum computing* (2023).
- [109] Z. Li and G. K.-L. Chan, *Repository: Active space model for iron sulfur clusters* (2017).
- [110] When considering samples from the uniform distribution over electronic configurations on the correct particle sector, the configuration recovery method cannot be applied since it only recovers configurations with the wrong particle number. Therefore, this scenario is equivalent to running QSCI (without configuration recovery) on those samples.

SUPPLEMENTARY INFORMATION: CHEMISTRY BEYOND EXACT SOLUTIONS ON A QUANTUM-CENTRIC SUPERCOMPUTER

CONTENTS

I. Conventions and notation	1
A. Hamiltonian	1
B. Molecular species and active spaces	2
C. Electron configurations and qubit mapping	2
II. Methods	3
A. HPC quantum estimator	3
1. Eigenstate solver	3
2. Self-consistent configuration recovery	4
3. Optimization of the circuit parameters	9
4. Orbital optimization	10
B. Wavefunction concentration and accuracy	12
1. Sample complexity	12
2. Assessing wavefunction concentration	14
3. A numerical study of the accuracy: Hubbard model	14
C. Energy-variance extrapolation	15
III. Experiments	16
A. Quantum circuits	16
1. Initialization of LUCJ parameters	17
2. Mapping to heavy-hex devices	18
3. Compiled circuits	19
4. Effect of the Jastrow terms in the accuracy	19
B. Quantum and classical computational resources	21
1. Quantum experiment details	21
2. Classical resource details	22
IV. Experimental results	23
A. Visualization of the wave function amplitudes	23
B. Scaling and runtime	25
C. Effect of orbital optimizations in the dissociation of N ₂ (cc-pVDZ)	26
D. Orbital occupation numbers for the low-energy spectrum of [2Fe-2S]	27
E. Quantum signal in the experiments	30
1. N ₂ experiments	30
2. [2Fe-2S] experiments	32
References	34

I. CONVENTIONS AND NOTATION

A. Hamiltonian

Our starting point is the Born-Oppenheimer Hamiltonian, written in second quantization using a basis of N_{MO} orthonormal orbitals $\{\varphi_p\}_{p=1}^{N_{\text{MO}}}$,

$$\hat{H} = E_0 + \sum_{\sigma} h_{pr} \hat{a}_{p\sigma}^{\dagger} \hat{a}_{r\sigma} + \sum_{\sigma\tau} \frac{(pr|qs)}{2} \hat{a}_{p\sigma}^{\dagger} \hat{a}_{q\tau}^{\dagger} \hat{a}_{s\tau} \hat{a}_{r\sigma}. \quad (\text{S1})$$

Indices p, r, q, s label spatial orbitals in the basis set, $\sigma, \tau \in \{\uparrow, \downarrow\}$ are spin indices, and the fermionic operator $\hat{a}_{p\sigma}^{\dagger}/\hat{a}_{p\sigma}$ creates/destroys an electron in orbital p with spin σ . We define the number operator $\hat{n}_{p\sigma} = \hat{a}_{p\sigma}^{\dagger} \hat{a}_{p\sigma}$ which describes

the number of electrons with spin σ on orbital p . For non-relativistic all-electron calculations, the quantities

$$\begin{aligned} E_0 &= \sum_{a < b}^{N_{nuc}} \frac{Z_a Z_b}{\|\mathbf{R}_a - \mathbf{R}_b\|} \\ h_{pr} &= \int d\mathbf{r} \varphi_p^*(\mathbf{r}) \left[-\frac{1}{2} \frac{\partial^2}{\partial \mathbf{r}^2} - \sum_{a=1}^{N_{nuc}} \frac{Z_a}{\|\mathbf{r} - \mathbf{R}_a\|} \right] \varphi_r(\mathbf{r}), \\ (pr|qs) &= \int d\mathbf{r}_1 \int d\mathbf{r}_2 \frac{\varphi_p^*(\mathbf{r}_1) \varphi_r(\mathbf{r}_1) \varphi_q^*(\mathbf{r}_2) \varphi_s(\mathbf{r}_2)}{\|\mathbf{r}_1 - \mathbf{r}_2\|}, \end{aligned} \quad (\text{S2})$$

describe the internuclear electrostatic interaction energy and the one-electron and two-electron parts of the Hamiltonian respectively (atomic units are used throughout, i.e. lengths and energies are measured in Bohr and Hartree units $a_B = \hbar^2/(m_e e^2)$ and $E_h = e^2/a_B$ respectively where $-e$ and m_e are the electron charge and mass). The symbols N_{nuc} , \mathbf{R}_a , and Z_a denote the total number of nuclei and their positions and atomic numbers respectively.

For relativistic and/or active-space calculations, the indices p, r, q, s label active-space orbitals, and the quantities E_0 , h_{pr} , and $(pr|qs)$ are modified to account for relativistic effects and/or the potential generated by the inactive-electron density.

In this work, we use orbitals from a restricted closed-shell Hartree-Fock calculation (also called molecular orbitals and denoted MOs) as the basis functions φ_p . Furthermore, we denote N_σ the number of spin- σ electrons in the exact ground state. Furthermore $N_\uparrow + N_\downarrow = N$, the total number of electrons in the exact ground state.

B. Molecular species and active spaces

The molecules simulated in this work are listed in Table I. For N_2 , we studied all non-core electrons and orbitals on quantum hardware. We computed the potential energy curve using (i) restricted Hartree-Fock (RHF) at 6-31G, and cc-pVDZ level [60–62] using PySCF [26, 63] and enforcing $D_{\infty h}$ symmetry and, after projecting the non-relativistic Born-Oppenheimer Hamiltonian in the space spanned by all non-core RHF orbitals with standard functionalities, with (ii) restricted and symmetry-preserving Moller-Plesset second-order perturbation theory (MP2), coupled cluster with singles and doubles (CCSD), complete active-space configuration interaction (CASCI), for 6-31G and selected configuration interaction, in its heat-bath flavor (HCI), for cc-pVDZ.

For [2Fe-2S] and [4Fe-4S], we employed active spaces [4, 45], spanned by Fe[3d] and S[3p] orbitals, derived from a localized Density Functional Theory calculation with BP86 functional [64, 65], TZP-DKH basis [66], and sf-X2C (spin-free exact two-component) Hamiltonian [67, 68] to include scalar relativistic effects. We computed approximations to the ground-state energy with restricted RHF, MP2, CCSD, and HCI. The RHF, MP2, CCSD, CISD, CASCI and HCI calculations were carried out with the PySCF library [26].

C. Electron configurations and qubit mapping

In this work, we represent many-electron states using qubit states with the Jordan-Wigner transformation, that maps an electronic configuration, i.e. a Slater determinant of the form

$$|\mathbf{x}\rangle = \prod_{p\sigma} (\hat{a}_{p\sigma}^\dagger)^{x_{p\sigma}} |\emptyset\rangle, \quad (\text{S3})$$

molecule	basis	(N, N_{MO})	D
N_2	6-31G	(10e,16o)	$1.91 \cdot 10^7$
N_2	cc-pVDZ	(10e,26o)	$4.32 \cdot 10^9$
[2Fe-2S]	TZP-DKH	(30e,20o)	$2.40 \cdot 10^8$
[4Fe-4S]	TZP-DKH	(54e,36o)	$8.86 \cdot 10^{15}$

Table I. Molecules studied in this work. For each molecule, we list the number of electrons and orbitals (N and N_{MO} respectively) studied on quantum hardware, along with the underlying basis set and the dimension $D = \binom{N_{\text{MO}}}{N/2}^2$ of the Hilbert space of $N_\uparrow = N_\downarrow = N/2$ electrons in N_{MO} spatial orbitals (not considering molecular point-group symmetries).

where $|\emptyset\rangle$ is the vacuum state (i.e., the state with zero electrons) and $x_{p\sigma} \in \{0, 1\}$, onto an element of the computational basis,

$$|\mathbf{x}\rangle = \bigotimes_{p\sigma} |x_{p\sigma}\rangle, \quad (\text{S4})$$

labeled by a bitstring $\mathbf{x} = (x_{N_{\text{MO}}-1\downarrow} \dots x_{0\downarrow} x_{N_{\text{MO}}-1\uparrow} \dots x_{0\uparrow})$. The first half of the bitstring is denoted by $\mathbf{x}_{\downarrow} = (x_{N_{\text{MO}}-1\downarrow} \dots x_{0\downarrow})$ and the second half of the bitstring is denoted by $\mathbf{x}_{\uparrow} = (x_{N_{\text{MO}}-1\uparrow} \dots x_{0\uparrow})$. The JW mapping employs $M = 2N_{\text{MO}}$ qubits and allows computing the number of spin- σ electrons for a given configuration \mathbf{x} as $N_{\mathbf{x}\sigma} = \sum_p x_{p\sigma}$. The total number of electrons, $N_{\mathbf{x}} = \sum_{\sigma} N_{\mathbf{x}\sigma}$, is the Hamming weight of \mathbf{x} . At the single-configuration level, any \mathbf{x} in the *right particle sector* must satisfy: $N_{\mathbf{x}\sigma} = N_{\sigma}$. An important example is the restricted Hatree-Fock (RHF) bitstring

$$|\mathbf{x}_{\text{RHF}}\rangle = |\underbrace{0\dots 0}_{N_{\text{MO}}-N_{\downarrow}} \underbrace{1\dots 1}_{N_{\downarrow}} \underbrace{0\dots 0}_{N_{\text{MO}}-N_{\uparrow}} \underbrace{1\dots 1}_{N_{\uparrow}}\rangle, \quad (\text{S5})$$

which by construction has $N_{\mathbf{x}_{\text{RHF}}\sigma} = N_{\sigma}$.

The Fock space is the vector space containing all possible electronic configurations for N_{MO} orbitals with all possible filling factors for each spin sector. The terms *determinant* and *electronic configuration* are used interchangeably in the manuscript.

II. METHODS

A. HPC quantum estimator

This section provides a detailed description of the HPC quantum estimator. We describe the subspace projection and diagonalization, the self-consistent configuration recovery scheme, a lower bound to the likelihood of recovering different families of electronic configurations, and different techniques to optimize circuit parameters to minimize the estimator energy.

We first introduce some notations. We use \mathcal{X} and $\tilde{\mathcal{X}}$ to denote a set of configurations sampled from probability distributions $P_{\Psi}(\mathbf{x}) = |\langle \mathbf{x} | \Psi \rangle|^2$ and $\tilde{P}_{\Psi}(\mathbf{x}) = \langle \mathbf{x} | \tilde{\rho} | \mathbf{x} \rangle$, where $\tilde{\rho}$ is a density operator corresponding to a noisy version of $|\Psi\rangle\langle\Psi|$. We denote $\mathcal{X}_N \subset \tilde{\mathcal{X}}$ the subset of configurations with the right particle number. The set of configurations recovered by the configuration recovery procedure is denoted by $\mathcal{X}_{\rightarrow N}$ and $\mathcal{X}_R = \mathcal{X}_N \cup \mathcal{X}_{\rightarrow N}$ the set of configurations output by self-consistent configuration recovery, and $\mathcal{S}^{(k)} \subset \mathcal{X}_R$ with $k = 1 \dots K$ a set of approximately d configurations sampled from \mathcal{X}_R . The wavefunction $|\psi^{(k)}\rangle$ is obtained by projection and diagonalization of \hat{H} in the subspace spanned by the configurations in $\mathcal{S}^{(k)}$.

1. Eigenstate solver

Given a set of d electronic configurations, the many-electron Hamiltonian is projected and diagonalized in the subspace spanned by the single-particle states defined by the electronic configurations, as proposed recently in [30]. We begin by considering a set of configurations $\mathcal{X} = \{\mathbf{x}_{(i)}\}$, all with the right-particle number for each spin sector. Sub-indices between parenthesis label configurations in a set and not configuration components. We use the generic label \mathcal{X} for a set of configurations with the right particle number. In practice, the configuration recovery procedure will execute the eigenstate solver on the sets \mathcal{X}_N and \mathcal{X}_R .

a. Conservation of spin.— In this study, we ensured that our calculations produced wavefunctions $|\psi^{(k)}\rangle$ that are eigenfunctions of Hamiltonian symmetries and in particular singlet states, i.e., eigenfunctions of total spin with eigenvalue 0, $\hat{S}^2|\psi^{(k)}\rangle = 0$. Conservation of symmetries is notoriously important because molecular eigenstates are joint eigenfunctions of \hat{H} and its symmetries, including \hat{S}_z , \hat{N} , molecular point-group symmetries (if any), and \hat{S}^2 . Conservation of \hat{S}_z and \hat{N} (and molecular point-group symmetries that are isomorphic to $\mathbb{Z}_2^{\times n}$) can be achieved in a relatively easy way, since the eigenfunctions of these operators are Slater determinants (SD) $|\mathbf{x}\rangle$, e.g.

$$\hat{N}|\mathbf{x}\rangle = (N_{\alpha\mathbf{x}} + N_{\beta\mathbf{x}})|\mathbf{x}\rangle, \quad \hat{S}_z|\mathbf{x}\rangle = (N_{\alpha\mathbf{x}} - N_{\beta\mathbf{x}})|\mathbf{x}\rangle. \quad (\text{S6})$$

Therefore, in this work, we ensure that the configurations used to span the SCI subspace all have the desired eigenvalue of \hat{S}_z and \hat{N} as part of the configuration recovery procedure in the presence of noise. Conservation of \hat{S}^2 is more difficult

to achieve in CI methods since the eigenfunctions of \hat{S}^2 are not SDs. The proper way of ensuring conservation of \hat{S}^2 is expanding $|\psi^{(k)}\rangle$ on a set of configuration state functions (CSFs), i.e. spin-symmetry-adapted linear combinations of SDs. The use of CSFs was common in early CI codes since, in addition to the obvious benefits of reducing the memory footprint of the CI vector, automatic conservation of total spin enhanced the stability of the CI iterations. Modern CI codes tend to employ SDs as opposed to CSFs, because the formation of the σ vector (the most memory- and rate-limiting step of CI algorithms) is considerably more efficient and easily parallelized in SD-based algorithms, and in part because system memory and disk are more plentiful than in previous machines. However, in SD-based CI codes, the conservation of total spin is no longer guaranteed. In the case of our HPC quantum estimator, sampling from a quantum computer may return sets of SDs that do not allow constructing eigenfunctions of total spin. For example, in a (2e,2o) system, one may sample the configuration $|1001\rangle$ (having a single spin-down excitation over the RHF state $|0101\rangle$) which is a linear combination of the open-shell singlet and triplet states, respectively $(|1001\rangle \pm |0110\rangle)/\sqrt{2}$. If the configuration $|0110\rangle$ is not sampled, one can construct neither eigenfunction of total spin, leading to spin contamination or redundancy (i.e. the configuration $|1001\rangle$ is involved in a CI calculation, but has coefficient 0 in the CI vector). In this work, to facilitate conservation of total spin, we relied on the following procedure: instead of collecting directly d i.i.d. samples from \mathcal{X} to make the batch $\mathcal{S}^{(k)}$, we collect $\sqrt{d}/2$ samples and identify all unique configurations \mathbf{x}^u (u for unique) of length $M/2$ obtained from $\mathbf{x}_{\uparrow(i)}$ and $\mathbf{x}_{\downarrow(i)}$ for $1 \leq i \leq \sqrt{d}/2$, forming the set $\mathcal{U}^{(k)} = \{\mathbf{x}^u\}$. The size of the set is upper bounded by $|\mathcal{U}^{(k)}| \leq \sqrt{d}$. From $\mathcal{U}^{(k)}$ we obtain the batch set $\mathcal{S}^{(k)}$ as:

$$\mathcal{S}^{(k)} = \left\{ \mathbf{x} \mid \mathbf{x} = \mathbf{x}_{(i)}^u \oplus \mathbf{x}_{(j)}^u \text{ for all } \mathbf{x}_{(i)}^u, \mathbf{x}_{(j)}^u \in \mathcal{U}^{(k)} \right\}. \quad (\text{S7})$$

The size of the set above is upper bounded by $|\mathcal{S}^{(k)}| \leq d$. This procedure facilitates total spin conservation (for example, from the configuration $|1001\rangle$ one can build the set $\{|1001\rangle, |1010\rangle, |0101\rangle, |0110\rangle\}$, which contains two closed-shell configurations and allows constructing an open-shell singlet state) but does not enforce it. Therefore, in combination with the sampling strategy mentioned above, we achieve the conservation of total spin by a soft constraint in the eigenstate solver, i.e. by adding a penalty term to mitigate spin contamination,

$$\left(H + \lambda [S^2 - s(s+1)]^2 \right) |\psi\rangle = E|\psi\rangle, \quad (\text{S8})$$

where λ can be understood as a Lagrange multiplier that penalizes contributions from $S^2 \neq s(s+1)$. In this work, we employed a soft constraint with $\lambda = 0.2$.

b. Projection and diagonalization.— For each subsampled set $\mathcal{S}^{(k)}$, the Hamiltonian is projected into the corresponding subspace spanned by the configurations in $\mathcal{S}^{(k)}$:

$$\hat{H}_{\mathcal{S}^{(k)}} = \hat{P}_{\mathcal{S}^{(k)}} \hat{H} \hat{P}_{\mathcal{S}^{(k)}}, \text{ with } \hat{P}_{\mathcal{S}^{(k)}} = \sum_{\mathbf{x} \in \mathcal{S}^{(k)}} |\mathbf{x}\rangle\langle\mathbf{x}|. \quad (\text{S9})$$

We then diagonalize this projected Hamiltonian (solving $\hat{H}_{\mathcal{S}^{(k)}}|\psi^{(k)}\rangle = E^{(k)}|\psi^{(k)}\rangle$), and its ground state forms an approximation to the ground state of \hat{H} . The approximate ground state $|\psi^{(k)}\rangle$ is defined by its amplitudes in the subspace:

$$|\psi^{(k)}\rangle = \sum_{\mathbf{x} \in \mathcal{S}^{(k)}} c_{\mathbf{x}}^{(k)} |\mathbf{x}\rangle. \quad (\text{S10})$$

The k^{th} estimate of the ground-state energy is given by:

$$E^{(k)} = \langle \psi^{(k)} | \hat{H}_{\mathcal{S}^{(k)}} | \psi^{(k)} \rangle. \quad (\text{S11})$$

2. Self-consistent configuration recovery

After a quantum state $\hat{\rho}$, corresponding to a noiseless state $|\Psi\rangle$, is prepared in our pre-fault-tolerant quantum processor, we measure it in the computational basis, obtaining the set of measurements

$$\tilde{\mathcal{X}} = \left\{ \mathbf{x} \mid \mathbf{x} \sim \tilde{P}_{\Psi} \right\}. \quad (\text{S12})$$

The class of noiseless states $|\Psi\rangle$ considered in this work are eigenstates to the total particle number operator and the total number operator for each spin species:

$$\sum_{p=1}^{N_{\text{MO}}} \sum_{\sigma} \hat{n}_{p\sigma} |\Psi\rangle = N |\Psi\rangle, \quad \sum_{p=1}^{N_{\text{MO}}} \hat{n}_{p\uparrow} |\Psi\rangle = N_{\uparrow} |\Psi\rangle, \quad \sum_{p=1}^{N_{\text{MO}}} \hat{n}_{p\downarrow} |\Psi\rangle = N_{\downarrow} |\Psi\rangle. \quad (\text{S13})$$

From the measurements on the quantum processors, we observe that there are a number of configurations in $\tilde{\mathcal{X}}$ whose $N_{\mathbf{x}\uparrow}$ and $N_{\mathbf{x}\downarrow}$ do not match the N_{\uparrow} and N_{\downarrow} of the ground state. In Table II we report typical values of the fraction of sampled configurations with the wrong particle number. Since the circuits we use to produce $|\Psi\rangle$ are particle-number preserving, we are certain that configurations with wrong particle numbers have been corrupted by noise. It is this subset of configurations that the configuration recovery scheme is applied to. The configurations in $\tilde{\mathcal{X}}$ whose $N_{\mathbf{x}\uparrow} = N_{\uparrow}$ and $N_{\mathbf{x}\downarrow} = N_{\downarrow}$ are not subject to the configuration recovery subroutine.

We then probabilistically flip bits and restore the correct particle number using information obtained from observables of the system. We use the spin-orbital occupancy averaged over all collected batches of subsamples \mathbf{n} , whose components are given by:

$$n_{p\sigma} = \frac{1}{K} \sum_{1 \leq k \leq K} \langle \psi^{(k)} | \hat{n}_{p\sigma} | \psi^{(k)} \rangle, \quad (\text{S14})$$

where $|\psi^{(k)}\rangle$ is defined in Eq. (S10). Consider a configuration \mathbf{x} with $N_{\mathbf{x}\uparrow}$ spin-up electrons, where $N_{\mathbf{x}\uparrow} > N_{\uparrow}$. From the set of occupied spin-orbitals in the first half of \mathbf{x} , $|N_{\mathbf{x}\uparrow} - N_{\uparrow}|$ bits are sampled to be flipped. If instead $N_{\mathbf{x}\uparrow} < N_{\uparrow}$, the bits to be flipped are instead sampled from the unoccupied spin-orbitals. The same procedure applies to the spin-down orbitals.

The probability of flipping bit $x_{p\sigma}$ depends on the distance between the value of the bit and the reference orbital occupancy $n_{p\sigma}$. The simplest approach would be to define the distribution proportional to $|x_{p\sigma} - n_{p\sigma}|$. However, this introduces an undesirable effect: if for some spin-orbital $p\sigma$, $n_{p\sigma} \approx 0.5$, then $|x_{p\sigma} - n_{p\sigma}| \approx 0.5$ as well, i.e., we assign roughly 50% probability weight to flip the bit, regardless of its initial value. On the other hand, the initial value $x_{p\sigma}$ will in general retain some correlation with the other values in the bitstring, even in the presence of noise. Hence a better approach is to de-weight the probability of flipping when $|x_{p\sigma} - n_{p\sigma}|$ is small by using a modified rectified linear unit (ReLU) function $w(|x_{p\sigma} - n_{p\sigma}|)$, defined as

$$w(y) = \begin{cases} \delta \frac{y}{h} & \text{if } y \leq h, \\ \delta + (1 - \delta) \frac{y-h}{1-h} & \text{if } y > h. \end{cases} \quad (\text{S15})$$

The parameter $h \in (0, 1)$ defines the location of the ‘‘corner’’ of the ReLU function, while the parameter $\delta \in [0, c]$ defines the value of the ReLU function at the corner. w becomes a true ReLU function when $\delta = 0$, and for values of $\delta > 0$ the ReLU is modified so that it is not identically zero except at $y = 0$: a plot is shown in Fig. S1. In the specific cases in this work, we chose the values $\delta = 0.01$ and $h = N/M$ (the filling factor) in all experiments.

We do not assume that we know \mathbf{n} *a priori*, and instead we compute it and improve it self-consistently, following the procedure:

1. Setup phase:
 - (a) Find the subset of configurations of $\tilde{\mathcal{X}}$ that live in the correct particle sector for both spin species, which we denote by \mathcal{X}_N : $\mathcal{X}_N \subset \tilde{\mathcal{X}}$.
 - (b) Obtain batches of samples $(\mathcal{S}^{(1)}, \dots, \mathcal{S}^{(K)})$ from \mathcal{X}_N as described in Sec. II A 1.
 - (c) Run the eigenstate solver on the batches and obtain approximate eigenstates $|\psi^{(1)}\rangle, \dots, |\psi^{(K)}\rangle$ (Eqs.S9 and S10).
 - (d) From the approximate eigenstates construct the first guess for \mathbf{n} , according to Eq. (S14).
2. Self-consistent iterations (repeat until stopping criterion is met):
 - (a) \mathbf{n} is used to correct the configurations with the wrong particle number in $\tilde{\mathcal{X}}$ (we give this subset the label $\mathcal{X}_{\rightarrow N}$). The resulting set of recovered configurations is labelled $\mathcal{X}_{\rightarrow N}$.
 - (b) From $\mathcal{X}_R = \mathcal{X}_N \cup \mathcal{X}_{\rightarrow N}$, batches of samples $(\mathcal{S}^{(1)}, \dots, \mathcal{S}^{(K)})$ are obtained as described in Sec. II A 1.
 - (c) Run the eigenstate solver on the batches and obtain approximate eigenstates $|\psi^{(1)}\rangle, \dots, |\psi^{(K)}\rangle$ (Eqs.S9 and S10).

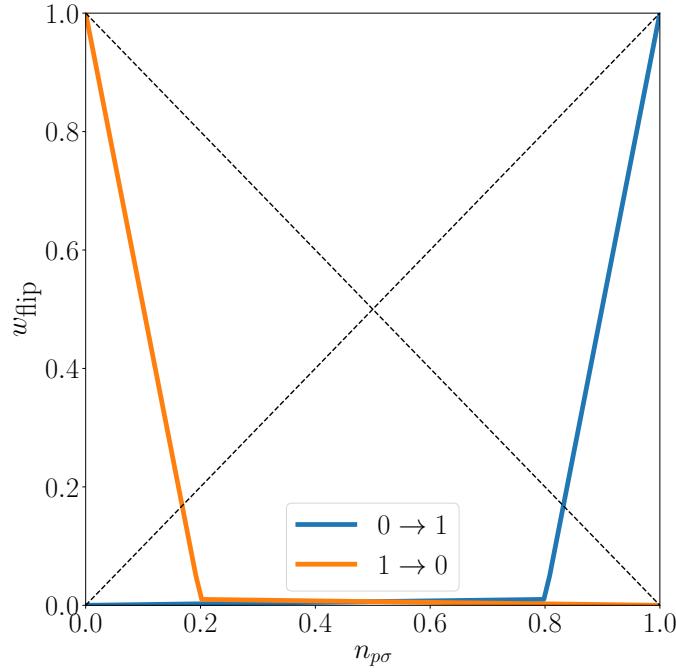


Figure S1. Probability weight associated with flipping a bit with value 0 or 1 as a function of $n_{p\sigma}$, following the definition in Eq. (S15). We set $\delta = 0.01$ and $h = 0.8$. The dashed back lines show the same probability weight associated to the naive approach of flipping according to $|x_{p\sigma} - n_{p\sigma}|$.

- (d) From the approximate eigenstates construct refined guess for \mathbf{n} , according to Eq. (S14).
- (e) If the stopping criterion is not met, go back to step 2.(a).

We direct the reader to Fig. 2 in the main text for a numerical emulation of the effect of the configuration recovery on the accuracy of the HPC quantum estimator in the presence of noise.

a. *An analytical lower bound to the probability of recovering families of configurations.*— Without loss of generality, we reorder the orbital labels such that $n_1 \leq n_2 \leq \dots \leq n_M$. Note that in the previous expression we have combined the spin-orbital multi index ($p\sigma$) into a single index p running from $1 \leq p \leq M$. In the following we assume that there exists a good reference configuration \mathbf{x}_r , obtained by assigning value 1 to the N bits of \mathbf{n} which have the largest magnitude, and zero otherwise. The average deviation of \mathbf{x}_r from \mathbf{n} is quantified by ε : $|\mathbf{x}_r - \mathbf{n}|_1 = \varepsilon M$. ε represents the average distance between each bit in \mathbf{x}_r and the corresponding entry in \mathbf{n} . The bits p for which $(x_r)_p = 1$ are referred to as the 1-sector of \mathbf{x}_r , while the bits for which $(x_r)_p = 0$ are referred to as the 0-sector of \mathbf{x}_r . For this analysis we take the probability of flipping a bit in \mathbf{x} proportional to $|x_p - n_p|$, which on average takes the value:

- ε ; if p is in the 1-sector of \mathbf{x}_r and $x_p = 1$.
- $1 - \varepsilon$; if p is in the 1-sector of \mathbf{x}_r and $x_p = 0$.
- ε ; if p is in the 0-sector of \mathbf{x}_r and $x_p = 0$.
- $1 - \varepsilon$; if p is in the 0-sector of \mathbf{x}_r and $x_p = 1$.

We use these average values for our analytical derivations.

We analyze the probability of recovering configurations using the configuration recovery scheme described above, as a function of their Hamming distance from \mathbf{x}_r . For this analysis, we consider a simple global depolarizing noise model, whose effect on the probability of sampled configurations is described by:

$$\tilde{P}_\Psi(\mathbf{x}) = \alpha P_\Psi(\mathbf{x}) + (1 - \alpha) \frac{1}{2^M}, \quad (\text{S16})$$

where $\alpha \in [0, 1]$ is a parameter that quantifies the amount of quantum signal. While we use for simplicity a global depolarizing noise model, we foresee similar derivations could be done for more realistic noise models.

The configurations that can be sampled from such a probability distribution fall into three categories: configurations with the right particle number that are in the support of the ideal distribution $P_\Psi(\mathbf{x})$, configurations with the right particle number that are not the support of $P_\Psi(\mathbf{x})$, and configurations with the wrong particle number. Configuration recovery does not apply to the first two categories. The third category, configurations with incorrect particle numbers, are guaranteed to have come from the noisy part of $\tilde{P}_\Psi(\mathbf{x})$. We now derive a lower bound on the probability of drawing a configuration with wrong number of particles from $\tilde{P}_\Psi(\mathbf{x})$ and then converting it (via the configuration recovery scheme) to a particular configuration $\mathbf{x}_{\text{target}}$ with the right particle number, whose Hamming distance to \mathbf{x}_r is $2b$.

Since $\mathbf{x}_{\text{target}}$ is Hamming distance $2b$ from \mathbf{x}_r , it contains b 1s in the 0-sector and b 0s in the 1-sector. There are two cases for initial configurations with wrong number of particles: those that have too many 1s or too many 0s.

1. Too many 1s. In this case only bits with the value 1 will be flipped by the configuration recovery. Consider the set of initial noisy configurations with Hamming weight $N + g + h$. Here, g is the number of 1s in the 0-sector of \mathbf{x}_r in the initial configuration that need to be flipped to reach $\mathbf{x}_{\text{target}}$, while h is the number of 1s in the 1-sector that need to be flipped to reach $\mathbf{x}_{\text{target}}$. All possible combinations of g and h are considered in this analysis. Starting from the initial noisy configuration, the probability of flipping one of the g 1s that need to be flipped to reach $\mathbf{x}_{\text{target}}$ (which we will call a “successful” bit-flip) is given by

$$\frac{g(1 - \varepsilon)}{(1 - \varepsilon)(b + g) + \varepsilon(N - b + h)}; \quad (\text{S17})$$

the denominator reflects the total number of 1s weighted by their probabilities of being flipped. After flipping the first 1 in the 0-sector, the probability of another successful bit-flip in the same sector is similarly

$$\frac{(g - 1)(1 - \varepsilon)}{(1 - \varepsilon)(b + g - 1) + \varepsilon(N - b + h)}. \quad (\text{S18})$$

Thus, the probability of flipping all g of the incorrect 1s in the 0-sector is

$$\prod_{i=0}^{g-1} \frac{(g - i)(1 - \varepsilon)}{(1 - \varepsilon)(b + g - i) + \varepsilon(N - b + h)}. \quad (\text{S19})$$

The same argument can be applied to the bit-flips performed in the 1-sector. The probability of a successful first bit-flip in the 1-sector is given by:

$$\frac{h\varepsilon}{(1 - \varepsilon)b + \varepsilon(N - b + h)}, \quad (\text{S20})$$

by essentially the same argument as above. The probability of a successful second flip is

$$\frac{(h - 1)\varepsilon}{(1 - \varepsilon)b + \varepsilon(N - b + h - 1)}. \quad (\text{S21})$$

One can apply the same iterative argument in this case. Hence, the overall probability of obtaining the target bitstring $\mathbf{x}_{\text{target}}$ via configuration recovery is

$$\begin{aligned} & \prod_{i=0}^{g-1} \left(\frac{(g - i)(1 - \varepsilon)}{(1 - \varepsilon)(b + g - i) + \varepsilon(N - b + h)} \right) \prod_{j=0}^{h-1} \left(\frac{(h - j)\varepsilon}{(1 - \varepsilon)b + \varepsilon(N - b + h - j)} \right) \\ &= \prod_{i=1}^g \left(\frac{i(1 - \varepsilon)}{(1 - \varepsilon)(b + i) + \varepsilon(N - b + h)} \right) \prod_{j=1}^h \left(\frac{j\varepsilon}{(1 - \varepsilon)b + \varepsilon(N - b + j)} \right). \end{aligned} \quad (\text{S22})$$

2. For the set of configurations with Hamming weight $N - g - h$ that came from g and h instances of $1 \rightarrow 0$ bit-flip errors in the one- and zero-sectors of the \mathbf{x}_r state (respectively), a symmetric argument applies. We omit the derivation for this case.

The above gives the probabilities of recovering configurations with excess 1s or 0s back to a particular bitstring $\mathbf{x}_{\text{target}}$ of Hamming distance $2b$ from \mathbf{x}_r . We now integrate this probability over different values of g and h to obtain

the probability of recovering a particular bitstring of Hamming distance $2b$ to \mathbf{x}_r , from a sample drawn from the uniform distribution. We obtain the lower bound

$$\begin{aligned} P_{\text{recovery}}(M, N, b, \varepsilon) &\geq \sum_{g=0}^{M-N-b} \sum_{h=0}^{N-b} \frac{\binom{M-N-b}{g} \binom{b}{h}}{2^M} \prod_{i=1}^g \left(\frac{i(1-\varepsilon)}{(1-\varepsilon)(b+i) + \varepsilon(N-b+h)} \right) \prod_{j=1}^h \left(\frac{j\varepsilon}{(1-\varepsilon)b + \varepsilon(M-N-b+j)} \right) \\ &+ \sum_{g=0}^{N-b} \sum_{h=0}^{M-N-b} \frac{\binom{N-b}{g} \binom{b}{h}}{2^M} \prod_{i=1}^g \left(\frac{i(1-\varepsilon)}{(1-\varepsilon)(b+i) + \varepsilon(M-N-b+h)} \right) \prod_{j=1}^h \left(\frac{j\varepsilon}{(1-\varepsilon)b + \varepsilon(M-N-b+j)} \right). \end{aligned} \quad (\text{S23})$$

The two terms correspond to the two cases above, with the combinatorial fractions in each giving the probabilities of obtaining a configuration equivalent to the particular bitstring $\mathbf{x}_{\text{target}}$ plus g random $0 \rightarrow 1$ bit-flips in the 0-sector of \mathbf{x}_r and h random $0 \rightarrow 1$ bit-flips in the 1-sector (first term), or g random $1 \rightarrow 0$ bit-flips in the one-sector of the HF state and h random $1 \rightarrow 0$ bit-flips in the zero-sector (second term).

Fig. S2 shows the lower bound in Eq. (S23) to the recovery probability, for $\varepsilon = 0.1$ and $N = 10$, as a function of the number of qubits M . We observe that the lower bound decays as an exponential of the number of qubits and that the recovery of configurations with small values of b is more likely than the recovery of configurations defined by large b values. We remark that the configuration recovery probability for configurations close to \mathbf{x}_r is substantially larger than the probability of obtaining the same samples by sampling from the uniform distribution, which would be the limit of no quantum signal. We argue that this feature helps in improving molecular energies over using simple post-selection over correct particle sector.

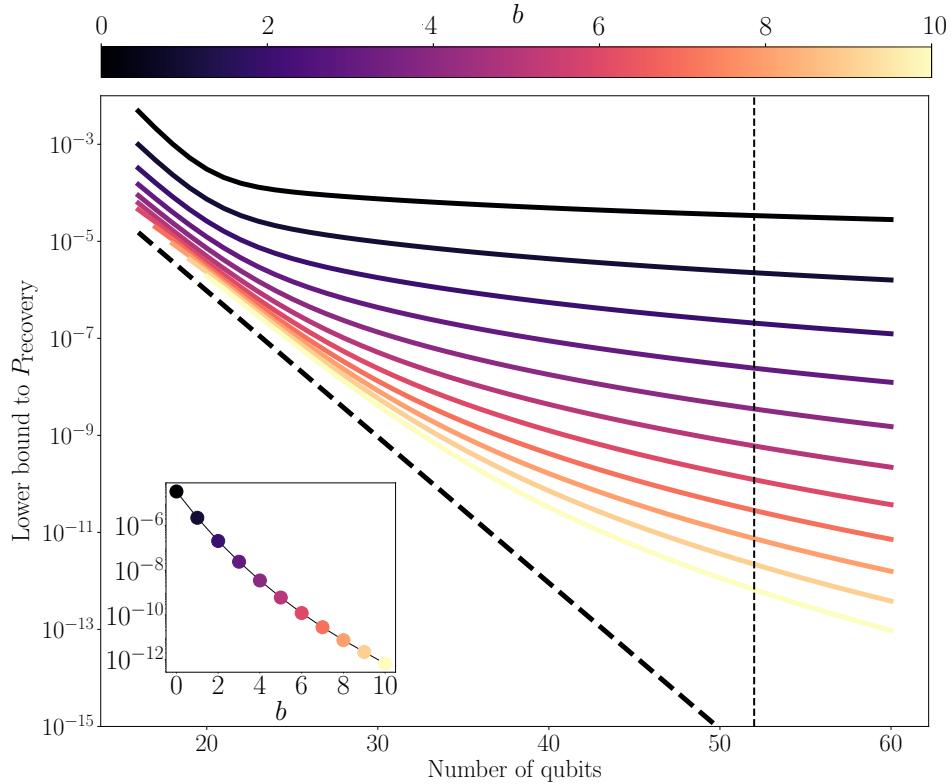


Figure S2. Lower bound to the probability of recovering a particular bitstring with Hamming distance $2b$ from a reference approximation \mathbf{x}_r bistring from a bistring sampled from the uniform distribution. The lower bound is shown as a function of the number of qubits. Different values of b are shown, as indicated by the colorbar. The dashed black curve shows $1/2^M$, which is the probability of obtaining the desired bitstring directly from the uniform distribution. The inset shows the lower bound as a function of b for $M = 52$ qubits.

3. Optimization of the circuit parameters

The optimization of circuit parameters to produce better distributions in the space of electronic configurations is an avenue of improvement of the HPC quantum estimator framework. We develop a formalism for parameter optimization using the subspace energy as the cost function, in contrast to previous experiments that rely on standard VQE optimization frameworks [30, 31]. The HPC quantum estimator is constructed by running a diagonalization procedure in K subspaces defined by different batches of sampled configurations, as described in Sec. II A 1. The number of subspaces considered K can go from $K = 1$ to $K = K_{\max}$, where $K_{\max} = \binom{D}{d}$. Recall that D is the dimensionality of the subspace of the Fock space spanned by configurations with the correct particle number and d is the dimension of the SCI subspace. Note that K_{\max} is double-combinatorial in the number of spin-orbitals and electrons. Therefore, computing the HPC quantum estimator (with or without configuration recovery) for all possible sets of d configurations is not efficient. Instead, we propose the use of a Monte Carlo estimator over batches of configurations $\mathcal{S}^{(k)}$ to optimize the energy of the HPC quantum estimator averaged over different sets of configurations.

Recall that each configuration \mathbf{x} is sampled with probability $P_\Psi(\mathbf{x}) = |\langle \mathbf{x} | \Psi \rangle|^2$, in the noiseless case. The following discussion applies to the noisy distribution \tilde{P}_Ψ . The method presented in this section can be applied also in conjunction with the configuration recovery technique. Since each configuration is i.i.d. sampled, the joint distribution that describes the probability of sampling the configurations in a batch $\mathcal{S}^{(k)}$ with d configurations is given by

$$P_\Psi(\mathcal{S}^{(k)}) = \prod_{\mathbf{x}_{(i)} \in \mathcal{S}^{(k)}} P_\Psi(\mathbf{x}_{(i)}). \quad (\text{S24})$$

We consider the situation where the quantum circuit that prepares $|\Psi\rangle$ is characterized by a set of variational parameters θ . The explicit dependence on variational parameters is denoted by Ψ_θ . A cost function is formulated for the circuit optimization:

$$\mathcal{E}(\theta) = \sum_{k=1}^{K_{\max}} P_{\Psi_\theta}(\mathcal{S}^{(k)}) E^{(k)}, \quad (\text{S25})$$

where $E^{(k)}$ is defined in Sec. II A 1, Eq. (S11) as the SCI ground-state energy for the configurations in $\mathcal{S}^{(k)}$. An estimator to the double combinatorially-large summation in $\mathcal{E}(\theta)$ can be obtained by the Monte-Carlo unbiased estimator:

$$\mathcal{E}(\theta) \approx \frac{1}{K} \sum_{k=1}^K E^{(k)} \quad (\text{S26})$$

with $K \ll K_{\max}$ and where the $E^{(k)}$ are obtained from $\mathcal{S}^{(k)}$ sampled according to $P_{\Psi_\theta}(\mathcal{S}^{(k)})$. The gradient of the cost function with respect to the variational parameters is given by:

$$\partial_\theta \mathcal{E}(\theta) = \sum_{k=1}^{K_{\max}} \left\{ \sum_{\mathbf{x}_{(i)} \in \mathcal{S}^{(k)}} [\partial_\theta P_{\Psi_\theta}(\mathbf{x}_{(i)})] \left(\prod_{\substack{\mathbf{x}_{(j)} \in \mathcal{S}^{(k)} \\ \mathbf{x}_{(j)} \neq \mathbf{x}_{(i)}}} P_{\Psi_\theta}(\mathbf{x}_{(j)}) \right) \right\} \cdot E^{(k)} \quad (\text{S27})$$

Multiplying each term in the sum from the product rule by $1 = P_{\Psi_\theta}(\mathbf{x}_{(i)}) / P_{\Psi_\theta}(\mathbf{x}_{(i)})$ we obtain:

$$\partial_\theta \mathcal{E}(\theta) = \sum_{k=1}^{K_{\max}} P_{\Psi_\theta}(\mathcal{S}^{(k)}) \left(\sum_{\mathbf{x}_{(i)} \in \mathcal{S}^{(k)}} \frac{[\partial_\theta P_{\Psi_\theta}(\mathbf{x}_{(i)})]}{P_{\Psi_\theta}(\mathbf{x}_{(i)})} \right) E^{(k)}, \quad (\text{S28})$$

whose Monte-Carlo estimator is given by:

$$\partial_\theta \mathcal{E}(\theta) \approx \frac{1}{K} \sum_{k=1}^K \left(\sum_{\mathbf{x}_{(i)} \in \mathcal{S}^{(k)}} \frac{[\partial_\theta P_{\Psi_\theta}(\mathbf{x}_{(i)})]}{P_{\Psi_\theta}(\mathbf{x}_{(i)})} \right) E^{(k)} \quad (\text{S29})$$

with $K \ll K_{\max}$ and where again the $E^{(k)}$ are obtained from $\mathcal{S}^{(k)}$ sampled according to $P_{\Psi_\theta}(\mathcal{S}^{(k)})$. The evaluation of $[\partial_\theta P_{\Psi_\theta}(\mathbf{x}_{(i)})]$ may be achieved by:

$$\partial_\theta P_{\Psi_\theta}(\mathbf{x}_{(i)}) = \partial_\theta (\langle \mathbf{x}_{(i)} | \Psi_\theta \rangle \langle \Psi_\theta | \mathbf{x}_{(i)} \rangle) = 2\text{Re} \{ \langle \mathbf{x}_{(i)} | \Psi_\theta \rangle \langle \partial_\theta \Psi_\theta | \mathbf{x}_{(i)} \rangle \}. \quad (\text{S30})$$

We remark that the limit where $K = 1$ is equivalent to the stochastic gradient-descent optimization technique.

For most classes of variational circuits, the gradient of the circuits $|\partial_\theta \Psi_\theta\rangle$ may be implemented by parameter-shift rules or similar techniques. The reader may have noticed that the estimator for the gradient in the cost function of Eq. (S29) is biased if the support of $[\partial_\theta P_{\Psi_\theta}]$ does not coincide with the support of P_{Ψ_θ} , as a direct consequence of writing $1 = P_{\Psi_\theta}(\mathbf{x}_{(i)}) / P_{\Psi_\theta}(\mathbf{x}_{(i)})$ for configurations outside the support of P_{Ψ_θ} , where that expression is ill-defined.

Another practical issue with the gradient-based optimization of $\mathcal{E}(\theta)$ is the requirement to evaluate $\langle \mathbf{x}_{(i)} | \Psi_\theta \rangle$ and $\langle \partial_\theta \Psi_\theta | \mathbf{x}_{(i)} \rangle$, which rely in the Hadamard test for their evaluation. The implementation of the Hadamard test requires controlled versions of the circuits that realize the variational states. The implementation of such controlled unitaries requires circuit depths beyond the reach of current quantum devices.

An alternative approach is to optimize $\mathcal{E}(\theta)$ by explicit-gradient-free methods like COBYLA [70] or simulated annealing. Gradient-free optimization of the circuit parameters is used in the numerical study of Sec. III A 4.

The optimization of the circuit parameters on quantum experiments to minimize the estimator energy will be the subject of future studies, which should consider the effect of noise in different gradient-free optimizers. Furthermore, the goal of this work is not to show that the optimization of the circuit parameters can improve the accuracy of the estimator, but to show that the quantum centric supercomputing estimator allows to tackle electronic structure problems described by an unprecedented number of qubits. Moreover, we show that the estimator run on circuits with fixed parameters already achieves good levels of accuracy.

4. Orbital optimization

We advise the reader to first visit Sec. II A 3, since the cost function that is optimized is similar in this scenario. Orbital or single-particle basis rotations are a common approach to improve the accuracy of variational calculations of electronic structure systems. Orbital optimizations can be applied in conjunction with a wide variety of electronic structure methods, including complete active space diagonalizations [71–74], the *density matrix renormalization group* [75–77], *variational Monte Carlo* [78] techniques, and quantum computing variational approaches [79–81]. Moreover, orbital optimizations are common practice on SCI-based approaches as well [82]. More generally, orbital optimizations make variational approaches invariant under orbital rotations. For notation clarity we replace in this section the notation of the two-body integral $(pr|qs)$ by the abbreviation h_{pqrs} .

The implementation of the orbital optimizations in conjunction with the HPC quantum estimator (with or without configuration recovery) follows the procedure presented in Ref. [78]. Orbital rotations consist of the application of the similarity transformation:

$$\hat{\tilde{H}} = \hat{U}^\dagger(\kappa) \hat{H} \hat{U}(\kappa), \quad (\text{S31})$$

where

$$\hat{U}(\kappa) = \exp \left(\sum_{pq} \kappa_{pq} \hat{a}_{p\sigma}^\dagger \hat{a}_{q\sigma} \right). \quad (\text{S32})$$

In this work, we choose the matrix that parametrizes the rotation to be real: $\kappa \in \mathbb{R}^{N_{\text{MO}} \times N_{\text{MO}}}$. To enforce unitarity in $\hat{U}(\kappa)$ we require that $\kappa_{pq} = -\kappa_{qp}$. According to Thouless's Theorem [83], the action of the similarity transformation defined by $\hat{U}(\kappa)$ transforms the creation operators according to:

$$\hat{a}_{p\sigma}^\dagger \mapsto \hat{U}(\kappa)^\dagger \hat{a}_{p\sigma}^\dagger \hat{U}(\kappa) = \sum_t \Omega_{tp} \hat{a}_{t\sigma}^\dagger, \quad (\text{S33})$$

where $\Omega = \exp(\kappa) \in \mathbb{R}^{N_{\text{MO}} \times N_{\text{MO}}}$. The Hamiltonian in the rotated basis can therefore be written in terms of the reference creation and annihilation operators:

$$\hat{\tilde{H}} = \sum_{pq} \tilde{h}_{pq} \hat{a}_{p\sigma}^\dagger \hat{a}_{q\sigma} + \sum_{pqrs} \frac{\tilde{h}_{pqrs}}{2} \hat{a}_{p\sigma}^\dagger \hat{a}_{q\tau}^\dagger \hat{a}_{s\tau} \hat{a}_{r\sigma} \quad (\text{S34})$$

where the one- and two-body integrals have been transformed according to the tensor transformations:

$$\begin{aligned} \tilde{h}_{pq} &= h_{tu} \Omega_{tp} \Omega_{uq}, \\ \tilde{h}_{pqrs} &= h_{tuvw} \Omega_{tp} \Omega_{uq} \Omega_{vr} \Omega_{ws}, \end{aligned} \quad (\text{S35})$$

using Einstein's summation convention.

A variational procedure is used to search for the single-particle basis that yields the optimal description of the ground state, given the collection of all possible K_{\max} bare variational trial states $|\psi_{\Theta}^{(k)}\rangle$. The variational procedure is similar to the method presented in Sec. II A 3. In this setting, $|\psi_{\Theta}^{(k)}\rangle$ is the SCI ansatz given as the linear combination of the d determinants in $\mathcal{S}^{(k)}$ (see Eq. (S10)). The set of variational parameters Θ is composed of the circuit parameters θ defining $|\Psi_{\theta}\rangle$ together with the wavefunction amplitudes $c_{\mathbf{x}}^{(k)}$ in the subspace defined by the SCI configurations. Each variational state is *dressed* by the same single-particle orbital rotation

$$|\psi_{\{\kappa, \Theta\}}^{(k)}\rangle = \hat{U}(\kappa) |\psi_{\Theta}^{(k)}\rangle. \quad (\text{S36})$$

The loss function to be optimized in the variational setting is defined by the average, over batches of configurations, of the Rayleigh quotient:

$$\mathcal{E}(\kappa, \Theta) = \sum_{k=1}^{K_{\max}} P_{\Psi_{\theta}}(\mathcal{S}^{(k)}) \langle \psi^{(k)} | \hat{U}^{\dagger}(\kappa) \hat{H} \hat{U}(\kappa) | \psi^{(k)} \rangle = \sum_{k=1}^{K_{\max}} P_{\Psi_{\theta}}(\mathcal{S}^{(k)}) \langle \psi^{(k)} | \hat{H}(\kappa) | \psi^{(k)} \rangle. \quad (\text{S37})$$

Gradient descent and its variants can be used to minimize both κ and θ in $\mathcal{E}(\kappa, \Theta)$, while the optimization of $c_{\mathbf{x}}^{(k)}$ is carried out by the diagonalization procedure (see Sec. II A 1). Note that the sum in the average contains double-combinatorially many terms. Consequently, a Monte-Carlo based estimator is used to obtain an unbiased estimate (see Sec. II A 3). Gradients with respect to θ have the same expression as in Eq. (S28) replacing \hat{H} by \hat{H} . Gradients with respect to the orbital rotations can be computed by the contraction of the bare one- and two-body reduced density matrices (1- and 2-RDMs) with the gradients of the one- and -two body integrals with respect to κ_{pq} in Eq. (S37):

$$\partial_{\kappa} \mathcal{E}(\kappa, \Theta) = \sum_{k=1}^{K_{\max}} P_{\Psi_{\theta}}(\mathcal{S}^{(k)}) \left[\sum_{\substack{pq \\ \sigma}} \tilde{h}'_{pq} \Gamma_{pq;\sigma}^{(k)} + \sum_{\substack{pqrs \\ \sigma}} \frac{\tilde{h}'_{pqsr}}{2} \left(\Gamma_{pqrs;\sigma,\sigma}^{(k)} + \Gamma_{pqsr;\sigma,-\sigma}^{(k)} \right) \right]. \quad (\text{S38})$$

In the previous expression the gradients of the integrals in the rotated basis are given by

$$\begin{aligned} \tilde{h}'_{pq} &= h_{tu} \frac{\partial}{\partial \kappa} (\Omega_{tp} \Omega_{uq}) \\ \tilde{h}'_{pqrs} &= h_{tuvw} \frac{\partial}{\partial \kappa} (\Omega_{tp} \Omega_{uq} \Omega_{vr} \Omega_{ws}), \end{aligned} \quad (\text{S39})$$

and the bare 1-RDMs and 2-RDMs are defined as

$$\begin{aligned} \Gamma_{pq;\sigma}^{(k)} &= \langle \psi^{(k)} | \hat{a}_{p\sigma}^{\dagger} \hat{a}_{q\sigma} | \psi^{(k)} \rangle \\ \Gamma_{pqsr;\sigma,\tau}^{(k)} &= \langle \psi^{(k)} | \hat{a}_{p\sigma}^{\dagger} \hat{a}_{q\tau}^{\dagger} \hat{a}_{s\tau} \hat{a}_{r\sigma} | \psi^{(k)} \rangle. \end{aligned} \quad (\text{S40})$$

The gradients of the integrals with respect to the rotation parameters are computed using the automatic differentiation (AD) tools of the software package Jax [84]. Since the evaluation of the gradients with respect to κ requires the evaluation of an intractable sum, we use an unbiased Monte-Carlo estimator for its evaluation:

$$\partial_{\kappa} \mathcal{E}(\kappa, \Theta) \approx \frac{1}{K} \sum_{k=1}^{K_{\max}} \left[\sum_{\substack{pq \\ \sigma}} \tilde{h}'_{pq} \Gamma_{pq;\sigma}^{(k)} + \sum_{\substack{pqrs \\ \sigma}} \frac{\tilde{h}'_{pqsr}}{2} \left(\Gamma_{pqrs;\sigma,\sigma}^{(k)} + \Gamma_{pqsr;\sigma,-\sigma}^{(k)} \right) \right], \quad (\text{S41})$$

with $K \ll K_{\max}$ and where the batches of configurations labelled by (k) are sampled according to $P_{\Psi_{\theta}}(\mathcal{S}^{(k)})$ (see Eq. (S24)).

For a full orbital optimization calculation, all sets of variational parameters κ , θ , and $c_{\mathbf{x}}^{(k)}$ should be updated. Since κ and θ sets rely on gradients for their optimization, they can be updated simultaneously. However, the wavefunction amplitudes are not updated by gradients. In this case, the optimization strategy consists of the alternation of a number of gradient-descent steps optimizing κ and θ followed by the eigenstate solver that now uses samples from an updated distribution according to the change in θ with a rotated Hamiltonian according to κ . This alternation is

repeated for a number of iterations N_{SCF} . In this work, we do not consider the optimization of circuit parameters θ and therefore the effect of the orbital optimization is not maximized, since the quantum circuit is not allowed to respond to the change in the Hamiltonian.

See Sec. IV C to see the effect of orbital optimizations in the accuracy of the HPC quantum estimator (with configuration recovery). The section shows results run on measurement outcomes obtained from the Heron device, to study the dissociation of N₂ at cc-pVDZ level of theory.

B. Wavefunction concentration and accuracy

1. Sample complexity

Under the assumption that the ground state is sufficiently concentrated, we show here that the energy obtained from the estimator converges to the ground-state energy exponentially fast in the number of samples. Suppose we sample from the ground state $|G\rangle = \sum_I c_I |I(\mathbf{x})\rangle$, where $I = 0, \dots, 2^M - 1$ label 2^M computational configurations and c_I are the corresponding coefficients. In other words, the symbol $I(\mathbf{x})$ denotes the integer with binary representation \mathbf{x} for a given binary string \mathbf{x} . To abbreviate the notation we drop the explicit dependence of \mathbf{x} on I . Let $P_I = |c_I|^2$ be the probability of each bitstring and, without loss of generality, assume the configurations are ordered by P_I : $P_0 \geq P_1 \geq P_2 \geq \dots$. To quantify the concentration of the configurations, we define parameters α_m and β_m for a given m such that

$$\sum_{I=0}^{m-1} P_I \geq \alpha_m, \quad \text{and} \quad P_1 \geq \dots \geq P_m \geq \beta_m. \quad (\text{S42})$$

The first condition imposes that the total probability of the first m configurations must be at least α_m while the second lower bounds the individual P_I by β_m . We first prove general statements using α_m and β_m and, after that, apply these statements to different distributions of P_x .

If we draw N_s samples from $|G\rangle$, we can show that the probability of *not* obtaining all of the first m configurations among the N_s samples decays exponentially with N_s . The probability of not seeing a given bitstring I in any of the N_s samples is $(1 - P_I)^{N_s}$. Therefore, the probability p_{fail} of having at least one of the first m configurations not appearing in the N_s samples is upper bounded by

$$\sum_{I=0}^{m-1} (1 - P_I)^{N_s} \leq \sum_{I=0}^{m-1} (1 - \beta_m)^{N_s} \leq m e^{-N_s \beta_m}. \quad (\text{S43})$$

Choosing N_s large enough so that $m \leq e^{N_s \beta_m / 2}$, the bound further simplifies to $p_{\text{fail}} \leq e^{-N_s \beta_m / 2}$. Therefore, if we would like the probability of failure to be at most η , it is sufficient to choose

$$N_s \geq \frac{2}{\beta_m} \log \frac{1}{\eta}. \quad (\text{S44})$$

Next, to understand how m should scale with the system size, we look at the error in the energy of the approximate ground state constructed from diagonalizing the Hamiltonian in the sampled basis. First, let $\mathcal{S} = \{|I\rangle : I = 0, \dots, m-1\}$ be the truncated basis consisting of exactly the first m configurations. Let

$$|G_m\rangle = \frac{1}{\sqrt{\mathcal{N}}} \sum_{I=0}^{m-1} c_I |x\rangle \quad (\text{S45})$$

be the normalized state constructed by truncating $|G\rangle$ to the subspace \mathcal{S} . Here, $\mathcal{N} = \sum_{I=1}^m P_I$ is a normalization constant. We have

$$\| |G\rangle - |G_m\rangle \| ^2 = 2 - \langle G | G_m \rangle - \langle G_m | G \rangle = 2 - 2 \sqrt{\sum_{I=0}^{m-1} P_I} \leq 2 - 2\sqrt{\alpha_m}. \quad (\text{S46})$$

Intuitively, if $\alpha_m \approx 1$, $|G\rangle$ is close to $|G_m\rangle$. In particular, it implies an upper bound on the energy of $|G_m\rangle$:

$$\begin{aligned} \langle G_m | \hat{H} | G_m \rangle &\leq \langle G | \hat{H} | G \rangle + 2\|\hat{H}\| \| |G\rangle - |G_m\rangle \|^2 \\ &\leq \langle G | \hat{H} | G \rangle + 4\|\hat{H}\|(1 - \sqrt{\alpha_m}). \end{aligned} \quad (\text{S47})$$

Now, let \tilde{H} be the $m \times m$ matrix that represents the Hamiltonian in the truncated subspace \mathcal{S} , i.e. the top left $m \times m$ block of H , and let $|\tilde{G}\rangle$ be the ground state of \tilde{H} . We have

$$\langle \tilde{G} | \hat{H} | \tilde{G} \rangle = \langle \tilde{G} | \tilde{H} | \tilde{G} \rangle \leq \langle G_m | \tilde{H} | G_m \rangle = \langle G_m | \hat{H} | G_m \rangle, \quad (\text{S48})$$

where the inequality follows from $|\tilde{G}\rangle$ being the ground state of \tilde{H} . Combining Eqs. (S47) and (S48), we have a bound on the difference between the ground-state energy of \tilde{H} and the original Hamiltonian H :

$$\langle \tilde{G} | \tilde{H} | \tilde{G} \rangle - \langle G | \hat{H} | G \rangle \leq 4\|\hat{H}\|(1 - \sqrt{\alpha_m}). \quad (\text{S49})$$

Earlier, we assumed that \mathcal{S} consists of exactly the first m bit strings. Note that $\langle \tilde{G} | \tilde{H} | \tilde{G} \rangle$ cannot increase if we add more configurations to \mathcal{S} . Therefore, combining with the earlier statement about the probability of seeing all m bit strings in N samples, we have that for any $\eta \in (0, 1)$, choosing

$$N_s \geq \max \left\{ \log \frac{2m}{\beta_m}, \frac{2}{\beta_m} \log \frac{1}{\eta} \right\} \quad (\text{S50})$$

guarantees

$$\langle \tilde{G} | \tilde{H} | \tilde{G} \rangle - \langle G | \hat{H} | G \rangle \leq 4\|\hat{H}\|(1 - \sqrt{\alpha_m}) \quad (\text{S51})$$

with probability at least $1 - \eta$. This result allows us to estimate the sufficient number of samples N_s given a distribution of the configurations in the ground state. For example, if $P_I \propto e^{-I}$ decays exponentially, we have $1 - \sqrt{\alpha_m} \propto e^{-m}$ and $\beta_m \propto e^{-m}$. At $m \approx c \log M$, where c is a parameter to be determined, choosing $N_s = \Omega(M^c)$ is sufficient to guarantee an energy error at most $\mathcal{O}(\|\hat{H}\|/M^c)$. We have the freedom to choose c large enough so that $\|\hat{H}\|/M^c \ll 1$.

In particular, if $\|\hat{H}\| = \mathcal{O}(M^4)$, we would choose $c > 4$. In short, the sufficient number of samples N_s scales only polynomially with the number of qubits M .

This favorable scaling persists even if $P_I \propto 1/I^\gamma$ decays algebraically for some constant $\gamma > 1$. In this case, we have $1 - \sqrt{\alpha_m} \propto 1/m^{\gamma-1}$ and $\beta_m \propto 1/m^\gamma$. Therefore, choosing $N_s = \Omega(m^{\gamma-1})$, we can guarantee the error is at most $\mathcal{O}(\|\hat{H}\|/m^\gamma)$. Again, we have the freedom to choose $m \propto M^c$ for some c large enough so that $\|\hat{H}\|/M^{c\gamma} \ll 1$. With this choice of m , the number of samples only needs to scale polynomially with M as $\mathcal{O}(M^{c(\gamma-1)})$.

Recall that in our procedure, further subsample from the set of samples \mathcal{S} . If the size of the subsamples is large enough, one can also prove that at least one of them contains all the m configurations. For simplicity, we assume that we will draw K subsamples uniformly from \mathcal{S} . This subsampling probability is the worst case for the success probability and, in practice, we would instead draw these subsamples based on their frequencies in the N_s samples. Assume that each subsample has d configurations. Additionally, we are interested in the limit $m \ll d \ll N_s$. We will calculate the probability that none of these K subsamples has all of the m correct configurations and show that this probability decays exponentially with K .

First, consider a subsample of d samples. There are $\binom{N_s}{d}$ ways to choose this subsample. Among them, $\binom{N_s-m}{d-m}$ contain the m configurations. So the probability of *not* having all m configurations is

$$p_{\text{fail}}^{(1)} \equiv 1 - \frac{d!(N_s-m)!}{N_s!(d-m)!} \leq 1 - \left(\frac{d-m}{N_s} \right)^m. \quad (\text{S52})$$

So the probability that none of the K subsamples has all of the m configurations is bounded by

$$p_{\text{fail}}^{(K)} \leq \left[1 - \left(\frac{d-m}{N_s} \right)^m \right]^K, \quad (\text{S53})$$

which indeed decays exponentially with K .

In the first example above where $P_I \propto e^{-I}$, if we choose $d \propto M^c$ such that $N_s/d = \kappa$ is a constant, we have

$$p_{\text{fail}}^{(K)} \lesssim \left(1 - \frac{1}{\kappa^{c \log M}} \right)^K = \left(1 - \frac{1}{M^{c \log \kappa}} \right)^K. \quad (\text{S54})$$

Therefore, a polynomially large number of subsamples $K \propto M^{c \log \kappa}$ would ensure that the failure probability is at most a constant. Although the scaling of d with M is the same as that of N_s , choosing a large κ significantly reduces the dimension of the matrices that we need to diagonalize. Instead of diagonalizing a large $N_s \times N_s$ matrix, we diagonalize K different $d \times d$ matrices, which can be performed in parallel.

2. Assessing wavefunction concentration

A-priori it is not possible to know if the ground-state wavefunction of a given system is concentrated. In this subsection, we present a number of tests that can be carried out to study the concentration of the ground state wavefunction, having access only to the eigenstates $|\psi^{(k)}\rangle$ produced by the estimator with or without configuration recovery.

Assume that the distribution over the space of electronic configurations generated by the wave function $|\Psi\rangle$ coincides with that of the ground state. The study of the convergence of the approximate ground state properties obtained from $|\psi^{(k)}\rangle$, as a function of d , can reveal if the ground state wavefunction is concentrated. When the ground state is not concentrated, the convergence of observables like the energy with d can be slow, or not converge at all. In the particular case of the energy, as the number of configurations d is increased, the value for the approximate ground-state energy decreases. If the approximate ground state energy has not converged when reaching the largest d amenable by the available computational resources, it is clear that the size of the support of the ground-state wavefunction exceeds the value of d .

Furthermore, when the ground state wave function is not concentrated, different batches of sampled configurations $\mathcal{S}^{(k)}$ will likely share only a few common configurations, resulting in a large variance between the properties extracted from the different $|\psi^{(k)}\rangle$.

The expectation values of observables being close to their extremal values can also be a good indicator of the concentration. For example, the closer the occupations are to a corner of the hypercube $[0, 1]^M$, the more concentrated the wavefunction is on the bitstring that corresponds to the corner.

We remark that while these tests can indicate the presence of a non-concentrated ground state wavefunction, they do not guarantee a definitive answer to whether the ground-ground state wavefunction was indeed concentrated.

3. A numerical study of the accuracy: Hubbard model

To illustrate the dependence of the accuracy of the SCI-based eigensolver as a function of the ground-state wavefunction concentration we consider the N_{MO} -site Hubbard model in a fully-connected lattice with random hopping amplitudes:

$$\hat{H} = -\frac{1}{\sqrt{L}} \sum_{\substack{p,q=1 \\ p \neq q}}^{N_{\text{MO}}} \sum_{\sigma} t_{pq} \hat{a}_{p\sigma}^{\dagger} \hat{a}_{q\sigma} + U \sum_{p=1}^{N_{\text{MO}}} \hat{n}_{p\uparrow} \hat{n}_{p\downarrow}. \quad (\text{S55})$$

The hopping amplitudes $t_{pq} = t_{qp}$ are independent random variables drawn from the Gaussian distribution with mean 0 and standard deviation 1, i.e. $\overline{t_{pq}} = 0$ and $\overline{t_{pq}^2} = t^2 = 1$. The results presented in this section are obtained from the average of results obtained from twenty disorder realizations. In the infinite-volume, this model is known to possess *self-averaging* properties [85]. However, on finite-size systems, there are sample-to-sample fluctuations. The numerical study in this section does not consider the effect of noise, so we run the estimator without configuration recovery. For the discussion that follows, we span the wavefunction amplitudes in terms of electronic configurations in the localized basis, i.e. the basis used to define the Hamiltonian in Eq. (S55), contrary to the molecular systems considered where the reference basis is that of Hartree-Fock orbitals.

We choose the Hubbard model for this analysis because the ground-state properties, and the wavefunction concentration, depend on a single parameter: U . Furthermore, the study of the Hubbard model with various flavors of Configuration Interaction approaches has been considered by previous works [86–89]. The onsite Hubbard repulsion U controls the nature of the ground state of the system. In the $N_{\text{MO}} \rightarrow \infty$ limit, and for $0 \leq U < U_c$, the ground state is a Fermi liquid metal. For $U > U_c$, the ground state is a Mott insulator. The metal to insulator transition in this case is of second order [90, 91]. *Dynamical Mean Field Theory* studies report a transition point of $U_c = 5.82\dots$ [92]. The transition on finite-size systems is harder to identify [93]. The change in the nature of the ground state is reflected in the concentration of the wavefunction amplitudes. In the Mott phase, electrons tend to localize and there is a preference for configurations with low double occupancy [93]. For $U = 0$ and small U , and given the disordered nature of the model, it is expected that all electronic configurations will be equivalent, especially after considering a number of disorder realizations. It is therefore expected that the eigenstate solver will show better accuracy for larger values of U .

Since the goal of this numerical experiment is to isolate the performance of the SCI eigensolver as a function of the wavefunction concentration, the electronic configurations used to run the SCI eigenstate solver are the $d = 10^3$ configurations of highest amplitude in the exact ground state. This choice removes any effects in the accuracy coming from the choice of the quantum circuit generating the configurations. This procedure is equivalent to the QSCI

proposal [30], where the configurations are sampled from the exact ground-state wavefunction. As a measure of concentration, we consider the entropy of the distribution defined by the wavefunction amplitudes:

$$S = - \sum_{\mathbf{x}} |c_{\mathbf{x}}|^2 \log_2 |c_{\mathbf{x}}|^2. \quad (\text{S56})$$

To measure the accuracy in the ground state we consider the relative error in the ground-state energy, a common metric in variational many-body calculations. The relative error is refined as: $|E_{\text{SCI}} - E_{\text{exact}}| / E_{\text{exact}}$. For the numerical experiments we consider $N_{\text{MO}} = 10$, and 31 equally-spaced values of U , between $U = 1$ and $U = 16$, both included.

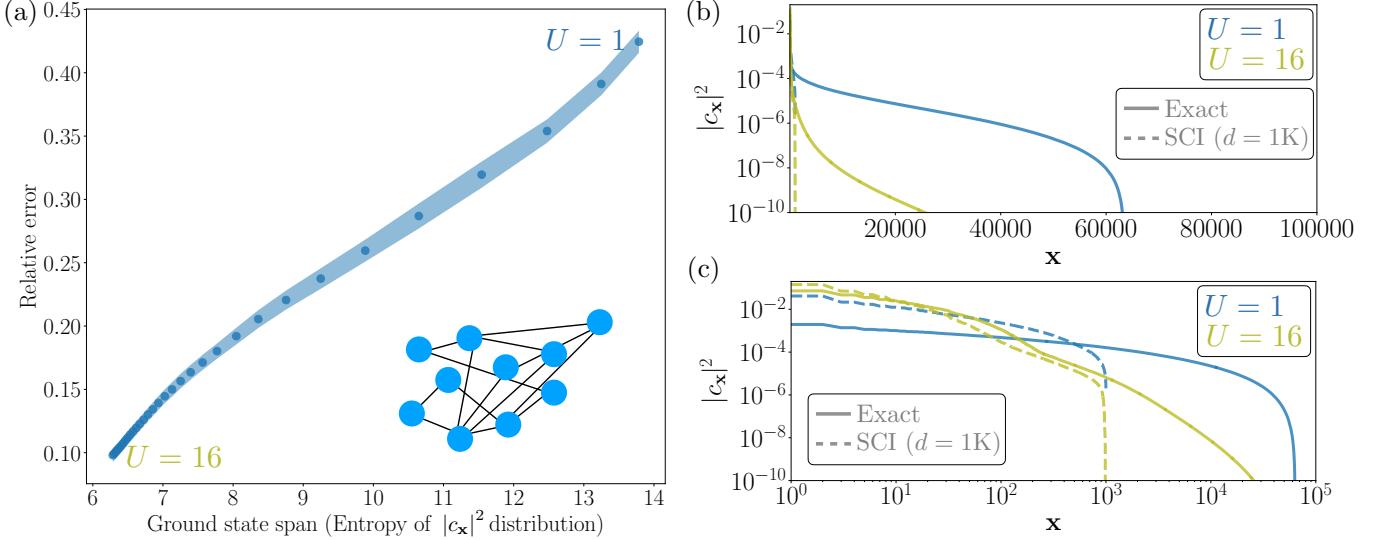


Figure S3. Numerical study on the dependence of the performance of SCI with the ground-state wavefunction concentration in the fully connected Hubbard model with random hopping amplitudes as depicted in the inset of panel (a). Quantities are averaged over twenty disorder realizations. (a) Relative error in the ground-state energy as a function of the exact ground-state wavefunction span, as defined in Eq. (S56). Different points correspond to different values of the onsite interaction between $U = 1$ and $U = 16$ as indicated. The shaded region indicates the standard error in the mean. (b) Wavefunction amplitudes in decreasing order of magnitude for two values of U as indicated in the legend. Solid lines show the exact ground amplitudes while the dashed lines show the approximate ground state amplitudes obtained by SCI with 10^3 configurations (see supplementary text). (c) Same as (b) with a logarithmic scale in the horizontal axis to better show the structure of the SCI-constructed amplitudes.

Fig. S3 (a) shows the relative error in the ground-state energy obtained by SCI as a function of the entropy of the distribution of the wavefunction amplitudes. The more concentrated the ground-state wavefunction (lower entropy), the lower the relative error becomes. As expected, the SCI-based eigensolver performs better for large U deep in the Mott phase. For small U in the metallic phase, the SCI trial wavefunction with $d = 10^3$ does not have enough expressive power to accurately represent the ground state of the system, whose amplitudes anti-concentrate.

Panels (b) and (c) in Fig. S3 show a comparison between the exact and SCI wavefunction amplitudes ordered by their magnitude for $U = 1$ (metal) and $U = 16$ (Mott insulator). For more concentrated wavefunctions ($U = 16$) the SCI amplitudes are in better agreement with the exact amplitudes. For $U = 1$ the number of relevant electronic configurations in the exact wavefunction greatly exceeds the $d = 10^3$ configurations that SCI is allowed to use, explaining the poor performance of SCI in the prediction of the ground-state energy as shown in Fig. S3 (a).

C. Energy-variance extrapolation

It is expected that the accuracy of the HPC quantum estimator (with or without configuration recovery) will increase as the number of configurations used for the subspace expansion d is increased. However, the convergence of the ground-state properties with d is not expected to follow any specific functional relation. Therefore, attempting to analyze the convergence as a function of d is not well motivated. Instead, the different eigenstate approximations obtained for different values of d and different batches of samples $\mathcal{S}^{(k)}$ are used for an energy-variance extrapolation [48,

[94–98]. Consider the approximate eigenstate $|\psi\rangle$, whose energy is given by $E = \langle\psi|\hat{H}|\psi\rangle$, and consider the exact eigenstate energy E_T . The difference between the approximate energy and E_T ,

$$\delta E = \langle\hat{H}\rangle - E_T, \quad (\text{S57})$$

vanishes linearly with the Hamiltonian variance divided by the square of the variational energy [48]

$$\frac{\Delta H}{E^2} = \frac{\langle\psi|\hat{H}^2|\psi\rangle - \langle\psi|\hat{H}|\psi\rangle^2}{E^2}. \quad (\text{S58})$$

This linear relation is satisfied as long as $|\psi\rangle$ is sufficiently close to an eigenstate of the Hamiltonian, as measured by the state fidelity. The least squares fit of a collection of energy-variance points yields an estimate of E_T , as the intersect of the fit with the ordinates. This point is the extrapolation of the estimate of the energy to the limit where $|\psi\rangle$ coincides with the exact eigenstate. Besides the energy extrapolation, the energy-variance analysis may also reveal the existence of multiple eigenstates close in energy to the ground state.

In the main text and in this supplementary material we apply the energy-variance analysis to two different sets of energy-variance pairs. The first one is the energy-variance pairs obtained by the HCI procedure where different energies and variances are obtained by changing the cutoff parameter that indirectly controls the number of determinants in the subspace projection and diagonalization, resulting in different levels of accuracy. The second set of energy-variance pairs are those obtained from the HPC quantum estimator for different numbers of configurations d as well as for different batches of sampled configurations $\mathcal{S}^{(k)}$.

III. EXPERIMENTS

This section begins by describing the quantum circuits that are employed to produce the configurations to which we apply the eigenstate solver. We describe how to set the circuit parameters from an efficient classical CCSD calculation, and show examples of the structure of the compiled circuits. We show numerical experiments comparing the performance of the circuits we use against the performance of a classically efficient reduction of the LUCJ circuits.

We also describe the setup for all of the experiments conducted in this work, highlighting the quantum and classical hardware used. We show the quantum device mappings, depth and number of gates in the circuits, and the number of measurement outcomes sampled. We also provide details about the classical computing resources.

A. Quantum circuits

In this work, we used the local-unitary cluster Jastrow (LUCJ) ansatz [40] to sample randomly distributed electronic configurations. LUCJ derives from the unitary cluster Jastrow (UCJ) ansatz, which has the form [41] of a product of L layers,

$$|\Psi\rangle = \prod_{\mu=1}^L e^{\hat{K}_\mu} e^{i\hat{J}_\mu} e^{-\hat{K}_\mu} |\mathbf{x}_{\text{RHF}}\rangle, \quad (\text{S59})$$

where

$$\hat{K}_\mu = \sum_{pq,\sigma} K_{pq}^\mu \hat{a}_{p\sigma}^\dagger \hat{a}_{q\sigma}, \quad \hat{J}_\mu = \sum_{pr,\sigma\tau} J_{p\sigma,q\tau}^\mu \hat{n}_{p\sigma} \hat{n}_{q\tau}. \quad (\text{S60})$$

In Eq. (S60), $p, q = 0 \dots N_{\text{MO}} - 1$ label molecular spatial orbitals and σ, τ label spin polarizations (α, β for spin-up and spin-down electrons respectively). $K_{pq}^\mu / J_{pq,\sigma\tau}^\mu$ has complex/real matrix elements and is anti-Hermitian/symmetric. The UCJ ansatz can be derived from a twice-factorized low-rank decomposition of the qUCCD ansatz [41, 99], and the L -product form is such that the exact full configuration interaction wavefunction can be obtained via Eq. (S59) [39, 41].

The local UCJ or LUCJ [40] introduces a “local” approximation of the UCJ ansatz, which makes the following modifications for opposite-spin and same-spin number-number terms:

$$\begin{aligned} \sum_{pq} J_{p\alpha,q\beta} \hat{n}_{p\alpha} \hat{n}_{q\beta} &\rightarrow \sum_{p \in S} J_{p\alpha,p\beta} \hat{n}_{p\alpha} \hat{n}_{p\beta} \\ \sum_{pq} J_{p\sigma,q\sigma} \hat{n}_{p\sigma} \hat{n}_{q\sigma} &\rightarrow \sum_{pq \in S'} J_{p\sigma,q\sigma} \hat{n}_{p\sigma} \hat{n}_{q\sigma}, \end{aligned} \quad (\text{S61})$$

where $\sigma = \alpha, \beta$ and the sets S, S' are such that the quantum circuit implementing $e^{i\hat{J}_\mu}$ has depth $O(1)$ and only comprises $O(|S| + |S'|)$ number-number ‘‘nn gates,’’ i.e. two-qubit unitaries of the form $U_{nn}(\varphi) = e^{-i\frac{\varphi}{4}(Z_p+Z_q-Z_pZ_q)}$, acting on adjacent qubits p, q in the topology of a certain device. For example, on a heavy-hex device, $S = \{4k, k = 0 \dots (N_{MO} - 1)/4\}$ and $S' = \{(p, p+1), p = 0 \dots N-2\}$ [40]. The circuits $e^{\pm\hat{K}_\mu}$, on the other hand, can be implemented by a Bogolyubov circuit acting on N_{MO} qubits and comprising $O(N_\alpha(N_{MO} - N_\alpha))$ gates and depth $O(N_{MO})$ [100–102]. Through the local approximation, the LUCJ Ansatz balances hardware friendliness and accuracy, the latter ultimately deriving from its connection to coupled cluster theory and adiabatic state preparation [40].

Unless otherwise specified, we use the truncated LUCJ circuit $|\Psi\rangle = e^{\hat{K}_2}e^{-\hat{K}_1}e^{i\hat{J}_1}e^{\hat{K}_1}|\mathbf{x}_{\text{RHF}}\rangle$, which is the result of considering the two-layer LUCJ circuit and removing the last orbital rotation and last Jastrow operations. The resulting state is implemented by a circuit whose depth is identical to the single-layer LUCJ circuit. The addition of the $e^{\hat{K}_2}$ operation to the circuit can have a large impact on the configurations generated by the circuit when the parameters are set from the t_2 tensor from a classical restricted closed-shell CCSD calculation (as described in the next section). For dynamically correlated species, the $J_{p\sigma,q\tau}^\mu$ parameters obtained from t_2 can have a small amplitude, resulting in the approximate cancellation of the $\exp(-\hat{K}_1)$ and $\exp(\hat{K}_1)$ terms in the ansatz. Without $\exp(\hat{K}_2)$, the resulting wavefunction can be over-concentrated around the Hartree-Fock configuration. Therefore, for dynamically correlated species the action of $\exp(\hat{K}_2)$ is to remove some of the excessive concentration of the $|\Psi\rangle$ wavefunction, when parameters are set from a restricted closed-shell CCSD calculation.

1. Initialization of LUCJ parameters

In our experiments, we parametrize the LUCJ circuits using the following procedure:

1. First, we carry out a classical restricted closed-shell CCSD calculation, yielding amplitudes $t_{1,ai}$ and $t_{2,aibj}$, where ij/ab labels occupied/unoccupied orbitals in the RHF state.
2. We reshape the t_2 tensor into the matrix $(t_2)_{ai,bj}$ and diagonalize it, $(t_2)_{ai,bj} = \sum_y \tau_y U_{ai,y} U_{bj,y}$, where the eigenvectors τ_y are sorted in decreasing order of absolute value.
3. We extend the unitaries to the following matrices,

$$\tilde{U}_{y,pr} = \delta_{pa}\delta_{ri}U_{ai,y}, \quad (\text{S62})$$

i.e. matrices where only the occupied/unoccupied block is non-zero.

4. We define the Hermitian operators

$$X_{\pm,y} = \frac{1 \mp i}{2} \left(\tilde{U}_y \pm i\tilde{U}_y^T \right) \quad (\text{S63})$$

and their eigenpairs

$$X_{\pm,y}V_{\pm,y} = g_{\pm,y}V_{\pm,y}. \quad (\text{S64})$$

5. We define the operators

$$\begin{aligned} (J_{2y})_{pr}^{\sigma\tau} &= \tau_y (g_{+,y})_p (g_{+,y})_r \\ (K_{2y})_{pr}^\sigma &= \log(V_{+,y})_{pr} \\ (J_{2y+1})_{pr}^{\sigma\tau} &= -\tau_y (g_{-,y})_p (g_{-,y})_r \\ (K_{2y+1})_{pr}^\sigma &= \log(V_{-,y})_{pr}. \end{aligned} \quad (\text{S65})$$

6. We retain the first L matrices J, K . This allows us to refine a non-local UCJ wavefunction [41, 99].
7. We zero out the entries of the J matrices leading to quantum gates acting on non-adjacent qubits. For a heavy-hex lattice, we retain the following elements:

$$\begin{aligned} (J_\mu)_{p,p+1}^{\alpha\alpha}, p &= 0 \dots N_{MO} - 2 \\ (J_\mu)_{p,p+1}^{\beta\beta}, p &= 0 \dots N_{MO} - 2 \\ (J_\mu)_{p,p}^{\alpha\beta}, p &= 0 \dots N_{MO} - 1, p \% 4 = 0. \end{aligned} \quad (\text{S66})$$

The final sparsification allows us to construct a LUCJ wavefunction. In a conventional LUCJ calculation, these parameters are the starting point of a variational optimization. For the hardware experiments reported in this study, we employed these parameters, without further optimization, to define a LUCJ circuit, that we used to sample randomly distributed configurations.

2. Mapping to heavy-hex devices

The choice of retaining the elements $(J_\mu)_{p,p}^{\alpha\beta}$ with $p = 0 \dots N_{\text{MO}} - 1$, $p \% 4 = 0$ when implementing LUCJ on a heavy-hex device has an important technical motivation: on such devices, assuming a number of qubits greatly exceeding N_{MO} , spin-up and spin-down orbitals can be mapped on two segments of adjacent qubits forming a “zig-zag” pattern and connected through an auxiliary qubit for $p = 0, 4, 8, \dots$ as shown in the three rightmost panels of Fig. S9. Such a qubit layout allows implementing LUCJ with a minimal overhead of **SWAP** gates (two per auxiliary qubit and layer of LUCJ). However, on current devices with up to 133 qubits, for $N_{\text{MO}} > 21$ one cannot couple $N_{\text{MO}}/4$ spin-orbitals with opposite spins through auxiliary qubits without incurring a significant overhead of **SWAP** gates, for the simple reason that a chain of 22 or more qubits is longer than the “diagonal” of the device. In such a situation, as shown in the rightmost panel of Fig. S9, the segments on which spin-up and spin-down qubits are mapped form two “tails” that are not connected by auxiliary qubits. This fact has two implications: (i) no more than 6 spin-orbitals with opposite spins can be coupled through auxiliary qubits, and (ii) if one retains the elements $(J_\mu)_{p,p}^{\alpha\beta}$ with $p = 0 \dots N_{\text{MO}} - 1$, $p \% 4 = 0$ and $p \leq 16$, the largest elements of J_μ may be discarded, yielding a lower-accuracy wavefunction. The first problem is a fundamental one, that can only be resolved with a substantially different mapping of fermionic degrees of freedom onto qubits and/or through the availability of larger devices. The second problem, on the other hand, has a simple solution, that we now describe. First, for any pair of spatial orbitals $p, r = 0 \dots N_{\text{MO}} - 1$, the orbital rotation

$$\hat{S}_{pr} = e^{-i\frac{\pi}{2}\sum_\sigma(\hat{a}_{p\sigma}^\dagger\hat{a}_{r\sigma} + \hat{a}_{r\sigma}^\dagger\hat{a}_{p\sigma})} \quad (\text{S67})$$

implements the permutation $S_{pr} \in \text{S}_{N_{\text{MO}}}$ exchanging orbitals p and r , in the sense that

$$\hat{S}_{pr}^\dagger \left(\sum_{qs,\tau} M_{qs} \hat{a}_{q\tau}^\dagger \hat{a}_{s\tau} \right) \hat{S}_{pr} = \sum_{qs,\tau} M'_{qs} \hat{a}_{q\tau}^\dagger \hat{a}_{s\tau}, \quad M'_{qs} = M_{S_{pr}(q), S_{pr}(s)}. \quad (\text{S68})$$

Three immediate implications of this fact are:

- $\hat{S}_{pr}^\dagger \hat{n}_{q\tau} \hat{S}_{pr} = \hat{n}_{S_{pr}(q)\tau}$.
- for any density-density operator $\hat{J}_\mu = \sum_{qs,\sigma\tau} (J_\mu)_{q,s}^{\sigma\tau} \hat{n}_{q\sigma} \hat{n}_{s\tau}$, one has

$$\hat{S}_{pr}^\dagger \hat{J}_\mu \hat{S}_{pr} = \sum_{qs,\tau} (J'_\mu)_{q,s}^{\sigma\tau} \hat{n}_{q\sigma} \hat{n}_{s\tau} = \hat{J}'_\mu, \quad (J'_\mu)_{q,s}^{\sigma\tau} = (J_\mu)_{S_{pr}(q), S_{pr}(s)}^{\sigma\tau}. \quad (\text{S69})$$

- that for any permutation $S \in \text{S}_{N_{\text{MO}}}$ there exists an orbital rotation $e^{\hat{K}_S}$ implementing the permutation S (this is true because permutations can be written as products of exchange permutations, an exchange permutation can be implemented by an orbital rotation, and orbital rotations are closed under multiplication).

Consider now the tensor $(J_\mu)_{q,s}^{\sigma\tau}$ resulting from the low-rank decomposition of the CCSD operator described in the previous Subsection. Let $p_0 \dots p_\ell$ be the ℓ elements of $(J_\mu)_{q,s}^{\alpha\beta}$ with the largest absolute values, and let $S \in \text{S}_{N_{\text{MO}}}$ be the permutation such that $S(p_0) = 0 \dots S(p_\ell) = 4\ell$. Then,

$$e^{-\hat{K}_S} \hat{J}_\mu e^{\hat{K}_S} = \hat{J}'_\mu \quad (\text{S70})$$

where the elements of $(J'_\mu)_{q,s}^{\alpha\beta}$ with the largest absolute values are at positions $0, \dots, 4\ell$. Before sparsifying the tensor \tilde{J}_μ , one can use the identity

$$e^{\hat{K}_\mu} e^{i\hat{J}_\mu} e^{-\hat{K}_\mu} = e^{\hat{K}_\mu} e^{\hat{K}_S} e^{i\hat{J}'_\mu} e^{-\hat{K}_S} e^{-\hat{K}_\mu} = e^{\hat{K}'_\mu} e^{i\hat{J}'_\mu} e^{-\hat{K}'_\mu} \quad (\text{S71})$$

to obtain a UCJ operator with transformed orbital rotations $e^{-\hat{K}'_\mu}$ and an opposite-spin density-density interaction whose largest elements in absolute value act on spatial orbitals $p = 0, 4, 8, \dots, 4\ell$. Sparsification of $(J'_\mu)_{q,s}^{\alpha\beta}$ then leads to retaining the ℓ dominant opposite-spin density-density interaction terms (as many as allowed by the size and topology of the available heavy-hex device) with a minimal overhead of **SWAP** gates.

In this work, we retained the elements $(J_\mu)_{p,p}^{\alpha\beta}$ with $p = 0 \dots N_{\text{MO}} - 1$, $p \% 4 = 0$ and $p \leq 16$. In future work, the procedure described here could be used to modify the LUCJ wavefunction and the resulting probability distribution for electronic configurations.

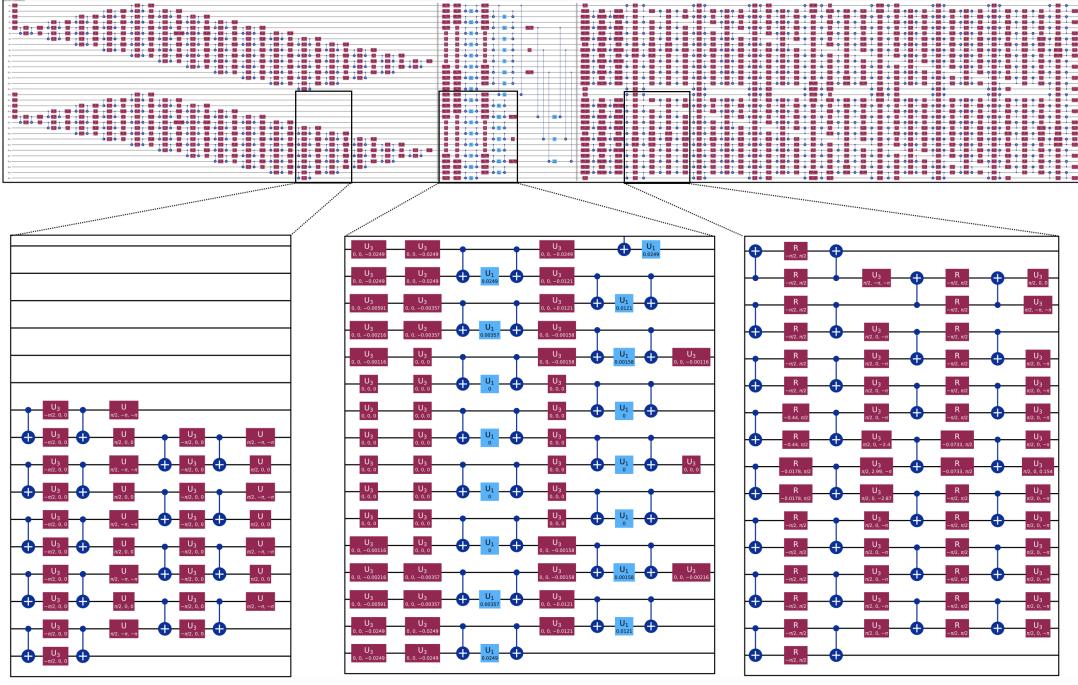


Figure S4. Top: LUCJ circuit used to simulate the ground state of $\text{N}_2/6\text{-}31\text{G}$ at the equilibrium bondlength, compiled into single-qubit (red, purple blocks) and CNOT (light blue symbols) gates. Bottom: zoomed-out views of gates from the first orbital rotation, density-density interaction, and second orbital rotation (left to right).

3. Compiled circuits

In Figures S4, S5, S6, and S7, we show the specific LUCJ circuits considered in this work. Each circuit is compiled into single-qubit

$$U_3(\theta, \phi, \lambda) = \begin{pmatrix} \cos(\theta/2) & -e^{i\lambda} \sin(\theta/2) \\ e^{i\phi} \cos(\theta/2) & e^{i(\phi+\lambda)} \cos(\theta/2) \end{pmatrix} \quad (\text{S72})$$

and two-qubit CNOT gates. For readability, the displayed circuits contain barriers separating orbital rotations and density-density interactions, which we removed in the final compilation to `ibm_nazca/ibm_torino` to reduce circuit depth. It is worth noting that the leftmost part of the quantum circuit does not implement the Bogolyubov transformation $\exp(\hat{K}_1)$, but prepares the Slater Determinant $\exp(\hat{K}_1)|\mathbf{x}_{\text{RHF}}\rangle$. In other words, it reproduces the action of $\exp(\hat{K}_1)$ on a specific Slater determinant, which allows for a drastic reduction in circuit size without information loss. Subsequent orbital rotations are not amenable to such a circuit simplification.

4. Effect of the Jastrow terms in the accuracy

In this section we study the impact of the two-body terms, $\exp(i\hat{J}_\mu)$, of the LUCJ ansatz in the performance of the HPC quantum estimator. We compare the case where the two-body parameters are zero, $J_{p\alpha,p\beta} = J_{p\alpha,p\alpha} = 0$, against the most general case where they can take arbitrary real values. Without the density-density interactions, the remaining circuit is a free-fermion evolution, i.e. the wavefunction amplitudes remain those of a single Slater determinant [83], so it can be efficiently simulated by a classical computer.

We present a numerical study of that emulates the procedure without taking into account the effect of noise. Thus, there is no configuration recovery. For the quantum circuit simulations, we used the software library `ffsim` [42]. `ffsim` performs the state vector simulation within a subspace of fixed particle number and total \hat{z} component of the spin, making the simulations much more efficient than a generic quantum circuit simulator.

Three circuits are considered: LUCJ($L = 1$), LUCJ($L = 4$), and the LUCJ circuit without two-body terms, which we refer to as ‘‘Determinant’’. We study the accuracy of the estimator in the dissociation of N_2 (6-31G). For the LUCJ

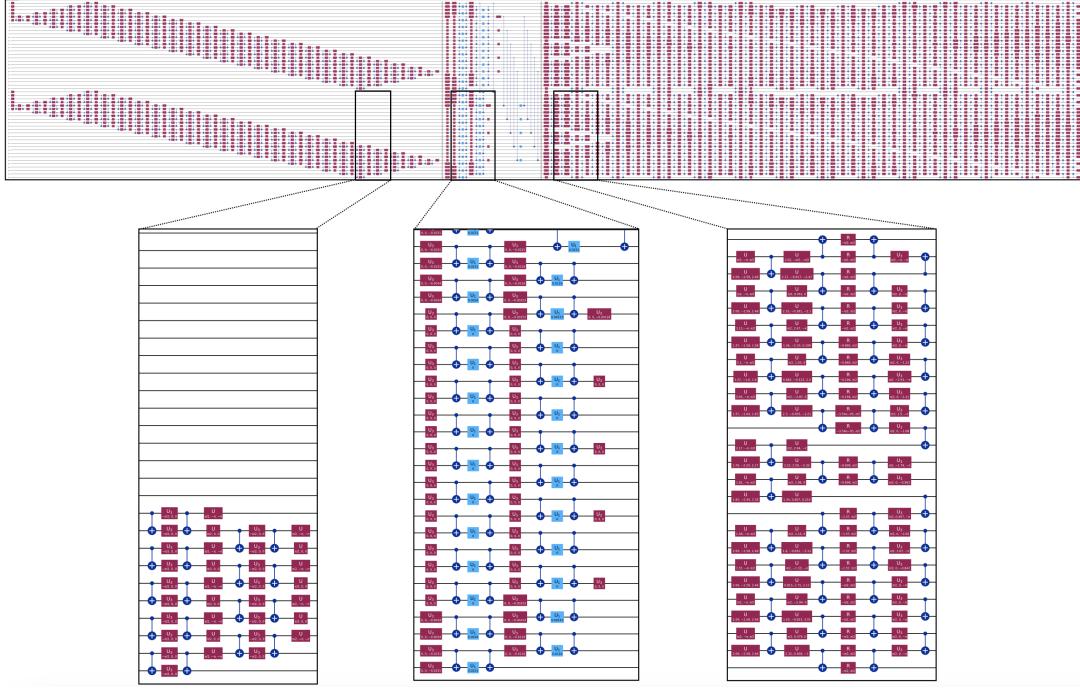


Figure S5. Top: LUCJ circuit used to simulate the ground state of N_2 /cc-pVDZ at the equilibrium bondlength, compiled into single-qubit (red, purple blocks) and CNOT (light blue symbols) gates. Bottom: zoomed-out views of gates from the first orbital rotation, density-density interaction, and second orbital rotation (left to right).

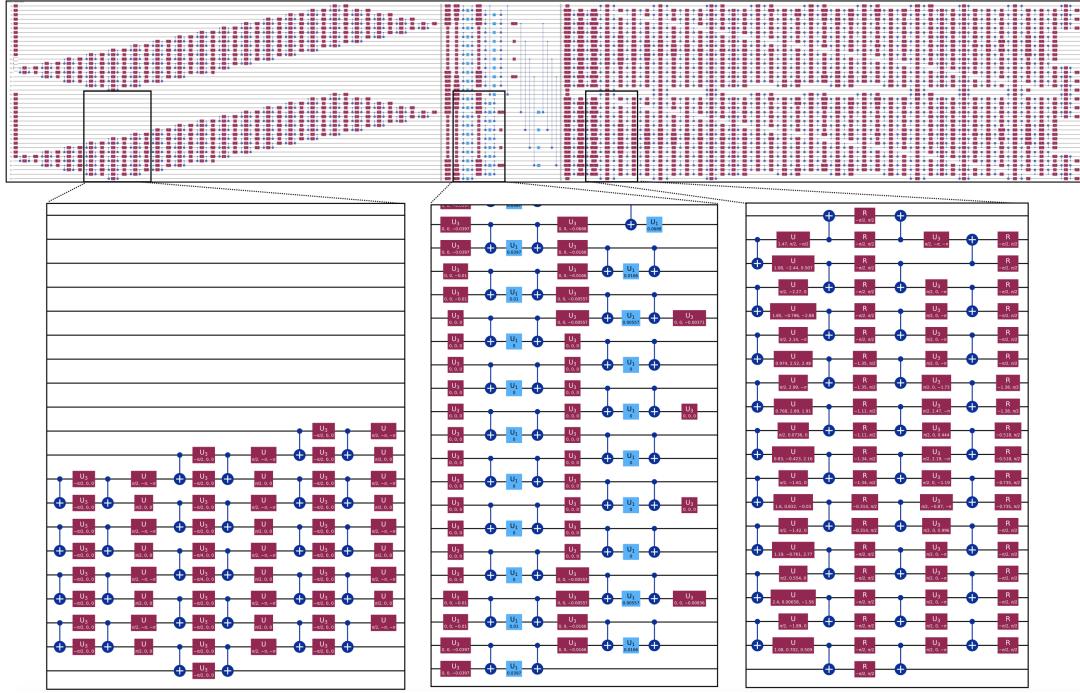


Figure S6. Top: LUCJ circuit used to simulate the ground state of the [2Fe-2S] cluster, compiled into single-qubit (red, purple blocks) and CNOT (light blue symbols) gates. Bottom: zoomed-out views of gates from the first orbital rotation, density-density interaction, and second orbital rotation (left to right).

circuits, we consider those with heavy-hex-like connectivity in the density-density interactions; see Eq. (S66). The circuits are optimized to minimize the estimator energy, as described in Sec. II A 3 using the COBYLA [70] optimizer.

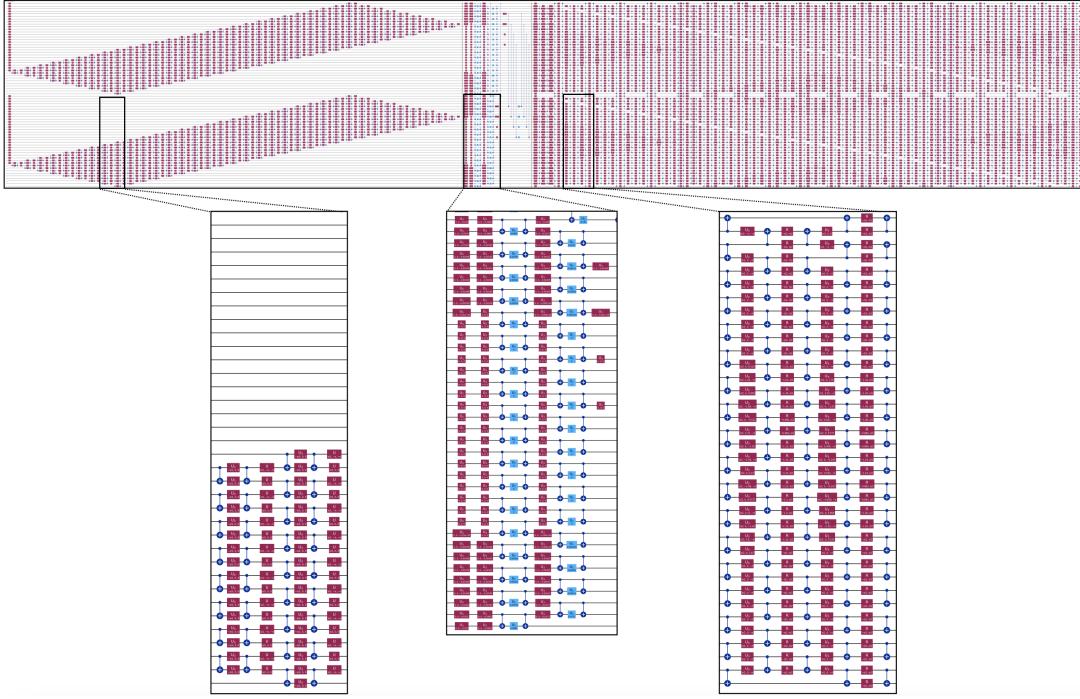


Figure S7. Top: LUCJ circuit used to simulate the ground state of the [4Fe-4S] cluster, compiled into single-qubit (red, purple blocks) and CNOT (light blue symbols) gates. Bottom: zoomed-out views of gates from the first orbital rotation, density-density interaction, and second orbital rotation (left to right).

At each optimization step we consider $K = 1$ batches of samples obtained from the set of sampled configurations \mathcal{X} , where $|\mathcal{X}| = 10^7$. The value of d at each point in the dissociation curve is chosen to coincide with the number of configurations resulting from a converged HCI calculation, carried out on PySCF, in the same system.

Fig. S8 (a) shows the potential energy surface obtained from the HPC quantum estimator run on configurations generated by LUCJ($L = 1$), LUCJ($L = 4$) and the determinant circuits. Panel (b) in Fig. S8 shows the energy error from the three different circuits. For most bond lengths, the error of the HPC quantum estimator is largest for the determinant circuit, showing that the density-density interactions play a crucial role for the generation of relevant electronic configurations. This claim is also supported by the decrease of the error (on average) when increasing the number of LUCJ layers. We note that the optimization of the LUCJ circuits takes more iterations to converge than the optimization of the determinant circuits.

B. Quantum and classical computational resources

1. Quantum experiment details

This section provides technical details about the experiments carried out on the quantum devices. The circuits that we use come from the truncation of two layers of the LUCJ circuits (as described in Sec. III A). For the quantum runs, the circuit parameters are not optimized and instead are obtained from a classical CCSD calculation, as described in Sec. III A. Note that the circuits themselves are still classically challenging to simulate.

Table II lists the different molecular species, together with the number of qubits used to encode the $|\Psi\rangle$ that generates samples of electronic configurations. The same table also provides details about the total circuit depths, and numbers of one- and two-qubit gates, as well as the IBM device that was used to run the experiments. The specific qubit layouts are also provided in Table II and depicted in Fig. S9.

In the experiments that study the dissociation of N_2 , the number of measurement outcomes collected is $|\tilde{\mathcal{X}}| = 1 \cdot 10^5$ and $9.8304 \cdot 10^4$ per point in the dissociation curve at 6-31G and cc-pVDZ level of theory respectively. In the experiments to study the ground-state properties of [2Fe-2S] and [4Fe-4S], the number of measurement outcomes collected is $|\tilde{\mathcal{X}}| = 2.4576 \cdot 10^6$. Out of the total number of measurement outcomes, a fraction of the configurations live

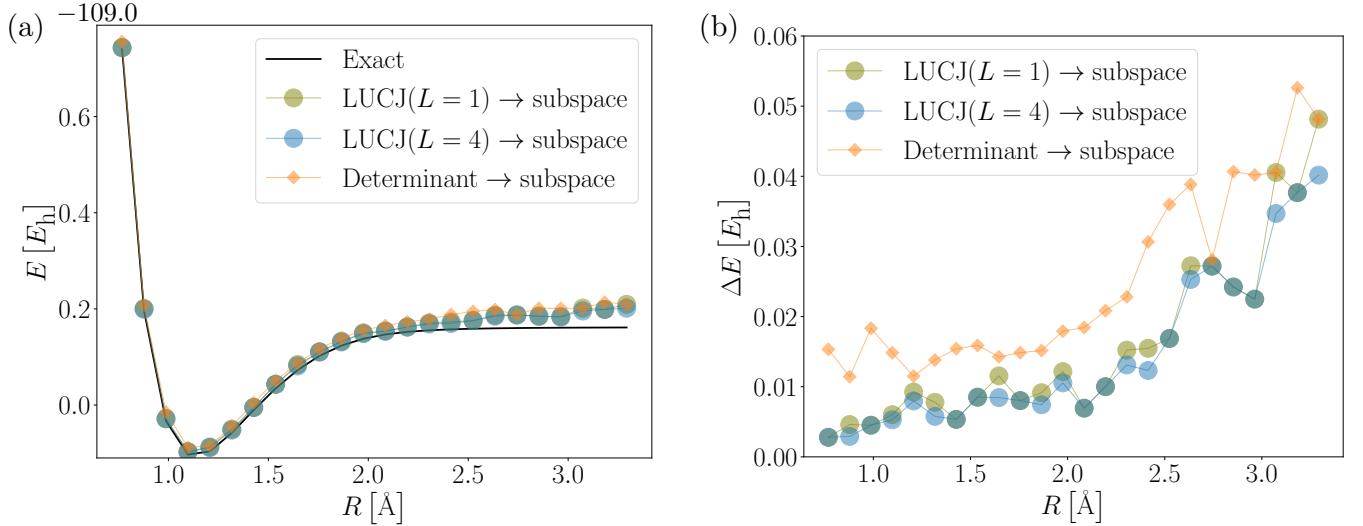


Figure S8. Numerical and noiseless study of the accuracy of the SCI-based HPC quantum estimator obtained from different levels of approximation of the LUCJ circuit and samples obtained from an Slater determinant (limit when the sampling from the LUCJ circuit can be efficiently simulated). The system considered is the N_2 dissociation in the 6-31G basis. The circuits are optimized to minimize the estimator energy. In the legend, L refers to the number of layers of the LUCJ ansatz. (a) ground-state energy as a function of the bond length. (b) Difference between the exact diagonalization energy and the HPC quantum estimator energy as a function of the bond length.

in the subspace of the Fock space with the correct particle number, defined as

$$p_N^{\text{hw}} = \frac{1}{N_s} \sum_{\ell=1}^{N_s} \delta_{N_{x_\ell \alpha}, N_\alpha} \delta_{N_{x_\ell \beta}, N_\beta}, \quad N_s = |\tilde{\mathcal{X}}|. \quad (\text{S73})$$

Table II shows the 95% confidence interval for that fraction. Recall that the measurement outcomes that correspond to configurations with the correct particle number are used in the setup phase of the configuration recovery procedure. Table II also shows the fraction of sampled configurations with the correct particle number if samples were collected from the uniform distribution over configurations of length M ,

$$p_N^{\text{unif}} = \binom{N_{\text{MO}}}{N_\alpha} \binom{N_{\text{MO}}}{N_\beta} 2^{-2N_{\text{MO}}}. \quad (\text{S74})$$

As shown in the Table, the two probabilities are statistically distinguishable for all the experiments we carried out. Our experiments used twirled readout error mitigation (ROEM) [103] to mitigate errors arising from qubit measurement, and dynamical decoupling (DD) [104–107] to mitigate errors arising from quantum gates. We employed the implementation of ROEM and DD available on the `Runtime` library of Qiskit [108], through the `Sampler` primitive. DD is implemented by sequences of X control pulses, whose effect is to protect qubits from decoherence due to low-frequency system-environment coupling. Here, we applied sequences of two X pulses (as in Ramsey echo experiments) to idle qubits.

2. Classical resource details

The SCI-based eigenstate solver used by HPC quantum estimator is the one implemented in PySCF [26, 63] for the N_2 (6-31G, cc-pVDZ), and [2Fe-2S] systems. For the [4Fe-4S] molecule we use the SCI-based eigenstate solver implemented in DICE [35, 38] because of its ability to run across multiple computer nodes.

For the [4Fe-4S] experiments, the eigenstate solver can be distributed across a number of classical nodes. The workflow for [4Fe-4S] is run on the supercomputer Fugaku. The supercomputer Fugaku has a total of 158,976 nodes with Armv8.2-A SVE 512 bit architecture. Each node has 82 cores and 32 GiB of memory. The largest calculation in this work uses 6400 nodes of the Fugaku supercomputer running in parallel.

system	q (JW, Tot.)	(d, CNOT, u)	device	layout	p_N^{hw} [95% c.i.]	p_N^{unif}	$ \tilde{\mathcal{X}} $
N_2 , (10e,16o)	(32, 36)	(148,762,1408)	ibm_nazca	[0→82]+[2→73]	(0.0069,0.0071)	0.0044	$1 \cdot 10^5$
N_2 , (10e,26o)	(52, 58)	(223,1792,3412)	ibm_torino	[0→121]+[2→127]	(0.0016,0.0017)	$9.6 \cdot 10^{-7}$	$9.8304 \cdot 10^4$
[2Fe-2S], (30e,20o)	(40, 45)	(173,1100,2070)	ibm_torino	[0→104]+[2→94]	(0.0044,0.0045)	0.00022	$2.4576 \cdot 10^6$
[4Fe-4S], (54e,36o)	(72, 77)	(301,3590,6980)	ibm_torino	[0→100]+[2→28]	($4.69 \cdot 10^{-5}$, $6.61 \cdot 10^{-5}$)	$1.88 \cdot 10^{-6}$	$2.4576 \cdot 10^6$

Table II. Details of hardware simulations. For each active space (column 1 from the left), we list the details of the LUCJ quantum circuit executed on hardware, specifically: its number of qubits (q) both for the Jordan-Wigner encoding (JW) and the total number of qubits used including auxiliary ones (Tot.) (column 2) and count of quantum operations (circuit depth d , number of CNOT and single-qubit u gates, column 3), the device used (column 4) and the qubit layout chosen (column 5, where $[a \rightarrow b]$ indicates the shortest path between qubits a and b in the device topology, also visible in Fig. S9). In columns 6 and 7 we report a 95% confidence interval for the fraction p_N^{hw} of configurations with the correct particle number and the corresponding value for configurations with uniform probability distribution (see the text for the definition of these quantities). In column 8, we list the number of noisy configurations sampled from quantum hardware.

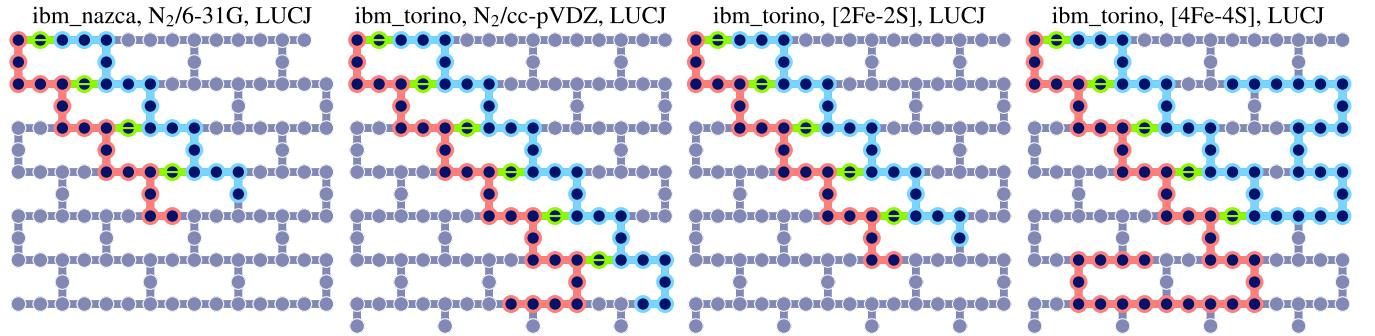


Figure S9. Left to right: schematics of the devices used to carry out experiments for the N_2 molecule with 6-31G basis, the N_2 molecule with cc-pVDZ basis, the [2Fe-2S] cluster, and the [4Fe-4S] cluster, using local unitary cluster Jastrow (LUCJ) quantum circuits. Qubits used in the calculation are shown in red (for qubits associated with α spin-orbitals), blue for qubits associated with β spin-orbitals), and green (for auxiliary qubits).

IV. EXPERIMENTAL RESULTS

This section contains the results from the experiments that are run on the quantum devices, as described in the previous section. We show a collection of investigations that are complementary to those shown in the main text.

We first visualize the prediction individual wave function components obtained by the HPC quantum estimator with configuration recovery in a system size where we have access to the exact ground state wave function amplitudes (N_2 in the 6-31G basis).

We then study the scaling and runtime of the estimator. In particular, we study the effect and cost that more intensive classical resources have on the quality of the estimator. We choose the N_2 (cc-pVDZ) and [2Fe-2S] molecules for this analysis.

We also study the effect of orbital optimizations in conjunction with the HPC quantum eigensolver in the quality of the predictions. We choose N_2 in the cc-pVDZ basis as the case of study.

The results reported in Figure 4 of the main text, for the [2Fe-2S] cluster show the resolution of three different eigenstates. In this section we provide additional information about their nature. In particular, we study the average orbital occupancy for the three eigenstates in the MO and localized bases.

Lastly, we investigate the amount of signal that can be extracted and amplified from our experiments. For this study we consider the dissociation of N_2 (cc-pVDZ) as well as the study of the low-energy spectrum of the [2Fe-2S] molecule. The same study for the [4Fe-4S] molecule is in the main text.

A. Visualization of the wave function amplitudes

The aim of this subsection is to show the accuracy of individual wavefunction amplitudes obtained with the HPC quantum estimator (with configuration recovery). Configuration recovery is run on measurement outcomes obtained from a quantum device. We consider the dissociation of N_2 in a basis set that is amenable to exact diagonalization

techniques but larger than a minimal basis set: the 6-31G basis. The exact diagonalization wavefunction serves as the reference to asses the accuracy of the amplitudes produced by the estimator. For each point of the dissociation curve we collect $|\tilde{\mathcal{X}}| = 100 \cdot 10^3$ measurement outcomes and run the configuration recovery procedure with $K = 10$ batches of samples. The configuration recovery method is run for ten self-consistent iterations. The value of the subspace dimension for the SCI eigenstate solver is $d = 4M$. For the potential energy surface we report $\min_k(E^{(k)})$. We show the amplitudes of the wavefunction from the batch of configurations whose energy is the lowest.

Panel (a) in Fig. S10 shows the potential energy surface of the molecule obtained by the estimator as well as the error in the estimation of the ground-state energy, as a function of the bond length. The error in the estimation of the ground-state energy remains below $10 mE_h$ for all points in the dissociation curve, thus showing good agreement with the exact results.

Panel (b) in Fig. S10 shows a comparison between the exact ground state wavefunction amplitudes and the wavefunction amplitudes obtained by the estimator, for different points in the dissociation curve. The agreement is exceptional for the larger wavefunction amplitudes and slowly deteriorates in the description of the tails of the wavefunction. We also observe that for larger bond lengths, where higher levels of static correlation are present, the agreement in the tails is worse than for bond lengths closer to the equilibrium geometry. This is to be expected for two reasons. First, for larger bond lengths the wavefunction amplitudes are less concentrated than close to equilibrium. As discussed in Sec. II B, the accuracy of the estimator using a SCI-based eigenstate solver deteriorates as the wavefunction becomes less concentrated. The second reason is that for bond lengths close to equilibrium, the wavefunction has a strong mean field character, but not for larger bond lengths. This directly impacts the shape of the reference spin-orbital occupancy \mathbf{n} that is used to recover configurations. When the wavefunction is mean-field dominated, \mathbf{n} takes the shape of a sharp step function shape with all spin-orbital average occupancies close to 0 or 1. One possible effect of static correlations is to soften the step function, moving average occupancies of some spin-orbitals away from 0 or 1. The recovery of configurations used in this work is more effective when the components of \mathbf{n} are close to 0 or 1.

In summary, the HPC quantum estimator with a SCI-based eigenstate solver and configuration recovery produces accurate representations of the ground state wavefunction (on experimental data), even at an individual amplitude level. Less concentrated wavefunctions tend to challenge the procedure more, both because the accuracy of the SCI solver is decreased, and because the configuration recovery becomes less effective due to softer profiles of \mathbf{n} .

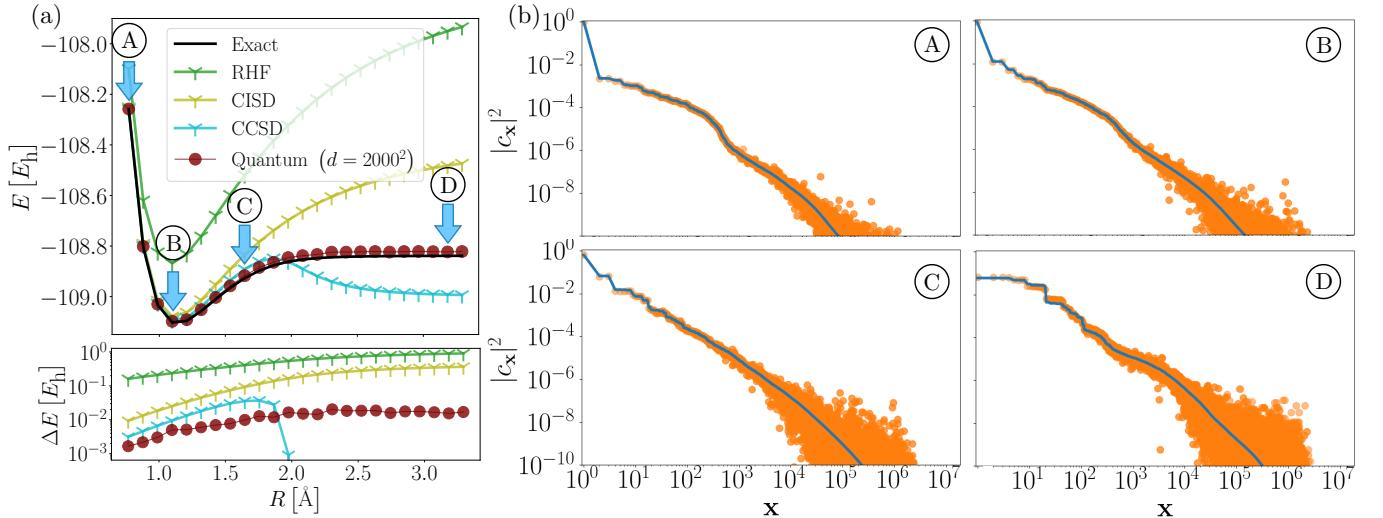


Figure S10. Visualization of the wavefunctions produced by the HPC quantum estimator in the dissociation of N_2 (6-31G), and the comparison against the exact wavefunctions. (a) Shows the energy and energy error as a function of the bond length for the estimator with configuration recovery run from measurement outcomes from the quantum device. Classical approximate methods are shown for reference. (b) Shows the wavefunction amplitudes $|c_x|^2$ for all possible electronic configurations \mathbf{x} . The configurations in the horizontal axis are ordered according to the magnitude of the amplitudes in the exact wavefunction. The blue solid lines correspond to the exact ground state amplitudes while the orange dots correspond to the wavefunction amplitudes obtained from the estimator with configuration recovery. Different points in the dissociation curve are shown and labelled A, B, C, D, following the same encoding as in panel (a).

B. Scaling and runtime

We begin this section by studying the improvement of the accuracy of the HPC quantum estimator with the number d of determinants used for the subspace expansion and diagonalization. It is clear that larger d yields more expressivity in the variational wavefunction, since we are increasing the possible size of the support of the state. It is also clear that increasing d increases the HPC quantum estimator runtime since both the projection and diagonalization processes become more computationally intensive. To show the improvement of the accuracy with d we consider the dissociation of N₂ (cc-pVDZ), where the measurement outcomes are generated by the quantum device. The estimator is applied to 10⁵ measurement outcomes and we take $K = 10$ batches of samples and report $\min_k(E^{(k)})$ for the potential energy surface for different values of d . Ten self-consistent iterations are considered for the recovery of configurations.

Fig. S11 shows the potential energy surfaces obtained by the estimator with different values of d , the number of determinants used for the subspace expansion and diagonalization. Increasing the value of d improves the results both qualitatively and quantitatively. For the smaller values of d , the potential energy surface shows unphysical oscillations when the bond length is large. For larger values of d , not only the energy values are lower, but also the oscillations in the energy as a function of the bond length are significantly decreased in amplitude. Improvement of the ground state approximation with an increasing amount of computational resources is also observed in the energy-variance analyses of the main text (Fig. 4), and the supplementary materials (Figs. S20 and S21).

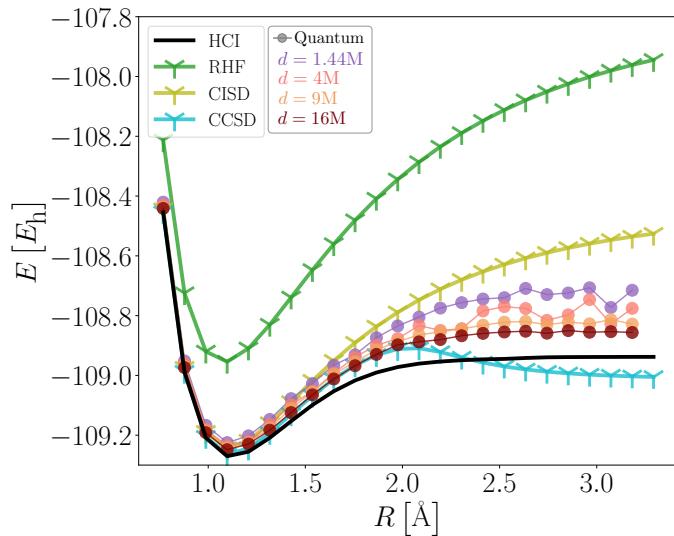


Figure S11. Improvement of the estimator accuracy (with configuration recovery) by increasing the number of determinants for the subspace expansion and diagonalization. Potential energy surface of the N₂ molecule (cc-pVDZ) obtained from the HPC quantum estimator estimator using measurements outcomes from the quantum device. The energies reported correspond to the lowest energy amongst the K batches of configurations, i.e. $\min_k(E^{(k)})$. Different subspace dimensions d are considered, as indicated in the legend. Different classical methods are shown for reference.

We analyze the vertical and horizontal scaling of the workflow, in the case where the subspace diagonalization is run on a single computer node, using the solver provided in the PySCF library. The [2Fe-2S] iron-sulfur cluster is used as the system of reference. Fig. S12 shows vertical scaling (increasing resources within a single node) of the eigenstate solver, and relative runtimes and schematics of horizontal scaling (adding nodes to cluster) of different blocks of the estimator (with configuration recovery). The eigenstate solver contributes the most to total runtime of the algorithm. Vertical scaling of the eigenstate solver shows noticeable decrease in runtime up to a point. Experiments showed that for [2Fe-2S], 30+ CPUs per node is the optimal configuration. A good strategy for classical scaling would be to find the optimal vertical configuration and then scale horizontally for better utilization of resources.

To conclude this section, we study the dependence of the runtime and accuracy of the estimator in the [4Fe-4S] system as a function of the amount of classical compute. The classical compute can be increased by increasing the size of the subspace dimension d , or by collecting more batches of configurations K . The DICE library allows to distribute the subspace diagonalization across a number of independent computer nodes. It is expected that more classical nodes will yield faster runtimes per diagonalization. For this experiment the Heron device Montecarlo is used, with $|\mathcal{X}| = 3163742$ collected measurement outcomes, and $p_N^{hw} = 5.4 \cdot 10^{-4}$.

Fig. S13 (a) shows the runtime of the diagonalization of a single batch of configurations as a function of the number

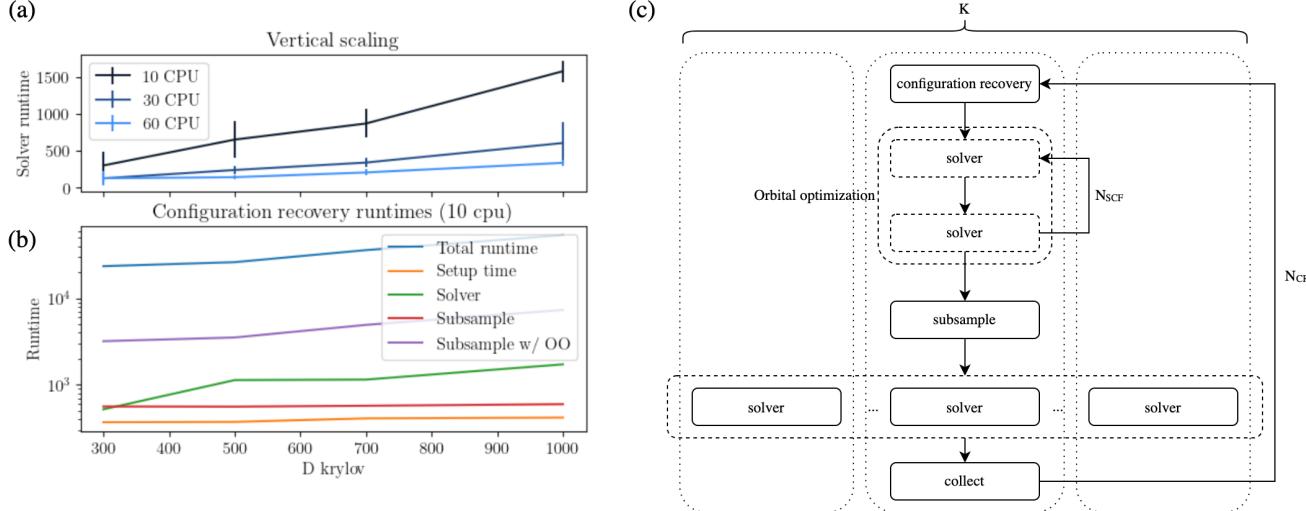


Figure S12. Classical-resource scaling of the HPC estimator with configuration recovery and orbital optimizations. (a) Vertical scaling of the solver (increase number of CPUs per node) for [2Fe-2S]. (b) Relative runtimes of blocks of the estimator for [2Fe-2S]. (c) Horizontal scaling of the estimator. Total runtime without horizontal scaling is $S \cdot (N_{SCF} + K) \cdot N_{CR}$. The total runtime with horizontal scaling is $S \cdot (N_{SCF} + 1) \cdot N_{CR}$, where S is the runtime of the solver, K is the number of nodes in a cluster, N_{CR} is the number of iterations, and N_{SCF} is the number of orbital optimization steps.

of classical nodes. Different subspace dimensions are considered. As expected, with an increased number of nodes, the runtime of the diagonalization decreases. After a certain number of nodes, the speedup saturates, due to the overhead in inter-node communications. This panel allows us to identify the number of classical nodes necessary to have a fixed runtime for the diagonalization for different subspace dimensions. To study improvement of the accuracy of the estimator with the number of batches of configurations K , we run the workflow with $K = 100$. From the $K = 100$, batches, we subsample the energy of smaller groups of batches and collect the minimum energy for each subsample. The subsampled minimum energies are averaged. Different sizes of the smaller groups of batches are considered until reaching the group of $K = 100$ batches of configurations. Panel (b) in Fig. S13 shows the average of the minimum energies as a function of the total number of nodes for different values of d , such that the runtime of each digonalization is fixed to 1.44 hours. The total number of classical nodes is given by the product of the size of the groups of batches of configurations and the number of nodes required to obtain a runtime of 1.44 hours per subspace projection and diagonalization.

C. Effect of orbital optimizations in the dissociation of N₂ (cc-pVDZ)

We apply orbital optimizations (OO) to improve the accuracy of the HPC quantum estimator, in the study of the dissociation of N₂ (cc-pVDZ). The estimator with configuration recovery and its orbital-optimized counterpart are applied to the same set of noisy measurement outcomes $\tilde{\mathcal{X}}$ obtained from the quantum device. We use $K = 10$ batches of samples in each case and $N_{SCF} = 10$ iterations of the alternation between the optimization of κ and running of the eigenstate solver to update $c_x^{(k)}$. After each $c_x^{(k)}$ update, 5000 iterations of gradient descent with momentum are used to optimize κ . Before applying orbital optimizations, the configuration recovery procedure is run for 10 iterations on the reference basis (molecular orbitals). Various sizes d of the batches are considered.

Fig. S14 compares the ground-state energy obtained by the estimator without and with orbital optimizations. We observe that close to the equilibrium bond length the effect of the orbital optimizations is negligible. However, upon dissociation, the estimator with orbital optimizations obtains noticeably lower ground state energies. For smaller values of d , the estimator without orbital optimizations shows unphysical oscillations in the potential energy surface. The orbital optimizations decrease the amplitude of those oscillations.

In summary, orbital optimizations allow finding the single-particle basis in which the estimator is most accurate, thus improving the quality of its predictions. Furthermore, the optimization of the circuit parameters θ would allow the circuit to respond to the change of basis to produce electronic configurations better suited for the new basis. As mentioned earlier, the optimization of the circuit parameters will be the subject of follow-up projects. Note that the

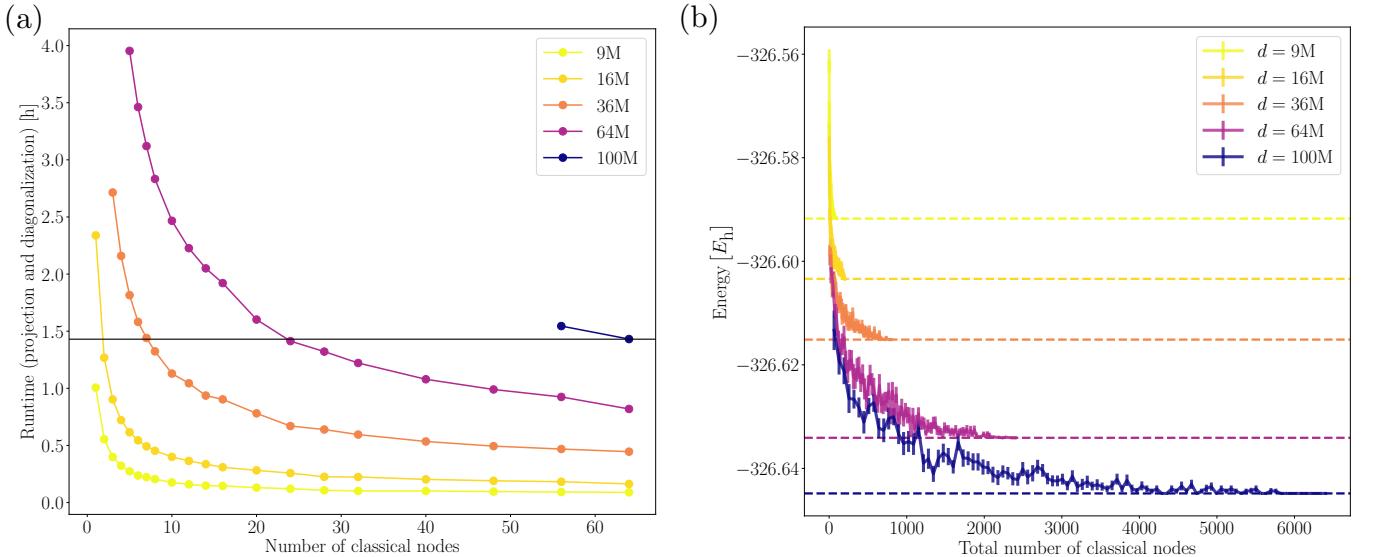


Figure S13. Accuracy and runtime of the HPC quantum estimator for the [4Fe-4S] iron-sulfur cluster. (a) Runtime of a single subspace projection and diagonalization as a function of the number of nodes that are used to distribute the calculation. Different colors correspond to different subspace dimensions. The smallest number of nodes per curve corresponds to the limit where the calculation runs out of memory on a single node. The subspace projection and diagonalization are handled by the DICE library. The horizontal line shows a fixed runtime of 1.44 hours for the different subspace sizes. (b) Statistical convergence of the HPC quantum estimator as the number of batches of samples K is increased. The vertical axis shows the the average minimum energy obtained by a fixed number of batches of samples. Each batch of samples is run on fixed number of nodes and the diagonalization for the different batches are run in parallel. The number of nodes for the diagonalization of a single batch is chosen to obtain a fixed runtime of 1.44 hours for all of the different subspace dimensions d . The horizontal axis shows the total number of nodes corresponding to different numbers of batches. The largest number of batches corresponds to $K = 100$.

unrestricted optimization of κ can break symmetries that may be desirable to preserve, so it is up to the user to decide whether to use orbital optimizations or not.

D. Orbital occupation numbers for the low-energy spectrum of [2Fe-2S]

The calculations reported in Figure 4 of the main text, for the [2Fe-2S] cluster, identified three distinct eigenstates that we labeled A, B, and C. In this section, we provide additional information about the nature of the three identified eigenstates. The purpose of this analysis is to illustrate the robust classification of the wavefunctions produced by our estimator into three families (A, B, and C) but also to shed light on the nature of the corresponding eigenstates and the ability of our method to accurately capture a multireference character in electronic eigenstates, pinpointing areas of possible future improvement. In particular, we show the spatial orbital occupancy for the three eigenstates that the HPC quantum estimator finds for [2Fe-2S] both in the MO basis and in a basis of spatially local orbitals. The MO basis helps highlight deviations from the mean-field states or lack thereof, and the basis of localized orbitals (identifiable with e.g. Fe 3d and S 2p orbitals) helps characterize our method's ability to capture local antiferromagnetic correlations between transition-metal atoms.

The spatial orbital occupancy is given by $n_p = \sum_{\sigma} n_{p\sigma}$. Therefore $n_p \in [0, 2]$. For Eigenstate A, the occupancy comes from the estimator run with $d = 9M$. For Eigenstates B and C, the occupancy comes from the estimator run with $d = 16M$. We cannot compare the occupancy for the same value of d since the description of eigenstates A and B require different numbers of configurations (see Sec. IV E 2 and Fig. S21). For Eigenstates B and C, we use the same number of determinants to make the comparison as even as possible. The number of batches of configurations is $K = 10$.

Fig. S15 shows the n_p profile for the three eigenstates identified from the energy-variance analysis, in the basis of MOs. The occupancy profile for Eigenstate A is the typical profile of an eigenstate with a strong mean-field character. The occupancy of the lowest-energy $N_{\sigma} = 15$ molecular orbitals is close to its maximum value, while the remaining orbitals have lower occupancy. The character of the occupancy profile for Eigenstate B is different from that of Eigenstate A. The occupancy of the third and fourth orbitals is depleted to give a more prevalent presence of

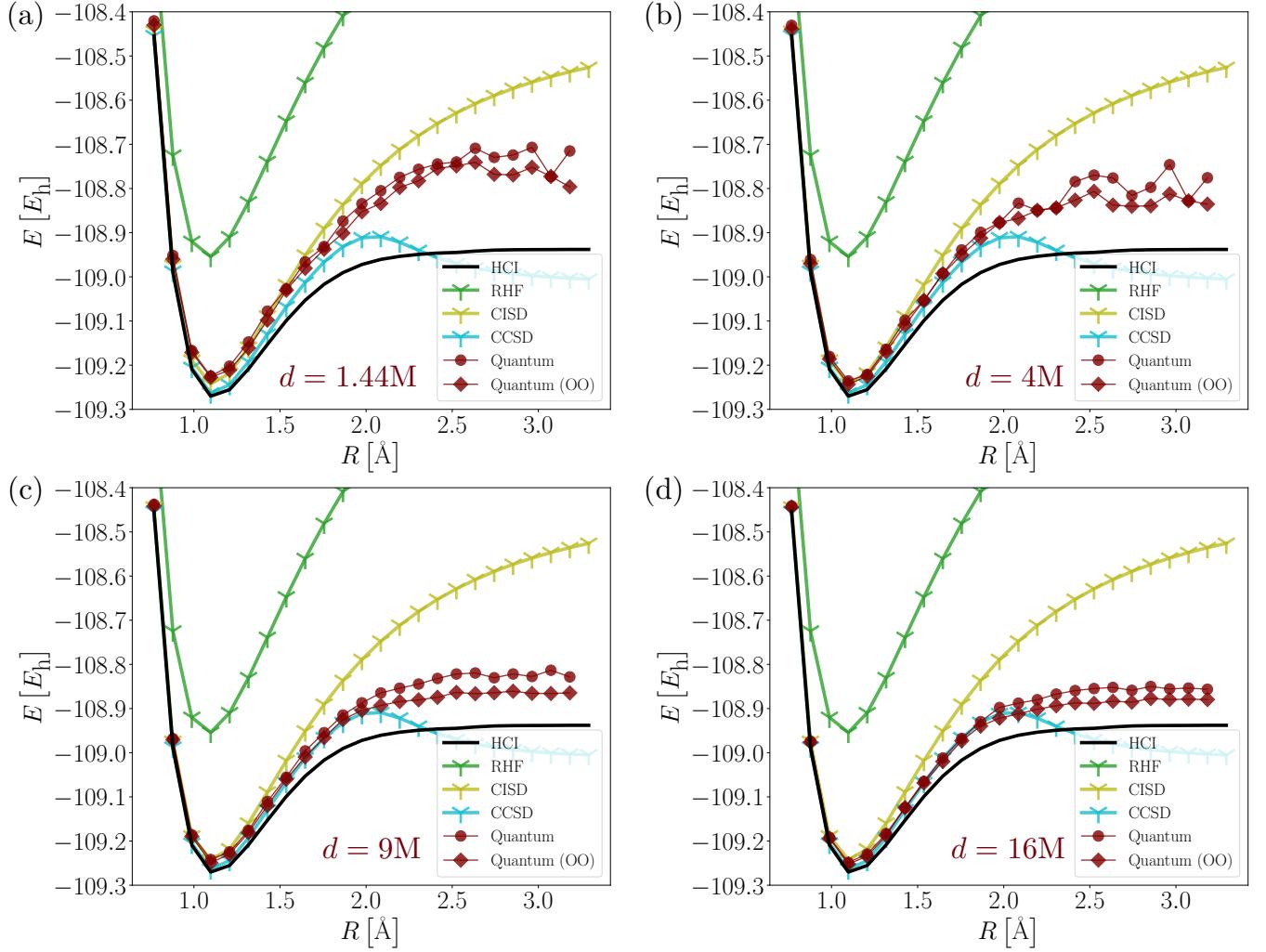


Figure S14. Effect of orbital optimizations (OO) in the performance of the HPC quantum estimator with configuration recovery in the dissociation of N_2 (cc-pVQZ). The ground-state energy is shown as a function of the bondlength. The estimator is run on measurement outcomes from a quantum device. The energies shown correspond to the lowest energy amongst the K batches of configurations, i.e. $\min_k (E^{(k)})$. Energies from different classical methods are shown for reference, as indicated in the legend. Results are shown for the estimator run with different numbers of configurations d , as indicated in each panel (a)-(d).

electrons in orbitals $p = 16$ and $p = 19$. This effect is even more noticeable for eigenstate C, where the occupancy of the third and fourth orbitals is depleted by half and the occupancy of orbital $p = 16$ is now over the half-filling factor.

To further understand the nature of Eigenstates B and C, we study the occupancy in the basis of localized orbitals $\tilde{\mathbf{n}}$ (see Sec. I B for details about the basis of localized orbitals). The reported occupancies are constructed as follows: first, we obtain the one-body density matrix for the approximate ground state obtained from each batch of configurations, expressed in the basis of MOs: $\Gamma_{pq,\sigma}^{(k)}$ (see Eq. S40). We then rotate the density matrix to the basis of localized orbitals with a similarity transformation denoted by Ω ,

$$\tilde{\Gamma}_{tu,\sigma}^{(k)} = \Gamma_{pq,\sigma}^{(k)} \Omega_{pt} \Omega_{qu}, \quad (\text{S75})$$

where $\tilde{\Gamma}_{tu,\sigma}^{(k)}$ is the one-body density matrix in the basis of localized orbitals. The occupancy of localized orbital t is given by

$$\tilde{n}_t = \frac{1}{K} \sum_{k=1}^K \sum_{\sigma} \tilde{\Gamma}_{tt,\sigma}^{(k)}. \quad (\text{S76})$$

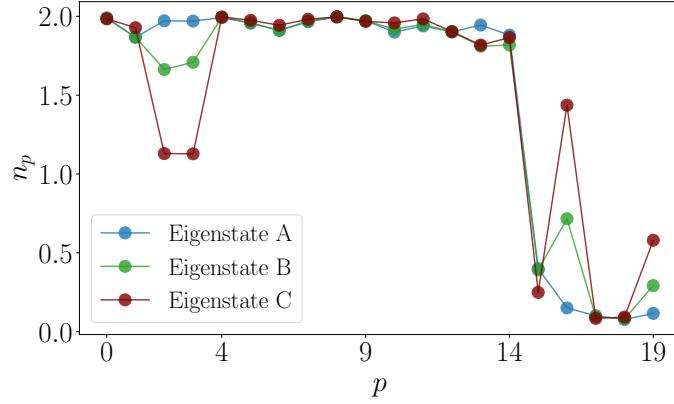


Figure S15. Orbital occupancy profile \mathbf{n} for the [2Fe-2S], produced by the HPC quantum estimator with configuration recovery. The increasing index p labels molecular orbitals of increasing energy. The occupancy profile is shown for the three eigenstates identified in the energy-variance analysis (see Fig. 4 (b) in the main text).

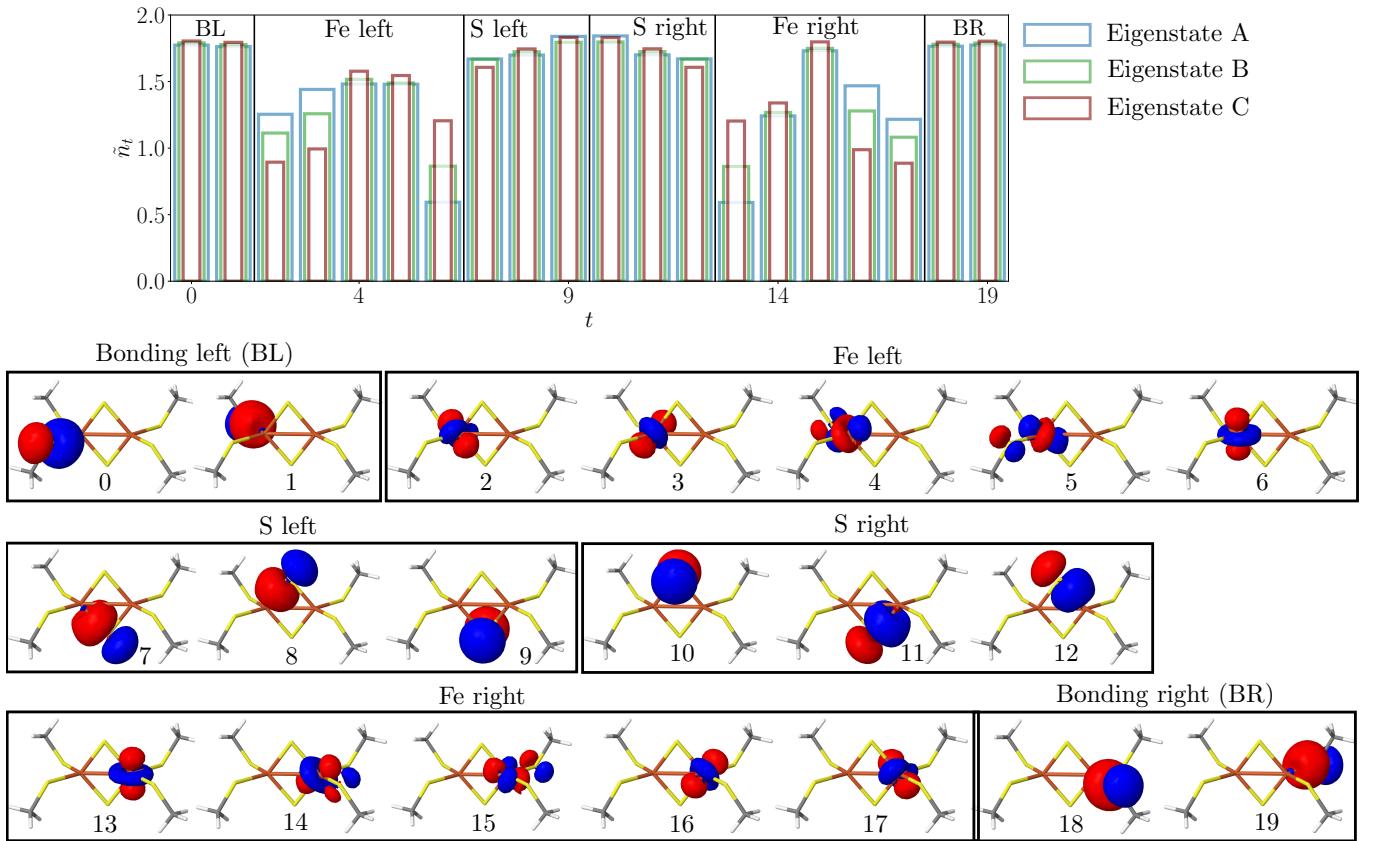


Figure S16. Orbital occupancy profile $\tilde{\mathbf{n}}$ for the [2Fe-2S], produced by the HPC quantum estimator with configuration recovery. The increasing index t labels localized orbitals as shown in the bottom of the panel. The occupancy profile is shown for the three eigenstates identified in the energy-variance analysis (see Fig. 4 (b) in the main text). The representation of the orbitals is obtained from the repository [109] from Ref. [45].

Fig. S16 shows a comparison of the value of $\tilde{\mathbf{n}}$ for Eigenstates A, B, and C. The reflection symmetry along the plane that separates the left and right sides of the molecule is broken for the three approximate eigenstates. This is highlighted by the observation that the occupancy of the equivalent orbitals for the two Fe atoms are not the same. We also observe that the occupancy of the bonding orbitals of the SCH_3 groups and the occupancy of the $3p$ orbitals of the S atoms is similar for the three eigenstates. The occupancy of the $3d$ orbitals of Fe has the largest differences

among the three different eigenstates. The more notable discrepancy in the d-orbitals of Fe may be a consequence of the fact that the low-energy physics of this molecule is dominated by the antiferromagnetic coupling of the electrons in the Fe 3d orbitals [4, 45].

E. Quantum signal in the experiments

In this section we assess the quality of the quantum signal for all our experiments, going beyond the [4Fe-4S] analysis presented in the main text.. Given the large circuit sizes of our experiments:

- [2Fe-2S]: 40 qubits (not including auxiliary), 1100 two-qubit gates, 3170 total gates,
- N_2 : 52 qubits (not including auxiliary), 1792 two-qubit gates, 5204 total gates,
- [4Fe-4S]: 72 qubits (not including auxiliary), 3590 two-qubit gates, 10570 total gates,

one should verify whether there is a useful signal coming out of the quantum circuits, comparing samples from \tilde{P}_Ψ and uniform random samples. We first investigate the fraction of sampled configurations that live in the correct particle sector p_s^{hw} and compare it to the fraction that would be obtained if the samples came from the uniform distribution p_s^{unif} . The values for p_s^{hw} alongside their 95% confidence interval, and p_s^{unif} are reported in Table II. For the 40-, 52-, and 72-qubit experiments, we observe that p_s^{hw} is ~ 20 , $\sim 10^3$, and ~ 40 times larger than p_s^{unif} respectively. This provides a first indication that configurations sampled from the quantum device contain a signal that is distinguishable from white noise.

The second set of tests that we conduct is the comparison of the ground-state energy obtained by the HPC quantum estimator with and without applying the configuration recovery procedure. Note that not applying configuration recovery makes the estimator equivalent to the QSCI framework [30]. Having reasonable energies (lower than or in the vicinity of the RHF energy) without applying configuration recovery reveals that the set \mathcal{X}_N already contains reasonable configurations to construct a non-random ground state.

The third set of tests involves comparing the performance of the HPC quantum estimator with configuration recovery run on configurations obtained from the quantum device and samples obtained from the uniform distribution. We consider the uniform distribution over all possible configurations in the Fock space, as well as the uniform distribution over configurations with the correct particle number, as in Fig. 4 (c) in the main text [110]. The results for these last two sets of tests are shown in the coming subsections for the 40-, 52- and 72-qubit experiments.

1. N_2 experiments

a. Effect of configuration recovery on the accuracy of the HPC quantum estimator.— In this section, we compare the accuracy of the estimator without configuration recovery (applied to $\mathcal{X}_N \subset \tilde{\mathcal{X}}$) to the accuracy of the estimator with configuration recovery (applied to $\tilde{\mathcal{X}}$). As for all N_2 experiments, $|\tilde{\mathcal{X}}| = 100 \cdot 10^3$. From the experimental measurements, we observe that the average (over the bondlengths considered in this study) size of the \mathcal{X}_N set is $|\mathcal{X}_N| = 165$. The maximum number of determinants that can be extracted to perform the projection and diagonalization is $d_N^{\text{hw}} = 110K$. Since typical values of d used for the N_2 molecule exceed d_N^{hw} we perform the projection and diagonalization using all the configurations obtained from \mathcal{X}_N (number of batches of configurations is $K = 1$). For the estimator run with configuration recovery we use $K = 10$. In both cases, we report the $\min_k(E^{(k)})$ energy for the potential energy surface. When running the estimator with configuration recovery we use 10 self-consistent recovery iterations.

Fig. S17 shows the comparison between the HPC quantum estimator run with and without configuration recovery. The ground-state energy is shown as a function of the bond length. The estimator run on \mathcal{X}_N , i.e. without configuration recovery, obtains energies comparable to RHF energies. A level of accuracy comparable to RHF is not desirable for correlated electronic structure methods. However, since no mitigation techniques have been used in the large circuits producing the samples, this is an indication that a useful quantum signal (mostly of mean-field character) is present in $\tilde{\mathcal{X}}$. As expected, the estimator with configuration recovery shows higher levels of accuracy as compared to not using configuration recovery, obtaining a qualitatively correct dissociation curve upon increasing d .

The accuracy of the energies obtained by the HPC quantum estimator without configuration recovery could be improved by obtaining more measurement outcomes and thus more configurations on the right particle sector. The typical values of d used for the estimator with configuration recovery are larger than d_N^{hw} . With current noise rates, obtaining a larger size of $|\mathcal{X}_N|$ to run the estimator without configuration recovery, requires the collection of many more measurement outcomes, at an efficiency where approximately only 1 in 1000 measured configurations has the right particle number (see Tab II for the precise fraction). Configuration recovery allows improving the accuracy of

the HPC quantum estimator without the collection of more measurement outcomes from the quantum device. This is desirable as it reduces the overall cost of the calculation.

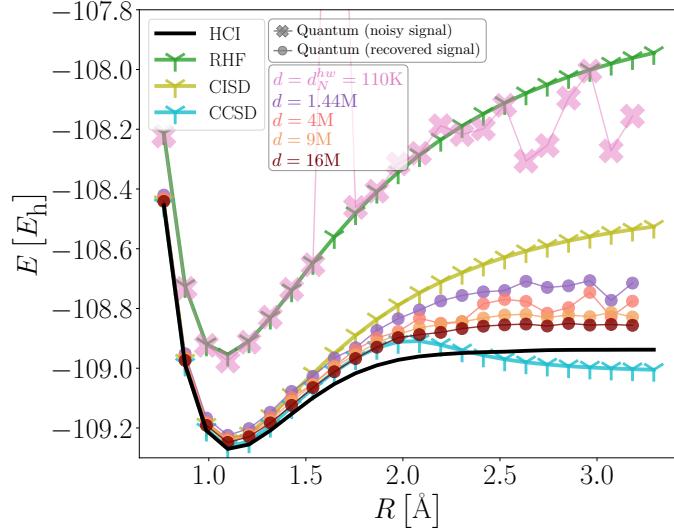


Figure S17. Comparison of the performance of the HPC quantum estimator, with and without configuration recovery, run on the same set of measurement outcomes from the quantum device. The dissociation of N_2 (cc-pVDZ) is considered. We compare the performance of the estimator with no configuration recovery (crosses) to the performance of the estimator with configuration recovery (points). Different numbers of configurations d are considered, labeled by different colors. d_N^{hw} refers to the number of configurations on the right particle sector of N electrons extracted from the measurement outcomes, used for the estimator without configuration recovery (see the text for more details). We note that with current noise rates, $d_N^{hw} < 1.44 \text{ M}$. Different classical methods are shown for reference, as indicated in the legend.

b. Comparison of the quality of the estimator with configuration recovery, applied to quantum data and configurations drawn from the uniform distribution.— We first compare the performance of the estimator with configuration recovery applied to measurement outcomes from the quantum device versus its application to configurations sampled from the uniform distribution in the right particle sector. The estimator is applied to 10^5 samples in both scenarios and we use the same value of d . In both cases, we take $K = 10$ batches of samples and report $\min_k (E^{(k)})$ for the potential energy surface. When running the estimator with configuration recovery we use 10 self-consistent iterations. Recall that the configuration recovery procedure recovers configurations that have the wrong particle number due to noise, thus the application of the estimator to configurations sampled from the uniform distribution in the right particle sector is equivalent to applying the SCI eigenstate solver to the random configurations. Panel (a) in Fig. S18 shows the comparison. We observe that the potential energy surface obtained from the measurement outcomes from the quantum device is both qualitatively and quantitatively more accurate than the one obtained from the uniform distribution over configurations in the right particle sector. The potential energy surface obtained from the configurations sampled from the uniform distribution shows a very prominent non-smooth behavior with energies worse than the RHF energies for a number of bond lengths.

The comparison with samples obtained from the configurations uniformly sampled over the full Fock space and the quantum samples is shown in Fig. S18 (b). We remark that 10^5 samples over the uniform distribution in the Fock space of electronic configurations are unlikely to reveal any configurations in the correct particle sector, since the probability of a uniformly random bitstring falling in the correct particle sector is $(\frac{N_{\text{MO}}}{N_{\uparrow}})(\frac{N_{\text{MO}}}{N_{\downarrow}})/2^M = 9.6 \cdot 10^{-7} < 10^{-5}$. Recall that the configurations on the right particle sector are required to start the configuration recovery procedure by providing the initial average occupancies. Therefore, we take $50 \cdot 10^6$ configurations sampled from the uniform distribution over configurations of length M for this comparison. Note that the use of more samples from the uniform distribution as compared to the number of measurement outcomes from quantum constitutes a comparison that favors the uniform distribution scenario.

In spite of this, Fig. S18 (b) shows a higher level of accuracy obtained by the estimator with configuration recovery run using measurement outcomes from the quantum device. These results indicate that the quantum device has a signal that is amplified by the configuration recovery procedure and used by the HPC quantum estimator to produce more accurate ground-state representations than running the same estimator on uniform noise.

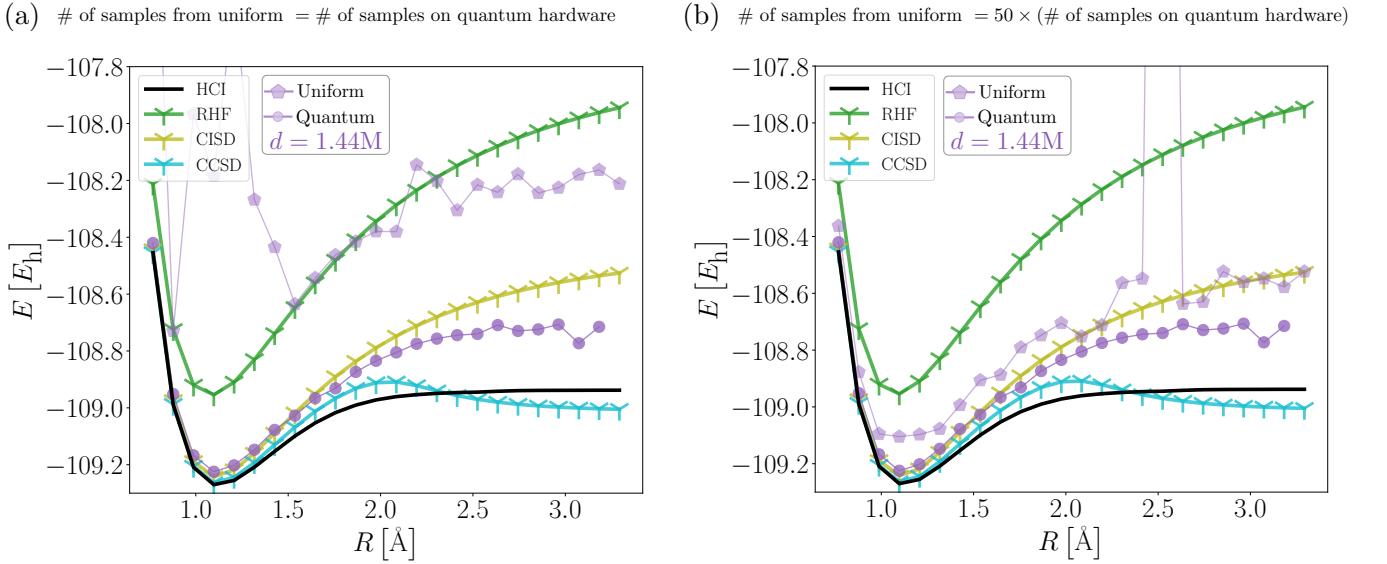


Figure S18. Comparison of the estimator with configuration recovery applied to measurement outcomes from the quantum device and configurations drawn from the uniform distribution, in the dissociation of N_2 (cc-pVDZ). Pentagons correspond to the energies obtained with the configurations sampled from the uniform distribution while dots correspond to the energies obtained from configurations sampled from a quantum device. Different classical methods are shown for reference, as indicated in the legend. (a) Uniform samples of configurations in the right particle sector. (b) Uniform samples of configurations over Fock space.

2. [2Fe-2S] experiments

a. *Effect of configuration recovery on the accuracy of the HPC quantum estimator.*— In this section, we compare the accuracy of the estimator without configuration recovery (i.e., applied to $\mathcal{X}_N \subset \tilde{\mathcal{X}}$) to the accuracy of the estimator with configuration recovery (applied to $\tilde{\mathcal{X}}$). As for all [2Fe-2S] experiments, $|\tilde{\mathcal{X}}| = 2.4576 \cdot 10^6$. In both cases, we take $K = 10$ batches of samples. When running the configuration recovery procedure, we use 10 self-consistent recovery iterations.

Fig. S19 shows the *Kernel Density Estimation* (KDE) of the $E^{(k)}$ distribution running the HPC quantum estimator both with and without configuration recovery for different values of d . The qualitative behavior of the KDE for the estimator with and without configuration recovery is similar. Upon increasing the value of d , the distributions shift towards lower values of the energy. However, the distributions of $E^{(k)}$ for the estimator with configuration recovery have lower energy than those without. Without configuration recovery, the best ground-state energy obtained is between the RHF and CISD energies, whereas when using the configuration recovery procedure, the energy is decreased even below the CCSD estimate. Again, note that the fact that the estimator without configuration recovery still has energies below RHF is a promising indication that the raw data from the device contains a signal that the configuration recovery procedure can amplify.

b. *Comparison of the quality of the estimator with configuration recovery, applied to quantum data and configurations drawn from the uniform distribution.*— We first compare the performance of the HPC quantum estimator with configuration recovery applied to measurement outcomes from the quantum device versus configurations drawn from the uniform distribution over configurations in the right particle sector. In both cases we consider $K = 10$ batches of electronic configurations and the estimator is applied to $2.4576 \cdot 10^6$ sampled configurations. The self-consistent recovery procedure is warm-started from the orbital occupancy vector \mathbf{n} obtained after running ten self-consistent iterations with $d = 250\text{K}$. Since the configuration recovery is warm-started, for $d > 250\text{K}$ we apply two self-consistent iterations. In both scenarios, the same values of d are considered. Recall that configuration recovery procedure recovers configurations that have the wrong particle number due to noise, thus the application of the estimator to configurations sampled from the uniform distribution in the right particle sector is equivalent to applying the SCI eigenstate solver to the random configurations. Fig. S20 shows a comparison of the energy-variance analysis obtained from the estimator on measurement outcomes from the quantum device and on uniformly random configurations with the right particle number. When the configurations come from measurement outcomes from a quantum device, the energy-variance analysis shows two clusters of points following a linear relation. The two clusters correspond to two eigenstates of the Hamiltonian. The extrapolated energy is in good agreement with the energy extrapolation from

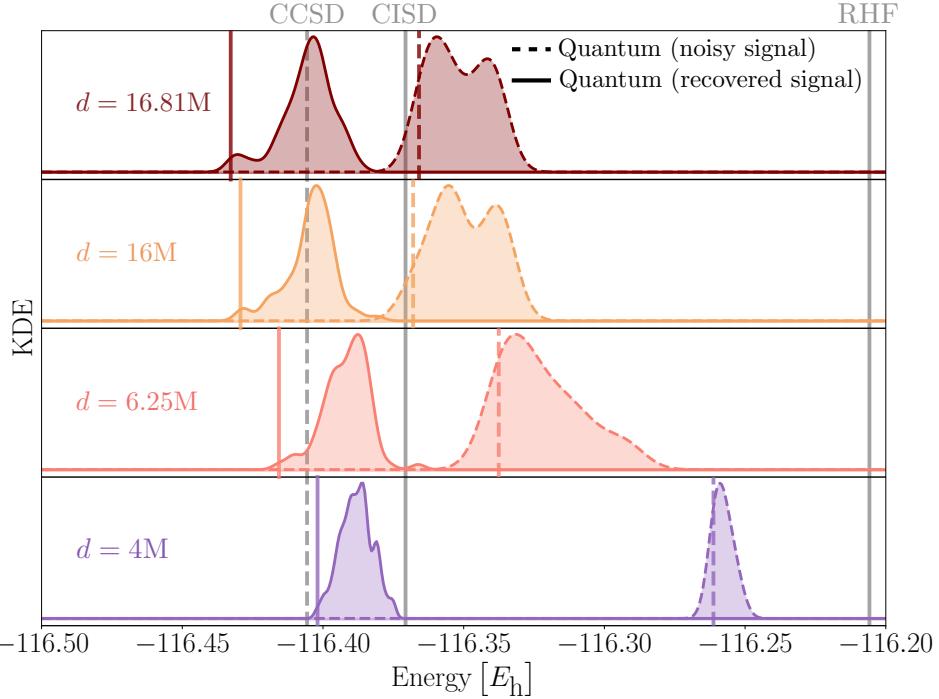


Figure S19. Comparison of the performance of the HPC quantum estimator, with and without configuration recovery, run on the same set of measurement outcomes for the ground state of [2Fe-2S]. The *kernel density estimation* (KDE) of the distribution of ground state energies obtained from the different batches of configurations is shown. The performance of the estimator with no configuration recovery (dashed lines) is compared to the performance of the estimator with configuration recovery (solid lines). Different panels correspond to different subspace dimensions d , as indicated in each panel. In each panel, the solid and dashed vertical lines matching the color of the KDE shows the lowest value of the energy amongst the different batches of configurations. The grey vertical lines show the value of the ground-state energy obtained from different classical methods, as indicated at the top of the panel.

HCI calculations. The energy-variance analysis on the estimator run on the random configurations with the right particle number shows a different behavior. The energy-variance pairs are not clustered as clearly as in the previous scenario. For the largest values of d considered, we observe that a first cluster begins to form, which coincides with the eigenstate of the highest energy that both HCI and the HPC quantum estimator applied to quantum measurement outcomes reveal. However, for these values of d the HPC estimator run on random samples is not capable of revealing the second eigenstate of lower energy that the estimator run on quantum data obtains.

We also compare the performance of the HPC estimator with configuration recovery run on the data obtained from the quantum device versus on configurations drawn from the uniform distribution over the full Fock space. In both cases, we consider $K = 10$ batches of electronic configurations and the estimator is applied to $2.4576 \cdot 10^6$ sampled configurations. The self-consistent recovery procedure is warm-started from the orbital occupancy vector \mathbf{n} obtained after running ten configuration recovery iterations with $d = 250K$. Since the configuration recovery procedure is warm-started, for $d > 250K$ we apply two self-consistent operations. Fig. S21 (a) shows the comparison of the energy-variance analyses in both scenarios for values of the subspace dimension up to $d = 16M$. The qualitative behavior is similar, showing a cluster of points following a line that coincides with Eigenstate A of the Hamiltonian. However, $d \approx 16M$ is enough for the estimator run on quantum samples to reveal the second cluster corresponding to Eigenstate B, while greater values of d are required in the case where the samples are obtained from the uniform distribution over the full Fock space. Being able to resolve eigenstates of lower energy with smaller values of d shows a more efficient description of the low-energy physics of the problem.

To quantify more precisely for which value of d the Eigenstate B cluster of energy-variance pairs appears, we study the evolution of $\Delta H/E^2$ as a function of d . A jump in $\Delta H/E^2$ reveals the discovery of a cluster in the energy-variance plane. Fig. S21 (b) shows the $\Delta H/E^2$ as a function of d in both scenarios. While a large enough value of d allows resolving Eigenstate B from the estimator with configuration recovery applied to the uniformly random samples, this transition requires the projection and diagonalization of the Hamiltonian in a subspace whose dimension is $d > 16M$. The observation of the same transition on the configurations sampled from the quantum device requires at most $d = 12.25M$ configurations. The difference in these dimensions is of at least $3.75M$ electronic configurations.

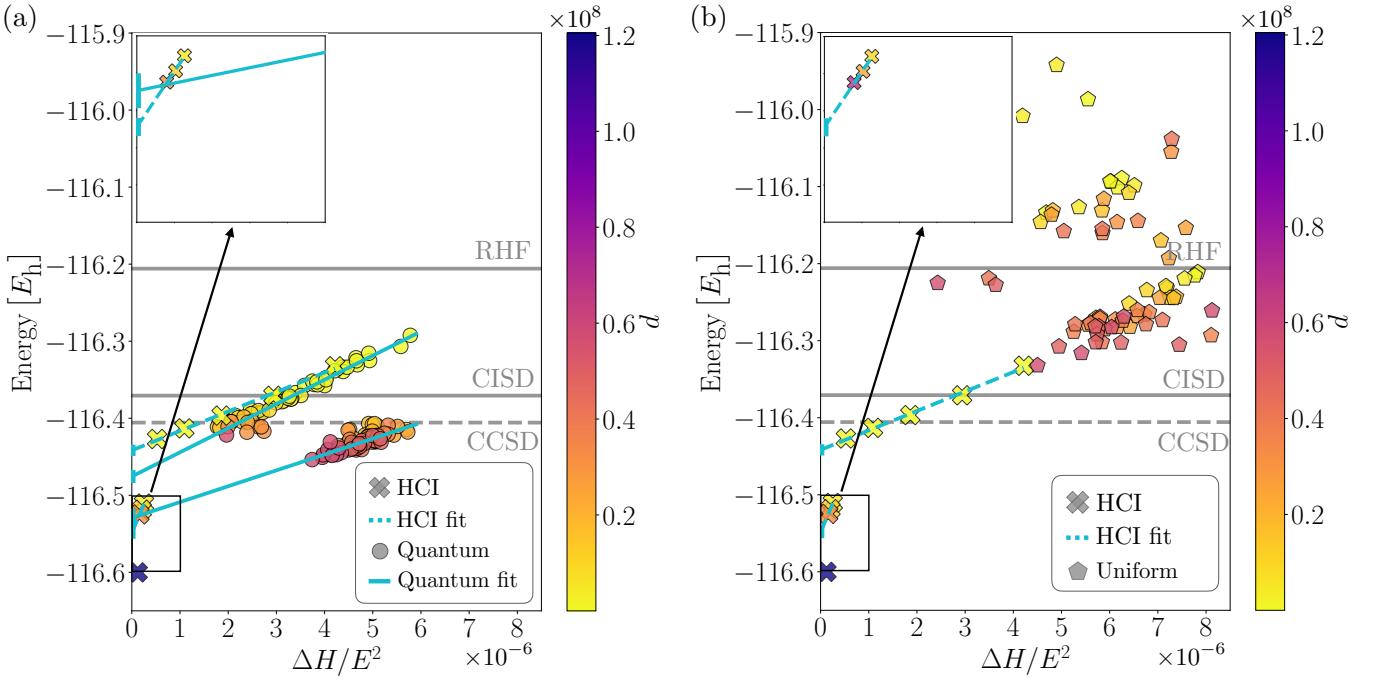


Figure S20. Energy-variance analysis in [2Fe-2S] of the approximate eigenstates obtained by running the estimator with configuration recovery on measurement outcomes from the quantum device (panel (a)) and from the uniform distribution over the set of configurations that have the right particle number (panel (b)). Crosses show the energy-variance relation of HCl for reference, alongside the fit to a dashed line. Dots and pentagons correspond to energy-variance points obtained from the estimator applied to different batches of configurations. The number of determinants in each batch is encoded in the color bar. The inset is a zoom into the HCl extrapolation of the second Eigenstate.

These results indicate that the quantum device provides a useful signal that is used by the HPC quantum estimator to procedure more accurate ground-state representations than using the same estimator on samples from uniform distribution.

-
- [1] R. P. Feynman, Simulating physics with computers, *Int. J. Theor. Phys.* **21**, 467 (1982).
- [2] A. Aspuru-Guzik, A. D. Dutoi, P. J. Love, and M. Head-Gordon, Simulated quantum computation of molecular energies, *Science* **309**, 1704 (2005).
- [3] D. Wecker, M. B. Hastings, and M. Troyer, Progress towards practical quantum variational algorithms, *Phys. Rev. A* **92**, 042303 (2015).
- [4] S. Sharma, K. Sivalingam, F. Neese, and G. K.-L. Chan, Low-energy spectrum of iron-sulfur clusters directly from many-particle quantum mechanics, *Nat. Chem* **6**, 927 (2014).
- [5] K. D. Vogiatzis, D. Ma, J. Olsen, L. Gagliardi, and W. A. De Jong, Pushing configuration-interaction to the limit: towards massively parallel MCSCF calculations, *J. Chem. Phys* **147**, 184111 (2017).
- [6] H. Gao, S. Imamura, A. Kasagi, and E. Yoshida, Distributed implementation of full configuration interaction for one trillion determinants, *J. Chem. Theory Comput.* (2024).
- [7] J. P. LeBlanc, A. E. Antipov, F. Becca, I. W. Bulik, G. K.-L. Chan, C.-M. Chung, Y. Deng, M. Ferrero, T. M. Henderson, C. A. Jiménez-Hoyos, *et al.*, Solutions of the two-dimensional Hubbard model: benchmarks and results from a wide range of numerical algorithms, *Phys. Rev. X* **5**, 041041 (2015).
- [8] M. Motta, D. M. Ceperley, G. K.-L. Chan, J. A. Gomez, E. Gull, S. Guo, C. A. Jiménez-Hoyos, T. N. Lan, J. Li, F. Ma, *et al.*, Towards the solution of the many-electron problem in real materials: equation of state of the hydrogen chain with state-of-the-art many-body methods, *Phys. Rev. X* **7**, 031059 (2017).
- [9] Y. Cao, J. Romero, J. P. Olson, M. Degroote, P. D. Johnson, M. Kieferová, I. D. Kivlichan, T. Menke, B. Peropadre, N. P. Sawaya, *et al.*, Quantum chemistry in the age of quantum computing, *Chem. Rev* **119**, 10856 (2019).
- [10] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, Quantum computational chemistry, *Rev. Mod. Phys* **92**, 015003 (2020).
- [11] B. Bauer, S. Bravyi, M. Motta, and G. Kin-Lic Chan, Quantum algorithms for quantum chemistry and quantum materials

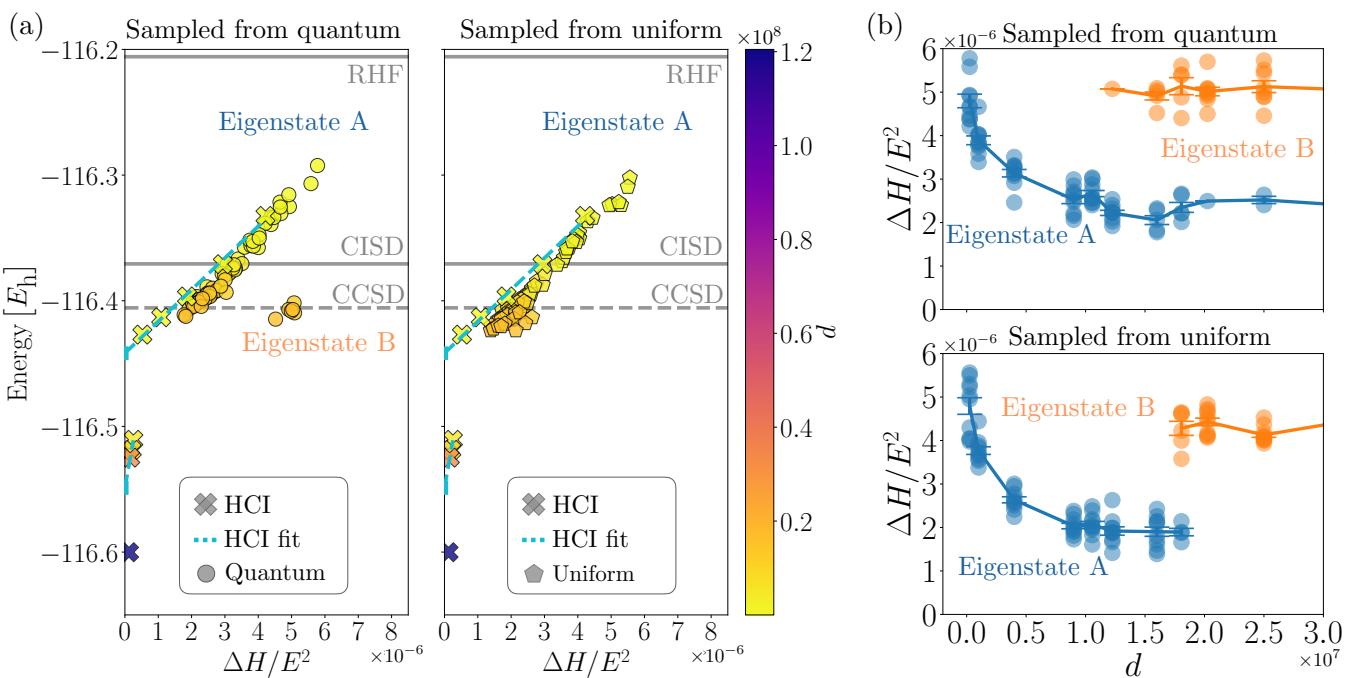


Figure S21. Comparison of the energy-variance analysis of the low-energy spectrum of [2Fe-2S], obtained from the estimator with configuration recovery applied to quantum data and configurations drawn from the uniform distribution. (a) The left panel shows energy-variance analysis of the low energy spectrum obtained by applying the estimator with configuration recovery to measurement outcomes from the quantum computer, while the right panel shows the same for samples from the uniform distribution on the Fock space. Crosses show the energy-variance relation of HCl for reference, alongside the fit to a (blue, dashed) line. Dots and pentagons correspond energy-variance points obtained from different batches of samples from the estimator. Subspace dimensions d are indicated by the colors of the points (corresponding to the colorbar), up to $d = 16M$, which is in the neighbourhood of the transition from Eigenstate A to B in the analysis applied to quantum samples. (b) Scatter plots of $\Delta H/E^2$ (horizontal axis on energy-variance plots) obtained from the estimator with configuration recovery on different batches of configurations, as a function of the number of determinants d in each batch. The solid lines show the mean value of $\Delta H/E^2$ as a function of d , and the error bars correspond to the standard error of the mean. Different groups of data points are color coded according to which eigenstate they come from. The grey region indicates the interval of d where the transition from Eigenstate A to Eigenstate B occurs. Top panel corresponds to the application of the estimator with configuration recovery to configurations sampled from a quantum device while the bottom panel corresponds to the application of the estimator with configuration recovery to configurations sampled from a uniform distribution over the Fock space.

science, *Chem. Rev.* **120**, 12685 (2020).

- [12] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, *Nature* **549**, 242 (2017).
- [13] Google AI Quantum and Collaborators, Hartree-Fock on a superconducting qubit quantum computer, *Science* **369**, 1084 (2020).
- [14] 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* **603**, 416 (2022).
- [15] M. Motta, G. O. Jones, J. E. Rice, T. P. Gujarati, R. Sakuma, I. Liepuoniute, J. M. Garcia, and Y.-y. Ohnishi, Quantum chemistry simulation of ground- and excited-state properties of the sulfonium cation on a superconducting quantum processor, *Chem. Sci.* **14**, 2915 (2023).
- [16] L. Zhao, J. Goings, K. Shin, W. Kyoung, J. I. Fuks, J.-K. Kevin Rhee, Y. M. Rhee, K. Wright, J. Nguyen, J. Kim, *et al.*, Orbital-optimized pair-correlated electron simulations on trapped-ion quantum computers, *npj Quantum Inf.* **9**, 60 (2023).
- [17] Google AI Quantum and Collaborators, Purification-based quantum error mitigation of pair-correlated electron simulations, *Nat. Phys.* **19**, 1787 (2023).
- [18] T. Weaving, A. Ralli, P. J. Love, S. Succi, and P. V. Coveney, Contextual subspace variational quantum eigensolver calculation of the dissociation curve of molecular nitrogen on a superconducting quantum computer (2023), [arXiv:2312.04392 \[quant-ph\]](https://arxiv.org/abs/2312.04392).
- [19] V. Verteletskyi, T.-C. Yen, and A. F. Izmaylov, Measurement optimization in the variational quantum eigensolver using a minimum clique cover, *J. Chem. Phys.* **152**, 124114 (2020).
- [20] W. J. Huggins, J. R. McClean, N. C. Rubin, Z. Jiang, N. Wiebe, K. B. Whaley, and R. Babbush, Efficient and noise

- resilient measurements for quantum chemistry on near-term quantum computers, *npj Quantum Inf* **7**, 1 (2021).
- [21] A. Dutt, W. Kirby, R. Raymond, C. Hadfield, S. Sheldon, I. L. Chuang, and A. Mezzacapo, Practical benchmarking of randomized measurement methods for quantum chemistry Hamiltonians, [arXiv:2312.07497](https://arxiv.org/abs/2312.07497) (2023).
- [22] A. Anand, P. Schleich, S. Alperin-Lea, P. W. Jensen, S. Sim, M. Díaz-Tinoco, J. S. Kottmann, M. Degroote, A. F. Izmaylov, and A. Aspuru-Guzik, A quantum computing view on unitary coupled cluster theory, *Chem. Soc. Rev* **51**, 1659 (2022).
- [23] M. B. Hastings, D. Wecker, B. Bauer, and M. Troyer, Improving quantum algorithms for quantum chemistry, *Quant. Info. Comput* **15**, 1–21 (2015).
- [24] J. Lee, D. W. Berry, C. Gidney, W. J. Huggins, J. R. McClean, N. Wiebe, and R. Babbush, Even more efficient quantum computations of chemistry through tensor hypercontraction, *Phys. Rev. X Quantum* **2**, 030305 (2021).
- [25] Y. Alexeev, M. Amsler, P. Baity, M. A. Barroca, S. Bassini, T. Battelle, D. Camps, D. Casanova, F. T. Chong, C. Chung, *et al.*, Quantum-centric supercomputing for materials science: A perspective on challenges and future directions, arXiv preprint [arXiv:2312.09733](https://arxiv.org/abs/2312.09733) (2023).
- [26] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, *et al.*, PySCF: the python-based simulations of chemistry framework, *WIREs Comput. Mol. Sci* **8**, e1340 (2018).
- [27] P. Jordan and E. Wigner, Über das paulische äquivalenzverbot, *Zeit. Phys* **47**, 631 (1928).
- [28] J. Ivanic and K. Ruedenberg, Identification of deadwood in configuration spaces through general direct configuration interaction, *Theor. Chem. Acc* **106**, 339 (2001).
- [29] See supplementary information.
- [30] K. Kanno, M. Kohda, R. Imai, S. Koh, K. Mitarai, W. Mizukami, and Y. O. Nakagawa, Quantum-selected configuration interaction: classical diagonalization of Hamiltonians in subspaces selected by quantum computers (2023), [arXiv:2302.11320 \[quant-ph\]](https://arxiv.org/abs/2302.11320).
- [31] Y. O. Nakagawa, M. Kamoshita, W. Mizukami, S. Sudo, and Y. ya Ohnishi, ADAPT-QSCI: Adaptive construction of input state for quantum-selected configuration interaction (2023), [arXiv:2311.01105 \[quant-ph\]](https://arxiv.org/abs/2311.01105).
- [32] F. A. Evangelista, Adaptive multiconfigurational wave functions, *J. Chem. Phys* **140**, 124114 (2014).
- [33] J. B. Schriber and F. A. Evangelista, Communication: An adaptive configuration interaction approach for strongly correlated electrons with tunable accuracy, *J. Chem. Phys* **144**, 161106 (2016).
- [34] A. A. Holmes, H. J. Changlani, and C. Umrigar, Efficient heat-bath sampling in Fock space, *J. Chem. Theory Comput* **12**, 1561 (2016).
- [35] A. A. Holmes, N. M. Tubman, and C. Umrigar, Heat-bath configuration interaction: An efficient selected configuration interaction algorithm inspired by heat-bath sampling, *J. Chem. Theory Comput* **12**, 3674 (2016).
- [36] N. M. Tubman, J. Lee, T. Y. Takeshita, M. Head-Gordon, and K. B. Whaley, A deterministic alternative to the full configuration interaction quantum Monte Carlo method, *J. Chem. Phys* **145**, 044112 (2016).
- [37] J. B. Schriber and F. A. Evangelista, Adaptive configuration interaction for computing challenging electronic excited states with tunable accuracy, *J. Chem. Theory Comput* **13**, 5354 (2017).
- [38] S. Sharma, A. A. Holmes, G. Jeanmairet, A. Alavi, and C. J. Umrigar, Semistochastic heat-bath configuration interaction method: Selected configuration interaction with semistochastic perturbation theory, *J. Chem. Theory Comput* **13**, 1595 (2017).
- [39] F. A. Evangelista, G. K.-L. Chan, and G. E. Scuseria, Exact parameterization of fermionic wave functions via unitary coupled cluster theory, *J. Chem. Phys* **151**, 244112 (2019).
- [40] M. Motta, K. J. Sung, K. B. Whaley, M. Head-Gordon, and J. Shee, Bridging physical intuition and hardware efficiency for correlated electronic states: the local unitary cluster Jastrow ansatz for electronic structure, *Chem. Sci* **14**, 11213 (2023).
- [41] Y. Matsuzawa and Y. Kurashige, Jastrow-type decomposition in quantum chemistry for low-depth quantum circuits, *J. Chem. Theory Comput* **16**, 944 (2020).
- [42] The ffsim developers, **ffsim: Faster simulations of fermionic quantum circuits** (2024).
- [43] P.-D. Fan and P. Piecuch, The usefulness of exponential wave function expansions employing one-and two-body cluster operators in electronic structure theory: The extended and generalized coupled-cluster methods, *Adv. Quantum Chem* **51**, 1 (2006).
- [44] I. W. Bulik, T. M. Henderson, and G. E. Scuseria, Can single-reference coupled cluster theory describe static correlation?, *J. Chem. Theory Comput* **11**, 3171 (2015).
- [45] Z. Li and G. K.-L. Chan, Spin-projected matrix product states: versatile tool for strongly correlated systems, *J. Chem. Theory Comput* **13**, 2681 (2017).
- [46] H. Beinert, R. H. Holm, and E. Munck, Iron-sulfur clusters: nature's modular, multipurpose structures, *Science* **277**, 653 (1997).
- [47] P. Venkateswara Rao and R. Holm, Synthetic analogues of the active sites of iron-sulfur proteins, *Chem. Rev* **104**, 527 (2004).
- [48] T. Kashima and M. Imada, Path-integral renormalization group method for numerical study on ground states of strongly correlated electronic systems, *J. Phys. Soc. Jpn* **70**, 2287 (2001).
- [49] V. Papaefthymiou, M. M. Millar, and E. Muennck, Moessbauer and EPR studies of a synthetic analog for the iron-sulfur Fe₄S₄ core of oxidized and reduced high-potential iron proteins, *Inorg. Chem* **25**, 3010 (1986).
- [50] Y. Kim, A. Eddins, S. Anand, K. X. Wei, E. Van Den Berg, S. Rosenblatt, H. Nayfeh, Y. Wu, M. Zaletel, K. Temme, *et al.*, Evidence for the utility of quantum computing before fault tolerance, *Nature* **618**, 500 (2023).

- [51] K. Shinjo, K. Seki, T. Shirakawa, R.-Y. Sun, and S. Yunoki, Unveiling clean two-dimensional discrete time quasicrystals on a digital quantum computer, arXiv preprint arXiv:2403.16718 (2024).
- [52] O. Reardon-Smith, M. Oszmaniec, and K. Korzekwa, Improved simulation of quantum circuits dominated by free fermionic operations, arXiv preprint arXiv:2307.12702 (2023).
- [53] S. Lee, J. Lee, H. Zhai, Y. Tong, A. M. Dalzell, A. Kumar, P. Helms, J. Gray, Z.-H. Cui, W. Liu, *et al.*, Evaluating the evidence for exponential quantum advantage in ground-state quantum chemistry, *Nat. Commun* **14**, 1952 (2023).
- [54] F. Becca and S. Sorella, *Quantum Monte Carlo approaches for correlated systems* (Cambridge University Press, 2017).
- [55] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, *Science* **355**, 602 (2017).
- [56] S. R. White and R. L. Martin, Ab initio quantum chemistry using the density matrix renormalization group, *J. Chem. Phys* **110**, 4127 (1999).
- [57] G. K.-L. Chan and S. Sharma, The density matrix renormalization group in quantum chemistry, *Annu. Rev. Phys. Chem* **62**, 465 (2011).
- [58] C. Hubig, J. Haegeman, and U. Schollwöck, Error estimates for extrapolations with matrix-product states, *Phys. Rev. B* **97**, 045125 (2018).
- [59] Z. Li, J. Li, N. S. Dattani, C. Umrigar, and G. K. Chan, The electronic complexity of the ground-state of the FeMo cofactor of nitrogenase as relevant to quantum simulations, *J. Chem. Phys* **150**, 024302 (2019).
- [60] W. J. Hehre, R. F. Stewart, and J. A. Pople, Self-consistent molecular-orbital methods. I. use of Gaussian expansions of Slater-type atomic orbitals, *J. Chem. Phys* **51**, 2657 (1969).
- [61] W. J. Hehre, R. Ditchfield, and J. A. Pople, Self-consistent molecular orbital methods. XII. Further extensions of Gaussian-type basis sets for use in molecular orbital studies of organic molecules, *J. Chem. Phys* **56**, 2257 (1972).
- [62] T. H. Dunning, Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen, *J. Chem. Phys* **90**, 1007 (1989).
- [63] Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt, N. A. Bogdanov, G. H. Booth, J. Chen, Z.-H. Cui, *et al.*, Recent developments in the PySCF program package, *J. Chem. Phys* **153**, 024109 (2020).
- [64] A. D. Becke, Density-functional exchange-energy approximation with correct asymptotic behavior, *Phys. Rev. A* **38**, 3098 (1988).
- [65] J. P. Perdew, Density-functional approximation for the correlation energy of the inhomogeneous electron gas, *Phys. Rev. B* **33**, 8822 (1986).
- [66] F. Jorge, A. Canal Neto, G. Camiletti, and S. Machado, Contracted Gaussian basis sets for Douglas-Kroll-Hess calculations: estimating scalar relativistic effects of some atomic and molecular properties, *J. Chem. Phys* **130** (2009).
- [67] W. Liu, Ideas of relativistic quantum chemistry, *Mol. Phys* **108**, 1679 (2010).
- [68] Z. Li, Y. Xiao, and W. Liu, On the spin separation of algebraic two-component relativistic Hamiltonians, *J. Chem. Phys* **137**, 154114 (2012).
- [69] H. L. Tang, V. Shkolnikov, G. S. Barron, H. R. Grimsley, N. J. Mayhall, E. Barnes, and S. E. Economou, Qubit-ADAPT-VQE: an adaptive algorithm for constructing hardware-efficient ansatzes on a quantum processor, *PRX Quantum* **2**, 020310 (2021).
- [70] M. J. D. Powell, A direct search optimization method that models the objective and constraint functions by linear interpolation, in *Advances in Optimization and Numerical Analysis*, edited by S. Gomez and J.-P. Hennart (Springer Netherlands, Dordrecht, 1994) pp. 51–67.
- [71] B. O. Roos, P. R. Taylor, and P. E. Sigbahn, A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach, *Chem. Phys* **48**, 157 (1980).
- [72] M. Head-Gordon and J. A. Pople, Optimization of wave function and geometry in the finite basis Hartree-Fock method, *J. Phys. Chem* **92**, 3063 (1988).
- [73] H.-J. Werner and P. J. Knowles, A second order multiconfiguration SCF procedure with optimum convergence, *J. Chem. Phys* **82**, 5053 (1985).
- [74] J. Olsen, The CASSCF method: A perspective and commentary, *Int. J. Quantum Chem* **111**, 3267 (2011).
- [75] D. Zgid and M. Nooijen, The density matrix renormalization group self-consistent field method: orbital optimization with the density matrix renormalization group method in the active space, *J. Chem. Phys* **128**, 144116 (2008).
- [76] A.-K. Wu, M. T. Fishman, J. Pixley, and E. Stoudenmire, Disentangling interacting systems with fermionic gaussian circuits: application to the single impurity Anderson model, arXiv:2212.09798 (2022).
- [77] S. Wouters and D. Van Neck, The density matrix renormalization group for ab initio quantum chemistry, *Eur. Phys. J. D* **68**, 272 (2014).
- [78] J. R. Moreno, J. Cohn, D. Sels, and M. Motta, Enhancing the expressivity of variational neural, and hardware-efficient quantum states through orbital rotations (2023), arXiv:2302.11588 [quant-ph].
- [79] I. O. Sokolov, P. K. Barkoutsos, P. J. Ollitrault, D. Greenberg, J. Rice, M. Pistoia, and I. Tavernelli, Quantum orbital-optimized unitary coupled cluster methods in the strongly correlated regime: can quantum algorithms outperform their classical equivalents?, *J. Chem. Phys* **152**, 124107 (2020).
- [80] 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. Research* **2**, 033421 (2020).
- [81] J. Bierman, Y. Li, and J. Lu, Improving the accuracy of variational quantum eigensolvers with fewer qubits using orbital optimization, *J. Chem. Theory Comput* **19**, 790 (2023).
- [82] J. E. Smith, B. Mussard, A. A. Holmes, and S. Sharma, Cheap and near exact casscf with large active spaces, *Journal of chemical theory and computation* **13**, 5468 (2017).

- [83] D. Thouless, Stability conditions and nuclear rotations in the Hartree-Fock theory, *Nuc. Phys.* **21**, 225 (1960).
- [84] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, **JAX: composable transformations of Python+NumPy programs** (2018).
- [85] D. Chowdhury, A. Georges, O. Parcollet, and S. Sachdev, Sachdev-Ye-Kitaev models and beyond: window into non-Fermi liquids, *Rev. Mod. Phys.* **94**, 035004 (2022).
- [86] B. Friedman and G. Levine, Configuration-interaction approach to the two-dimensional Hubbard model near half-filling, *Phys. Rev. B* **55**, 9558 (1997).
- [87] E. Louis, F. Guinea, M. P. López Sancho, and J. A. Vergés, Configuration-interaction approach to hole pairing in the two-dimensional Hubbard model, *Phys. Rev. B* **59**, 14005 (1999).
- [88] L. R. Schwarz, G. H. Booth, and A. Alavi, Insights into the structure of many-electron wave functions of mott-insulating antiferromagnets: The three-band Hubbard model in full configuration interaction quantum Monte Carlo, *Phys. Rev. B* **91**, 045139 (2015).
- [89] W. Dobrátz, H. Luo, and A. Alavi, Compact numerical solutions to the two-dimensional repulsive Hubbard model obtained via nonunitary similarity transformations, *Phys. Rev. B* **99**, 075119 (2019).
- [90] M. J. Rozenberg, G. Kotliar, and X. Y. Zhang, Mott-Hubbard transition in infinite dimensions. II, *Phys. Rev. B* **49**, 10181 (1994).
- [91] A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions, *Rev. Mod. Phys.* **68**, 13 (1996).
- [92] S.-S. B. Lee, J. von Delft, and A. Weichselbaum, Doublon-holon origin of the subpeaks at the Hubbard band edges, *Phys. Rev. Lett.* **119**, 236402 (2017).
- [93] C. Gauvin-Ndiaye, J. Tindall, J. R. Moreno, and A. Georges, Mott transition and volume law entanglement with neural quantum states (2023), arXiv:2311.05749 [cond-mat.str-el].
- [94] Y. Kwon, D. M. Ceperley, and R. M. Martin, Effects of backflow correlation in the three-dimensional electron gas: Quantum Monte Carlo study, *Phys. Rev. B* **58**, 6800 (1998).
- [95] M. Imada and T. Kashima, Path-integral renormalization group method for numerical study of strongly correlated electron systems, *J. Phys. Soc. Jpn.* **69**, 2723 (2000).
- [96] Y. Nomura, A. S. Darmawan, Y. Yamaji, and M. Imada, Restricted Boltzmann machine learning for solving strongly correlated quantum systems, *Phys. Rev. B* **96**, 205152 (2017).
- [97] S. Sorella, Generalized Lanczos algorithm for variational quantum Monte Carlo, *Phys. Rev. B* **64**, 024512 (2001).
- [98] T. Mizusaki and M. Imada, Extrapolation method for shell model calculations, *Phys. Rev. C* **65**, 064319 (2002).
- [99] M. Motta, E. Ye, J. R. McClean, Z. Li, A. J. Minnich, R. Babbush, and G. K.-L. Chan, Low rank representations for quantum simulation of electronic structure, *npj Quantum Inf.* **7**, 1 (2021).
- [100] M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani, Experimental realization of any discrete unitary operator, *Phys. Rev. Lett.* **73**, 58 (1994).
- [101] W. R. Clements, P. C. Humphreys, B. J. Metcalf, W. S. Kolthammer, and I. A. Walmsley, Optimal design for universal multiport interferometers, *Optica* **3**, 1460 (2016).
- [102] Z. Jiang, K. J. Sung, K. Kechedzhi, V. N. Smelyanskiy, and S. Boixo, Quantum algorithms to simulate many-body physics of correlated fermions, *Phys. Rev. Appl.* **9**, 044036 (2018).
- [103] P. D. Nation, H. Kang, N. Sundaresan, and J. M. Gambetta, Scalable mitigation of measurement errors on quantum computers, *Phys. Rev. X Quantum* **2**, 040326 (2021).
- [104] L. Viola and S. Lloyd, Dynamical suppression of decoherence in two-state quantum systems, *Phys. Rev. A* **58**, 2733 (1998).
- [105] A. Kofman and G. Kurizki, Universal dynamical control of quantum mechanical decay: modulation of the coupling to the continuum, *Phys. Rev. Lett.* **87**, 270405 (2001).
- [106] M. J. Biercuk, H. Uys, A. P. VanDevender, N. Shiga, W. M. Itano, and J. J. Bollinger, Optimized dynamical decoupling in a model quantum memory, *Nature* **458**, 996 (2009).
- [107] S. Niu and A. Todri-Sanial, Effects of dynamical decoupling and pulse-level optimizations on IBM quantum computers, *IEEE Trans. Quantum Eng.* **3**, 1 (2022).
- [108] Qiskit contributors, **Qiskit: An open-source framework for quantum computing** (2023).
- [109] Z. Li and G. K.-L. Chan, **Repository: Active space model for iron sulfur clusters** (2017).
- [110] When considering samples from the uniform distribution over electronic configurations on the correct particle sector, the configuration recovery method cannot be applied since it only recovers configurations with the wrong particle number. Therefore, this scenario is equivalent to running QSCI (without configuration recovery) on those samples.