

Quantum-classical auxiliary field quantum Monte Carlo with matchgate shadows on trapped ion quantum computers

Luning Zhao,^{*} Joshua J. Goings[✉], Willie Aboumrad, Andrew Arrasmith[✉], Lazaro Calderin[✉], Spencer Churchill[✉], Dor Gabay[✉], Thea Harvey-Brown[✉], Melanie Hiles, Magda Kaja[✉], Matthew Keesan[✉], Karolina Kulesz, Andrii Maksymov, Mei Maruo[✉], Mauricio Muñoz, Bas Nijholt[✉], Rebekah Schiller[✉], Yvette de Sereville, Amy Smidutz, Felix Tripier, Grace Yao[✉], Trishal Zaveri[✉], Coleman Collins[✉], Martin Roetteler, and Evgeny Epifanovsky[✉]
IonQ Inc, College Park, Maryland 20740, USA

Arseny Kovyrshin[✉], Lars Tornberg[✉], and Anders Broo[✉]
Data Science and Modelling, Pharmaceutical Sciences, R&D, AstraZeneca Gothenburg, Pepparedsleden 1, Molndal SE-431 83, Sweden

Jeff R. Hammond[✉]
NVIDIA Helsinki Oy, Porkkalankatu 1, Helsinki 00180, Finland

Zohim Chandani[✉], Pradnya Khalate[✉], and Elica Kyoseva[✉]
NVIDIA Corporation, 2788 San Tomas Expressway, Santa Clara, California 95051, USA

Yi-Ting Chen[✉], Eric M. Kessler, Cedric Yen-Yu Lin, Gandhi Ramu, and Ryan Shaffer[✉]
AWS Quantum Technologies, Seattle, Washington 98121, USA

Michael Brett[✉], Benchen Huang[✉], Maxime R. Hugues, and Tyler Y. Takeshita[✉]
AWS Worldwide Specialist Organization, Seattle, Washington 98170, USA



(Received 15 August 2025; accepted 17 June 2026; published 15 July 2026)

We demonstrate an end-to-end workflow to model chemical reaction barriers with the quantum-classical auxiliary field quantum Monte Carlo (QC-AFQMC) algorithm with quantum tomography using matchgate shadows. The workflow operates within an accelerated quantum supercomputing environment with the IonQ Forte quantum computer and NVIDIA graphic processing units (GPUs) on Amazon Web Services. We present several algorithmic innovations and an efficient GPU-accelerated execution that achieves a several orders of magnitude speedup over the state-of-the-art implementation of QC-AFQMC. We apply the algorithm to simulate the oxidative addition step of the nickel-catalyzed Suzuki-Miyaura reaction using 24 qubits of IonQ Forte with 16 qubits used to represent the trial state, plus 8 additional ancilla qubits for error mitigation, resulting in the largest QC-AFQMC with matchgate shadow experiments ever performed on quantum hardware. We achieve a $9\times$ speedup in collecting matchgate circuit measurements, and our distributed-parallel postprocessing implementation attains a $656\times$ time-to-solution improvement over the prior state of the art. Chemical reaction barriers for the model reaction evaluated with active-space QC-AFQMC are within the uncertainty interval of ± 4 kcal/mol from the reference CCSD(T) (coupled cluster singles and doubles with perturbative triples) result when matchgates are sampled on the ideal simulator and within 10 kcal/mol from reference when measured on quantum processing unit. This work marks a step toward practical quantum chemistry simulations on quantum devices while identifying several opportunities for further development.

DOI: [10.1103/n1tf-8kr7](https://doi.org/10.1103/n1tf-8kr7)

I. INTRODUCTION

Computationally intensive tasks have become integral to drug discovery and development (DDD), driving advancement

across the entire pharmaceutical R&D value chain. There is a strong motivation to find ways to accomplish traditionally experimental tasks faster and cheaper with the aid of computer simulations. Quantum computing is an emerging technology that enables the study of a certain class of problems thought to be intractable with classical computation. The possibility of realizing these theoretical speedups in practice has sparked research in a range of application areas, including pharmaceutical R&D. Therefore, it is unsurprising to see the recent development of quantum algorithms for many problems within DDD intended for near-term [1–5] and future fault-tolerant [6,7] quantum computing devices. While the

^{*}Contact author: zhao@ionq.co

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

early drug discovery phase has been the primary target for computational methods, in this work we focus on scenarios arising in the drug development phase that require accurate electronic structure modeling. Specifically, we consider finding a synthetic pathway of desired purity and yield using transition metal (TM) catalysis during the drug development phase [8].

Some of the most significant chemical reactions in medicinal chemistry include amide bond formation, nucleophilic aromatic substitution, and Suzuki-Miyaura cross-coupling [9]. The first two reactions typically involve only organic compounds, and density functional theory (DFT) [10] calculations have sufficient resolution to give useful mechanistic insights. However, the Suzuki-Miyaura cross-coupling reactions include catalytic TM complexes that exhibit complicated electronic structure. In particular, unpaired d or f electrons lead to dense low-lying electronic states, resulting in strong electron correlation. Additionally, the open-shell character of many transition metal complexes leads to states with different spin symmetries, shaping their reactivity, stability, and underlying reaction mechanisms. As a result, multiple reaction spin channels emerge, allowing reactions to proceed through distinct spin states with different activation barriers. Moreover, because high-spin and low-spin states in transition metal complexes often lie close in energy, the spin state can shift during a reaction [11], a phenomenon known as spin crossover.

The intricate electronic structure of TM complexes powers diverse catalytic mechanisms. At the same time, it makes it difficult to model their chemistry reliably using approximate single-reference methods since the wave function often exhibits a multireference character, and strong electron correlation plays a central role [12]. Single-reference methods such as DFT and even the coupled cluster singles and doubles with perturbative triples [CCSD(T)] [13], which is regarded as the gold standard in terms of accuracy, are not able to reliably model these systems. In fact, CCSD(T) often exhibits well-known breakdowns for strongly correlated systems, including complexes with polynuclear transition metal cores, systems undergoing metal to insulator phase transitions, and multielectron excited states. (It may still yield reasonable accuracy for mononuclear TM complexes [14].) Methods such as the complete active space self-consistent field method (CASSCF) [15] are capable of treating strong correlation, but their computational cost grows exponentially with the size of the active space. The density matrix renormalization group method (DMRG) [16,17], originally developed for one-dimensional and quasi-one-dimensional systems, offers polynomial scaling in computational cost for such systems and has demonstrated strong performance on chemically relevant systems, including TM complexes [12,18,19]. While DMRG enables significantly larger active-space computations compared to conventional approaches and is amenable to GPU acceleration [20–22], its performance at large scales increasingly depends on available physical memory and memory bandwidth [20,23].

This gives rise to one of the most promising applications of quantum computers pertaining to the DDD pipeline—modeling strongly correlated catalytic systems. In recent years, a variety of different algorithms have been developed to solve the electronic structure problem using quantum

computers. Within these algorithms, quantum phase estimation (QPE) [24,25] could in principle provide exact solutions to quantum chemistry problems in polynomial time, but even with recent innovations such as qubitization [26,27] and low-rank tensor decomposition [28], it demands error rates far beyond the capability of today's noisy quantum computers, and likely unachievable without robust error correction. The variational quantum eigensolver (VQE) [29–33] has been considered one of the most promising near-term algorithms. However, it requires an often prohibitive number of measurements on quantum computers [34], and the predicted chemical properties are highly sensitive to device noise. Both factors limit its ability to efficiently deliver usable results on current-generation devices.

In this work, we focus on an alternative algorithm, the quantum-classical auxiliary field quantum Monte Carlo (QC-AFQMC) [35] algorithm, which builds upon the classical AFQMC method [36–40]. AFQMC is a projector Monte Carlo approach that solves for the ground state of the electronic Hamiltonian with imaginary time propagation. Although AFQMC is formally exact, it relies on an approximate trial state to control the fermionic phase problem through the phaseless constraint introducing a bias proportional to the trial state's deviation from the true ground state. Classical methods can generate multideterminant trial wave functions (CASSCF and selected configuration interaction (CI)) or high-quality single-reference states from coupled cluster theory, but these approaches face intrinsic barriers: Multideterminant trials scale exponentially in cost, while single-reference methods fail in strongly correlated regimes. QC-AFQMC overcomes both by using quantum computers to prepare correlated trial states that capture multireference character without explicit enumeration. Early results [35,41] show QC-AFQMC reaching chemical accuracy with polynomial measurement cost, on both weakly and strongly correlated systems with multiple-bond breaking, demonstrating notable noise resilience compared to VQE and sidestepping the steep overhead of QPE.

The core innovation in QC-AFQMC is not simply its hybrid structure, but the way it divides tasks: the quantum computer prepares the correlated trial state and performs shadow tomography measurements, after which all imaginary time propagation and observable estimation are carried out on the classical computer. Unlike VQE, which requires ongoing quantum-classical feedback for each parameter update, QC-AFQMC isolates quantum measurements to the start, avoiding the iterative quantum-classical loop. Despite this conceptual elegance, several studies [41–44] have argued that the classical postprocessing of QC-AFQMC is prohibitively expensive, making the method impractical for systems of real scientific interest [41–44]. This work serves as a demonstration to the contrary. We significantly enhance the practicality of QC-AFQMC by reducing its time to solution through several key advances: tuning the performance of quantum hardware by increasing the throughput of job submission and execution, algorithmic improvements (inspired by some recent developments of AFQMC [45]) leading to a computational cost reduction in postprocessing, and GPU-accelerated implementations of routines required for postprocessing via the NVIDIA CUDA Toolkit leveraging

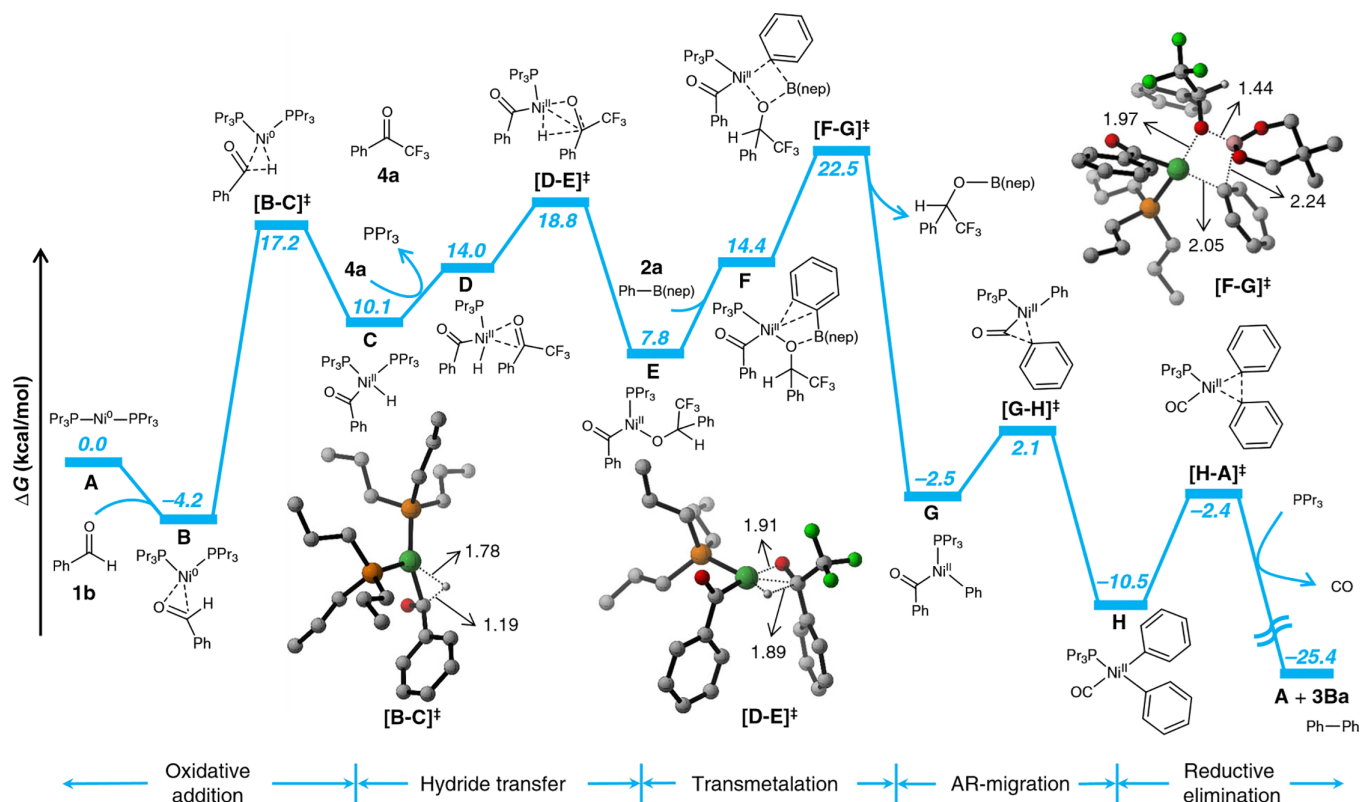


FIG. 1. Reaction mechanism and energy profile of the nickel-catalyzed deformylative Suzuki-Miyaura cross-coupling reaction. Our work focuses on modeling the oxidative addition step denoted by $\mathbf{B} \rightarrow [\mathbf{B-C}]^\ddagger \rightarrow \mathbf{C}$. (Figure reproduced from Ref. [47] under the terms of the Creative Commons CC BY license).

cuBLAS, cuSOLVER, and cuTENSOR. Overall, with these improvements we have been able to reduce the computational cost of QC-AFQMC energy evaluation and imaginary time propagation from $\mathcal{O}(N^{8.5})$ to $\mathcal{O}(N^{5.5})$ and the cost of imaginary time propagation from $\mathcal{O}(N^{7.5})$ to $\mathcal{O}(N^{4.5})$, where N is the number of spin orbitals or qubits. Executing the steps of QC-AFQMC via CUDA-Q on IonQ hardware hosted on Amazon Braket [46] combined with postprocessing on Amazon Web Services (AWS), we establish an efficient, end-to-end workflow for applying QC-AFQMC to complex chemical problems.

This paper demonstrates the QC-AFQMC workflow by modeling a key step in the Suzuki-Miyaura cross-coupling reaction (Fig. 1), which uses a nickel-based catalytic complex as a cost-effective alternative to palladium [47]. Specifically, we focus on the prediction of the reaction barrier for the first step: oxidative addition of aryl halide to nickel bis(tripropylphosphine). This step begins with the complex denoted by $\mathbf{B}$ in Fig. 1, proceeds through a transition state $[\mathbf{B-C}]^\ddagger$, and ends with the intermediate $\mathbf{C}$. These complexes exhibit an electronic structure primarily defined by weak correlation with limited multiconfigurational character, which makes it suitable for validating the QC-AFQMC workflow against reliable single-reference classical benchmarks such as CCSD(T).

The paper is organized as follows. Section II introduces the classical AFQMC algorithm, outlining its strengths and limitations, and then presents QC-AFQMC and the classical shadow sampling technique at its core. Once the theoretical

framework is established, Sec. III describes the computational details: the chosen chemical reaction, active space selection, algorithmic improvements, and key aspects of implementation and execution. Section IV reports QC-AFQMC energies for the species in the reaction and compares them to results from other methods. We then discuss time to solution, separating quantum execution from classical postprocessing. The paper concludes with a discussion of open challenges and future directions.

II. METHODS

A. The electronic structure problem

The chemical reactivity of a molecular system can be modeled by connecting reactants, products, and relevant transition states on a potential energy surface (PES), which is the energy of the molecule as a function of its atomic geometric configuration. Each point of the PES represents a solution to the ground-state energy problem, which can be determined computationally from first principles by solving the electronic Schrödinger equation:

$$H|\Psi_g\rangle = E_g|\Psi_g\rangle, \quad (1)$$

in which $H = H_1 + H_2$ is the electronic Hamiltonian consisting of one- and two-body contributions, $|\Psi_g\rangle$ is the ground-state wave function, and E_g is the energy.

Solving the Schrödinger equation exactly on a classical computer has exponential computational and memory complexity with respect to system size, which is prohibitive for

practical applications. This motivated the development of many physics-informed approximations and heuristics that trade approximation error (usually not controllable) for time to solution. We have discussed some of the most widely used approximate methods such as CCSD(T), CASSCF, and DMRG in Sec. I. A notable class of classical electronic structure methods relies on projector quantum Monte Carlo (QMC). These techniques [48,49] employ stochastic walkers to represent the system's wave function, evolving them through imaginary time propagation to generate ground-state samples. The precise form of projector QMC is highly adaptable, determined by factors such as the walker's Hilbert space, the propagator's mathematical structure, and the nature of walker interactions. This flexibility gives rise to a range of QMC variants [50–52], including the AFQMC algorithm, which we will now explore in detail.

B. Classical AFQMC

The AFQMC method [38–40,53] is a powerful tool for modeling the ground state of molecules and materials. In AFQMC, the wave function is expressed as a linear combination of walkers:

$$|\Psi\rangle = \sum_l w_l |\phi_l\rangle, \quad (2)$$

where each walker $|\phi_l\rangle$ is a Slater determinant with its corresponding weight w_l . One can obtain the ground state by evolving the wave function in imaginary time τ

$$|\Psi(\tau)\rangle = e^{-\tau H} \sum_l w_l |\phi_l\rangle. \quad (3)$$

As the system evolves, contributions from excited states decay exponentially, and eventually only the ground state survives. This evolution can be expressed as a product of short-time propagators:

$$|\Psi(\tau)\rangle = [e^{-\Delta\tau H}]^N \sum_l w_l |\phi_l\rangle, \quad \Delta\tau = \frac{\tau}{N}. \quad (4)$$

To carry this out in practice, each propagator must be efficiently approximated and applied to the set of walkers. The electronic Hamiltonian that defines the propagator is given by

$$H = H_1 + H_2 = \sum_{pq} t_{pq} a_p^\dagger a_q + \frac{1}{2} \sum_{pqrs} g_{pqrs} a_p^\dagger a_q^\dagger a_s a_r, \quad (5)$$

where t and g are one- and two-electron integrals, and $a_p^\dagger(a_p)$ are the creation (annihilation) operators for the p th orbital. Using Cholesky decomposition, the Hamiltonian can be rewritten as

$$H = v_0 - \frac{1}{2} \sum_\gamma v_\gamma^2, \quad (6)$$

where v_0 is the modified one-body term and $v_\gamma = i\mathcal{L}_\gamma$, with $\mathcal{L}_\gamma$ denoting the Cholesky vector associated with the two-body term:

$$\mathcal{L}_\gamma = \sum_{pq} L_{pq}^\gamma a_p^\dagger a_q, \quad g_{pqrs} = \sum_\gamma L_{pr}^\gamma L_{sq}^{\gamma*}. \quad (7)$$

In QC-AFQMC, the Cholesky decomposition is done with a modified algorithm [52], and its cost scaling is $\mathcal{O}(N^4)$.

Now, using the second-order Trotter-Suzuki decomposition, the imaginary time propagator becomes

$$e^{-\Delta\tau H} \approx e^{-\frac{\Delta\tau}{2} v_0} \prod_\gamma e^{\frac{\Delta\tau}{2} v_\gamma^2} e^{-\frac{\Delta\tau}{2} v_0}. \quad (8)$$

The Hubbard-Stratonovich transformation [54,55] converts the term involving v_γ^2 into an integral over an auxiliary field, resulting in the exponentials of one-body operators coupled to this auxiliary field:

$$e^{\frac{\Delta\tau}{2} v_\gamma^2} = \int \frac{dx_\gamma}{\sqrt{2\pi}} e^{-\frac{x_\gamma^2}{2}} e^{\sqrt{\Delta\tau} x_\gamma v_\gamma}. \quad (9)$$

Substituting into Eq. (8), we obtain

$$e^{-\Delta\tau H} \approx \int d\mathbf{x} p(\mathbf{x}) B(\mathbf{x}), \quad (10)$$

in which the operator $B(\mathbf{x})$ is defined as

$$B(\mathbf{x}) = e^{-\frac{\Delta\tau}{2} v_0} \prod_\gamma e^{\sqrt{\Delta\tau} x_\gamma v_\gamma} e^{-\frac{\Delta\tau}{2} v_0}, \quad (11)$$

and $p(\mathbf{x})$ is a multidimensional Gaussian distribution over the auxiliary field variable $\mathbf{x}$.

According to Eqs. (10) and (11), to propagate the walkers, one could first sample $\mathbf{x}$ from $p(\mathbf{x})$, and then construct and apply the operator $B(\mathbf{x})$ to each walker. Since $B(\mathbf{x})$ is the exponential of a one-body operator, its action on a Slater determinant simply yields another determinant:

$$|\phi'_l(\mathbf{x})\rangle = B(\mathbf{x}) |\phi_l\rangle. \quad (12)$$

Finally, with the propagated walkers that sample from the ground state, AFQMC energy is a normalized weighted sum of the local energies,

$$E = \frac{\sum_l w_l E_L(\phi_l)}{\sum_l w_l}, \quad (13)$$

in which the local energy is computed as

$$E_L(\phi) = \frac{\langle \Psi_T | H | \phi \rangle}{\langle \Psi_T | \phi \rangle}. \quad (14)$$

C. The phase problem

The approach described above is usually referred to as the “free-projection” version of AFQMC. Although it is formally exact, in reality it suffers from the phase problem [36,37,56], which causes instabilities in the energy evaluation. In order to mitigate the phase problem, a trial wave function $|\Psi_T\rangle$ is introduced to guide the propagation. The trial wave function should be a close approximation to the exact ground state, and the propagator in Eq. (10) is modified with [37]

$$\int \langle \Psi_T | \phi'(\mathbf{x} - \bar{\mathbf{x}}) \rangle p(\mathbf{x} - \bar{\mathbf{x}}) B(\mathbf{x} - \bar{\mathbf{x}}) \frac{1}{\langle \Psi_T | \phi \rangle} d\mathbf{x}, \quad (15)$$

in which a constant shift $\bar{\mathbf{x}}$ is also introduced, which does not affect the equality of Eq. (10). The optimal choice of $\bar{\mathbf{x}}$ will be determined later. The above equation could be written as

$$\int p(\mathbf{x}) w(\mathbf{x}, \phi) B(\mathbf{x} - \bar{\mathbf{x}}) d\mathbf{x}, \quad (16)$$

where

$$w(\mathbf{x}, \phi) = \frac{\langle \Psi_T | \phi'(\mathbf{x} - \bar{\mathbf{x}}) \rangle}{\langle \Psi_T | \phi \rangle} e^{\mathbf{x} \cdot \bar{\mathbf{x}} - (\bar{\mathbf{x}} \cdot \bar{\mathbf{x}}/2)}. \quad (17)$$

In each step, an $\bar{\mathbf{x}}$ is first decided for each walker, and the walker is propagated to a new walker with $B(\mathbf{x} - \bar{\mathbf{x}})$. Here, $w(\mathbf{x}, \phi)$ is the importance sampling function, which could be taken into account by allowing every walker to have a weight factor and update it during the propagation [38,41].

The optimal choice of $\bar{\mathbf{x}}$ is determined by minimizing the fluctuation of $w(\mathbf{x}, \phi)$ with respect to $\mathbf{x}$, and one obtains [37]

$$\bar{x}_\gamma = -\sqrt{\Delta\tau} \frac{\langle \Psi_T | v_\gamma | \phi \rangle}{\langle \Psi_T | \phi \rangle}, \quad (18)$$

which is usually referred to as force bias in AFQMC. With this choice of $\bar{\mathbf{x}}$, one could further approximate the weight factor as $w(\mathbf{x}, \phi) \approx \exp[-\Delta\tau E_L(\phi)]$ [37], in which $E_L(\phi)$ is the local energy defined in Eq. (14).

If the trial wave function is exact, then one can show that the local energy is real valued and equal to the exact energy. In this case, all the weights also become real valued and positive, which mitigates the phase problem. However, for approximate trial wave functions, the local energy is a complex number. Therefore, a common approximation is to replace it with its real part. $w(\mathbf{x}, \phi) \approx \exp[-\Delta\tau \text{Re}\{E_L(\phi)\}]$, and the overall weight factor is

$$w(\mathbf{x}, \phi) \approx \exp[-\Delta\tau \text{Re}\{E_L(\phi)\}] \times \max\left(0, \cos\left[\arg\left(\frac{\langle \Psi_T | B(\mathbf{x} - \bar{\mathbf{x}}) | \phi \rangle}{\langle \Psi_T | \phi \rangle}\right)\right]\right), \quad (19)$$

in which an additional factor is introduced to prevent abrupt phase changes during the random walk [38].

D. QC-AFQMC

The introduction of the weight factor in Eq. (19) is an approximation to the imaginary time propagation, and it leads to a bias in energy predictions. This is similar to the fixed node approximation [48] in diffusion Monte Carlo. The quality of the approximation depends on the quality of the trial state, and, more specifically, on the quality of the nodal structure of the trial state. In fact, one could show that if the trial state is the exact ground state, then the weight factor is no longer an approximation, and the predicted energy is also exact. On the other hand, if the trial state is a poor approximation to the ground state, it will introduce a significant bias to the predicted energy [43]. It remains an area of active research to identify the specific characteristics of an approximate trial state that lead to bias reduction.

Unfortunately, only a few trial state options are efficient to prepare on classical computers. The most widely used trial state is the single-determinant Hartree-Fock (HF) state. However, single-determinant wave functions are poor approximations to the ground state for strongly correlated systems, and using them as trial states leads to large errors in energy. In such cases, one needs to use multideterminant trial states [57,58], but the size of the multideterminant expansion grows exponentially with system size, and in practice further approximations need to be introduced to make the approach

efficient, such as truncating the multideterminant expansion to a predefined, limited number of terms, which then leads to errors in the predicted energy.

The development of the quantum computer offers a different perspective for AFQMC. Efficient quantum state preparation techniques [31,59,60] for the ground state of molecules and materials open up a possibility of using them as the trial state for AFQMC. Huggins *et al.* [35] introduced the quantum-classical AFQMC method, in which one prepares the trial state on a quantum computer and uses it to guide the imaginary time propagation in AFQMC. Furthermore, Huggins *et al.* demonstrated that even with a simple perfect pairing wave function prepared on the quantum computer, the QC-AFQMC method yields more accurate energy estimates in both molecules and materials compared with classical AFQMC using single-determinant trial states. The QC-AFQMC approach is compatible with any quantum state preparation technique, including ansätze used in the VQE approach, such as those based on unitary coupled cluster theory [61,62], hardware-efficient circuits [31], and Clifford+T ansatz [60], as well as other states that one could prepare nonvariationally.

E. Classical shadows and the evaluation of overlaps

In order for a quantum state to be used as the trial state in AFQMC, one needs to efficiently evaluate its overlap with the walker states: $\langle \Psi_T | \phi \rangle$. Since both the trial state and the walker state could be prepared on the quantum computer, the most straightforward way for evaluating the overlap is to use the Hadamard test. However, this would require a large number of measurements and back-and-forth communications between the quantum and classical computers, and it also greatly increases the number of qubits, circuit depth (controlled operations), and coherent execution time. These issues make the Hadamard test a very inefficient choice for QC-AFQMC. To tackle this issue, the initial study by Google [35] proposed to use classical shadow techniques. The classical shadow method was first proposed by Huang *et al.* [63,64]. This method tries to store a “classical copy” of the quantum state by measuring the quantum state in a number of random bases. These measurement results are classical shadows of the quantum state, and by inverting the measurement channel, one could reconstruct the quantum state’s density matrix and use it to compute reduced quantities such as expectation values, fidelity, and nonlinear functions. This method is especially advantageous if one wants to predict many properties of a quantum state simultaneously, as it has been shown that one could achieve so with very few measurements that only scale logarithmically with the number of properties.

QC-AFQMC is a scenario well suited for classical shadows, as one needs to evaluate the overlap between the trial state $|\Psi\rangle$ and many different walkers simultaneously. However, although classical shadow is efficient in terms of the number of measurements one needs, to compute properties one needs to invert the measurement channel classically, and whether such an inversion can be done efficiently depends on both the randomized basis and the properties of interest. The initial study of QC-AFQMC used randomized Clifford basis, which turned out to be very inefficient in the classical

postprocessing to compute overlaps in AFQMC, with a cost that scales exponentially with system size [42]. However, a follow-up study shows that such an exponential bottleneck could be reduced to polynomial by switching to a basis generated by randomized matchgate matrices [41,44,65]. The first demonstration of the matchgate shadow-based QC-AFQMC on quantum hardware was demonstrated in 2024 by a team from AWS and the University of Chicago [41] on a system with four qubits.

In order to compute the overlap $\langle \Psi_T | \phi \rangle$ with matchgate shadows, we follow the same procedure described in Wan *et al.* [65]. For the full derivation of working equations, we point the reader to the following literature regarding the matchgate shadow technique [41,65,66] and we briefly state the main results for completeness. The protocol consists of two phases: we first prepare a trial state $|\Psi_T\rangle$ via VQE and construct a new state $|\Psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle^{\otimes N} + |\Psi_T\rangle)$. In this work, the VQE parameters have been optimized on a noiseless quantum simulator. The reason for preparing such a state is because overlaps $\langle \Psi_T | \phi \rangle$ could be formulated as its expectation values with respect to the observable $|\phi\rangle\langle 0|$, defined by $|\phi\rangle$. Second, we perform shadow measurements by composing the state $|\Psi\rangle$ circuit with Gaussian circuits constructed from orthogonal matrices $\mathbf{Q}_p \in \mathbb{R}^{2N \times 2N}$. These matrices can be derived from either Haar-random orthogonal matrices or random signed permutation matrices, with the latter ensuring unit determinant by construction. In our study, signed permutation matrices are used. For each measurement $|b\rangle$ (a single binary string), we obtain a covariance matrix $\mathbf{C}_{|b\rangle}$ that contributes to the classical shadow representation. The overlap computation exploits the mathematical structure of matchgate circuits through efficient Pfaffian evaluation [67]. For a fermionic state represented by determinant p , we construct an antisymmetric matrix:

$$\mathbf{A}_{p|b\rangle}(z) = \mathbf{C}_{|0\rangle}^{(s)} + z \cdot \mathbf{B}_{p,|b\rangle}^{(s)}. \quad (20)$$

Here, $\mathbf{C}_{|0\rangle}^{(s)}$ represents the selected submatrix of the vacuum state covariance matrix. Index (s) indicates that this is a submatrix by selecting certain columns and rows of the original full matrix. z is an auxiliary variable we will use later to perform polynomial interpolation, and $\mathbf{B}_{p,|b\rangle}^{(s)}$ combines the shadow measurement with the walker structure through the transformation:

$$\mathbf{B}_{p,|b\rangle} = \mathbf{W}^* \mathbf{M}_\phi \mathbf{Q}_p^T \mathbf{C}_{|b\rangle} \mathbf{Q}_p \mathbf{M}_\phi^T \mathbf{W}^\dagger, \quad (21)$$

where $\mathbf{W}$ encodes the fermionic structure of the system through carefully constructed block rotations and $\mathbf{M}_\phi$ is the matrix that defines the walker $|\phi\rangle$. The definitions of $\mathbf{W}$, $\mathbf{C}$, and $\mathbf{M}_\phi$ could be found in the Supplemental Material [68]. The overlap between trial state and walker $|\phi\rangle$ is computed by first finding the coefficients of the polynomial defined by the Pfaffian of $\mathbf{A}_{p,|b\rangle}(z) = \sum_x^l c_{p,|b\rangle,x} z^x$, which could be calculated by evaluating the Pfaffian at Chebyshev nodes [44]

$$z_k = \cos\left(\frac{2k+1}{2N}\pi\right), \quad (22)$$

and performing polynomial interpolation to obtain the coefficients. The final contribution to the overlap for this shadow

takes the form

$$\langle \Psi_T | \phi \rangle = \frac{1}{N_{\text{shadow}}} \sum_p o_p \quad \text{and} \quad o_p = \frac{i^{\eta/2}}{2^{N-\eta/2}} \sum_{x=0}^{\ell} c_{p|b\rangle,x} \binom{2N}{2x} / \binom{N}{x}, \quad (23)$$

where $\ell = N - \lfloor \eta/2 \rfloor$ defines the polynomial degree and η is the number of Fermions in the system. The coefficients $c_{p,|b\rangle,x}$ are determined through polynomial interpolation at the Chebyshev nodes. As been shown in Ref. [65], to bound the error of the computed overlap by ϵ , the required number of measurements scales as $\mathcal{O}(\sqrt{N} \ln N / \epsilon^2)$. The overall cost for the above algorithm for computing overlaps scales as $\mathcal{O}(N^{4.5})$ per walker, as computing matrix Pfaffian scales as $\mathcal{O}(N^3)$, and one needs to do it for every grid point and shadow. The number of grid points (Chebyshev nodes) and shadows scale as $\mathcal{O}(N)$ and $\mathcal{O}(N^{0.5})$, respectively, so the overall cost for overlaps is $\mathcal{O}(N^{4.5})$.

F. Force bias and local energy

Despite achieving efficient overlap evaluation through quantum measurements, QC-AFQMC still faces another computational bottleneck: without algorithmic improvements, computing the local energy and force-bias terms would scale as $\mathcal{O}(N^{8.5})$ and $\mathcal{O}(N^{7.5})$ per walker. Each step requires evaluating the local energy:

$$E_L(\phi) = \frac{\langle \Psi_T | H | \phi \rangle}{\langle \Psi_T | \phi \rangle} \quad (24)$$

and the force-bias terms:

$$\bar{x}_\gamma = -\sqrt{\Delta\tau} \frac{\langle \Psi_T | v_\gamma | \phi \rangle}{\langle \Psi_T | \phi \rangle}. \quad (25)$$

The most straightforward way of evaluating the local energy and force bias is to enumerate every term in the Hamiltonian and the v_γ operator, apply each term to $|\phi\rangle$ to obtain a new walker, and evaluate the overlap of the new walker with the trial state. Since the cost for computing one overlap scales as $\mathcal{O}(N^{4.5})$, computing local energies scales as $\mathcal{O}(N^{8.5})$ as there are $\mathcal{O}(N^4)$ terms in the Hamiltonian. For force bias, although there are only $\mathcal{O}(N^2)$ terms in each v_γ , one needs to compute it for every Cholesky vector, and since there are $\mathcal{O}(N)$ Cholesky vectors, the overall cost for force bias scales as $\mathcal{O}(N^{7.5})$.

However, a more efficient algorithm has been developed. As noted by Jiang *et al.* [45], since these quantities represent variations of the overlap under state transformations, they can be computed through (algorithmic) differentiation of the overlap expressions. For example, the force bias could be written as

$$\begin{aligned} \frac{\langle \Psi_T | v_\gamma | \phi \rangle}{\langle \Psi_T | \phi \rangle} &= i \frac{\langle \Psi_T | \sum_{pq} L_{pq}^\gamma a_p^\dagger a_q | \phi \rangle}{\langle \Psi_T | \phi \rangle} \\ &= i \frac{\partial}{\partial \lambda} \ln \langle \Psi_T | e^{\lambda \sum_{pq} L_{pq}^\gamma a_p^\dagger a_q} | \phi \rangle \big|_{\lambda=0} \\ &= i \frac{\partial}{\partial \lambda} \ln \langle \Psi_T | \tilde{\phi}(\lambda) \rangle \big|_{\lambda=0}, \end{aligned} \quad (26)$$

and it becomes the derivative of the trial wave function overlap with a modified walker, with respect to an auxiliary variable λ .

The one-body part of the energy shares the same form as the force bias, and it could be computed in the same way. The two-body part of the energy is

$$\begin{aligned} & \sum_{\gamma} \frac{\langle \Psi_T | \sum_{pqrs} L_{pq}^{\gamma} L_{rs}^{\gamma} a_p^{\dagger} a_q a_r^{\dagger} a_s | \phi \rangle}{\langle \Psi_T | \phi \rangle} \\ &= \sum_{\gamma} \frac{\partial^2}{\partial \lambda_1 \partial \lambda_2} \langle \Psi_T | e^{\lambda_1 \sum_{pq} L_{pq}^{\gamma} a_p^{\dagger} a_q} e^{\lambda_2 \sum_{rs} L_{rs}^{\gamma} a_r^{\dagger} a_s} | \phi \rangle \big|_{\lambda_1=0, \lambda_2=0} / \\ & \quad \times \langle \Psi_T | \phi \rangle \\ &= \sum_{\gamma} \frac{\partial^2}{\partial \lambda_1 \partial \lambda_2} \langle \Psi_T | \tilde{\phi}(\lambda_1, \lambda_2) \rangle \big|_{\lambda_1=0, \lambda_2=0} / \langle \Psi_T | \phi \rangle, \quad (27) \end{aligned}$$

which becomes the second-order partial derivative of the overlap.

These derivatives could be computed using numerical differentiations, and it transforms the computational complexity from $\mathcal{O}(N^{8.5})$ per walker [41,44] to $\mathcal{O}(N^{5.5})$ [45] for local energies, and from $\mathcal{O}(N^{7.5})$ to $\mathcal{O}(N^{5.5})$ [45] for computing all force bias. The efficiency stems from the scaling properties: the overlap and one-body energy terms maintain $\mathcal{O}(1)$ scaling relative to the base overlap computation, while the two-body energy and the number of force-bias terms scale as $\mathcal{O}(N_{\gamma})$ relative to the overlap, where N_{γ} is the number of Cholesky vectors.

We further notice that by using algorithmic differentiation instead of finite difference, one could further reduce the computational cost, especially for force bias. Notice that the derivative of the matrix Pfaffian is

$$\frac{\partial \text{Pf}(\mathbf{A}(\lambda))}{\partial \lambda} = \frac{\text{Pf}(\mathbf{A}(\lambda))}{2} \text{Tr} \left(\mathbf{A}(\lambda)^{-1} \frac{\partial \mathbf{A}}{\partial \lambda} \right) \bigg|_{\lambda=0}. \quad (28)$$

since we only need the derivative at $\lambda = 0$, the matrix Pfaffian and inverse only need to be computed once for each time step, and they could be reused later. For force bias, only the trace needs to be recomputed for each Cholesky vector. The cost for computing the trace of product of inverse and matrix derivative scales as $\mathcal{O}(N^2)$ for each Cholesky vector, so the overall cost for computing all force bias becomes $\mathcal{O}(N^{4.5}) + \mathcal{O}(N_{\gamma})\mathcal{O}(N^2)\mathcal{O}(N^{1.5}) \sim \mathcal{O}(N^{4.5})$, one order of magnitude reduction from numerical differentiations.

The local energy requires the second-order derivatives of the matrix Pfaffian,

$$\begin{aligned} & \frac{\partial^2 \text{Pf}(\mathbf{A}(\lambda_1, \lambda_2))}{\partial \lambda_1 \partial \lambda_2} \\ &= \frac{\text{Pf}(\mathbf{A}(\lambda_1, \lambda_2))}{2} \left\{ \text{tr} \left[\mathbf{A}^{-1}(\lambda_1, \lambda_2) \frac{\partial^2 \mathbf{A}(\lambda_1, \lambda_2)}{\partial \lambda_1 \partial \lambda_2} \right] \right. \\ & \quad - \text{tr} \left[\mathbf{A}^{-1}(\lambda_1, \lambda_2) \frac{\partial \mathbf{A}(\lambda_1, \lambda_2)}{\partial \lambda_1} \mathbf{A}^{-1}(\lambda_1, \lambda_2) \frac{\partial \mathbf{A}(\lambda_1, \lambda_2)}{\partial \lambda_2} \right] \\ & \quad + \frac{1}{2} \text{tr} \left[\mathbf{A}^{-1}(\lambda_1, \lambda_2) \frac{\partial \mathbf{A}(\lambda_1, \lambda_2)}{\partial \lambda_1} \right] \\ & \quad \times \text{tr} \left[\mathbf{A}^{-1}(\lambda_1, \lambda_2) \frac{\partial \mathbf{A}(\lambda_1, \lambda_2)}{\partial \lambda_2} \right] \bigg\} \bigg|_{\lambda_1=0, \lambda_2=0}. \quad (29) \end{aligned}$$

again, since we only need the derivative at $\lambda_1 = \lambda_2 = 0$, the matrix Pfaffian and inverse only need to be computed once for each time step. Unlike force bias, the middle term now requires explicit matrix-matrix multiplication, and algorithmic differentiation provides the same big- $\mathcal{O}$ scaling as numerical differentiation. However, using algorithmic differentiation converts the computational bottleneck from computing matrix Pfaffian to matrix products, which could be easily accelerated with GPUs. The summary of the big- $\mathcal{O}$ scaling improvements and computational bottlenecks with the techniques introduced above is shown in Table I.

G. Virtual correlation energy

In the QC-AFQMC algorithm, the trial wave function is prepared on quantum computers, with a one-to-one mapping from orbitals to qubits. However, for large-scale chemical systems such as the TM complexes, even a small basis set produces hundreds of orbitals. In contrast, most of today's quantum computers only have a few tens of qubits. To tackle this issue, the most common option is to reduce the system to an active space. Since the active space only contains a fraction of the orbitals in the full space, it becomes feasible to prepare its wave function on today's quantum computers.

However, accurate modeling of electron correlations inside the active space is not enough to obtain quantitatively accurate predictions to chemical properties, and one also has to take into account the electron correlations outside the active space. This presents a distinct challenge in QC-AFQMC. Huggins *et al.* [35] introduced the virtual correlation energy (VCE) technique to address this through orbital space partitioning, enabling quantum resources to focus on the active space while treating remaining correlation effects with the trial state of a single Slater determinant. The original proposal had a significant overhead in evaluating the local energy using overlaps, resulting in costs $\mathcal{O}(N^4)$ more expensive than the overlap evaluation. Building on the improvements by Jiang *et al.* [45], our implementation eliminates this overhead, making the method significantly more efficient for large basis sets.

Suppose one has already prepared the trial wave function in the active space $|\Psi_{T,a}\rangle$ (with N orbitals), then we can write the trial wave function in the full space (with N_B orbitals) as

$$|\Psi_T\rangle = |\Xi_c\rangle \otimes |\Psi_{T,a}\rangle \otimes |0_v\rangle, \quad (30)$$

where $|\Xi_c\rangle$ represents frozen-core orbitals as a Slater determinant and $|0_v\rangle$ denotes the virtual orbital vacuum state. One could show that for a walker $|\phi\rangle$ that lives in the full space, its overlap with $|\Psi_T\rangle$ could be computed as

$$\langle \Psi_T | \phi \rangle = \det(\Sigma_c R) \langle \Psi_{T,a} | \tilde{\phi}_a \rangle / (\det(U^{\dagger}) \det(V)), \quad (31)$$

where Σ_c , $\det(U^{\dagger})$, and $\det(V)$ capture the contributions from the core space through matrix operations and R is the renormalization factor. This formulation enables QC-AFQMC to run in the full space with the trial wave function defined in an active space, and therefore capture electron correlations outside the active space. VCE allows for basis set convergence while concentrating quantum resources on the strongly correlated active space. A detailed derivation of Eq. (31) could be found in the Supplemental Material [68].

TABLE I. Big- $\mathcal{O}$ scaling and bottlenecks for computing AFQMC overlaps, force biases, and local energies with matchgate shadows using different techniques.

Algorithm	Overlap	Force bias	Local energy	Bottleneck
Original (Huang <i>et al.</i> [41]) ^a	$\mathcal{O}(N^{4.5})$	$\mathcal{O}(N^{7.5})$	$\mathcal{O}(N^{8.5})$	Pfaffian evaluations
Finite differences (Jiang <i>et al.</i> [45]) ^b	$\mathcal{O}(N^{4.5})$	$\mathcal{O}(N^{5.5})$	$\mathcal{O}(N^{5.5})$	Pfaffian evaluations
Analytic Pfaffian derivatives (this work) ^c	$\mathcal{O}(N^{4.5})$	$\mathcal{O}(N^{4.5})$	$\mathcal{O}(N^{5.5})$	Matrix products

Scalings assume $N_\gamma = \mathcal{O}(N)$ and omit logarithmic factors.

^aThis force-bias scaling is not stated explicitly in Huang *et al.*; it is estimated from term-by-term enumeration of the N_γ Cholesky one-body operators v_γ .

^bOverlap-derivative approach of Jiang *et al.* evaluated here by finite differences; Jiang *et al.* also noted an $\mathcal{O}(N_\gamma)$ speedup from algorithmic differentiation.

^cThe analytic Pfaffian derivatives and contraction restructuring shift the bottleneck from Pfaffian evaluations to matrix products.

III. COMPUTATIONAL DETAILS

A. Workflow

The QC-AFQMC computation proceeds in multiple steps executed on a variety of hardware, as depicted in the chart in Fig. 2. In this section, we discuss each step in detail.

(a) *Analyze input molecule.* The job begins with the specification of a molecular system and its geometry, which then undergoes a classical simulation at the self-consistent field level to obtain a set of molecular orbitals and the molecular Hamiltonian. At this stage, we also determine the size of the active space. See Secs. IIIB and IIIC for details.

(b) *Optimize trial state.* With the Hamiltonian defined, the job enters the state preparation stage. In this work, we obtain the QC-AFQMC trial state using VQE with the unitary pair coupled cluster doubles (upCCD) ansatz [69]. Using CUDA-Q, we construct the kernel for the trial VQE state and variationally optimize the VQE parameters, which is executed on an ideal simulator.

(c) *Collect shadows.* Next, the trial state is sampled using matchgate shadows on a quantum computer. Matchgate shadow circuits are constructed as CUDA-Q kernels and submitted for execution on the IonQ Forte quantum processing unit (QPU) from the Amazon Braket Hybrid Job environment. Upon collecting the measurements, the workflow aggregates them to prepare for postprocessing.

(d) *Compute properties.* Finally, the workflow finishes with performing AFQMC imaginary time propagation as a postprocessing step to evaluate the energy of the molecule.

All electronic structure calculations were performed using the PySCF [70] framework, interfaced with LibXC [71] for density functional theory evaluations. Orbital entanglement analysis employed the Block2 density matrix renormalization group software [17], accessed through the open-source ActiveSpaceFinder package [72]. Quantum Monte Carlo calculations utilized iPie [52,73], with our matchgate shadow implementation building upon routines from the open-source symmetry-adjusted-classical-shadows repository [66,74,75]. More implementation details on the

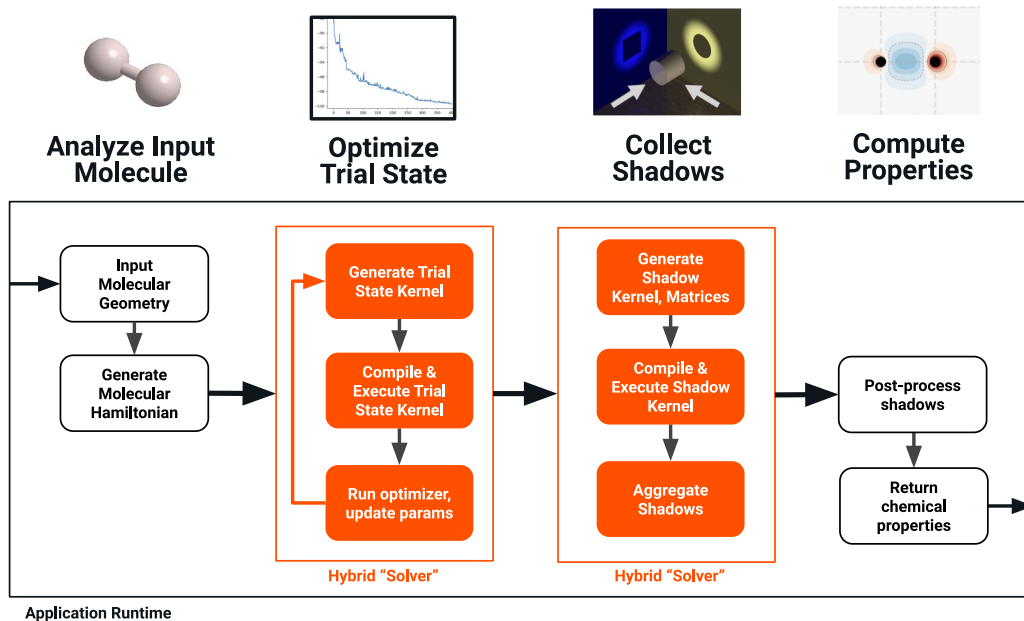


FIG. 2. Diagram of the computational workflow for QC-AFQMC. The hybrid steps indicated in orange are executed using a combination of QPU and CPUs. White boxes indicate classical computations using a combination of CPUs and GPUs.

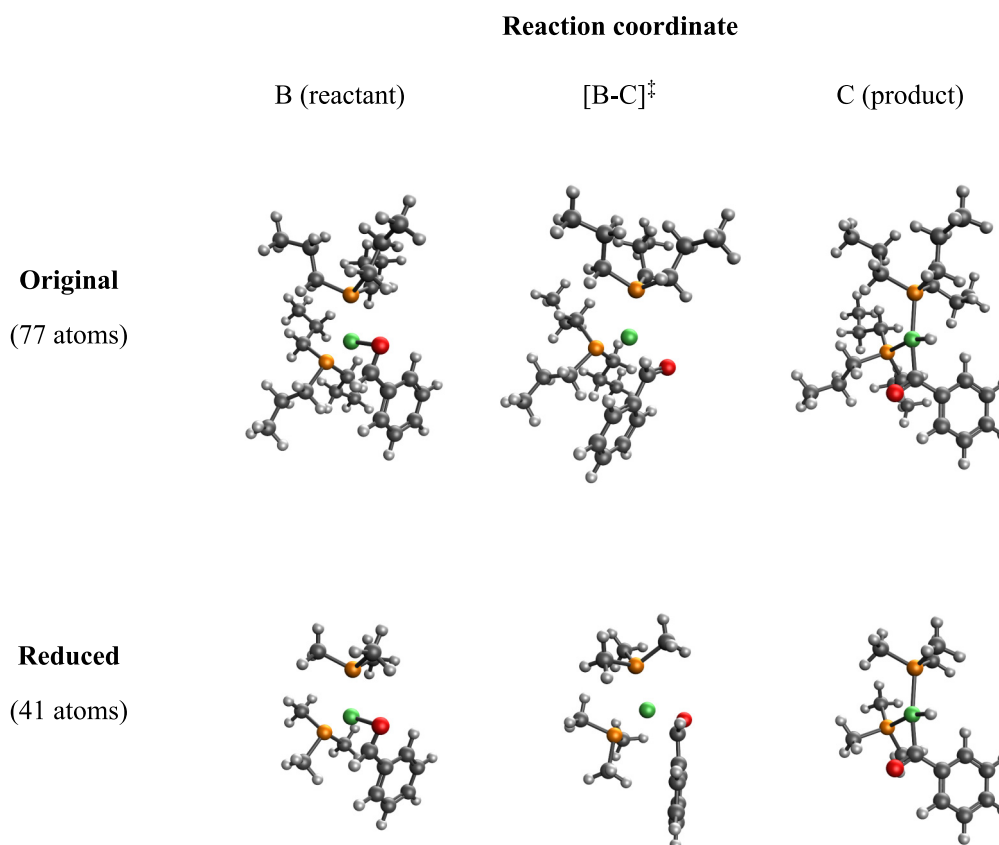


FIG. 3. Three-dimensional structures of the molecular system at each reaction coordinate (B, $[B-C]^\ddagger$, C) and truncation level. Atoms are colored as follows: nickel (green), phosphorus (orange), oxygen (red), carbon (gray), and hydrogen (white).

QC-AFQMC workflow can be found in the Supplemental Material [68].

B. Model chemical reaction

For this demonstration, we selected the oxidative addition step of the nickel-catalyzed deformylative Suzuki-Miyaura cross-coupling reaction shown in Fig. 1. We used the molecular geometries of the complexes B, $[B-C]^\ddagger$, and C from Ref. [47]. To make the demonstration computationally tractable, we applied two approximations that reduce the computational effort while retaining relevant properties. First, we truncated the molecular structures from 77 atoms to 41 atoms, which reduces the computational cost of postprocessing. On top of that, we selected an active space to reduce quantum computing requirements.

The truncation of molecular structures was carried out following established protocols in computational organometallic chemistry [76], beginning from the original, 77-atom reference system. Peripheral substituents that do not participate in the reaction mechanism were excised via selective cleavage of aliphatic C–C bonds at sp^3 -hybridized centers, with $-\text{CH}_2\text{CH}_2\text{CH}_3$ groups replaced by $-\text{CH}_3$ substituents. This strategy preserves the integrity of the P coordination sphere while avoiding the disruption of critical C–P bonds. The original and reduced molecules are shown in Fig. 3, and atomic cartesian coordinates could be found in the Supplemental

Material [68]. In this study, we will be using the reduced 41-atom systems in QC-AFQMC calculations.

To validate this molecular reduction, we first conducted density functional theory calculations using four distinct functionals (ω B97X [77], B3LYP [78–81], PBE0 [82–84], and M06-2X [85]) with the Slater-type orbital approximated by three Gaussians (STO-3G) basis set [86–90] (which is sufficient for the purposes of this demonstration; we discuss the scalability of QC-AFQMC with the basis set in Sec. IV). The reaction energy profiles (Figs. S1–S4 in the Supplemental Material [68]) show remarkable consistency across all truncation levels we considered. Energetic differences (Table SI in the Supplemental Material [68]) were on the order of the expected statistical error margins of AFQMC (~ 1 – 2 kcal/mol).

C. Active-space determination

Having verified that our reduced models preserve the energetics across different DFT functionals, we then used tools from quantum information theory to analyze the electronic structure, which provides quantitative indicators of orbital correlation. This framework employs two complementary measures derived from DMRG wave functions: single-orbital entropy and orbital-orbital mutual information. Single-orbital entropy quantifies the quantum entanglement between individual orbitals and their environment—approaching a theoretical maximum of $\ln(4) \approx 1.4$ at complete

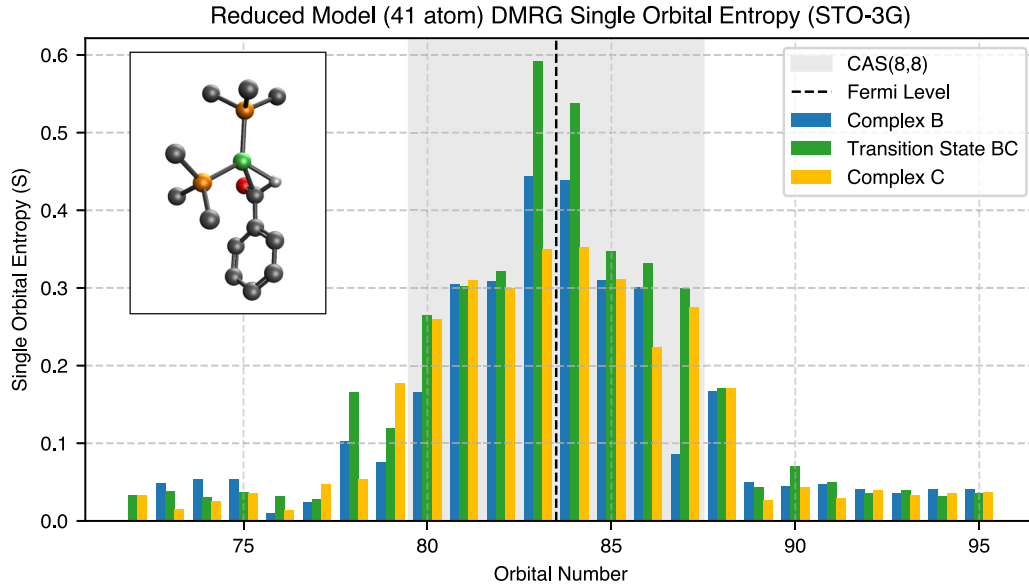


FIG. 4. Single-orbital entropy for the reduced model (41 atoms). The major entangled orbitals remain largely unchanged, validating the truncation approach.

entanglement. Previous studies [91,92] have established that orbitals exhibiting entropy above 0.14 (10% of maximum) reliably indicate strong static correlation effects.

Applying this analysis to the molecules in Fig. 3 revealed striking patterns. As one could see in Fig. 4 (additional plots available in the Supplemental Material [68]), for the reduced systems, single-orbital entropy profiles across the reaction coordinate (initial complex B, transition state [B-C][‡], and product C) identified a pronounced cluster of frontier orbitals (orbital numbers 80–87, shown in Fig. 4) maintaining high entanglement throughout. These orbitals also exhibit strong mutual correlation patterns in their network structure (Fig. S7 in the Supplemental Material [68]), with the profiles remaining consistent across all truncation levels—providing independent validation of our reduction strategy.

The correlation network analysis, derived from two-electron reduced density matrix (RDM) elements (Fig. S7 in the Supplemental Material [68]), quantifies these orbital-orbital interactions through edge weights representing correlation strengths. The resulting network topology naturally suggests a hierarchy of active spaces based on established metrics [91]: from a primary (12e, 11o) space capturing orbitals with entropy $S > 0.1$, through an intermediate (8, 8) space for orbitals exceeding $S > 0.2$, to a minimal (6, 6) space for the most strongly correlated orbitals ($S > 0.3$). From this hierarchy, we selected the (8, 8) active space, shown in Fig. 4, as it provides a balanced description of the key correlations for the demonstrations in this study. Additional correlations are captured by the virtual correlation energy procedure, as discussed in Sec. II G.

These systematically constructed active spaces serve as trial states for QC-AFQMC calculations, with residual dynamical correlation handled efficiently through the VCE technique [35]. This balanced treatment proves particularly valuable for nickel-catalyzed C-H activation reactions, where accurate correlation treatment directly impacts mechanis-

tic understanding. The integration of quantum information metrics with QC-AFQMC methodology thus establishes a robust framework for investigating electronic structure reorganization in transition metal catalysis, especially in cases where traditional single-reference methods prove inadequate.

D. VQE trial state

In our study, we use the VQE algorithm with the upCCD ansatz as the trial state for QC-AFQMC. The upCCD ansatz is

$$|\Psi\rangle = e^{T-T^\dagger} |\Phi_0\rangle, \quad T = \sum_{ia} t_i^a a_{a\alpha}^\dagger a_{a\beta}^\dagger a_{i\beta} a_{i\alpha}, \quad (32)$$

in which $|\Phi_0\rangle$ is the HF state. $a_{p\alpha}^\dagger$ ($a_{p\beta}^\dagger$) and $a_{p\alpha}$ ($a_{p\beta}$) are the fermionic creation and annihilation operators in the p th spin-up (down) orbital. For compactness, $|\Psi\rangle$ will refer to the active-space trial state $|\Psi_{T,a}\rangle$ in Eq. (30).

As we have shown before [69], the upCCD ansatz is one of the most efficient VQE ansatz. It could be compiled with shallow circuits that contain only $O(N^2)$ entangling gates, and only a constant number of measurements are needed to compute energy. In this work, we obtain the optimal parameters of the upCCD ansatz by minimizing the energy with an ideal quantum simulator using the COBYLA optimizer implemented in CUDA-Q. For such a small system, the VQE simulation only takes a few minutes. However, we also note here that for a large-scale system, simulating VQE with an ideal simulator becomes infeasible, and parameter optimizations on quantum computers are also very challenging. Fortunately, there are methods [93,94] to obtain the optimal or close-to-optimal ansatz parameters classically without doing parameter optimizations on quantum computers. For upCCD, the most straightforward way is to use the optimal parameters of the classical pair coupled cluster doubles (pCCD) ansatz. Solving optimal parameters of pCCD only scales cubically

with the system size, which makes it a scalable way of preparing upCCD trial states for QC-AFQMC.

E. Execution of matchgate circuits on QPU

(a) *Quantum hardware.* The experimental demonstration was performed on the IonQ Forte QPU executed via NVIDIA CUDA-Q [95–97] on Amazon Braket [98]. In these systems, trapped $^{171}\text{Yb}^+$ ions serve as the qubits, with quantum information encoded in two hyperfine levels of the ground state. Ions are generated via laser ablation and selective ionization before being loaded into a surface linear Paul trap in a compact integrated vacuum package. Qubit states are manipulated by illuminating individual ions with pulses of 355 nm light that drive two-photon Raman transitions, thereby enabling the implementation of arbitrary single-qubit rotations and ZZ-type entangling two-qubit gates.

IonQ Forte integrates acousto-optic deflectors that allow for independent steering of each laser beam to its respective ion, substantially reducing beam alignment errors across the ion chain. This optical architecture, combined with a robust control system and control software suite that automates calibration and optimizes gate execution, has enabled the realization of larger qubit registers with enhanced gate fidelities. Consequently, IonQ Forte establishes a scalable and high-fidelity platform for quantum information processing, marking a notable progression in trapped ion QPU technology.

Several specific improvements were made to the control software and firmware suite to maximize the circuit execution rate (*throughput*) of circuits for this workload. This performance tuning took advantage of shadow circuits' specific and somewhat uncommon structure. Notably, each shadow circuit is pregenerated, only run for a single shot, and consists of a common trial state, followed by classical shadows sourced from a common pool. This allowed for aggressive caching of waveforms and other real-time instructions that were common among the shadow circuits within the real-time control subsystem, as well as enhanced pipelining and parallelization efforts up and down the control stack to ensure that the now-much-faster circuit execution was never blocked by classical overhead. Specific results of this performance tuning are discussed in Sec. IV C.

(b) *Error mitigation techniques.* Classical shadow techniques enjoy the benefit of responding predictably to Markovian, invertible, gate-independent quantum error channels, in the sense that expectation values are simply rescaled [75,99,100]. For computing overlaps, as done in QC-AFQMC, the rescaling of expectation values cancels out, making the technique robust to these simple errors [35]. This robustness does not hold, however, for cases where the error channels are context dependent or noninvertible [101]. On our platform, one particularly important such channel is a spontaneous emission event that can happen during the Raman ZZ gates. This error channel results in the ion entering a state where it is insensitive to all further gates [102,103], meaning that the placement of the gate where the error occurs drastically changes its impact. Further, given that the error represents a complete loss of information, the resulting shadow channel is noninvertible.

While this type of spontaneous emission error has the potential to inject nontrivial bias into a shadow evaluation of overlaps, it is fairly easy to detect and remove via postselection. To that end, we use leakage error detection gadgets [104] attached at the end of the circuit. Each gadget requires two ZZ gates and an ancillary qubit. To reduce the overhead, we attach the gadgets on half of the qubits, alternating even and odd, thus requiring eight additional qubits. If any ancilla is flagged, the corresponding measurement is discarded.

F. GPU-accelerated postprocessing

We implement key linear algebra routines in the QC-AFQMC postprocessing with NVIDIA GPUs on AWS ParallelCluster. The construction of the **A** and **B** matrices defined in Eqs. (20) and (21), along with the derivative of **A** used in force bias and local energy evaluations, and matrix and vector products needed to compute Pfaffian derivatives in Eqs. (28) and (29), are implemented with NVIDIA's cuBLAS and cuTENSOR linear algebra packages. cuBLAS allows for highly efficient matrix multiplications on GPUs. cuTENSOR extends those capabilities to multidimensional arrays. The polynomial fitting needed to compute overlaps and their derivatives is implemented with NVIDIA's cuSOLVER package, which allows for efficient linear system solutions on GPUs.

IV. RESULTS

A. Particle number and effect of hardware noise

We first prepared the trial states of **B**, $[\text{B-C}]^\ddagger$, and **C** (Fig. 3), which represent the reactant, transition state, and product, respectively, along the reaction coordinate for the oxidative addition of aldehyde to Ni(0) in the nickel-catalyzed Suzuki-Miyaura cross-coupling. For each molecule, we obtained a restricted Hartree-Fock solution followed by VQE parameter optimization with the upCCD ansatz in an active space. We followed the procedure described in Sec. III B to define an (8, 8) active space for all three systems, which corresponds to selecting orbitals with single-orbital entropy $S > 0.2$.

To sample the trial state, we set up matchgate shadow tomography on the ideal simulator with 58 482 randomly generated shadows. On Forte QPU, we used 60 365, 58 375, and 61 118 matchgate shadows for system **B**, $[\text{B-C}]^\ddagger$, and **C**, respectively, following the error mitigation procedure described in Sec. III E. The difference in the number of shadows is due to a slightly different rate of rejection at the error mitigation step.

We first assessed the quality of the collected matchgate shadow measurements by evaluating the total number of particles in the trial state by taking the trace of the one-particle reduced density matrix, which was computed from the measured matchgates. The details of how RDMs are computed could be found in the Supplemental Material [68]. Because we used the (8, 8) active space in our models, the ground truth number of electrons is known (8), making it a convenient metric to validate the measurements. The results from the ideal simulator and the Forte QPU for all three molecules are shown in Table II. The particle numbers measured from

TABLE II. Particle number of the trial state evaluated from matchgate shadow measurements computed for the three molecular structures. There are eight active-space electrons in the trial state, which is the ground truth value.

Molecule	Ideal simulator		Forte QPU	
	Shadows	Particle number	Shadows	Particle number
B	58 482	7.998	60 365	10.663
[B-C] [‡]	58 482	8.086	58 375	10.598
C	58 482	7.932	61 118	10.514

the ideal simulator of all three molecules are close to 8, as expected, with the remaining discrepancy with the integer value due to the finite sample of matchgates. The measured particle numbers on Forte QPU are around 10.5, which deviate from the ideal value due to hardware noise. The amount of difference between the particle number measured on QPU and the ideal simulator could also serve as an indicator of the noise level.

B. Reaction energy

We then applied QC-AFQMC to model the ground state of the three molecules using both simulated and QPU measurements. The imaginary time propagation of AFQMC, with force bias and energies computed using the algorithm described in Sec. II F, ran with 1280 walkers and 1500 time steps of 0.01 a.u. for a total of 15 a.u. imaginary time. We plot the energy as a function of the imaginary time in Fig. 5. A comparison of the final converged QC-AFQMC energies, including their 1σ statistical uncertainties, for both ideally

simulated matchgates and those collected on the Forte QPU, is shown in Table III. We also compare QC-AFQMC with a baseline of electronic structure methods such as HF, VQE, and CCSD(T). To provide a direct comparison with its classical counterpart, we also ran classical phaseless AFQMC (ph-AFQMC) using a single-determinant restricted Hartree-Fock (RHF) trial state and the same postprocessing protocol as QC-AFQMC (50-block burn-in, outlier cleaning, reblocking).

For all three molecules, the initial energies of QC-AFQMC closely match the VQE results, which is expected since the initial step represents a sample of the trial state. As imaginary time evolution proceeds, QC-AFQMC recovers between 1.24 and 1.31 hartree of correlation energy. This reflects the method's ability to capture correlation effects within the active space through the trial state, while also accounting for electron correlation outside the active space via VCE. As shown in Fig. 5, we occasionally observe spikes in the measured energy with a magnitude of almost 2 hartrees above the converged value. We attribute these anomalies to intermittent numerical instabilities in the VCE evaluation. Specifically, one needs to differentiate VCE-related quantities, such as matrices U , V , and R , to compute force bias and local energies with VCE. We used finite differences, which proved unstable for small step sizes. Fortunately, the spikes quickly subside and do not affect overall convergence. Moreover, such a problem is easy to resolve by performing differentiation of VCE analytically, which will be tackled in a future study. For our final energy estimates, we discard the first 50 blocks for equilibration, then remove any remaining outliers (outliers are defined to be points that are at least 200 mhartree higher or lower in energy than its adjacent points) from the next 100 blocks, and use the rest for reblocking analysis to obtain the mean energy and its error bars [105]. Unless noted otherwise, values in

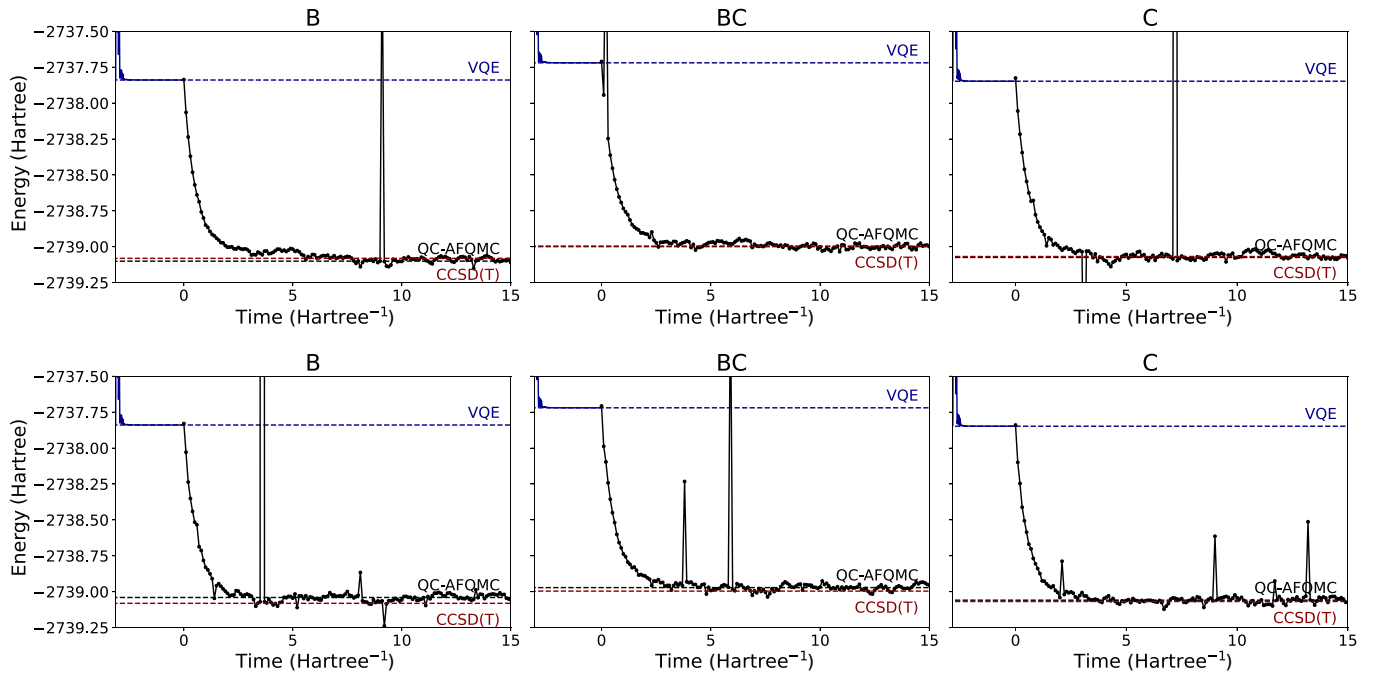


FIG. 5. Convergence of the QC-AFQMC energy vs imaginary time from matchgate shadow measurements on the ideal simulator (top) and the Forte QPU (bottom). The dashed lines indicate the VQE/upCCD (blue) and CCSD(T) (red) reference energies, as well as the converged QC-AFQMC energy (black).

TABLE III. Total electronic energies computed for the three molecular structures. All energies are in hartrees.

Method	B	[B-C] [‡]	C
RHF	−2737.810 245	−2737.686 838	−2737.822 695
CCSD ^a	−2739.045 232	−2738.952 254	−2739.040 565
CCSD(T) ^a	−2739.081 251	−2738.996 283	−2739.068 637
VQE/upCCD	−2737.839 079	−2737.719 325	−2737.847 562
ph-AFQMC ^b			
Initial	−2737.809	−2737.685	−2737.821
Converged	−2739.074(3)	−2738.975(5)	−2739.096(5)
QC-AFQMC (ideal simulator)			
Initial	−2737.836	−2737.710	−2737.825
Converged	−2739.085(4)	−2738.994(4)	−2739.064(4)
QC-AFQMC (Forte QPU)			
Initial	−2737.829	−2737.707	−2737.839
Converged	−2739.040(4)	−2738.972(3)	−2739.059(3)

^aCCSD and CCSD(T) energies computed with 29 frozen orbitals (frozen-core approximation).

^bUsing a single RHF Slater determinant as a trial state.

parentheses denote $\pm 1\sigma$ statistical uncertainties from AFQMC sampling estimated by reblocking. Systematic effects (i.e., basis set incompleteness, solvation model, relativistic approximations, etc.) are not included.

Unlike the particle number via one-particle RDM, the energy is much more resilient to hardware noise. Because hardware noise causes an overestimate of the number of electrons by more than 2, one might intuitively expect that it could result in errors in the energy on the order of several hartrees. However, the largest error we observed in energy between the ideal simulator and the QPU is only a few tens of mhartree. This disparity arises from fundamental differences in the mathematical formalisms used to evaluate energies and RDMs. Energy is computed as ratios of overlaps, a structure known to be robust to noise due to error cancellation, as pointed out by Huang *et al.* [41]. Conversely, individual one-particle RDM elements share a similar formalism as single overlaps, which do not benefit from the same error cancellation mechanism.

Table IV shows the reaction barrier computed with various methods. As a reference, we use CCSD(T), which is widely regarded as reliable for systems with moderate correlation and has recently demonstrated good accuracy for spin ener-

getics in mononuclear iron complexes [14]. As can be seen in Table IV, among single-reference methods, RHF severely overestimates the reaction barrier, while DFT (which includes the treatment of solvent) exhibits its typical behavior of underestimating the reaction barrier.

As a classical baseline, we performed ph-AFQMC with a single-determinant RHF trial, applying the same protocol. In total energies, ph-AFQMC is close to CCSD(T): $\Delta = +7.3$ mhartree (B), $+21.3$ mhartree (TS), and -27.4 mhartree (C), where $\Delta = E^{\text{ph}} - E^{\text{CCSD(T)}}$. Despite this, the barriers deviate substantially from CCSD(T): $+8.7$ and $+30.6$ kcal/mol for $B \rightarrow [B-C]^{\ddagger}$ and $C \rightarrow [B-C]^{\ddagger}$, respectively. Consequently, the B versus C minima ordering follows HF ($C < B$) rather than CCSD(T) ($B < C$). One explanation for such a contradiction is that the RHF trial state fails to provide balanced descriptions of the three systems, which then leads to imbalanced approximation qualities of the nodal structures. The quality of the nodal structure determines the imaginary time propagation and the converged energy of QC-AFQMC, which eventually causes the difference in relative energies between ph-AFQMC and CCSD(T). We used a single-determinant RHF trial as a conservative classical baseline and to isolate trial-quality effects. While multideterminant classical trials (e.g., selected-CI or CASSCF) can reduce phaseless-constraint bias, their size typically grows exponentially with active-space size, reintroducing the scaling issue discussed previously.

QC-AFQMC based on the ideal matchgate samples produces a reaction barrier profile that closely resembles that from CCSD(T) with barriers within 4 kcal/mol from the reference. Compared with the ph-AFQMC results, it demonstrates the importance of using high-quality trial state in AFQMC for more accurate predictions of chemical properties. In contrast, QC-AFQMC based on noisy Forte QPU samples differs by up to 10 kcal/mol compared to CCSD(T). Furthermore, in this case the noisy matchgate shadows cause the energy barriers to shift in the opposite directions, changing the picture qualitatively: unlike CCSD(T) and ideal QC-AFQMC, which place molecule B lower in energy than C, noisy QC-AFQMC predicts C to be lower in energy. This result highlights the importance of developing further error mitigation techniques

TABLE IV. Reaction barriers of the oxidative addition reaction estimated with RHF, CCSD(T), and QC-AFQMC. All energies are in kcal/mol.

Method	B $\rightarrow$ [B-C] [‡]	C $\rightarrow$ [B-C] [‡]
DFT ^a	21.4	7.1
RHF	77.4	85.3
CCSD(T)	53.3	45.4
VQE/upCCD	75.1	80.5
ph-AFQMC ^b	62(4)	76(5)
QC-AFQMC (ideal simulator)	57(4)	44(4)
QC-AFQMC (Forte QPU)	43(3)	55(3)

^aFree energies in solution [solvation model based on density (SMD) (1,4-dioxane)] evaluated with M06/Def2-TZVPP// ω B97xD/Def2-TZVP(Ni)/Def2-SVP(nonmetal) [47].

^bUsing a single RHF Slater determinant as a trial state.

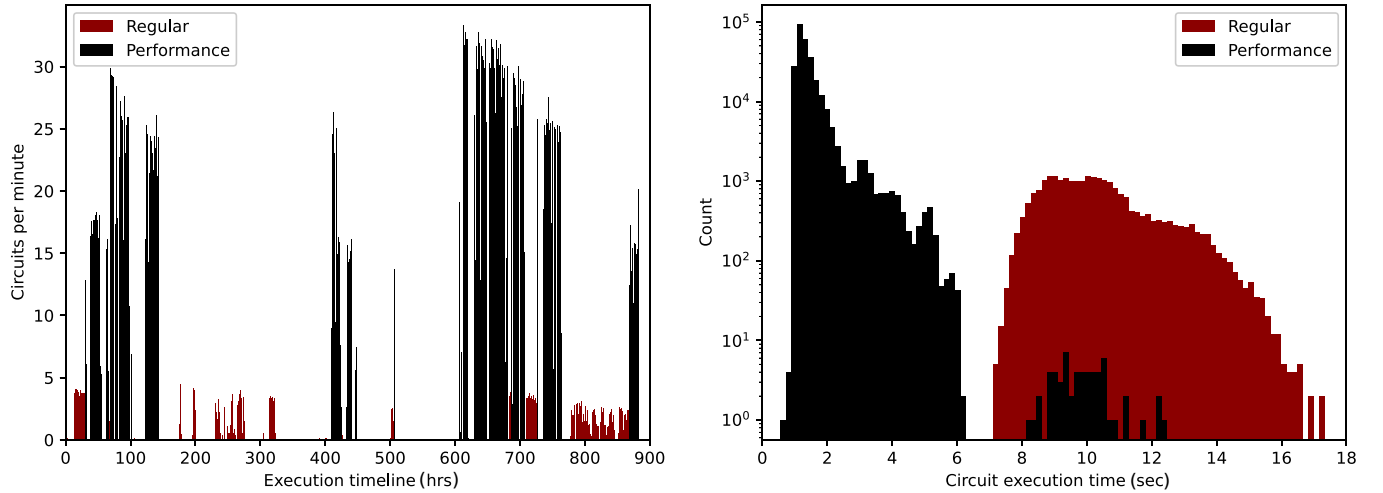


FIG. 6. Rates (per minute) of processing matchgate shadow circuits on Forte QPU over the entire execution timeline (left panel). Distribution of matchgate shadow circuit execution times (seconds) before and after performance tuning of Forte QPU for the workload (right panel).

[75] for QC-AFQMC on noisy hardware. The uncertainty interval in QC-AFQMC reaction barriers (up to ± 4 kcal/mol) is due to the statistical sampling of AFQMC. It can be improved by increasing the number of AFQMC walkers or the length of imaginary time propagation (with a proportional increase in computational effort in postprocessing), as well as by enhancing the quality of the trial state.

Using larger basis sets, ideally approaching the complete basis set limit, is expected to improve electron correlation treatment and may slightly increase barrier heights. Additionally, environmental effects must also be considered [14,106], as interactions with solvents and additives can significantly influence spin-state energetics. As the nuclear charge of TM increases, relativistic effects, such as kinematic effects and spin-orbit coupling, become more pronounced. These effects can be addressed with varying levels of accuracy and approximation using, for example, Zeroth-Order Regular Approximation [107], Douglas-Kroll-Hess method [108], and relativistic effective core potentials [109].

C. Time to solution with Forte QPU

1. Quantum processing time

In the course of collecting matchgate shadow measurements on Forte QPU, we executed 300 983 circuits in total, including a number of trial runs and results that were later discarded in error mitigation postselection. Those circuits were executed over a period of several weeks, during which this workload was mixed with other shared workloads on the system. Additionally, during certain periods of time, the QPU was performing technical tasks such as calibration and characterization.

Figure 6 (left panel) shows the rate of executing the circuits on Forte over the entire execution timeline. Nonpertinent workloads are not shown in the chart. During the run, we switched between the regular and performance-tuned processing regimes, which now allows us to directly compare the rate of execution, thus evaluating the effect of problem-specific enhancements implemented on Forte QPU's control software

and firmware (see Sec. III E for further details). The right panel of Fig. 6 depicts the distribution of individual circuit execution times with and without performance tuning. While this execution time captures all the essential execution steps, it does not count overheads such as circuit submission and waiting in the queue.

It is now possible to summarize the timing results in Fig. 6 and distill them to an execution speedup achieved through the performance tuning of Forte QPU. Table V shows the median circuit execution time with and without performance tuning permitting direct comparison. While a matchgate shadow circuit executes, on average, in 9.9 sec in the regular regime, this time reduces to 1.1 sec with all the performance improvements activated, yielding a $9\times$ speedup. Had we run this entire workload without any interruption and without the performance improvements, it would have taken us almost 35 days to complete. In the performance regime, all circuit measurements would have completed in under 4 days. This demonstrates the importance of performance tuning (and, ultimately, codesign) of the QPU to specific workloads.

There are many avenues for further reducing execution time. Developing new, economical trial state ansätze and optimizing matchgate circuit generation could make preparation more efficient. Improved error mitigation strategies would allow for faster execution rates. Advances in classical shadow techniques tailored to electronic structure might yield more sample-efficient shadows, lowering the total number of required circuits. Finally, faster QPU hardware and parallel measurement scheduling across multiple devices would further accelerate the overall workflow.

2. High performance computing (HPC) postprocessing time

The second key contributor to the time to solution of QC-AFQMC is the classical postprocessing step, which converts matchgate measurements to QC-AFQMC energies. We executed the postprocessing workload on AWS ParallelCluster equipped with NVIDIA H100 and H200 GPUs. Because of difficulties obtaining a large allocation of GPU-enabled in-

TABLE V. Evaluation of the speedup factor due to tuning the performance of Forte QPU for the QC-AFQMC workload.

Regime	Median execution time (sec)	Number of circuits	Projected total time (days) ^a	Speedup
Regular	9.9	24 101	34.5	
Performance	1.1	276 882	3.8	9×

^aAssuming the workload ran uninterrupted in this regime.

stances at once, postprocessing took place in stages across various resources. Despite the fragmented execution, normalized timings exhibit remarkable consistency (Table VI): using P5-type instances with NVIDIA H100, full processing takes 0.29 GPU h per shadow (179 858 shadows in total across the three molecules); on P5en-type instances with higher-performance NVIDIA H200s, processing time reduces to 0.22 GPU h per shadow. This translates to about 17 400 and 13 200 GPU h per molecule on H100 and H200, respectively.

We will now compare the timings attained by the improved postprocessing algorithm with the performance of the state-of-the-art QC-AFQMC implementation by Huang *et al.* [41] to quantify computational speedups. There are several major differences in the implementation and problem setup between the two demonstrations. These are shown in a side-by-side comparison in Table VII. For example, size of the problem, number of qubits, number of shadows and walkers, accounting for the active space, and hardware used are all different between the two studies.

Huang *et al.* provide run-time projections for a series of increasingly larger systems, including water, benzene, and the chromium dimer, based on a simple performance model fitted to observed timings. For example, the projected run-time per step for the benzene molecule with 12 qubits on a million CPU cores is 6 h. We will use these projections in order to compare the performance of the two algorithms, making appropriate adjustments for the implementation differences. Because it is difficult to properly account for VCE, Table VIII shows two comparisons, with and without VCE processing time. Ignoring the VCE correction puts this work at a disadvantage, assuming that it is free for Huang *et al.*'s algorithm, which yields a lower-end estimate of the speedup at 656. Naively correcting for VCE adds to the cost of Huang *et al.*'s algorithm, but it is likely a significant overestimate; therefore, we treat the resulting speedup of 4.58×10^7 as a higher-end estimate. The time per step of our 16-qubit imaginary time propagation is only 1.8 min, achieved with 320 NVIDIA H200 GPUs (benchmarked separately). Comparing this with a 6 h projection for a 12-qubit system and one million CPU cores,

our work achieves several orders of magnitude improvement, significantly enhancing the practicality of the QC-AFQMC algorithm.

The details are shown in Table VIII, where we compare the amount of time needed to compute one local energy with a single shadow measurement between the two implementations. In Huang *et al.*, it takes 1 min to process 16 000 shadow circuits for each walker (roughly 10 unique measurement outcomes per circuit with 1024 shots, yielding 160 000 shadow measurements in total), so that the time per shadow measurement is 3.75×10^{-4} sec. Projecting this to a 16-qubit system (an 8-electron, 8-orbital active space) with big- $\mathcal{O}$ scaling as $\mathcal{O}(N_q^8)$, we obtain the projected time of 24.58 sec. The $\mathcal{O}(N_q^8)$ scaling arises from two factors: first, there are $\mathcal{O}(N_q^4)$ terms in the Hamiltonian; and second, it costs $\mathcal{O}(N_q^4)$ to process one shadow measurement for each term. We then further project it to include VCE, and the big- $\mathcal{O}$ scaling of VCE becomes $\mathcal{O}(N_B^4 N_q^4)$, in which N_B is the total number of orbitals in the full orbital space. With this taken into account, for a system with 130 spatial orbitals (the same as the systems studied in this work), we arrive at the final projection of $t_1 = 1.71 \times 10^6$ sec per step per shadow. The algorithm implemented in this work solves the problem of the same size but with 179 858 shadows with $t_2 = 7.5 \times 10^{-4}$ sec per step per shadow on an NVIDIA H200 GPU, which translates to $t_2 = 3.7 \times 10^{-2}$ sec of CPU time under the assumption of $50\times$ GPU-over-CPU acceleration. As shown in Fig. 7, the ratio between the classical processing duration for Ni complexes is $t_1/t_2 \simeq 2 \times 10^9$. Even when adjusted for GPU acceleration, the ratio is still on the order of 10^7 speedup. This dramatic computational cost reduction is primarily driven by the reduced asymptotic algorithmic cost achieved through automatic differentiation and the use of Cholesky decomposition of Hamiltonian. As shown in Fig. 7, the combination of the algorithmic improvement and the usage of GPUs put the postprocessing of even larger chemical systems in practical regime.

Finally, we analyze the scalability of QC-AFQMC with the size of the basis set. For a fixed active space, the number of qubits is independent of the basis set size, and so is the quan-

TABLE VI. Details of computational resources used for GPU-accelerated QC-AFQMC postprocessing on AWS for all three molecules and 1280 walkers.

Molecule	AWS instance type	Instances	GPUs	Median time per block (min)	Projected total ^a GPU hours per shadow
B	P5 (H100)	10	80	86.6	0.29
[B-C] [‡]	P5 (H100)	5	40	166.6	0.29
[B-C] [‡]	P5 (H100)	20	160	44.1	0.30
C	P5en (H200)	5	40	135.4	0.22
C	P5 (H100)	20	160	44.4	0.29

^aFor 150 blocks of QC-AFQMC.

TABLE VII. Side-by-side comparison of the demonstration in this work with the current state-of-the-art result by Huang *et al.* [41].

	Ref. [41]	This work
Molecular system	Nitrogen vacancy center in diamond	Ni catalytic complexes
Number of basis functions		130
Number of fields		1169
Active space	(4, 3)	(8, 8)
Number of qubits	4	16
Frozen orbitals	Included in core energy	Included via VCE
Trial state	VQE/UCCSD	VQE/upCCD
QPU	IonQ Aria	IonQ Forte
Number of shadows	4000	60 365 (B) 58 375 ([B-C] [‡]) 61 118 (C)
Number of measurements per shadow	10	1
Number of walkers	4800	1280
Imaginary time step	0.4 hartree ⁻¹	0.01 hartree ⁻¹
Number of time steps	400	1500 (150 blocks, 10 steps each)
Total propagation time	160 hartree ⁻¹	15 hartree ⁻¹
Postprocessing resources	4800 CPU cores	40–160 NVIDIA H100 and H200 GPUs

tum part of time to solution. However, the situation is different for classical postprocessing. As discussed in Sec. II F, the cost of computing force bias and local energy scales as $\mathcal{O}(N^4)$ and $\mathcal{O}(N^5)$ per shadow measurement. To be more specific, the more precise scaling is $\mathcal{O}(N_B N_q^3)$ and $\mathcal{O}(N_B N_q^4)$, respectively, in which N_B is the size of the basis set and N_q is the number of qubits. Therefore, the cost of postprocessing scales linearly with the size of the basis set.

V. CONCLUSIONS

We demonstrate the operation of an end-to-end workflow based on the QC-AFQMC method to model the reaction barrier of a step of the nickel-catalyzed Suzuki-Miyaura reaction. Operating in a high-performance hybrid quantum-classical compute environment, the workflow includes state preparation

on a quantum circuit simulator, sampling of the trial state on the IonQ Forte QPU via the matchgate shadow protocol, and execution of the QC-AFQMC imaginary time propagation on GPU-accelerated cluster on AWS. While previous studies of QC-AFQMC concluded that the method is too computationally demanding to be of practical use, we demonstrate that it in fact could be applied to problems as large as catalytic transition metal complexes. This is achieved through the use of the active-space approximation and virtual correlation energy. The accuracy of estimating the reaction barriers of the chemical reactions with QC-AFQMC + VCE based on ideally sampled matchgates is within the uncertainty interval of AFQMC (± 4 kcal/mol) relative to the reference CCSD(T) method. With QPU sampled matchgates, the discrepancy increases to 10 kcal/mol, calling for advanced error mitigation techniques.

TABLE VIII. Comparison of the projections of time to compute one step (force bias and local energy) with one walker and one shadow measurement between the state of the art algorithm by Huang *et al.* [41] (hydrogen molecule) and this work. All time units are in seconds.

	Ref. [41]	This work
Baseline: 4 qubits, (2, 2) space, 160 000 shadows	60	
Baseline: time per shadow measurement	3.75×10^{-4}	
Baseline extrapolated to 16 qubits assuming $\mathcal{O}(N_q^8)$ scaling	24.58	
This work: 16 qubits, (8, 8) space, 179 858 shadows, 1 block, VCE, H200 GPU		746.83
This work: time per shadow		4.15×10^{-3}
This work: adjust for one force bias per block ^a		7.49×10^{-4}
Assume that GPU provides a $50\times$ speedup over CPU ^b		3.74×10^{-2}
Speedup factor over baseline ignoring VCE (lower-end estimate)		656
Adjustment for VCE ^c assuming $\mathcal{O}(N_B^4 N_q^4)$ scaling	1.71×10^6	
Speedup factor over baseline (higher-end estimate)		4.58×10^7

^aIn this study, energy was evaluated every ten time steps. Energy computation alone takes 3.71×10^{-4} GPU sec per shadow.

^bTypical observed speedup on a single NVIDIA H200 GPU over a single core of AMD Rome CPU in QC-AFQMC postprocessing.

^cVCE assumes the full system has 130 spatial orbitals.

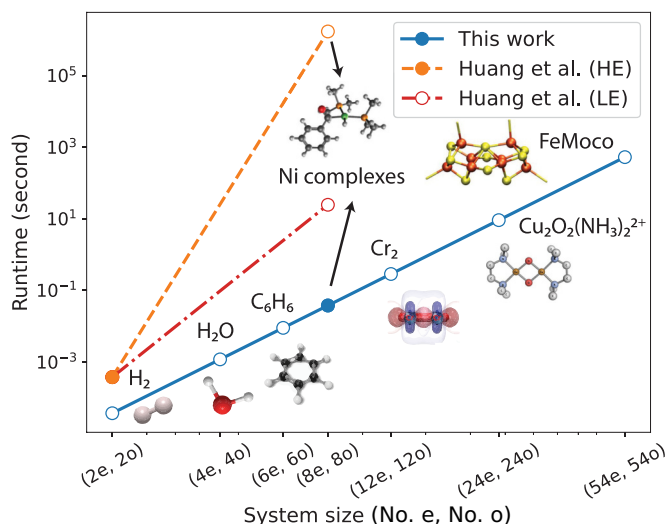


FIG. 7. Scaling of classical postprocessing run-time in QC-AFQMC with respect to the size of the active space, showing the run-time per time step per shadow. The data are from QPU experiments (solid circles), as well as from estimated run-time (hollow circles). Ni complexes are the molecular systems studied in this work. The estimated run-time for this work is calculated by scaling the data from the QPU experiment with $\mathcal{O}(N_q^5)$. The estimated run-time for the work of Huang *et al.* is calculated by scaling the data from the QPU experiment with $\mathcal{O}(N_B^4 N_q^4)$. The orange dashed line scales to a full space of 130 spatial orbitals (higher-end estimate, HE) while the red dash-dot line scales to 8 spatial orbitals (lower-end estimate, LE) at problem size (8e, 8o).

We show the importance of tuning the system performance of a QPU for the workload of a specific application. By applying performance optimizations to the control software of the QPU, we achieve a $9\times$ speedup in collecting matchgate circuit measurements compared to a baseline without

application-specific tuning. In the QC-AFQMC imaginary time propagation step, we demonstrate the use of analytic differentiation for computing the energy and force bias, leading to a reduction in the complexity of the algorithm from $\mathcal{O}(N^{8.5})$ to $\mathcal{O}(N^{5.5})$, which translates into a $656\times$ time-to-solution improvement over a baseline projected from the state-of-the-art implementation of the method by Huang *et al.* [41].

This work represents a significant step toward practical quantum chemistry simulations on quantum computing devices and offers opportunities for further improvements. Although the species in the chemical reaction we studied in this work exhibit an electronic structure primarily defined by weak correlations, QC-AFQMC is also expected to provide reliable modeling of strongly correlated systems, for which several challenges will arise. For example, it is more difficult to prepare accurate trial states, and one might need to go beyond the simple upCCD ansatz used in this work. Possible choices include the full unitary coupled cluster ansatz and its approximate variants, such as the localized unitary cluster Jastrow [110] ansatz widely used in sample-based quantum diagonalization [93]. State preparation methods beyond unitary coupled cluster, such as the Clifford + T ansatz [60], and methods based on Lindbladians [111] in early fault-tolerant regime could also be promising avenues to explore. Another challenge is the possibility of vanishing overlaps, as pointed out by Mazzola and Carleo [42]. Moreover, convergence of energy for strongly correlated systems requires one to propagate the walkers for longer imaginary time, further increasing the postprocessing time. Future work should examine these directions, along with advanced and efficient classical shadow techniques, and its combination of error mitigation and correction.

DATA AVAILABILITY

There are no publicly available research data or software supporting this manuscript. Requests for further information or data should be sent to the authors.

- [1] V. K. Mulligan, H. Melo, H. I. Merritt, S. Slocum, B. D. Weitzner, A. M. Watkins, P. D. Renfrew, C. Pelissier, P. S. Arora, and R. Bonneau, Designing peptides on a quantum computer, *bioRxiv* (2019), doi: <https://doi.org/10.1101/752485>.
- [2] A. Robert, P. K. Barkoutsos, S. Woerner, and I. Tavernelli, Resource-efficient quantum algorithm for protein folding, *npj Quantum Inf.* **7**, 38 (2021).
- [3] M. Fingerhuth, T. Babej, and C. Ing, A quantum alternating operator ansatz with hard and soft constraints for lattice protein folding, *arXiv:1810.13411*.
- [4] A. M. Smaldone and V. S. Batista, Quantum-to-classical neural network transfer learning applied to drug toxicity prediction, *J. Chem. Theory Comput.* **20**, 4901 (2024).
- [5] S. Mensa, E. Sahin, F. Tacchino, P. K. Barkoutsos, and I. Tavernelli, Quantum machine learning framework for virtual screening in drug discovery: A prospective quantum advantage, *Mach. Learn.: Sci. Technol.* **4**, 015023 (2023).
- [6] J. Allcock, A. Vangone, A. Meyder, S. Adaszewski, M. Strahm, C.-Y. Hsieh, and S. Zhang, The prospects

- of Monte Carlo antibody loop modelling on a fault-tolerant quantum computer, *Front. Drug Discov.* **2**, 908870 (2022).
- [7] F. M. Weidner, J. D. Schwab, S. Wölk, F. Rupprecht, N. Ikonomi, S. D. Werle, S. Hoffmann, M. Kühl, and H. A. Kestler, Leveraging quantum computing for dynamic analyses of logical networks in systems biology, *Patterns* **4**, 100705 (2023).
- [8] A. Kovyshin, L. Tornberg, J. Crain, S. Mensa, I. Tavernelli, and A. Broo, Prioritizing quantum computing use cases in the drug discovery and development pipeline, *Drug Discov. Today* **30**, 104323 (2025).
- [9] J. Boström, D. G. Brown, R. J. Young, and G. M. Keserü, Expanding the medicinal chemistry synthetic toolbox, *Nat. Rev. Drug Discov.* **17**, 709 (2018).
- [10] R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules* (Oxford University Press, New York, 1989).
- [11] J. N. Harvey, Spin-forbidden reactions: Computational insight into mechanisms and kinetics, *WIREs Comput. Mol. Sci.* **4**, 1 (2014).

- [12] M. Reiher, A theoretical challenge: Transition-metal compounds, *CHIMIA* **63**, 140 (2009).
- [13] K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, A fifth-order perturbation comparison of electron correlation theories, *Chem. Phys. Lett.* **157**, 479 (1989).
- [14] M. Radoń, G. Drabik, M. Hodorowicz, and J. Szklarzewicz, Performance of quantum chemistry methods for a benchmark set of spin-state energetics derived from experimental data of 17 transition metal complexes (SSE17), *Chem. Sci.* **15**, 20189 (2024).
- [15] B. O. Roos, P. R. Taylor, and P. E. M. Siegbahn, A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach, *Chem. Phys.* **48**, 157 (1980).
- [16] S. R. White, Density matrix formulation for quantum renormalization groups, *Phys. Rev. Lett.* **69**, 2863 (1992).
- [17] H. Zhai, H. R. Larsson, S. Lee, Z.-H. Cui, T. Zhu, C. Sun, L. Peng, R. Peng, K. Liao, J. Tölle, J. Yang, S. Li, and G. K.-L. Chan, Block2: A comprehensive open source framework to develop and apply state-of-the-art DMRG algorithms in electronic structure and beyond, *J. Chem. Phys.* **159**, 234801 (2023).
- [18] K. H. Marti, I. M. Ondík, G. Moritz, and M. Reiher, Density matrix renormalization group calculations on relative energies of transition metal complexes and clusters, *J. Chem. Phys.* **128**, 014104 (2008).
- [19] G. K.-L. Chan, and S. Sharma, The density matrix renormalization group in quantum chemistry, *Annu. Rev. Phys. Chem.* **62**, 465 (2011).
- [20] A. Menczer, M. van Damme, A. Rask, L. Huntington, J. Hammond, S. S. Xantheas, M. Ganahl, and Ö. Legeza, Parallel implementation of the density matrix renormalization group method achieving a quarter petaflops performance on a single DGX-H100 GPU node, *J. Chem. Theory Comput.* **20**, 8397 (2024).
- [21] Ö. Legeza, A. Menczer, Á. Ganyecz, M. A. Werner, K. Kapás, J. Hammond, S. S. Xantheas, M. Ganahl, and F. Neese, Orbital optimization of large active spaces via AI-accelerators, *J. Chem. Theory Comput.* **21**, 6545 (2025).
- [22] C. Xiang, W. Jia, W.-H. Fang, and Z. Li, Distributed multi-GPU *ab initio* density matrix renormalization group algorithm with applications to the P-cluster of nitrogenase, *J. Chem. Theory Comput.* **20**, 775 (2024).
- [23] H. Liu, G. H. Low, D. S. Steiger, T. Häner, M. Reiher, and M. Troyer, Prospects of quantum computing for molecular sciences, *Mater. Theory* **6**, 11 (2022).
- [24] A. Aspuru-Guzik, A. D. Dutoi, P. J. Love, and M. Head-Gordon, Simulated quantum computation of molecular energies, *Science* **309**, 1704 (2005).
- [25] B. P. Lanyon, J. D. Whitfield, G. G. Gillett, M. E. Goggin, M. P. Almeida, I. Kassal, J. D. Biamonte, M. Mohseni, B. J. Powell, M. Barbieri, A. Aspuru-Guzik, and A. G. White, Towards quantum chemistry on a quantum computer, *Nat. Chem.* **2**, 106 (2010).
- [26] G. H. Low and I. L. Chuang, Hamiltonian simulation by qubitization, *Quantum* **3**, 163 (2019).
- [27] R. Babbush, C. Gidney, D. W. Berry, N. Wiebe, J. McClean, A. Paler, A. Fowler, and H. Neven, Encoding electronic spectra in quantum circuits with linear T complexity, *Phys. Rev. X* **8**, 041015 (2018).
- [28] 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, *PRX Quantum* **2**, 030305 (2021).
- [29] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, A variational eigenvalue solver on a photonic quantum processor, *Nat. Commun.* **5**, 4213 (2014).
- [30] P. J. J. O'Malley *et al.*, Scalable quantum simulation of molecular energies, *Phys. Rev. X* **6**, 031007 (2016).
- [31] 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 (London)* **549**, 242 (2017).
- [32] A. I. Google Quantum and Collaborators, Hartree-Fock on a superconducting qubit quantum computer, *Science* **369**, 1084 (2020).
- [33] Y. Nam, J.-S. Chen, N. C. Pienti, K. Wright, C. Delaney, D. Maslov, K. R. Brown, S. Allen, J. M. Amini, J. Apisdorf, *et al.*, Ground-state energy estimation of the water molecule on a trapped-ion quantum computer, *npj Quantum Inf.* **6**, 33 (2020).
- [34] J. F. Gonthier, M. D. Radin, C. Buda, E. J. Duskocil, C. M. Abuan, and J. Romero, Measurements as a roadblock to near-term practical quantum advantage in chemistry: Resource analysis, *Phys. Rev. Res.* **4**, 033154 (2022).
- [35] 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 (London)* **603**, 416 (2022).
- [36] S. Zhang, J. Carlson, and J. E. Gubernatis, Constrained path Monte Carlo method for fermion ground states, *Phys. Rev. B* **55**, 7464 (1997).
- [37] S. Zhang and H. Krakauer, Quantum Monte Carlo method using phase-free random walks with Slater determinants, *Phys. Rev. Lett.* **90**, 136401 (2003).
- [38] M. Motta and S. Zhang, *Ab initio* computations of molecular systems by the auxiliary-field quantum Monte Carlo method, *WIREs Comput. Mol. Sci.* **8**, e1364 (2018).
- [39] J. Lee, H. Q. Pham, and D. R. Reichman, Twenty years of auxiliary-field quantum Monte Carlo in quantum chemistry: An overview and assessment on main group chemistry and bond-breaking, *J. Chem. Theory Comput.* **18**, 7024 (2022).
- [40] J. Shee, J. L. Weber, D. R. Reichman, R. A. Friesner, and S. Zhang, On the potentially transformative role of auxiliary-field quantum Monte Carlo in quantum chemistry: A highly accurate method for transition metals and beyond, *J. Chem. Phys.* **158**, 140901 (2023).
- [41] B. Huang, Y.-T. Chen, B. Gupt, M. Suchara, A. Tran, S. McArdle, and G. Galli, Evaluating a quantum-classical quantum Monte Carlo algorithm with matchgate shadows, *Phys. Rev. Res.* **6**, 043063 (2024).
- [42] G. Mazzola and G. Carleo, Exponential challenges in unbiasing quantum Monte Carlo algorithms with quantum computers, *arXiv:2205.09203*.
- [43] M. Amsler, P. Deglmann, M. Degroote, M. P. Kaicher, M. Kiser, M. Kühn, C. Kumar, A. Maier, G. Samsonidze, A. Schroeder, M. Streif, D. Vodola, and C. Wever, Quantum-enhanced quantum Monte Carlo: An industrial view, *arXiv:2301.11838*.

- [44] M. Kiser, A. Schroeder, G.-L. R. Anselmetti, C. Kumar, N. Moll, M. Streif, and D. Vodola, Classical and quantum cost of measurement strategies for quantum-enhanced auxiliary field quantum Monte Carlo, *New J. Phys.* **26**, 033022 (2024).
- [45] T. Jiang, B. O’Gorman, A. Mahajan, and J. Lee, Unbiasing fermionic auxiliary-field quantum Monte Carlo with matrix product state trial wavefunctions, *Phys. Rev. Res.* **7**, 013038 (2025).
- [46] Bracket contributors, *Amazon Bracket: A Fully Managed Quantum Computing Service Provided by AWS* (Amazon Web Services, Seattle, WA, 2019).
- [47] L. Guo, W. Srimontree, C. Zhu, B. Maity, X. Liu, L. Cavallo, and M. Rueping, Nickel-catalyzed Suzuki-Miyaura cross-couplings of aldehydes, *Nat. Commun.* **10**, 1957 (2019).
- [48] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Quantum Monte Carlo simulations of solid, *Rev. Mod. Phys.* **73**, 33 (2001).
- [49] P. R. C. Kent *et al.*, QMCPACK: Advances in the development, efficiency, and application of auxiliary field and real-space variational and diffusion quantum Monte Carlo, *J. Chem. Phys.* **152**, 174105 (2020).
- [50] C. J. Umrigar, Observations on variational and projector Monte Carlo methods, *J. Chem. Phys.* **143**, 164105 (2015).
- [51] L. R. Schwartz, A. Alavi, and G. H. Booth, Projector quantum Monte Carlo method for nonlinear wave functions, *Phys. Rev. Lett.* **118**, 176403 (2017).
- [52] F. Malone, A. Mahajan, J. Spencer, and J. Lee, Ipie: A Python-based auxiliary-field quantum Monte Carlo program with flexibility and efficiency on CPUs and GPUs, *J. Chem. Theory Comput.* **19**, 109 (2022).
- [53] S. Zhang, *Ab initio* electronic structure calculations by auxiliary-field quantum Monte Carlo, *Handbook of Materials Modeling*, edited by W. Andreoni and S. Yip (Springer International Publishing, Cham, 2020), pp. 123–149.
- [54] R. L. Stratonovich, On a method of calculating quantum distribution functions, *Sov. Phys. Dokl.* **2**, 416 (1957).
- [55] J. Hubbard, Calculation of partition functions, *Phys. Rev. Lett.* **3**, 77 (1959).
- [56] S. Zhang, Auxiliary-field quantum Monte Carlo for correlated electron systems, in *Emergent Phenomena in Correlated Matter; Lecture Notes of the Autumn School Correlated Electrons* (Forschungszentrum Jülich, Jülich, Germany, 2013), Chap. 15.
- [57] A. Mahajan, J. Lee, and S. Sharma, Selected configuration interaction wave functions in phaseless auxiliary field quantum Monte Carlo, *J. Chem. Phys.* **156**, 174111 (2022).
- [58] A. Mahajan and S. Sharma, Taming the sign problem in auxiliary-field quantum Monte Carlo using accurate wave functions, *J. Chem. Theory Comput.* **17**, 4786 (2021).
- [59] A. Cowtan, W. Simmons, and R. Duncan, A generic compilation strategy for the unitary coupled cluster ansatz, [arXiv:2007.10515](https://arxiv.org/abs/2007.10515).
- [60] A. Gu, H.-Y. Hu, D. Luo, T. L. Patti, N. C. Rubin, and S. F. Yelin, Zero and finite temperature quantum simulations powered by quantum magic, *Quantum* **8**, 1422 (2024).
- [61] 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).
- [62] I. Khan, M. Tudorovskaya, J. Kirsopp, D. M. Ramo, P. Warrier, D. Papanastasiou, and R. Singh, Chemically aware unitary coupled cluster with *ab initio* calculations on system model H1: A refrigerant chemicals application, *J. Chem. Phys.* **158**, 214114 (2023).
- [63] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, *Nat. Phys.* **16**, 1050 (2020).
- [64] H.-Y. Huang, R. Kueng, G. Torlai, V. V. Albert, and J. Preskill, Provably efficient machine learning for quantum many-body problems, *Science* **377**, 682 (2022).
- [65] K. Wan, W. J. Huggins, J. Lee, and R. Babbush, Matchgate shadows for fermionic quantum simulation, *Commun. Math. Phys.* **404**, 629 (2023).
- [66] A. Zhao, N. C. Rubin, and A. Miyake, Fermionic partial tomography via classical shadows, *Phys. Rev. Lett.* **127**, 110504 (2021).
- [67] M. Wimmer, Efficient numerical computation of the Pfaffian for dense and banded skew-symmetric matrices, *ACM Trans. Math. Software* **38**, 1 (2012).
- [68] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/n1tf-8kr7> for matrix definitions, virtual correlation energy derivation, reduced density matrix formulas, CUDA-Q and workflow implementation details, DFT reaction energy profiles, single-orbital entropy analysis, orbital entanglement networks, and molecular coordinates, which includes Refs. [96,97].
- [69] L. Zhao, J. Goings, K. Wright, J. Nguyen, J. Kim, S. Johri, K. Shin, W. Kyoung, J. I. Fuks, J.-K. K. Rhee, and Y. M. Rhee, Orbital-optimized pair-correlated electron simulations on trapped-ion quantum computers, *npj Quantum Inf.* **9**, 60 (2023).
- [70] 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).
- [71] S. Lehtola, C. Steigemann, M. J. T. Oliveira, and M. A. L. Marques, Recent developments in libxc—a comprehensive library of functionals for density functional theory, *SoftwareX* **7**, 1 (2018).
- [72] HQS Quantum Simulations, Active Space Finder, <https://github.com/HQSquantumsimulations/ActiveSpaceFinder> (2021).
- [73] T. Jiang, M. K. A. Baumgarten, P.-F. Loos, A. Mahajan, A. Scemama, S. F. Ung, J. Zhang, F. D. Malone, and J. Lee, Improved modularity and new features in ipie: Toward even larger AFQMC calculations on CPUs and GPUs at zero and finite temperatures, *J. Chem. Phys.* **161**, 162502 (2024).
- [74] A. Zhao, Symmetry-adjusted-classical-shadows: Classical shadows codes for quantum simulations of fermions and qubits, with error-mitigation routines for systems with symmetries, GitHub, 2021, <https://github.com/zhao-andrew/symmetry-adjusted-classical-shadows>.
- [75] A. Zhao and A. Miyake, Group-theoretic error mitigation enabled by classical shadows and symmetries, *npj Quantum Inf.* **10**, 57 (2024).
- [76] *Computational Studies in Organometallic Chemistry, Structure and Bonding*, 1st ed., edited by S. A. Macgregor and O. Eisenstein (Springer International Publishing, Cham, Switzerland, 2016).

- [77] J.-D. Chai and M. Head-Gordon, Systematic optimization of long-range corrected hybrid density functionals, *J. Chem. Phys.* **128**, 084106 (2008).
- [78] S. H. Vosko, L. Wilk, and M. Nusair, Accurate spin-dependent electron liquid correlation energies for local spin density calculations: A critical analysis, *Can. J. Phys.* **58**, 1200 (1980).
- [79] C. Lee, W. Yang, and R. G. Parr, Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density, *Phys. Rev. B* **37**, 785 (1988).
- [80] A. D. Becke, Density-functional thermochemistry. III. The role of exact exchange, *J. Chem. Phys.* **98**, 5648 (1993).
- [81] P. J. Stephens, F. J. Devlin, C. S. Ashvar, C. F. Chabalowski, and M. J. Frisch, Theoretical calculation of vibrational circular dichroism spectra, *Faraday Discuss.* **99**, 103 (1994).
- [82] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized gradient approximation made simple, *Phys. Rev. Lett.* **77**, 3865 (1996).
- [83] M. Ernzerhof and G. E. Scuseria, Assessment of the Perdew-Burke-Ernzerhof exchange-correlation functional, *J. Chem. Phys.* **110**, 5029 (1999).
- [84] C. Adamo and V. Barone, Toward reliable density functional methods without adjustable parameters: The PBE0 model, *J. Chem. Phys.* **110**, 6158 (1999).
- [85] Y. Zhao and D. G. Truhlar, The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other functionals, *Theor. Chem. Acc.* **120**, 215 (2008).
- [86] W. J. Pietro and W. J. Hehre, Molecular orbital theory of the properties of inorganic and organometallic compounds. 3. STO-3G basis sets for first- and second-row transition metals, *J. Comput. Chem.* **4**, 241 (1983).
- [87] W. J. Pietro, E. S. Blurock, R. F. Hout, Jr., W. J. Hehre, D. J. DeFrees, and R. F. Stewart, Molecular orbital theory of the properties of inorganic and organometallic compounds. 2. STO-NG basis sets for fourth-row main-group elements, *Inorg. Chem.* **20**, 3650 (1981).
- [88] W. J. Pietro, B. A. Levi, W. J. Hehre, and R. F. Stewart, Molecular orbital theory of the properties of inorganic and organometallic compounds. 1. STO-NG basis sets for third-row main-group elements, *Inorg. Chem.* **19**, 2225 (1980).
- [89] 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).
- [90] 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).
- [91] C. J. Stein and M. Reiher, Automated selection of active orbital spaces, *J. Chem. Theory Comput.* **12**, 1760 (2016).
- [92] C. J. Stein and M. Reiher, AUTOCAS: A program for fully automated multi-configurational calculations, *J. Comput. Chem.* **40**, 2216 (2019).
- [93] J. Robledo-Moreno, M. Motta, H. Haas, A. Javadi-Abhari, P. Jurcevic, W. Kirby, S. Martiel, K. Sharma, S. Sharma, T. Shirakawa, I. Sitdikov, R.-Y. Sun, K. J. Sung, M. Takita, M. C. Tran, S. Yunoki, and A. Mezzacapo, Chemistry beyond the scale of exact diagonalization on a quantum-centric supercomputer, *Sci. Adv.* **11**, eadu9991 (2025).
- [94] M. R. Hirsbrunner, D. Chamaki, J. W. Mullinax, and N. M. Tubman, Beyond MP2 initialization for unitary coupled cluster quantum circuits, *Quantum* **8**, 1538 (2024).
- [95] NVIDIA, CUDA-Q (2025), <https://nvidia.github.io/cuda-quantum/latest/index.html>.
- [96] H. Bayraktar, A. Charara, D. Clark, S. Cohen, T. Costa, Y.-L. L. Fang, Y. Gao, J. Guan, J. Gunnels, A. Haidar, *et al.*, cuQuantum SDK: A high-performance library for accelerating quantum science, in *2023 IEEE International Conference on Quantum Computing and Engineering (QCE)* (IEEE, Bellevue, WA, 2023), Vol. 1, pp. 1050–1061.
- [97] NVIDIA, CUDA-QX (2025), <https://nvidia.github.io/cudaqx/>.
- [98] J.-S. Chen, E. Nielsen, M. Ebert, V. Inlek, K. Wright, V. Chaplin, A. Maksymov, E. Páez, A. Poudel, P. Maunz, and J. Gamble, Benchmarking a trapped-ion quantum computer with 30 qubits, *Quantum* **8**, 1516 (2024).
- [99] S. Chen, W. Yu, P. Zeng, and S. T. Flammia, Robust shadow estimation, *PRX Quantum* **2**, 030348 (2021).
- [100] D. E. Koh and S. Grewal, Classical shadows with noise, *Quantum* **6**, 776 (2022).
- [101] R. Brieger, M. Heinrich, I. Roth, and M. Kliesch, Stability of classical shadows under gate-dependent noise, *Phys. Rev. Lett.* **134**, 090801 (2025).
- [102] R. Ozeri, W. M. Itano, R. Blakestad, J. Britton, J. Chiaverini, J. D. Jost, C. Langer, D. Leibfried, R. Reichle, S. Seidelin, *et al.*, Errors in trapped-ion quantum gates due to spontaneous photon scattering, *Phys. Rev. A* **75**, 042329 (2007).
- [103] I. Moore, W. Campbell, E. Hudson, M. Boguslawski, D. Wineland, and D. Allcock, Photon scattering errors during stimulated Raman transitions in trapped-ion qubits, *Phys. Rev. A* **107**, 032413 (2023).
- [104] R. Stricker, D. Vodola, A. Erhard, L. Postler, M. Meth, M. Ringbauer, P. Schindler, T. Monz, M. Müller, and R. Blatt, Experimental deterministic correction of qubit loss, *Nature (London)* **585**, 207 (2020).
- [105] H. Flyvbjerg and H. G. Petersen, Error estimates on averages of correlated data, *J. Chem. Phys.* **91**, 461 (1989).
- [106] H. Hayashi, S. Maeda, and T. Mita, Quantum chemical calculations for reaction prediction in the development of synthetic methodologies, *Chem. Sci.* **14**, 11601 (2023).
- [107] E. van Lenthe, E. J. Baerends, and J. G. Snijders, Relativistic regular two-component Hamiltonians, *J. Chem. Phys.* **99**, 4597 (1993).
- [108] A. Wolf, M. Reiher, and B. A. Hess, The generalized Douglas-Kroll transformation, *J. Chem. Phys.* **117**, 9215 (2002).
- [109] M. Dolg, Relativistic effective core potentials, in *Relativistic Electronic Structure Theory, Theoretical and Computational Chemistry*, edited by P. Schwerdtfeger (Elsevier, Amsterdam, 2002), Chap. 14, Vol. 11, pp. 793–862.
- [110] 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).
- [111] Z. Ding, C.-F. Chen, and L. Lin, Single-ancilla ground state preparation via Lindbladians, *Phys. Rev. Res.* **6**, 033147 (2024).