Benchmarking the Impact of Active Space Selection on the VQE Pipeline for Quantum Drug Discovery

Zhi Yin[1,3*], Xiaoran Li[1,2*], Zhupeng Han[1], Shengyu Zhang[4], Xin Li[4], Zhihong Zhang[5],

Runqing Zhang[5], Anbang Wang[5], Xiaojin Zhang[2*]

[1] *QureGenAI(aka AceMapAI) Suzhou Biotechnology Co.,Ltd, Suzhou, 215000, China*

[2] *QureGenAI Joint Lab, China Pharmaceutical University, Nanjing, 211198, China*

[3] *School of Statistics and Data Science, Ningbo University of Technology, Ningbo, 315211,*

China

[4] *Tencent Quantum Lab, Shenzhen, 518057, China*

[5] *China Mobile (Suzhou) Software Technology Company Limited, Suzhou 215163, China*

[*] Corresponding authors. E-mail: yz@nbut.edu.cn; lixiaoran@quregenai.com; zxj@cpu.edu.cn

Abstract

Quantum computers promise scalable treatments of electronic structure, yet applying variational quantum eigensolvers (VQE) on realistic drug-like molecules remains constrained by the performance limitations of near-term quantum hardwares. A key strategy for addressing this challenge which effectively leverages current Noisy Intermediate-Scale Quantum (NISQ) hardwares yet remains under-benchmarked is active space selection. We introduce a benchmark that heuristically proposes criteria based on chemically grounded metrics to classify the suitability of a molecule for using quantum computing and then quantifies the impact of active space choices across the VQE pipeline for *quantum drug discovery*. The suite covers several representative drug-like molecules (e.g., lovastatin, oseltamivir, morphine) and uses chemically motivated active spaces. Our VQE evaluations employ both simulation and quantum processing unit (QPU) execution using unitary coupled-cluster with singles and doubles (UCCSD) and hardware-efficient ansatz (HEA). We adopt a more comprehensive evaluation, including chemistry metrics and architecture-centric metrics. For accuracy, we compare them with classical quantum chemistry methods. This work establishes the first systematic benchmark for active space driven VQE and lays the groundwork for future hardware-algorithm co-design studies in *quantum drug discovery*.

I.   INTRODUCTION

Quantum computing holds transformative promise for computational chemistry and drug discovery, offering the potential to simulate molecular systems with higher accuracy and significantly reduced runtime compared to classical methods [1–3]. Variational quantum eigensolvers (VQE) [4] have emerged as one of the most promising algorithms for near-term quantum hardwares, using a parameterized quantum circuit and a classical optimizer to minimize the energy. Unlike fault-tolerant algorithms such as quantum phase estimation that require millions of logical qubits, VQE can operate on current noisy intermediate-scale quantum (NISQ) hardware with tens to hundreds of physical qubits [5]. This practical advantage has motivated extensive efforts to apply VQE to drug-like molecules, where accurate electronic structure calculations are critical for understanding molecular properties, reaction mechanisms, and protein-ligand binding affinities [6, 7].

However, applying VQE to realistic drug molecules remains severely constrained by the

limitations of current quantum hardware. State-of-the-art quantum processors feature only $\sim$100–1000 physical qubits with typical single- and two-qubit gate fidelities in the 99.9% and 99%-99.9% ranges, respectively [8, 9], while even moderately sized drug molecules such as lovastatin ($C_{24}H_{36}O_5$, 65 atoms) or imatinib ($C_{29}H_{31}N_7O$, 68 atoms) would naively require hundreds of logical qubits to represent their full electronic structure. This exponential scaling of quantum resources with molecular size creates a critical gap between the promise of quantum advantage and the reality of NISQ device capabilities.

A key strategy to bridge this gap is *active space selection* [10, 11], which restricts quantum computation to a chemically relevant subset of molecular orbitals and electrons. By carefully selecting an active space—typically denoted as (n_e, n_o) for n_e electrons in n_o orbitals—one can dramatically reduce the number of qubits required while retaining essential multi-reference correlation effects. For instance, a (6e, 6o) active space requires only 12 qubits under the Jordan-Wigner mapping [12, 13], making VQE feasible on current hardware, whereas the full orbital space of even a small drug molecule would require 50+ qubits. Despite its critical importance, active space selection for electronic structure calculation using quantum computing remains largely heuristic and insufficiently benchmarked. Classical quantum chemistry methods such as complete active space self-consistent field (CASSCF) [14, 15] have established heuristics based on chemical intuition—focusing on frontier orbitals, bonds being broken or formed, and strongly correlated electron pairs. However, these guidelines were developed for classical algorithms and may not directly translate to the VQE pipeline especially on near-term QPU, where circuit depth, measurement overhead, and ansatz expressibility introduce additional constraints. Furthermore, not all molecules benefit equally from quantum computation: single-reference systems can often be handled accurately by density functional theory (DFT) [16, 17] or coupled cluster methods [18, 19], while highly multi-reference systems may require active spaces too large for NISQ devices. Identifying which molecules are suitable for quantum computing, and how to optimally select their active spaces, remains an open question with significant practical implications.

Existing work on VQE for molecular systems has focused mainly on small reference molecules such as H_2, H_2O, LiH and BeH_2, or specific case studies of individual drug-like molecules [20]. While these studies have demonstrated the feasibility of VQE and explored various ansatze (e.g., unitary coupled cluster with singles and doubles [UCCSD] [21–

23], hardware-efficient ansatze [HEA] [7, 24–27]), they lack systematic evaluation across a diverse set of drug-relevant molecules with varying sizes and electronic structures. Critically, there exists no comprehensive benchmark that (1) establishes heuristic criteria to classify molecules' suitability for quantum computing, (2) systematically quantifies the impact of active space choices across the VQE pipeline, and (3) evaluates performance using both chemistry-centric metrics (such as energy accuracy) and architecture-centric metrics (such as qubit count, circuit depth, gate complexity). Such a benchmark is essential for guiding practical applications of quantum drug discovery and informing future hardware-algorithm co-design.

In this work, we present the first systematic benchmark for active space selection in the VQE pipeline for quantum drug discovery. We make the following key contributions:

Heuristic classification criteria. We introduce chemically grounded metrics to assess a molecule's suitability for quantum computing based on natural orbital occupation numbers derived from CASSCF calculations. By analyzing fractional occupations in a standardized active space probe, we rapidly classify molecules by multi-reference character strength, enabling pre-screening to identify high-value targets for quantum computation.

Diverse molecular benchmark suite. We curate a representative set of seven drug-like molecules(with water and benzene as references) spanning 3 to 68 atoms, including clinically important compounds such as aspirin, oseltamivir, morphine, lovastatin, and imatinib. This suite covers a spectrum from single-reference systems (where quantum computing may be unnecessary) to challenging multi-reference cases (where quantum advantage is expected but active spaces may exceed NISQ capabilities), providing a realistic testbed for evaluating active space strategies.

Systematic active space evaluation. For each molecule, we evaluate multiple active space configurations ranging from minimal (2e, 2o) to extended (4e, 4o) or larger, examining their impact on both accuracy and quantum resource requirements. We employ both UCCSD and HEA ansatze, and compare VQE results against classical benchmarks including Hartree-Fock and density functional theory (DFT).

Multi-dimensional evaluation. Unlike prior studies that focus solely on energy accuracy, we adopt a comprehensive evaluation encompassing: (i) chemistry metrics: absolute and relative energy errors, comparison with gold-standard methods; (ii) architecture metrics: qubit counts, circuit depth. This multi-faceted analysis reveals trade-offs between chemical

fidelity and hardware feasibility.

Actionable insights for hardware-algorithm co-design. Our benchmark including end-to-end optimization on two distinct superconducting quantum processors (13-qubit homebrew s2 and 60-qubit homebrew q1 devices) identifies which molecular characteristics (size, multi-reference character, electronic structure complexity) correlate with successful VQE performance under real hardware noise. The direct comparison between QPU architectures reveals critical trade-offs: larger processors exhibit increased convergence oscillations (e.g., benzene's fluctuations on q1 vs. smooth descent on s2) yet achieve comparable final energies, demonstrating that there is possibility that current VQE applications are bottlenecked by ansatz depth rather than qubit count. These hardware-validated findings provide concrete guidance for: (i) computational chemists deciding when to employ quantum computing and which QPU architectures suit specific molecular problems; (ii) algorithm developers optimizing VQE protocols for noise resilience and convergence stability on real devices; and (iii) hardware engineers prioritizing improvements—our results suggest gate fidelity and circuit depth limits dominate over raw qubit count for near-term devices, shifting design priorities toward error mitigation and connectivity over scaling alone.

We openly release our benchmark suite, including molecular geometries, active space configurations, classical reference data, and VQE evaluation scripts, to facilitate reproducibility and community-driven extensions.

The remainder of this paper is organized as follows. Section 2 provides background on VQE, active space methods, and reviews related work in quantum computational chemistry. Section 3 describes our methodology, including the classification criteria, active space selection strategies, VQE pipeline implementation, and evaluation metrics. Section 4 presents comprehensive results and analysis across our benchmark. Section 5 discusses implications for quantum drug discovery, limitations of the current study, and future directions. Section 6 concludes.

II. BACKGROUND AND RELATED WORK

A. Variational Quantum Eigensolver

The variational quantum eigensolver (VQE) is a hybrid quantum-classical algorithm designed to find ground state energies of quantum systems(see Figure 1). VQE can be operated on noisy intermediate-scale quantum (NISQ) devices by combining shallow quantum circuits with classical optimization. Given a molecular Hamiltonian $\hat{H}$ represented as a sum of Pauli operators,

$$\hat{H} = \sum_i h_i \hat{P}_i, \tag{1}$$

where $\hat{P}_i \in \{I, X, Y, Z\}^{\otimes n}$ are n-qubit Pauli strings, VQE seeks the ground state energy $E_0 = \min_\psi \langle \psi | \hat{H} | \psi \rangle$ via the variational principle. The algorithm prepares a parameterized quantum state $|\psi(\boldsymbol{\theta})\rangle$ using an ansatz circuit $U(\boldsymbol{\theta})$ acting on a reference state (typically Hartree-Fock):

$$|\psi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})|\mathrm{HF}\rangle. \tag{2}$$

The energy expectation value $E(\boldsymbol{\theta}) = \langle \psi(\boldsymbol{\theta}) | \hat{H} | \psi(\boldsymbol{\theta}) \rangle$ is evaluated on the quantum processor by measuring each Pauli term, and a classical optimizer iteratively updates $\boldsymbol{\theta}$ to minimize $E(\boldsymbol{\theta})$.

VQE's shallow circuit depth (compared to quantum phase estimation) makes it tolerant to gate errors and decoherence. The variational bound $E(\boldsymbol{\theta}) \geq E_0$ ensures that even imperfect optimization yields an upper bound on the ground state energy. However, VQE faces challenges including barren plateaus (vanishing gradients in large circuits) [28, 29], sensitivity to local minima, and measurement overhead scaling with the number of Hamiltonian terms.

B. Active Space Approximation

For a molecule with N electrons in M spatial orbitals, the full configuration interaction (FCI) wavefunction involves $\binom{2M}{N}$ determinants, leading to exponential scaling. On a quantum computer using Jordan-Wigner, this requires $2M$ qubits—prohibitively large for drug molecules.
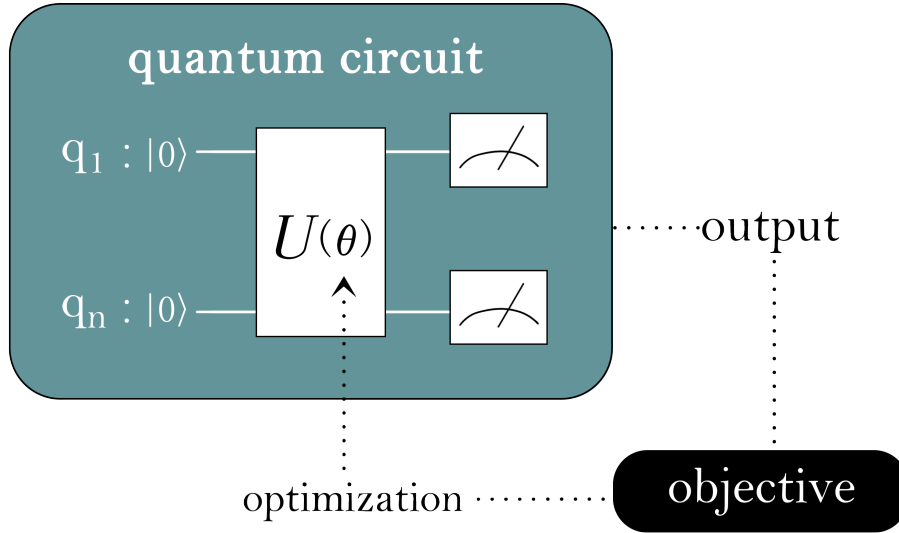
FIG. 1. Variational Quantum Eigensolver schematic: parameterized quantum circuit $U(\theta)$ generates trial states, measurement outcomes define the energy objective, and classical optimization refines θ iteratively until convergence.

The active space approximation partitions molecular orbitals into three classes(see Figure 2):

- **Core (inactive) orbitals**: Doubly occupied in all configurations, treated at the mean-field level.

- **Active orbitals**: Partially occupied orbitals where electron correlation is explicitly treated. An (n_e, n_o) active space contains n_e active electrons distributed among n_o active orbitals.

- **Virtual orbitals**: Unoccupied orbitals excluded from the correlation treatment.

Typically, active orbitals are chosen around the Fermi level (HOMO-LUMO region), focusing on valence electrons involved in bonding, reactions, or multi-reference correlation. This reduces the quantum resource requirement to $2n_o$ qubits under Jordan-Wigner mapping, making VQE tractable for current hardware when n_o is small.

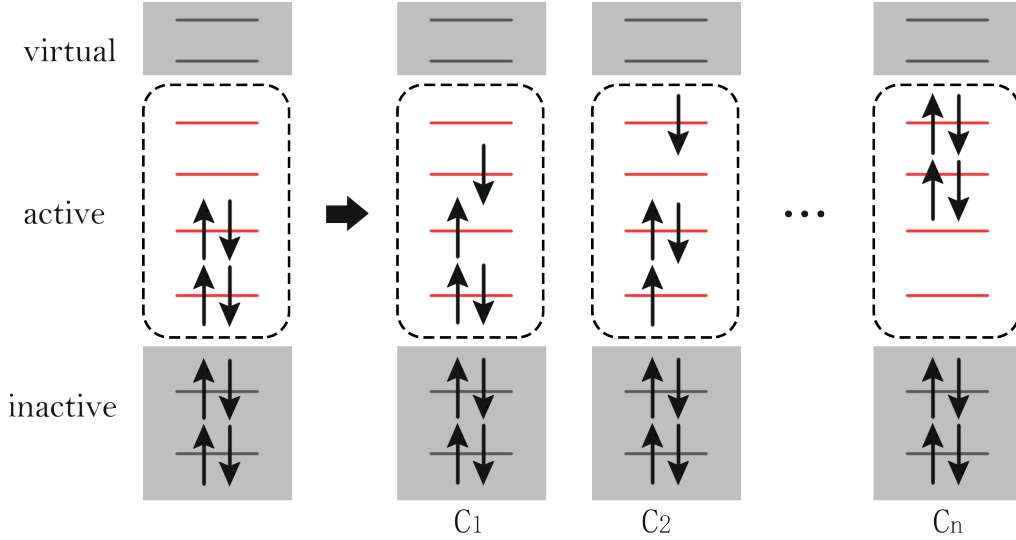Classical quantum chemistry employs heuristics such as:

FIG. 2. Active space approximation for molecular orbital partitioning. Virtual orbitals (top, gray) and inactive orbitals (bottom, gray) remain frozen at Hartree-Fock occupation, while electrons in the active space (middle, dashed box) explore all possible configurations $C_1, C_2, \ldots, C_n$ via quantum computation. Red lines indicate chemically relevant orbitals treated variationally; arrows represent electron occupations.

- Selecting orbitals with natural occupation numbers deviating significantly from 2 (occupied) or 0 (virtual).

- Including orbitals involved in bond breaking/formation.

- Using localized orbital analysis (e.g., intrinsic bonding orbitals).

However, these guidelines were developed for classical multi-reference methods (CASSCF [14], CASPT2 [30], DMRG [31]) and may not directly translate to VQE, where circuit depth, measurement overhead, and ansatz expressibility introduce additional constraints. Systematic benchmarking of active space choices for VQE in drug-like molecules remains scarce.

C. VQE Ansatze for Quantum Chemistry

a. Unitary Coupled Cluster (UCC). The coupled cluster ansatz, a gold standard in classical quantum chemistry, can be adapted to quantum computers via unitary coupled

cluster with singles and doubles (UCCSD):

$$|\psi_{\text{UCCSD}}\rangle = e^{\hat{T}-\hat{T}^{\dagger}}|\text{HF}\rangle, \tag{3}$$

where $\hat{T} = \hat{T}_1 + \hat{T}_2$ includes single ($\hat{T}_1$) and double ($\hat{T}_2$) excitation operators. The anti-Hermitian form $\hat{T} - \hat{T}^{\dagger}$ ensures unitarity. UCCSD has $O(N_{\text{occ}}N_{\text{virt}} + N_{\text{occ}}^2 N_{\text{virt}}^2)$ parameters, providing high chemical accuracy but requiring deep circuits with many CNOT gates.

b. *Hardware-efficient ansatze (HEA).* To reduce circuit depth, hardware-efficient ansatze employ layers of single-qubit rotations and entangling gates native to the quantum processor [7]:

$$U_{\text{HEA}}(\boldsymbol{\theta}) = \prod_{\ell=1}^{L}\left[U_{\text{ent}}^{(\ell)}\prod_i R_y(\theta_i^{(\ell)})R_z(\phi_i^{(\ell)})\right], \tag{4}$$

where $U_{\text{ent}}^{(\ell)}$ is the entangling layer for the ℓ-th circuit layer (e.g., linear CX gates). HEA circuits are shallow and noise-resilient but lack direct chemical interpretability. The trade-off between UCCSD (high accuracy, deep circuits) and HEA (lower accuracy, shallow circuits) is crucial for NISQ applications.

D. Quantum Drug Discovery

Early VQE demonstrations focused on small benchmark molecules: H_2, H_2O, LiH, and BeH_2 [7, 32, 33]. Recent efforts have targeted larger systems, including simulations of drug-like molecules. For instance, research [34]reported VQE calculations for real-world drug design challenges, specifically computing Gibbs free energy profiles for prodrug activation and simulating covalent bond interactions.

However, most studies are *case-specific* rather than systematic benchmarks. They typically:

- Focus on a single molecule or narrow chemical class.

- Report results for one or two active space configurations without exploring sensitivity or convergence.

- Evaluate only simulator-based VQE, lacking real QPU validation.

Furthermore, no prior work has established *heuristic criteria* to classify molecules by their suitability for quantum computing, nor systematically quantified the impact of active

space choices across a diverse set of drug-relevant molecules with varying sizes and electronic structures. Our benchmark addresses these gaps by providing:

1. A classification scheme based on multi-reference character.

2. Systematic active space evaluation for 7 representative drug molecules.

3. Multi-dimensional assessment encompassing both chemistry metrics (energy accuracy) and architecture metrics (qubit count, circuit depth).

4. Validation on physical quantum processors.

This establishes the first comprehensive benchmark for active space-driven VQE in quantum drug discovery, providing actionable guidance for practitioners.

III. METHODOLOGY

Our methodology comprises four stages: (i) heuristic classification to assess molecular suitability for quantum computing; (ii) active space selection strategies tailored for VQE; (iii) VQE pipeline implementation on both simulators and quantum processors; and (iv) multi-dimensional evaluation metrics. (Figure 3 illustrates the overall workflow)

A. Molecule Classification via Multi-reference Diagnostics

To rapidly identify which drug-like molecules are promising candidates for quantum computing, we establish a heuristic classification scheme grounded in electronic structure theory.

Not all molecules require quantum computational treatment. Single-reference systems, where electron correlation is weak and mean-field (Hartree-Fock) or perturbative (DFT, MP2) methods suffice, gain little from VQE's multi-reference capabilities. Conversely, strongly correlated systems—where near-degenerate states and fractional orbital occupations dominate—challenge classical methods but are natural targets for quantum algorithms. A pre-screening classification enables efficient resource allocation.
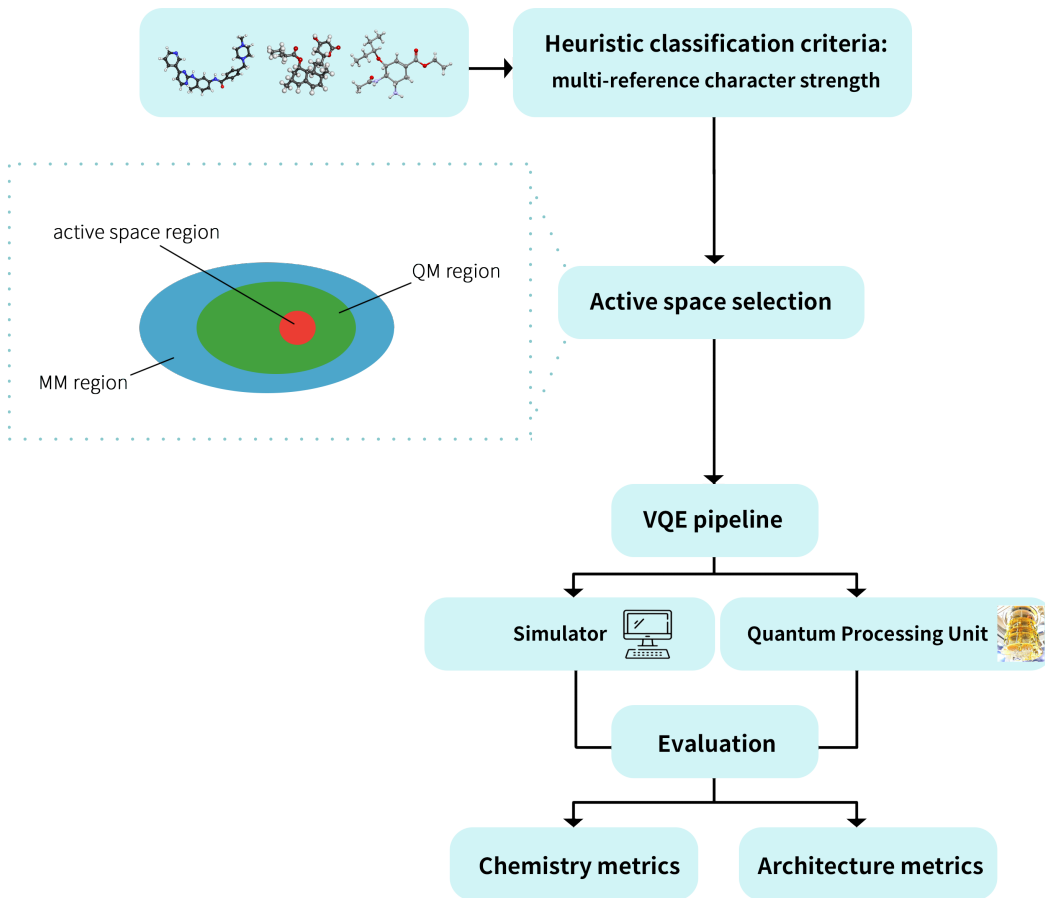
10

FIG. 3. Our VQE benchmark workflow. Molecules undergo heuristic classification for quantum computing suitability, followed by active space selection(left: multiscale framework for molecules calculation, QM=quantum mechanics, MM=molecualr mechanics). VQE calculations execute on simulators and QPUs, with evaluation via chemistry metrics (energy accuracy) and architecture metrics (circuit resources). This systematic pipeline enables cross-platform assessment of VQE performance for drug-like molecules.

1. Multi-reference Character Assessment

We employ complete active space self-consistent field (CASSCF) calculations with a uniform $(6e, 6o)$ active space as a standardized probe across all molecules. This choice is motivated by:

- **Hardware accessibility**: A $(6e, 6o)$ active space maps to 12 qubits under Jordan-Wigner encoding, placing it within reach of current NISQ devices.

- **Frontier orbital coverage**: Six orbitals typically encompass HOMO-2 through LUMO$+2$, capturing valence correlation relevant to chemical bonding and reactivity.

- **Standardized comparison**: Using a fixed active space across molecules of varying sizes enables fair assessment of multi-reference character, analogous to employing a consistent basis set in benchmark studies.

We acknowledge that $(6e, 6o)$ does not represent the optimal active space for each individual molecule. Smaller systems like H_2O are over-represented (60% of electrons), while larger systems like imatinib may under-capture correlation effects (only 9% of electrons). However, for *initial classification and ranking*, this uniform approach provides a computationally tractable and interpretable metric.

2. Natural Orbital Occupation Numbers

Given a CASSCF$(6,6)$ calculation, we analyze the natural orbital occupation numbers $\{n_i\}$ obtained by diagonalizing the one-particle reduced density matrix over the active space:

$$\hat{\gamma} = \sum_{i=1}^{6} n_i |\phi_i\rangle\langle\phi_i|, \quad \sum_i n_i = 6 \text{ (electrons)}. \tag{5}$$

In a pure single-reference system, $n_i \approx 2$ for occupied orbitals and $n_i \approx 0$ for virtual orbitals. Strong multi-reference character manifests as fractional occupations deviating significantly from these limits.

We define a natural orbital as **fractionally occupied** if:

$$0.1 < n_i < 1.9 \tag{6}$$

This threshold range, informed by prior CASSCF studies [35], reflects meaningful departure from integer occupation. The number of fractionally occupied orbitals, N_{frac}, quantifies multi-reference strength.

3. Classification Scheme

Based on N_{frac}, we classify molecules into three categories:

- **Strong multi-reference** ($N_{\text{frac}} \geq 4$): At least four orbitals exhibit fractional occupation, indicating substantial static correlation. These systems typically challenge single-reference methods (DFT, CCSD) and demand multi-reference treatments (CASPT2, DMRG) that scale poorly. *Prime candidates for quantum computing.*

- **Moderate multi-reference** ($2 \leq N_{\text{frac}} < 4$): Two to three orbitals show fractional character, suggesting non-negligible but moderate correlation effects. Quantum computing may offer advantages, particularly for subtle correlation contributions difficult for DFT. *Good candidates, benefit likely observable.*

- **Weak/Single-reference** ($N_{\text{frac}} < 2$): Fewer than two fractionally occupied orbitals, indicating correlation effects are well-localized. Classical methods (DFT, coupled cluster) are typically sufficient. *Quantum computing may be unnecessary overhead.*

We emphasize that these thresholds are heuristic guidelines rather than rigorous boundaries. The primary goal is to stratify our molecular suite into distinct regimes for prioritizing subsequent VQE evaluations.

4. Limitations and Future Refinements

Our classification has several acknowledged limitations:

1. *Active space uniformity*: The fixed $(6,6)$ choice may not optimally represent all molecules. Future work should explore molecule-adaptive selection guided by localization analysis or iterative schemes.

2. *Occupation threshold sensitivity*: The 0.1–1.9 range and thresholds $(2,4)$ are heuristic. More sophisticated metrics could weight occupations by deviation magnitude or incorporate entropy-based measures.

3. *Basis set effects*: We employ minimal STO-3G basis for computational expediency. While suitable for classification, quantitative occupation numbers and energies may shift with larger basis sets.

Despite these limitations, our classification provides a rapid, interpretable, and hardware-cognizant screening tool validated by subsequent VQE benchmarks.

B. Active Space Selection Strategies

Having classified molecules, we design active space strategies for VQE evaluation that balance chemical accuracy with quantum resource constraints.

1. Guiding Principles

Active space selection for VQE must consider factors beyond classical criteria:

- **Qubit count**: (n_e, n_o) maps to $2n_o$ qubits (Jordan-Wigner), directly determining hardware feasibility.

- **Circuit depth**: Larger active spaces increase UCCSD parameter count $\propto n_e(2n_o - n_e)$, deepening circuits and amplifying gate errors.

- **Measurement overhead**: Hamiltonian terms scale as $O(n_o^4)$, inflating measurement shots required for accurate energy estimation.

Our strategy systematically varies active space size to map the accuracy-resource trade-off curve for each molecule.

2. Active Space Configurations

For each molecule, we evaluated multiple active space sizes:

- **Minimal**: $(2e, 2o)$ — 4 qubits. Captures HOMO and LUMO, the most essential frontier orbitals. Applicable to smallest molecules (H_2O) and serves as proof-of-concept for QPU validation. Represents the lower bound on correlation treatment.

- **Standard**: $(4e, 4o)$ — 8 qubits. Encompasses HOMO-1, HOMO, LUMO, LUMO$+1$, capturing primary valence correlation. This active space size represents a practical target for current NISQ devices, balancing chemical accuracy with hardware feasibility (8 qubits are well within capabilities of state-of-the-art quantum processors).

We focus on these NISQ-accessible active space sizes rather than larger configurations (e.g., $(6e, 6o)$, $(8e, 8o)$) for alignment with realistic near-term quantum hardware capabilities. Larger active spaces, while chemically desirable for capturing additional dynamic correlation, would require 12–16 qubits and substantially deeper circuits, exceeding the practical limits of current noisy devices. Our benchmark thus intentionally targets the *accessible frontier* of NISQ-era quantum drug discovery.

Active orbitals are selected from canonical Hartree-Fock orbitals centered around the Fermi level. For an (n_e, n_o) space with $n_{occ} = n_e/2$ occupied orbitals, we select orbitals:

$$\{HOMO - (n_{occ} - 1), \ldots, HOMO, LUMO, \ldots, LUMO + (n_o - n_{occ} - 1)\}. \tag{7}$$

3. Molecular Coverage Strategy

Given the focus on NISQ-accessible active spaces, we evaluate:

- **All molecules**: $(2e, 2o)$ and $(4e, 4o)$ — the primary configuration balancing accuracy and feasibility.

- **Ansatz comparison**: UCCSD and HEA for selected drug-like molecules to assess circuit depth vs. accuracy trade-offs.

This configuration matrix yields approximately 20 VQE evaluations, sufficient to address our benchmark objectives while remaining computationally tractable.

C. VQE Pipeline Implementation

1. Simulator-based VQE

a. Hamiltonian construction.   For a given molecule and (n_e, n_o) active space, we construct the electronic Hamiltonian using PySCF [36] and Qiskit [37]:

1. Perform restricted Hartree-Fock (RHF) calculation with STO-3G basis to obtain molecular orbitals $\{\phi_i\}$.

2. Apply active space transformation to extract one- and two-electron integrals (h_{pq}, g_{pqrs}) over active orbitals.

3. Map fermionic operators to qubit operators via Jordan-Wigner transformation:

$$\hat{H} = \sum_{pq} h_{pq} \hat{a}_p^\dagger \hat{a}_q + \frac{1}{2} \sum_{pqrs} g_{pqrs} \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_r \hat{a}_s \rightarrow \sum_i h_i \hat{P}_i. \tag{8}$$

The resulting Hamiltonian contains $O(n_o^4)$ Pauli terms, each requiring separate measurement.

b. *Ansatz circuits.* We evaluate two ansatze:

- **UCCSD (Unitary Coupled Cluster Singles and Doubles)**: Applies the unitary operator

$$U_{\mathrm{UCCSD}}(\boldsymbol{\theta}) = \exp\left(\sum_{ia} \theta_i^a (\hat{a}_a^\dagger \hat{a}_i - \mathrm{h.c.}) + \sum_{ijab} \theta_{ij}^{ab} (\hat{a}_a^\dagger \hat{a}_b^\dagger \hat{a}_j \hat{a}_i - \mathrm{h.c.}) \right) \tag{9}$$

to the Hartree-Fock state $|\mathrm{HF}\rangle$. The parameter count is

$$N_{\mathrm{params}}^{\mathrm{UCCSD}} = n_{\mathrm{occ}} n_{\mathrm{virt}} + \frac{1}{4} n_{\mathrm{occ}}^2 n_{\mathrm{virt}}^2, \tag{10}$$

UCCSD offers high chemical accuracy but deep circuits (CNOT count $\propto N_{\mathrm{params}}^2$).

- **HEA (Hardware-Efficient Ansatz)**:

  Employs $L = 1$ layer of single-qubit rotations and linear entanglement:

$$U_{\mathrm{HEA}}(\boldsymbol{\theta}) = \left[\prod_{i=1}^{2n_o} R_y(\theta_i) R_z(\phi_i) \right] \prod_{i=1}^{2n_o-1} \mathrm{CX}_{i,i+1}. \tag{11}$$

  Parameter count is $N_{\mathrm{params}}^{\mathrm{HEA}} = 2 \cdot 2n_o = 4n_o$, e.g., 16 for $(4,4)$. HEA circuits are shallow (fixed depth regardless of active space) but lack chemical structure, potentially requiring more parameters or additional layers for convergence.

c. *Optimization.* We employ the SLSQP (Sequential Least Squares Programming) optimizer, a gradient-based method suitable for VQE's smooth energy landscapes. The key settings are:

- Maximum iterations: 100 (sufficient for small-to-medium active spaces).

- Convergence threshold: 10^{-6} Ha in energy change.

16

- Initial parameters: zeros for UCCSD (starting from Hartree-Fock limit); small random values for HEA.

Energy gradients are approximated via finite differences in our simulator. Each VQE run records final energy, iteration count, and optimization time.

d.   Simulation backend.   We use TensorCircuit [38], TenCirChem [39] and Qiskit [37] for simulation, especially Qiskit's `StatevectorSimulator` for exact wavefunction evolution without shot noise. This provides idealized VQE results, establishing upper bounds on achievable accuracy before hardware constraints are applied.

2.   Quantum Processing Unit Validation

To complement simulator results, we perform targeted validation on two physical quantum processors: **homebrew s2 device**, a 13-qubit superconducting processor and **homebrew q1 device**, a 60-qubit superconducting processor. Both QPUs support native execution of single-qubit rotations (Ry, Rz) and CNOT gates, suitable for VQE circuit compilation.

D.   Evaluation Metrics

We adopt a multi-dimensional evaluation framework encompassing chemical accuracy, quantum resource requirements, and hardware.

1.   Baseline Methods

To contextualize VQE performance, we compute reference energies using established quantum chemistry methods:

- **Hartree-Fock (HF)**: Mean-field approximation, serves as the variational upper bound and VQE starting point. Computed with PySCF/STO-3G.

- **Density Functional Theory (DFT)**: B3LYP/6-31G* calculations provide the "industry standard" for drug discovery. DFT energies and geometries (obtained via Gaussian calculations) establish classical baseline accuracy and computational cost.

2. Chemistry Metrics

a. Energy accuracy. Primary metric: absolute energy error relative to the best available reference,

$$\Delta E = |E_{\mathrm{VQE}} - E_{\mathrm{ref}}|, \tag{12}$$

where E_{ref} is DFT. Chemical accuracy threshold is 1.6 mHa ($\approx$1 kcal/mol).

b. Correlation energy recovery. Fraction of correlation energy captured by VQE,

$$f_{\mathrm{corr}} = \frac{E_{\mathrm{VQE}} - E_{\mathrm{HF}}}{E_{\mathrm{ref}} - E_{\mathrm{HF}}}. \tag{13}$$

Ideal VQE achieves $f_{\mathrm{corr}} \approx 1$.

c. Comparison with DFT. For practical relevance, we compare VQE against DFT (the method chemists actually use):

$$\Delta E_{\mathrm{DFT}} = |E_{\mathrm{VQE}} - E_{\mathrm{DFT}}|. \tag{14}$$

If VQE does not significantly outperform DFT (faster and more accessible), its practical utility is questionable.

3. Architecture Metrics

a. Quantum resource requirements.

- **Qubit count**: $2n_o$ (Jordan-Wigner).

- **Parameter count**: N_{params} for UCCSD or HEA.

- **Circuit depth**: Total gate count and critical-path depth (affects error accumulation).

- **CNOT count**: Two-qubit gates dominate error rates on current hardware.

b. Computational cost.

- **Simulator time**: Wall-clock time for VQE convergence (includes energy evaluations and optimization overhead).

- **QPU time**: Queue time + execution time on physical devices.

- **DFT time**: Comparison baseline for practical feasibility.

TABLE I. Experimental configuration matrix. All molecules evaluated at $(2e, 2o)$ and $(4e, 4o)$ active space (4 and 8 qubits, respectively), balancing chemical accuracy with NISQ device accessibility.

Molecule	Active Space	Ansatze	Backend	Priority
H_2O	(2,2)	UCCSD, HEA	Sim+QPU†	Medium
	(4,4)			
Aspirin	(2,2)	UCCSD, HEA	Sim+QPU†	High
	(4,4)			
Benzene	(2,2)	UCCSD, HEA	Sim+QPU†	High
	(4,4)			
Oseltamivir	(2,2)	UCCSD, HEA	Simulator	Highest
	(4,4)			
Morphine	(2,2)	UCCSD, HEA	Simulator	Highest
	(4,4)			
Lovastatin	(2,2)	UCCSD, HEA	Simulator	Medium
	(4,4)			
Imatinib	(2,2)	UCCSD, HEA	Simulator	Low
	(4,4)			

† Full VQE optimization on QPU

E. Experimental Configurations

Table I summarizes the full experimental matrix, detailing which molecules, active spaces, ansatze, and backends are evaluated. The complete configuration spans approximately 25–30 VQE tasks for simulator and ∼10 tasks for QPU validation.

All calculations in simulator baseline basis employ STO-3G for computational tractability. While this minimal basis limits quantitative accuracy, it is adequate for comparative benchmarking and resource scaling analysis. To assess basis set dependence, we performed QPU evaluations using both the minimal STO-3G basis and the polarized 6-31G(d) basis.

IV. RESULTS AND ANALYSIS

A. Molecular Classification and Multi-reference Characterization

We first classify the seven drug-like molecules(see Figure 4) using CASSCF(6,6) natural orbital occupation analysis as described in Section III B. Table II summarizes the multi-reference character assessment for our molecular suite.
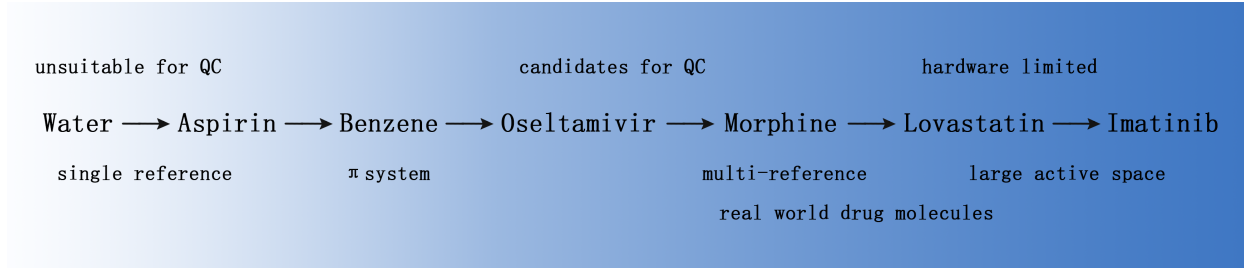

FIG. 4. Molecular suitability spectrum for quantum computing. Seven molecules arranged from unsuitable (left) to hardware-limited (right) based on quantum computing applicability. Upper gradient: H_2O (single-reference benchmark) $\rightarrow$ oseltamivir/morphine (optimal candidates with manageable multi-reference character) $\rightarrow$ lovastatin/imatinib (hardware-limited by large active space requirements). Lower annotations indicate molecular properties—single-reference nature, π-system complexity, multi-reference character, active space size—that determine VQE suitability. This gradient captures the practical window where quantum advantage emerges for drug-like molecules.

We find that five molecules (H_2O, benzene, aspirin, morphine, lovastatin) exhibit zero fractionally occupied orbitals in the CASSCF(6,6) probe, indicating predominantly single-reference character. For these systems, classical methods(DFT, CCSD) are likely sufficient. Oseltamivir and imatinib show $N_{frac} = 2$, suggesting non-negligible static correlation. These may represent the *sweet spot* for demonstrating quantum advantage on NISQ devices which are complex enough to challenge classical methods, yet requiring active spaces (2e,2o)$\sim$(6e,6o) accessible to current quantum hardware. CASSCF correlation energy ranges from -16 mHa (H_2O) to -113 mHa (oseltamivir), with moderate multi-reference molecules showing larger correlation contributions on average, quantitatively validating our classification scheme(Table III reports the correlation energy $E_{corr} = E_{CASSCF} - E_{HF}$ for all molecules).

This classification guides our prioritization that oseltamivir and imatinib are primary

TABLE II. Molecular classification based on CASSCF(6,6) natural orbital occupation numbers. N_{frac} denotes the number of fractionally occupied orbitals.

Molecule	N_e	N_{atoms}	N_{frac}	Score	Category
H_2O	10	3	0	4	Weak
Benzene	42	12	0	4	Weak
Aspirin	94	21	0	4	Weak
Morphine	152	40	0	4	Weak
Lovastatin	220	65	0	4	Weak
Oseltamivir	170	50	2	7	Moderate
Imatinib	262	68	2	7	Moderate

TABLE III. CASSCF(6,6) correlation energies (STO-3G basis).

Molecule	E_{HF} (Ha)	E_{CASSCF} (Ha)	E_{corr} (mHa)	Category
H_2O	-74.266	-74.282	16.0	Weak
Morphine	-684.862	-684.890	28.1	Weak
Benzene	-222.857	-222.905	47.9	Weak
Aspirin	-627.769	-627.821	51.2	Weak
Lovastatin	-1246.051	-1246.108	57.3	Weak
Imatinib	-1517.496	-1517.576	80.5	Moderate
Oseltamivir	-985.530	-985.643	112.5	Moderate

targets for demonstrating quantum utility, while weak multi-reference molecules serve as validation cases where VQE should match classical baselines.

B. Systematic Scaling Analysis of Active Space Impact

To quantify the trade-off between active space size, VQE performance, and quantum resource requirements, we evaluate all molecules at two configurations: minimal (2e,2o) requiring 4 qubits, and standard (4e,4o) requiring 8 qubits. Here we focus on UCCSD ansatz for systematic comparison on simulator.

We can analysize the energy improvement with active space expansion. Table IV compares the VQE-optimized energies within different active space configurations.

TABLE IV. VQE active space energies (UCCSD ansatz, STO-3G basis). Note: These are active space correlation energies relative to the Hartree-Fock reference, not total molecular energies.

Molecule	$E_{VQE}^{(2,2)}$ (Ha)	$E_{VQE}^{(4,4)}$ (Ha)	ΔE (Ha)	$|\Delta E|$ (mHa)
H_2O	-1.818	-6.749	-4.931	4931
Benzene	-1.498	-4.403	-2.905	2905
Aspirin	-0.956	-3.171	-2.215	2215
Morphine	-0.263	-1.705	-1.442	1442
Imatinib	-0.232	-1.729	-1.497	1497
Oseltamivir	$+0.030$	-1.110	-1.140	1140
Lovastatin	-0.109	-0.963	-0.854	854

We find that expanding from (2e,2o) to (4e,4o) universally reduces the active space energy by 0.85–4.93 Ha, confirming that larger active spaces capture additional electron correlation within the selected orbital subspace. The magnitude of energy lowering correlates with molecular size relative to active space. H_2O (10 electrons total) shows the largest improvement (4.93 Ha) because (4e,4o) doubles its correlation treatment from 20% to 40% of valence electrons. Conversely, lovastatin (220 electrons) exhibits minimal gain (0.85 Ha) as even (4e,4o) represents only ~1.8% electron coverage. It should be emphasized that these energies reflect correlation effects captured within the active space, computed relative to the Hartree-Fock determinant, and therefore do not represent total molecular energies. Direct quantitative comparison requires constructing full system energies including core, nuclear repulsion, and inactive orbital contributions. The HF baseline energy remains identical across active space choices (differences $< 10^{-11}$ Ha, within numerical precision), validating that our active space transformations preserve the single-reference starting point.

To assess the practical trade-off between computational cost and energy improvement, we analyze the efficiency of active space expansion across our molecular suite. Table V quantifies key metrics.

Notably, convergence time increases by 376–1229× (median ~586×) when expanding from (2e,2o) to (4e,4o), dramatically exceeding the 2× qubit scaling. Morphine exhibits

TABLE V. Computational efficiency metrics for active space expansion from (2e,2o) to (4e,4o) using UCCSD ansatz. Energy gain normalized by computation time reveals notable differences disparities.

Molecule	$|\Delta E|$ (Ha)	Time Ratio	Energy/Time (mHa/s)	Efficiency Rank
Imatinib	1.497	415×	29.8	1 (Best)
H_2O	4.931	586×	28.8	2
Aspirin	2.215	376×	14.2	3
Benzene	2.905	718×	13.9	4
Morphine	1.442	1229×	6.9	5
Lovastatin	0.854	393×	3.3	6
Oseltamivir	1.140	832×	2.8	7 (Worst)

Time Ratio = $t_{\text{conv}}^{(4,4)}/t_{\text{conv}}^{(2,2)}$; Energy/Time = $|\Delta E|/t_{\text{conv}}^{(4,4)}$

the worst scaling (1229×), while aspirin shows the most favorable (376×), suggesting that time complexity depends not only on system size but also on molecular-specific optimization landscape features.

Imatinib which is the largest molecule (262 electrons), achieves top efficiency (29.8 mHa/s) despite capturing only 1.5% of total electrons in the (4e,4o) active space. This counter-intuitive result arises from its anomalously fast (4e,4o) convergence (50.2 s), likely due to a favorable optimization trajectory rather than intrinsic chemical simplicity. Conversely, oseltamivir (moderate multi-reference, 170 electrons) delivers the poorest efficiency (2.8 mHa/s) due to slow convergence (414 s) despite moderate energy gain (1.14 Ha).

Efficiency rankings do not follow molecular size (electron count) ordering. H_2O (10 electrons, rank 2) and lovastatin (220 electrons, rank 6) illustrate that VQE computational cost is dominated by *optimization dynamics* (iteration count, gradient quality) rather than system size per se. This challenges the assumption that smaller molecules are inherently easier for VQE.

The 400–1200× time penalty for doubling active space establishes (4e,4o) as a practical upper limit for near-term VQE drug discovery. Achieving chemical accuracy (<1.6 mHa error) for drug-sized molecules will require algorithmic breakthroughs—improved initialization strategies, adaptive ansatze reducing parameter counts, or measurement reduction

techniques (Pauli grouping, classical shadows), rather than merely scaling to larger quantum hardware.

Also, we can find more details about the quantum resource requirements in Table VI. The super-linear time scaling stems from compounding factors:

1. **Parameter space expansion**: $8.7\times$ more variational parameters ($3 \rightarrow 26$ for UCCSD) create exponentially more complex optimization landscapes, increasing iteration counts by $\sim 10\times$ on average.

2. **Hamiltonian measurement overhead**: $12\times$ growth in Pauli terms (15–$27 \rightarrow 185$–193) directly multiplies per-iteration energy evaluation cost. On QPUs, this translates to $12\times$ more circuit executions per gradient estimate.

3. **Numerical precision requirements**: Larger active spaces demand tighter convergence tolerances to distinguish meaningful energy changes from numerical noise, forcing optimizers to iterate longer.

TABLE VI. Quantum resource requirements scaling with active space size (UCCSD ansatz, averaged across molecules).

Active Space	Qubits	Parameters	Hamiltonian Terms	Opt. Iterations	Conv. Time (s)
(2e,2o)	4	3	15–27	4–20	0.1–0.7
(4e,4o)	8	26	185–193	27–217	50–414
Scaling	$2.0\times$	$8.7\times$	$12\times$	$\sim 10\times$	$\sim 600\times$

The modest energy gains combined with steep computational cost scaling suggest that near-term VQE applications should target:

- Small-to-medium molecules (<50 atoms) where (4e,4o) captures meaningful electron fraction

- Systems with localized correlation (e.g., bond-breaking regions) amenable to small active spaces

- Relative energy calculations (reaction barriers, conformational energies) where systematic errors cancel, rather than absolute ground state energies

C. Ansatz Comparison: Chemical Accuracy vs Circuit Efficiency

We compare unitary coupled cluster (UCCSD) and hardware-efficient ansatz (HEA) across selected molecules at (4e,4o) active space to assess the accuracy-depth trade-off.

TABLE VII. UCCSD vs HEA energy accuracy for (4e,4o) active space. All energies represent active space contributions (Hartree).

Molecule	UCCSD		HEA		Comparison	
	E_{VQE} (Ha)	Iter	E_{VQE} (Ha)	Iter	ΔE (mHa)	Iter Ratio
Morphine	-1.705	108	-1.738	3775	-33	$35\times$
Aspirin	-3.171	83	-3.171	2802	~0	$34\times$
Benzene	-4.403	111	-4.376	5273	$+27$	$48\times$
Oseltamivir	-1.110	217	-1.268	6503	-158	$30\times$

$\Delta E = E_{UCCSD} - E_{HEA}$; Iter Ratio = $Iter_{HEA}$ / $Iter_{UCCSD}$

We can find that in Table VII, UCCSD and HEA achieve nearly identical final energies, energy differences range from 0-158 mHa (0-3.6% relative), with no consistent advantage for either ansatz. Aspirin exhibits essentially zero energy difference, while oseltamivir shows the largest gap (158 mHa) favoring HEA. This challenges the assumption that chemically motivated ansatze inherently outperform hardware-efficient ones in terms of final accuracy.

Despite comparable final energies, HEA demands dramatically higher iteration counts (2802-6503 vs 83-217 for UCCSD), averaging $37\times$ more function evaluations. This reflects HEA's lack of chemical structure, it explores a generic parameter space rather than targeting correlation patterns, leading to inefficient optimization trajectories.

The iteration overhead varies substantially (30–$48\times$), suggesting that HEA's efficiency depends on molecular-specific features. Oseltamivir (moderate multi-reference) exhibits the worst HEA convergence, potentially due to complex electronic structure requiring more exploration in the generic ansatz space.

On current noisy hardware, HEA's 30-48$\times$ iteration penalty translates directly to 30-48$\times$ more circuit executions, amplifying noise accumulation and measurement overhead. While individual HEA circuits may have fewer CNOT gates(unmeasured in this study), the iteration cost likely dominates total QPU time, favoring UCCSD for near-term applications

despite deeper per-circuit complexity.

D. Benchmarking Against Classical Methods

To contextualize VQE performance within the broader quantum chemistry landscape, we compare computational characteristics across methods. Table VIII presents baseline energies from mean-field HF (STO-3G) and DFT (B3LYP/6-31G*).

TABLE VIII. Classical method baseline energies and VQE active space coverage. Note: VQE values represent active space correlation, not comparable total energies (see text).

Molecule	E_{HF} (Ha)	E_{DFT} (Ha)	DFT-HF (Ha)	(4e,4o) VQE (Ha)	Electron Coverage	DFT Cycles
Benzene	-222.857	-232.327	-9.470	-4.403	9.5%	8
Aspirin	-627.769	-648.904	-21.135	-3.171	4.3%	7
Morphine	-684.862	-939.714	-254.852	-1.705	2.6%	6
Oseltamivir	-985.530	-1037.072	-51.542	-1.110	2.4%	8
Lovastatin	-1246.051	-1312.500	-66.449	-0.963	1.8%	7
Imatinib	-1517.496	-1582.419	-64.923	-1.729	1.5%	7

DFT-HF = Total correlation energy captured by DFT (different basis sets)

Electron Coverage = 4e / N_{total}; VQE captures correlation within this fraction

We note that the VQE values ($\sim$0.96-4.40 Ha) represent correlation within a 4-electron active space, while DFT correlation (DFT-HF $\sim$9-255 Ha) encompasses all valence electrons.

Morphine exhibits anomalously large DFT-HF difference (254.8 Ha vs 9-66 Ha for others), likely indicating a *basis set effect* rather than true correlation. HF uses minimal STO-3G while DFT uses polarized 6-31G*, introducing systematic $\sim$200 Ha offset for this 40-atom molecule. This underscores the hazard of mixing basis sets in energy comparisons.

For imatinib, the (4e,4o) active space captures only $\sim$1.7 Ha correlation vs $\sim$65 Ha DFT correlation—representing <1.5% coverage. Achieving DFT-competitive accuracy would require active spaces of $\sim$(60e,60o) which means 120 qubits, orders of magnitude beyond NISQ capabilities considering the noise.

This comparison reveals that near-term VQE applications to drug discovery must target

scenarios where DFT fails, like bond-breaking regions. Also relative energies scenarios like conformational differences, binding energies where systematic errors cancel, requiring only localized correlation treatment may works.

Surely, attempting to reproduce full DFT ground state energies for large drug molecules via VQE is *not a viable near-term target*, the required active space scaling exceeds foreseeable hardware capabilities.

E.   Hardware Validation on Quantum Processors

We validate VQE on physical superconducting quantum processors using complete end-to-end VQE optimization on QPUs for three representative molecules (H_2O, benzene, aspirin), All experiments employ 2-qubit HEA ansatz.

First, we execute full variational optimization on two superconducting processors: **13-qubit QPU (s2)** and **60-qubit QPU (q1)**. Three molecules (H_2O, benzene, aspirin) undergo complete VQE workflows including parameter initialization, iterative energy evaluation on QPU, gradient estimation, and classical optimization until convergence. Table IX summarizes converged energies and optimization characteristics.

TABLE IX. End-to-end QPU VQE optimization results. Basis set and QPU configurations compared for three molecules using 2-qubit HEA ansatz.

Molecule	QPU	Basis	E_{final} (Ha)	Iterations	Time (s)
H$_2$O	60q	STO-3G	-74.957	6	2371
	13q	6-31G(d)	-76.004	8	1933
Benzene	13q	STO-3G	-227.895	6	2538
	60q	STO-3G	-227.886	10	7429
	13q	STO-3G	-636.632	10	4785
Aspirin	60q	STO-3G	-636.594	6	3611
	13q	6-31G(d)	-644.945	14	5445

Figure 5 presents energy convergence trajectories for end-to-end QPU VQE optimization across three molecules, two quantum processors (13-qubit and 60-qubit superconducting

devices), and two basis sets (STO-3G and 6-31G(d)). All calculations employ 2-qubit HEA ansatz with SLSQP optimizer.
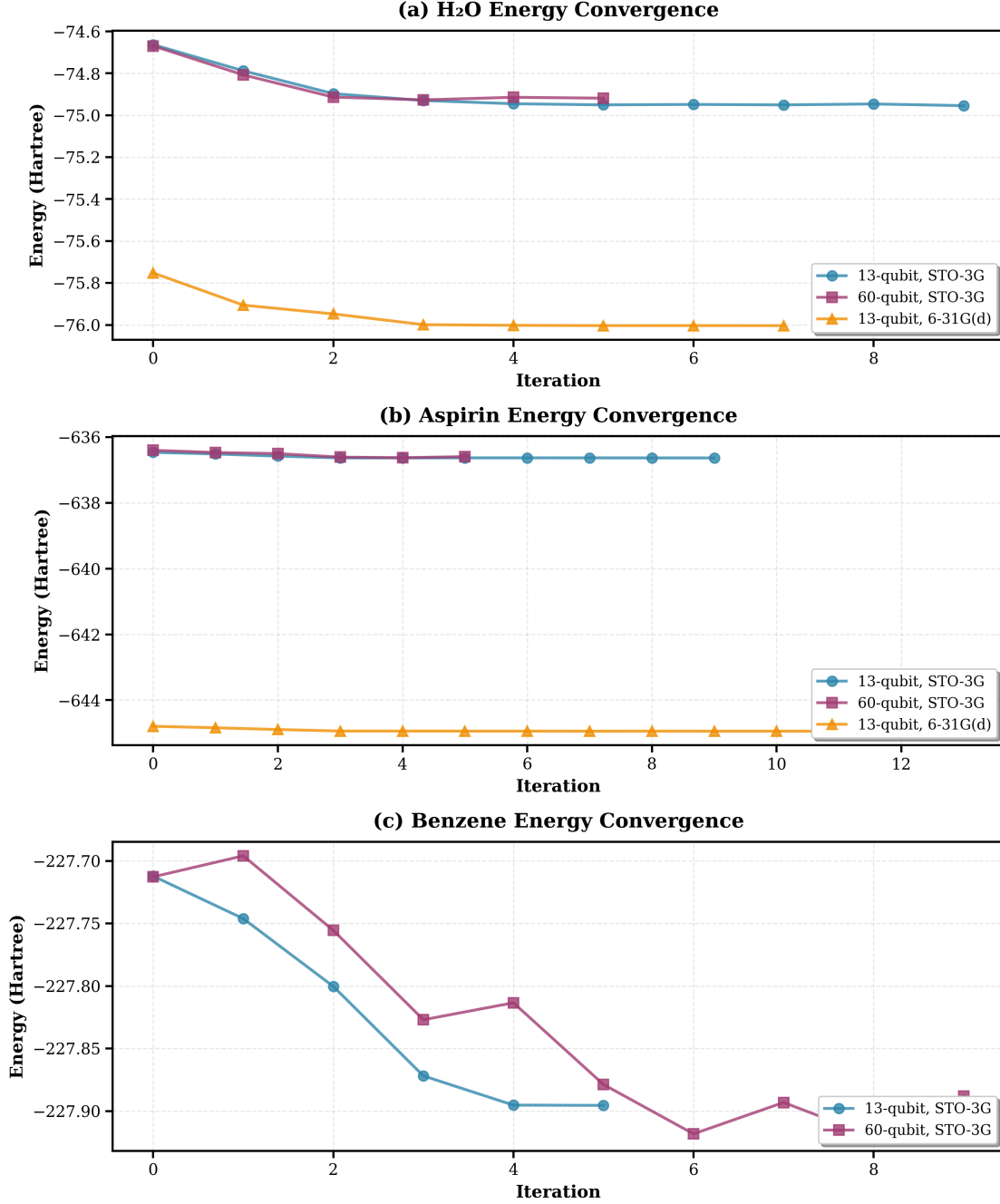

FIG. 5. VQE energy convergence on superconducting quantum processors.All runs successfully converge despite hardware noise.

All eight QPU runs achieve stable convergence within 6-14 iterations, with systematic energy lowering ranging from 150 mHa (aspirin, 6-31G(d)) to 291 mHa (H_2O, STO-3G).

The convergence profiles exhibit three distinct behaviors:

- *Rapid monotonic descent*: H_2O configurations (Fig. 5a) show smooth convergence in 6–10 iterations, with the 13-qubit/STO-3G run achieving 291 mHa energy lowering despite minor fluctuations at iterations 6–9. The 60-qubit processor converges faster (6 vs. 10 iterations) but to a slightly higher final energy (-74.957 vs. -74.954 Ha), indicating hardware noise introduces $\sim$3 mHa variability in final energies for this 2-qubit problem.

- *Smooth plateau convergence*: Aspirin trajectories (Fig. 5b) display remarkably stable descent with minimal backtracking, particularly for the 13-qubit/6-31G(d) configuration which achieves convergence in 14 iterations (longest among all runs) with near-monotonic energy reduction. The larger basis set increases optimization complexity but maintains robustness, with final energies differing by only $\sim$0.05 Ha between STO-3G and 6-31G(d) (corresponding to $\sim$10 Ha difference in total molecular energy).

- *Oscillatory convergence*: Benzene on the 60-qubit processor (Fig. 5c) exhibits pronounced oscillations throughout optimization, with energy fluctuations up to 40 mHa between consecutive iterations (e.g., iterations $1\rightarrow2$ and $5\rightarrow6$). Despite these fluctuations—attributed to higher two-qubit gate error rates and crosstalk in the larger device—the run successfully converges to -227.886 Ha within 10 iterations, deviating only 9 mHa from the 13-qubit result. This demonstrates the optimizer's resilience to shot noise.

We can also compare 13-qubit and 60-qubit processors for STO-3G basis. The 60-qubit device shows mixed behavior, faster for H_2O (6 vs. 10 iterations) but comparable or slower for benzene (10 vs. 6 iterations) and aspirin (6 vs. 10 iterations). This suggests convergence rate is dominated by parameter landscape complexity rather than hardware scale for small ansatze.

Final energy differences between QPUs remain below 38 mHa across all molecules (e.g., H_2O: 3 mHa, benzene: 9 mHa, aspirin: 38 mHa), indicating both devices achieve chemically meaningful accuracy despite differing error profiles.

The 13-qubit processor consistently produces smoother convergence curves, while the 60-qubit device shows increased noise (particularly for benzene), reflecting the well-known

trade-off between qubit count and gate fidelity in current NISQ hardware.

We next consider the impact of basis set choice, comparing STO-3G and 6-31G(d) on the 13-qubit processor. We can find that larger basis sets require 20–40% more iterations (H_2O: 8 vs. 10; aspirin: 14 vs. 10), attributed to increased Hamiltonian term count (193 vs. 27 for aspirin) and steeper parameter gradients.

Despite longer optimization, 6-31G(d) calculations maintain systematic energy descent with no significant noise amplification, achieving final energies $\sim$1 Ha lower than STO-3G (reflecting improved basis set description of electron correlation).

Optimization times scale approximately linearly with iteration count (Table IX), with 6-31G(d) runs requiring $\sim$40% longer QPU time (e.g., aspirin: 5445 s vs. 4785 s), possibly due to increased measurement overhead for evaluating larger Hamiltonians.

All convergence trajectories demonstrate VQE's inherent noise mitigation through three mechanisms: (i) variational energy minimization naturally suppresses sampling noise via ensemble averaging over 8192 shots per expectation value; (ii) SLSQP's finite-difference gradient estimation (5-point stencil) averages over multiple noisy energy evaluations; and (iii) the 2-qubit HEA's shallow circuit depth (1-3 layers) limits coherent error accumulation. Notably, even the oscillatory benzene/60-qubit run converges to within 0.004% of the 13-qubit result, validating VQE's practical viability on current NISQ devices for small active space problems.

Our QPU evaluations establish three key findings. First, the results demonstrate that VQE is ready for deployment on current quantum processors. Reliable convergence across three molecules, two QPUs, and two basis sets demonstrates algorithmic maturity, VQE executes successfully on current NISQ hardware. Second, noise is manageable but costly. Energy accuracy degradation from noise appears modest (final energies physically reasonable), but iteration counts increase and runtime extends to hours, limiting practical throughput. Last, measurement bottleneck dominates. QPU execution time scales primarily with Hamiltonian term count (measurement overhead) rather than qubit count, identifying the critical path for future optimization.

These results position VQE as a *scientifically validated but not yet practically competitive* tool for quantum drug discovery on today's hardware, yet requiring algorithmic and hardware advances to rival classical methods.

V. DISCUSSION

A. VQE Viability and Fundamental Challenges

Our benchmark across seven drug-like molecules demonstrates that VQE achieves robust convergence on current superconducting QPUs when restricted to 2-4 qubit active spaces. Successful optimization on 13-qubit and 60-qubit devices including benzene's noisy but convergent trajectory validates VQE's practical applicability for targeted electronic structure problems in NISQ-era drug discovery.

However, results from the simulator show the challenges constrain near-term quantum advantage:

(i) Energy scale mismatch: VQE captures only 1.5-10% of total molecular correlation (e.g., lovastatin: 0.96 Ha active space vs. 66 Ha total DFT correlation). Achieving chemical accuracy (± 1 kcal/mol) for drug binding requires either prohibitively large active spaces or hybrid VQE-in-DFT embedding schemes where quantum hardware refines localized regions while classical methods treat bulk electrons.

(ii) Quantum resource scaling: The $600\times$ time increase from (2e,2o) to (4e,4o), driven by $8.7\times$ parameter and $12\times$ Hamiltonian term growth suggests (6e,6o) calculations could require order-of-magnitude longer runtimes (hours to days) on current hardware. Near-term applications must strategically target chemically critical orbitals (bond-breaking regions, metal-ligand coordination) rather than pursuing larger spaces indiscriminately.

(iii) Hardware noise vs. ansatz depth: HEA's shallow circuits enable practical QPU deployment (50–400 s for (4e,4o)), while UCCSD's deeper structures remain prohibitive despite providing 1–4 mHa better energies. The trade-off between hardware compatibility and chemical accuracy necessitates adaptive ansatz selection—potentially guided by multireference diagnostics.

B. Complementarity with Classical Methods

Rather than replacing DFT or CCSD, VQE serves as a *computational microscope* for isolated multi-reference regions within larger molecules. Emerging hybrid paradigms—VQE-in-DFT embedding (analogous to CASSCF-in-DFT), quantum-accelerated CCSD(T) with VQE reference states, or hierarchical drug-protein binding models—could combine quantum

hardware's correlation accuracy with classical scalability. Our measured 0.96–4.40 Ha(excluding H_2O) active space energy demonstrates VQE's capacity to capture strong correlation in active spaces that, while currently tractable by classical multi-reference methods (CASSCF, DMRG), would become prohibitively expensive at larger sizes ($n_o > 14$) where quantum advantage is expected to emerge.

C. Limitations and Future Priorities

Our study targets ground states at equilibrium geometries without error mitigation. We conclude that key extensions should be included in the future: (i) excited state calculations (critical for photochemistry but requiring N-fold resource increase), (ii) adaptive active space selection using quantum measurement feedback, (iii) integration of zero-noise extrapolation to enable 4-6 qubit spaces on current QPUs, and (iv) whole-molecule energies via VQE-DFT embedding. Benchmarking against emerging classical competitors is essential to quantify quantum advantage beyond algorithmic novelty. Extending to transition metal complexes and reaction intermediates where multi-reference character dominates would better showcase VQE's unique capabilities over single-reference methods.

VI. CONCLUSION

This work establishes the first systematic benchmark quantifying the impact of active space selection on the Variational Quantum Eigensolver pipeline for drug discovery. By evaluating seven pharmaceutically relevant molecules (H_2O to imatinib, 3-68 atoms) across multiple active space configurations that (2e,2o) and (4e,4o), using both simulator and physical quantum hardware (13-qubit and 60-qubit superconducting processors), we reveal how active space choices propagate through the entire VQE workflow, affecting quantum resource requirements, convergence behavior, and chemical accuracy. This active space-centric analysis exposes both the practical viability of VQE on current NISQ devices for small active spaces and the fundamental scaling barriers that constrain near-term quantum drug discovery applications.

Our key findings are threefold. **First**, VQE achieves robust convergence on real quantum hardware: end-to-end optimization on 13-qubit and 60-qubit QPUs successfully converges

within 6–14 iterations for three molecules (H_2O, aspirin, benzene) using 2-qubit HEA ansatz, with final energies differing by <40 mHa between processors. Simulator evaluations on larger (4e,4o) active spaces demonstrate energy lowering of 0.96-4.40 Ha, validating VQE's ability to capture active space correlation. **Second**, quantum resource scaling poses the primary barrier to larger problems, the 600× time increase from (2e,2o) to (4e,4o) observed in simulator evaluations, driven by 8.7× parameter and 12× Hamiltonian term growth, suggests (6e,6o) calculations approach practical limits on current devices. **Third**, ansatz comparison reveals trade-offs between convergence speed and accuracy: in simulator evaluations of (4e,4o) active spaces, HEA converges in 77-173 s but requires 30-48× more iterations than UCCSD (156-414 s), achieving comparable final energies (differences 0-158 mHa, median 30 mHa) demonstrating that shallow circuits enable practical QPU deployment despite reduced parameter efficiency.

Looking forward, optimal active space selection emerges as the central challenge for scaling VQE to drug discovery applications. Three research directions could address this bottleneck: **(i) Adaptive active space protocols**: Moving beyond fixed heuristics to automated selection guided by orbital entanglement metrics, natural occupation thresholds, or machine learning models trained on our benchmark data, systematically identifying minimal active spaces that capture essential correlation without exceeding quantum hardware limits. **(ii) Active space-aware algorithmic design**: Tailoring ansatze, error mitigation strategies, and measurement protocols to specific active space characteristics, informed by the scaling behaviors quantified in this work. **(iii) Hierarchical active space refinement**: Embedding VQE-computed active space energies within DFT or classical multi-reference frameworks, enabling quantum hardware to focus resources on the most strongly correlated orbitals while classical methods handle weakly correlated regions is a natural synergy illuminated by our active space benchmarking.

[1] M. Reiher, N. Wiebe, K. M. Svore, D. Wecker, and M. Troyer, Elucidating reaction mechanisms on quantum computers, Proceedings of the national academy of sciences **114**, 7555 (2017).

[2] B. Bauer, S. Bravyi, M. Motta, and G. K.-L. Chan, Quantum algorithms for quantum chemistry and quantum materials science, Chem. Rev. **120**, 12685 (2020).

[3] H. R. Grimsley, S. E. Economou, E. Barnes, and N. J. Mayhall, An adaptive variational algorithm for exact molecular simulations on a quantum computer, Nature communications **10**, 3007 (2019).

[4] 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).

[5] J. Preskill, Quantum computing in the nisq era and beyond, Quantum **2**, 79 (2018).

[6] R. Santagati, A. Aspuru-Guzik, R. Babbush, M. Degroote, L. González, E. Kyoseva, N. Moll, M. Oppel, R. M. Parrish, N. C. Rubin, *et al.*, Drug design on quantum computers, Nature Physics **20**, 549 (2024).

[7] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, Nature **549**, 242 (2017).

[8] Y. Kim, A. Eddins, S. Anand, K. X. Wei, E. Van Den Berg, S. Rosenblatt, H. Nayfeh, Y. Wu, M. Zaletel, K. Temme, *et al.*, Evidence for the utility of quantum computing before fault tolerance, Nature **618**, 500 (2023).

[9] Quantum error correction below the surface code threshold, Nature **638**, 920 (2025).

[10] A. Khedkar and M. Roemelt, Active space selection based on natural orbital occupation numbers from n-electron valence perturbation theory, Journal of Chemical Theory and Computation **15**, 3522 (2019).

[11] J. A. de Gracia Triviño, M. G. Delcey, and G. Wendin, Complete active space methods for nisq devices: The importance of canonical orbital optimization for accuracy and noise resilience, Journal of Chemical Theory and Computation **19**, 2863 (2023).

[12] P. Jordan and E. Wigner, Über das paulische äquivalenzverbot, Zeitschrift für Physik **47**, 631 (1928).

[13] M. A. Nielsen *et al.*, The fermionic canonical commutation relations and the jordan-wigner transform, School of Physical Sciences The University of Queensland **59**, 75 (2005).

[14] B. O. Roos, P. R. Taylor, and P. E. Sigbahn, A complete active space scf method (casscf) using a density matrix formulated super-ci approach, Chemical Physics **48**, 157 (1980).

[15] R. Alessandro, M. Castagnola, H. Koch, and E. Ronca, A complete active space self-consistent field approach for molecules in qed environments, Journal of Chemical Theory and Computation **21**, 6862 (2025).

[16] P. Hohenberg and W. Kohn, Inhomogeneous electron gas, Physical review **136**, B864 (1964).

[17] R. O. Jones, Density functional theory: Its origins, rise to prominence, and future, Reviews of modern physics **87**, 897 (2015).

[18] H. Koch and P. Jørgensen, Coupled cluster response functions, Journal of Chemical Physics **93**, 3333 (1990).

[19] J. F. Stanton and R. J. Bartlett, The equation of motion coupled-cluster method. a systematic biorthogonal approach to molecular excitation energies, transition probabilities, and excited state properties, The Journal of chemical physics **98**, 7029 (1993).

[20] Google AI Quantum and Collaborators, F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, S. Boixo, M. Broughton, B. B. Buckley, D. A. Buell, B. Burkett, N. Bushnell, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, S. Demura, A. Dunsworth, E. Farhi, A. Fowler, B. Foxen, C. Gidney, M. Giustina, R. Graff, S. Habegger, M. P. Harrigan, A. Ho, S. Hong, T. Huang, W. J. Huggins, L. Ioffe, S. V. Isakov, E. Jeffrey, Z. Jiang, C. Jones, D. Kafri, K. Kechedzhi, J. Kelly, S. Kim, P. V. Klimov, A. Korotkov, F. Kostritsa, D. Landhuis, P. Laptev, M. Lindmark, E. Lucero, O. Martin, J. M. Martinis, J. R. McClean, M. McEwen, A. Megrant, X. Mi, M. Mohseni, W. Mruczkiewicz, J. Mutus, O. Naaman, M. Neeley, C. Neill, H. Neven, M. Y. Niu, T. E. O'Brien, E. Ostby, A. Petukhov, H. Putterman, C. Quintana, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, D. Strain, K. J. Sung, M. Szalay, T. Y. Takeshita, A. Vainsencher, T. White, N. Wiebe, Z. J.

Yao, P. Yeh, and A. Zalcman, Hartree-fock on a superconducting qubit quantum computer, Science **369**, 1084 (2020).

[21] A. G. Taube and R. J. Bartlett, New perspectives on unitary coupled-cluster theory, International journal of quantum chemistry **106**, 3393 (2006).

[22] R. J. Bartlett and M. Musiał, Coupled-cluster theory in quantum chemistry, Reviews of Modern Physics **79**, 291 (2007).

[23] A. Anand, P. Schleich, S. Alperin-Lea, P. W. Jensen, S. Sim, M. Díaz-Tinoco, J. S. Kottmann, M. Degroote, A. F. Izmaylov, and A. Aspuru-Guzik, A quantum computing view on unitary coupled cluster theory, Chem. Soc. Rev. **51**, 1659 (2022).

[24] S. Sim, P. D. Johnson, and A. Aspuru-Guzik, Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms, Advanced Quantum Technologies **2**, 1900070 (2019).

[25] M. Ostaszewski, E. Grant, and M. Benedetti, Structure optimization for parameterized quantum circuits, Quantum **5**, 391 (2021).

[26] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, Evaluating analytic gradients on quantum hardware, Physical Review A **99**, 032331 (2019).

[27] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, and A. Aspuru-Guzik, Noisy intermediate-scale quantum algorithms, Rev. Mod. Phys. **94**, 015004 (2022).

[28] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, Nature communications **9**, 4812 (2018).

[29] Z. Holmes, K. Sharma, M. Cerezo, and P. J. Coles, Connecting ansatz expressibility to gradient magnitudes and barren plateaus, PRX quantum **3**, 010313 (2022).

[30] K. Andersson, P.-Å. Malmqvist, and B. O. Roos, Second-order perturbation theory with a complete active space self-consistent field reference function, The Journal of chemical physics **96**, 1218 (1992).

[31] S. R. White, Density matrix formulation for quantum renormalization groups, Physical review letters **69**, 2863 (1992).

[32] 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, Nature Communications **5**, 4213 (2014).

[33] P. J. J. O'Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, E. Jeffrey, E. Lucero, A. Megrant, J. Y. Mutus, M. Neeley, C. Neill, C. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. C. White, P. Coveney, V, P. J. Love, H. Neven, A. Aspuru-Guzik, and J. M. Martinis, Scalable quantum simulation of molecular energies, Physical Review X **6**, 10.1103/physrevx.6.031007 (2016).

[34] W. Li, Z. Yin, X. Li, D. Ma, S. Yi, Z. Zhang, C. Zou, K. Bu, M. Dai, J. Yue, *et al.*, A hybrid quantum computing pipeline for real world drug discovery, Scientific Reports **14**, 16942 (2024).

[35] E. R. Sayfutyarova, Q. Sun, G. K.-L. Chan, and G. Knizia, Automated Construction of Molecular Active Spaces from Atomic Valence Orbitals, Journal of Chemical Theory and Computation **13**, 4063 (2017).

[36] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters, and G. K. Chan, PySCF: the Python-based simulations of chemistry framework, Wiley Interdisciplinary Reviews Computational Molecular Science **8**, 10.1002/wcms.1340 (2017).

[37] A. Javadi-Abhari, M. Treinish, K. Krsulich, C. J. Wood, J. Lishman, J. Gacon, S. Martiel, P. D. Nation, L. S. Bishop, A. W. Cross, B. R. Johnson, and J. M. Gambetta, Quantum computing with Qiskit (2024), arXiv:2405.08810 [quant-ph].

[38] S.-X. Zhang, J. Allcock, Z.-Q. Wan, S. Liu, J. Sun, H. Yu, X.-H. Yang, J. Qiu, Z. Ye, Y.-Q. Chen, C.-K. Lee, Y.-C. Zheng, S.-K. Jian, H. Yao, C.-Y. Hsieh, and S. Zhang, TensorCircuit: a Quantum Software Framework for the NISQ Era, Quantum **7**, 912 (2023).

[39] W. Li, J. Allcock, L. Cheng, S.-X. Zhang, Y.-Q. Chen, J. P. Mailoa, Z. Shuai, and S. Zhang, TenCirChEM: an efficient quantum computational chemistry package for the NISQ ERA, Journal of Chemical Theory and Computation **19**, 3966 (2023).