

COMPUTATIONAL APPROACHES TO MOLECULAR SIMULATIONS: CLASSICAL
GEOMETRY OPTIMIZATION AND QUANTUM-CENTRIC ELECTRONIC STRUCTURE
METHODS

By

Akhil Shajan

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Chemistry—Doctor of Philosophy

2025

ABSTRACT

Molecular simulations are essential in computational chemistry, with classical and quantum methods offering complementary approaches to studying molecular structures and interactions. This thesis explores two distinct projects: classical geometry optimization (Chapter 2) and quantum computing for electronic structure simulations (Chapters 3 and 4).

Chapter 2 investigates geometry optimization using various open-source optimizers interfaced with the QUICK program. By analyzing structures from the Baker test set, we compare internal and Cartesian coordinates, evaluate quasi-Newton strategies, and assess a machine learning-based Gaussian Process Regression (GPR) optimizer. Among the tested methods, ASE/Berny and ASE/Sella achieve the fastest convergence, making them suitable for large-scale applications.

Chapter 3 presents quantum-centric simulations of noncovalent interactions using sample-based quantum diagonalization (SQD). We compute the potential energy surfaces of water and methane dimers on quantum hardware, achieving deviations within 1 kcal/mol from high-accuracy classical methods. A 54-qubit simulation explores the current limitations of quantum methods in modeling hydrophobic interactions.

Chapter 4 extends quantum-centric simulations to larger systems by integrating SQD with density matrix embedding theory (DMET). This hybrid approach efficiently reduces quantum resource requirements while maintaining accuracy. Simulations of an 18-atom hydrogen ring and cyclohexane conformers on the `ibm_cleveland` quantum device demonstrate the feasibility of quantum embedding techniques for extended molecular systems.

By advancing classical optimization strategies and demonstrating scalable quantum simulations, this thesis contributes to the development of computational tools for accurate and efficient molecular modeling.

Copyright by
AKHIL SHAJAN
2025

This thesis is dedicated to my parents, sister, and Shanu.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to God for guiding me throughout this academic journey. His infinite wisdom and support have been my constant source of strength, helping me overcome challenges and reach this significant milestone.

I am profoundly indebted to my advisor, Professor Kenneth M. Merz Jr., whose exceptional mentorship, expertise, and unwavering support were instrumental in the completion of this work. Professor Merz's dedication to my development, both as a researcher and as an individual, has shaped my academic and professional path. His trust in my abilities gave me the freedom to pursue my interests, while his guidance kept me focused and driven. I am truly grateful for the opportunities I've had to learn from him, not just in terms of research, but also in life.

I would like to extend my heartfelt thanks to my committee members, Professor Angela Wilson, Professor Warren F. Beck, and Professor James Geiger. Your insightful feedback, thought-provoking questions, and constructive suggestions have played a crucial role in refining this dissertation. I am thankful for your time, effort, and support throughout my PhD journey.

A special note of gratitude goes to my friends, Pushpender, Anand, Anshu, Soumik, Atanu, Darshika, Ajay, Nila, Daya, Megha, and Majma. You have been my companions through thick and thin, always offering support and encouragement when needed most. Whether we were discussing research, enjoying moments of relaxation, or sharing laughs, you made this journey memorable and enjoyable. Thank you for always being there.

I would also like to thank my childhood friends, Robin, Ashish, Aby, and John, for the lifelong friendships and support we've shared. The bond we formed in our early years has stayed with me, and I am deeply grateful for the memories and the encouragement you've always given me.

I am deeply grateful to my family, especially my papa and mom, for their endless love, sacrifices, and unwavering belief in me. Their encouragement has been my anchor through the ups and downs of this journey. To my sister, thank you for your constant love and understanding. The strength and support of my family have been my greatest source of inspiration. To my partner, Shanu, your patience, love, and unwavering support have been invaluable. Thank you for standing by me,

especially during the challenging moments when I doubted myself. I feel truly blessed to have you by my side.

This research was made possible by the generous support of the National Science Foundation (NSF) through the CSSI Elements grant OAC-1835144 and CSSI Frameworks Grant OAC-2209717, as well as by the National Institutes of Health (NIH), under Grant Number GM130641. I am sincerely thankful for the funding that enabled me to pursue my research and contribute to the scientific community.

I would also like to express my deepest appreciation for my colleagues at the Merz lab: Danil, Susanta, Zhen, Hongni, Mithony, Abhishek and Shubamoi. Your intellectual contributions, collaborative spirit, and camaraderie made this journey richer and more fulfilling. The discussions, teamwork, and friendships I've had with all of you have been an essential part of my success. Thank you for your invaluable insights and for making the lab environment so welcoming.

Lastly, to everyone who has supported me along the way, whether through direct involvement or by providing encouragement from afar, I thank you. This journey has been a collective effort, and I am grateful to have had so many wonderful people accompany me along the path.

TABLE OF CONTENTS

CHAPTER 1	INTRODUCTION.....	1
1.1	The Many-Body Problem in Quantum Chemistry.....	2
1.2	Classical Methods for Electronic Structure Calculations	3
1.3	Quantum Computing for Electronic Structure Calculations	7
	BIBLIOGRAPHY	14
CHAPTER 2	GEOMETRY OPTIMIZATION: A COMPARISON OF DIFFERENT OPEN-SOURCE GEOMETRY OPTIMIZERS.....	18
2.1	Introduction.....	18
2.2	Overview of Open-Source Geometry Optimization Software.....	22
2.3	Methodology	26
2.4	Results and discussion.....	28
2.5	Conclusion	32
	BIBLIOGRAPHY	35
	APPENDIX.....	39
CHAPTER 3	ACCURATE QUANTUM-CENTRIC SIMULATIONS OF SUPRAMOLECULAR INTERACTIONS.....	42
3.1	Introduction.....	42
3.2	Methods and Computational Details	45
3.3	Results.....	50
3.4	Conclusions and Outlook	56
	BIBLIOGRAPHY	58
	APPENDIX.....	70
CHAPTER 4	TOWARDS QUANTUM-CENTRIC SIMULATIONS OF EXTENDED MOLECULES: SAMPLE-BASED QUANTUM DIAGONALIZATION ENHANCED WITH DENSITY MATRIX EMBEDDING THEORY.....	73
4.1	Introduction.....	73
4.2	Methods.....	76
4.3	Results.....	82
4.4	Conclusion	87
	BIBLIOGRAPHY	89
	APPENDIX.....	97

CHAPTER 1

INTRODUCTION

Electronic structure theory forms the foundation of modern theoretical chemistry, enabling the accurate prediction of molecular and material properties across diverse scientific fields, including chemistry, materials science, and drug discovery¹⁻³. At its core, this theory aims to solve the electronic Schrödinger equation to determine how electrons propagate in a system, providing insights into molecular stability, reactivity, and spectra. In reaction mechanism studies, electronic structure calculations have been instrumental in elucidating transition states and activation energies, as demonstrated in computational studies of enzymatic catalysis⁴. Similarly, in materials science, first-principles calculations based on density functional theory (DFT) have guided the design of new semiconductors and battery materials by predicting electronic conductivity and defect formation energies^{5,6}. In drug discovery, quantum mechanical approaches, such as free-energy perturbation (FEP) and quantum mechanics/molecular mechanics (QM/MM) methods, have significantly improved the prediction of binding affinities, aiding in the rational design of inhibitors for kinase and protease targets⁷.

Despite its success, solving the Schrödinger equation for many-electron systems remains a formidable challenge due to the complexity of electron-electron interactions and the exponential computational scaling associated with wavefunction-based methods^{8,9}. While exact methods like Full Configuration Interaction (FCI) provide benchmark-quality results, their factorial scaling renders them impractical for all but the smallest molecules¹⁰. To address these limitations, modern electronic structure methods employ sophisticated approximations, such as coupled cluster (CC) theory for high-accuracy correlated wavefunctions¹¹ and density matrix embedding theory (DMET) for treating strongly correlated materials¹². These advancements have enabled the accurate simulation of reaction pathways, such as CO₂ reduction catalysts, and the discovery of novel organic semiconductors^{4,6}. Thus, electronic structure theory remains a cornerstone of scientific discovery, bridging fundamental quantum mechanics with practical applications in chemistry, materials science, and pharmaceutical research.

1.1 The Many-Body Problem in Quantum Chemistry

The Schrödinger equation governs the behavior of quantum systems and is fundamental to quantum mechanics and quantum chemistry. In its time-independent form, the equation is given Eq. (1.1)

$$\hat{H}\Psi = E\Psi \quad (1.1)$$

where $\hat{H}$ is the Hamiltonian operator, representing the total energy of the system, Ψ is the wavefunction, which contains all the information about the system, and E is the total energy associated with the system. Solving this equation for many-electron systems is extremely challenging because of the intricate interactions between the electrons.

For a molecule with N electrons and M nuclei, the Hamiltonian operator in atomic units (*a.u.*) can be represented by Eq. (1.2)

$$\hat{H} = -\sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{A=1}^M \frac{1}{2M_A} \nabla_A^2 - \sum_i \sum_A \frac{Z_A}{r_{iA}} + \sum_{i<j}^N \frac{1}{r_{ij}} + \sum_{A<B}^M \frac{Z_A Z_B}{R_{AB}} \quad (1.2)$$

where the terms represent the kinetic energy of electrons, the kinetic energy of nuclei, the electron-nucleus Coulomb interaction, the electron-electron repulsion, and the nucleus-nucleus interaction, respectively. The Born-Oppenheimer (BO) approximation¹³ enables the separation of nuclear and electronic motion by exploiting the significant mass difference between nuclei and electrons. Since electrons move much faster than nuclei, this approximation is valid for many applications, allowing the electronic Schrödinger equation to be solved independently at fixed nuclear positions. The resulting electronic Hamiltonian, given in Eq. (1.3), defines the potential energy surface (PES), which governs nuclear motion and molecular dynamics.

$$\hat{H} = -\sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_i \sum_A \frac{Z_A}{r_{iA}} + \sum_{i<j}^N \frac{1}{r_{ij}} \quad (1.3)$$

However, even with this approximation, solving the full wavefunction for a multi-electron system remains computationally demanding due to the need to account for all possible electronic configurations and interactions. The Schrödinger equation, however, can be solved exactly only for a one-electron system.¹

The computational scaling problem arises from the fact that the number of parameters required to describe the wavefunction grows exponentially with the number of electrons. As a result, solving the full wavefunction exactly for very large systems is intractable, making it necessary to develop approximations and efficient algorithms to tackle these systems.

1.2 Classical Methods for Electronic Structure Calculations

Classical computational approaches for solving the electronic Schrödinger equation are essential tools in quantum chemistry, providing approximate yet powerful methods to investigate molecular properties, reaction dynamics, and fundamental chemical interactions. These methods rely on a variety of approximations to efficiently handle the complexities of many-electron systems, as solving the full electronic wavefunction exactly is only possible for the simplest systems, such as the hydrogen like atom with a single electron. Broadly, classical methods can be categorized into wavefunction-based techniques and density-based approaches. Wavefunction-based methods, such as variational and perturbation theories, approximate the many-electron wavefunction by constructing systematically improvable solutions. Density-based methods, like density functional theory (DFT), reformulate the problem in terms of the electron density rather than the wavefunction, significantly reducing computational cost while capturing essential quantum effects.

1.2.1 Wavefunction-Based Methods

Wavefunction-based methods aim to approximate the exact many-body wavefunction of a quantum system by solving the electronic Schrödinger equation with varying levels of accuracy and computational cost. These methods play a crucial role in quantum chemistry, providing systematic ways to incorporate electron correlation effects, which are essential for accurate energy predictions. The most widely used approaches include:

1.2.1.1 Hartree-Fock (HF) Approximation

The Hartree-Fock (HF) method serves as the fundamental starting point for most wavefunction-based electronic structure calculations^{1,2}. It approximates the many-electron wavefunction as a single Slater determinant, ensuring the proper antisymmetry required by the Pauli exclusion principle. In HF theory, each electron is treated as moving in an average potential generated by

all other electrons, leading to an effective mean-field approximation. This reduces the complex many-body problem into a set of self-consistent, single-particle equations.

Despite its efficiency, HF suffers from a major limitation: it completely neglects electron correlation, meaning it fails to capture the instantaneous interactions between electrons. This omission leads to significant errors in systems where correlation effects are critical, such as transition metal complexes, dispersion-dominated interactions, and strongly correlated materials. While HF provides a qualitatively useful starting point, more sophisticated post-HF methods are required for quantitatively accurate results.¹⁴

1.2.1.2 Configuration Interaction (CI) Method

Configuration Interaction (CI) improves upon HF theory by incorporating electron correlation explicitly through a linear expansion of the wavefunction⁸. In this approach, the wavefunction of the system is expressed as a superposition of multiple Slater determinants, constructed by exciting electrons from the HF reference state:

$$|\Psi_{\text{CI}}\rangle = c_0|\Phi_{\text{HF}}\rangle + \sum_{ia} c_i^a |\Phi_i^a\rangle + \sum_{ijab} c_{ij}^{ab} |\Phi_{ij}^{ab}\rangle + \dots \quad (1.4)$$

where the coefficients c_0, c_i^a, c_{ij}^{ab} , etc., are determined by solving the electronic Schrödinger equation within the chosen basis set.

The most accurate form of CI is Full Configuration Interaction (FCI), which includes all possible excitations within a given basis set, yielding exact solutions within that basis². However, FCI is computationally intractable for large systems due to its factorial scaling of $O(K^N)$, where K is the number of basis functions and N is the number of electrons. As a result, truncated CI approaches, such as CI with Single and Double excitations (CISD), are commonly employed. While these methods improve accuracy relative to HF, they suffer from a fundamental limitation: CI methods are not size-extensive, meaning their accuracy does not scale properly with system size⁸.

In terms of computational achievements, a notable FCI calculation was performed on the propane molecule (C_3H_8) using the STO-3G basis set, involving an active space of 26 electrons

in 23 orbitals, corresponding to a Hilbert space of approximately 1.3 trillion determinants. This represents one of the largest FCI calculations reported to date.¹⁵

1.2.1.3 Coupled-Cluster (CC) Theory

Coupled-Cluster (CC) theory is one of the most powerful and widely used wavefunction-based methods, offering a systematically improvable approach to electron correlation¹⁶. Unlike CI, which employs a linear expansion, CC uses an exponential ansatz to construct the correlated wavefunction:

$$|\Psi_{\text{CC}}\rangle = e^{\hat{T}}|\Phi_{\text{HF}}\rangle, \quad (1.5)$$

where $\hat{T}$ is the cluster operator, defined as:

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \dots, \quad (1.6)$$

with $\hat{T}_n$ representing n -electron excitations.

The most commonly used approximation, Coupled-Cluster with Single and Double excitations (CCSD), includes cluster operators from only one- and two-electron excitations. To account for higher-order effects, the CCSD(T) method introduces a perturbative treatment of triple excitations. This method is often referred to as the “gold standard” of quantum chemistry, as it provides highly accurate results for a broad range of chemical systems while remaining computationally feasible for moderately sized molecules¹⁷. However, CC methods still suffer from steep computational scaling, making them impractical for very large systems.

Moreover in systems characterized by strong electron correlation, single-reference coupled-cluster methods like CCSD(T) often fail to provide accurate results. This failure is particularly evident in cases where multiple electronic configurations contribute significantly to the wavefunction, rendering a single-reference approach inadequate. A notable example is the symmetric dissociation of hydrogen rings, such as the H_6 molecule. Studies have shown that as the H–H bond distances increase, methods like CCSD, CCSDt, and CCSDT exhibit erratic behavior, deviating significantly from Full Configuration Interaction (FCI) benchmarks. For instance, at H–H

separations of 2.5 Å, these methods can show errors relative to FCI exceeding 200 millihartrees, indicating a complete breakdown in the strongly correlated regime.¹⁸

1.2.2 Density Functional Theory (DFT)

While wavefunction-based approaches provide systematically improvable accuracy, their computational cost scales steeply with system size, limiting their applicability to small and medium-sized molecules. An alternative approach, *Density Functional Theory (DFT)*, circumvents the need to explicitly compute the many-body wavefunction by reformulating the problem in terms of the *electron density*. According to the Hohenberg-Kohn theorems, the ground-state electron density uniquely determines all properties of a molecular system¹⁹, allowing the total energy to be expressed as a functional of the density:

$$E[\rho] = T[\rho] + V_{\text{ext}}[\rho] + E_{\text{xc}}[\rho] \quad (1.7)$$

where $T[\rho]$ represents the kinetic energy, $V_{\text{ext}}[\rho]$ accounts for external potential interactions, and $E_{\text{xc}}[\rho]$ is the *exchange-correlation functional*, which encapsulates complex many-body effects.

Due to its significantly lower computational cost compared to wavefunction-based methods, DFT has become the most widely used electronic structure method in chemistry, materials science, and condensed matter physics²⁰. The accuracy of DFT calculations depends on the choice of the exchange-correlation functional, with common approximations including:

- *Local Density Approximation (LDA)*: Assumes that the exchange-correlation energy at each point in space depends only on the local electron density²¹.
- *Generalized Gradient Approximation (GGA)*: Incorporates the gradient of the density, improving accuracy for molecular and solid-state systems^{22,23}.
- *Hybrid Functionals*: Incorporate a fraction of exact exchange from HF theory, enhancing accuracy for a broad range of chemical systems^{24,25}.

Despite its computational efficiency, Density Functional Theory (DFT) has several inherent limitations, primarily due to the unknown exact form of the exchange-correlation (XC) functional, which is often derived semi-empirically²⁶. While DFT performs well for many systems, it fails

in strongly correlated cases where multiple electronic configurations are important, such as the dissociation of H_2 , where it incorrectly predicts the dissociation energy curve²⁷. Additionally, DFT suffers from delocalization and self-interaction errors, leading to inaccuracies in stretched bonds and charge-transfer systems²⁸.

1.3 Quantum Computing for Electronic Structure Calculations

Quantum computing provides a novel framework for addressing electronic structure problems by leveraging the principles of quantum mechanics. While classical computers store information in bits (0s and 1s), quantum computers use qubits, which can exist in superposition states. However, the final measurement in a quantum computation still yields discrete classical outcomes (0 or 1), with probabilities determined by quantum amplitudes, allowing for more efficient encoding of wavefunction information.

Near-term quantum hardware offers advantages primarily in representing and processing correlated electronic states more compactly than classical methods. For instance, representing an N -qubit quantum state requires storing 2^N complex amplitudes on a classical computer, whereas a quantum device inherently encodes this information within its qubits. However, current quantum devices are limited by noise and hardware constraints, restricting practical applications to variational approaches, such as the Variational Quantum Eigensolver (VQE), rather than exact diagonalization via Quantum Phase Estimation (QPE). While QPE theoretically provides exact eigenvalues of the molecular Hamiltonian, its implementation requires deep circuits and error correction, making it impractical for near-term quantum processors.

1.3.1 Quantum Algorithms for Electronic Structure

Advancements in quantum computing have led to the development of specialized quantum algorithms designed to tackle the computational challenges inherent in Configuration Interaction (CI) methods for large molecules. While CI-based classical approaches, such as Full Configuration Interaction (FCI), provide exact solutions within a given basis set, their factorial scaling with system size renders them infeasible for all but the smallest molecules^{1,2}. Quantum algorithms offer an alternative by leveraging superposition and entanglement to encode and manipulate electronic

wavefunctions more efficiently, potentially providing a path to scalable electronic structure calculations^{29,30}. Below, we discuss some of the most prominent quantum algorithms developed to compute molecular electronic energies.

1.3.1.1 Quantum Phase Estimation (QPE)

Quantum Phase Estimation (QPE) is a fully quantum algorithm designed to extract the eigenvalues of a Hamiltonian with high precision³¹. It works by encoding the phase of an eigenstate of the Hamiltonian into a quantum register, allowing the eigenvalue to be determined with exponential precision relative to the number of qubits used. QPE follows these general steps:

1. *State Preparation*: The algorithm begins with a quantum state that approximates an eigenstate of the Hamiltonian, often derived from a classical mean-field method, such as Hartree-Fock (HF).
2. *Quantum Fourier Transform (QFT)*: Ancilla qubits undergo controlled-unitary operations to encode the phase information of the Hamiltonian. The accuracy of QFT-based algorithms depends on the number of terms in the Fourier expansion, with truncation leading to errors in phase estimation and energy calculations^{29,31}.
3. *Measurement and Phase Extraction*: The output register is measured, and a classical post-processing step extracts the energy eigenvalue of the system with high accuracy.

One of the key advantages of QPE is its ability to provide eigenvalues with an exponentially small error, making it an ideal candidate for applications requiring high precision. However, despite its theoretical appeal, QPE has significant practical limitations:

- *High Circuit Depth*: The algorithm requires deep quantum circuits consisting of numerous controlled-unitary operations, imposing stringent coherence time requirements on quantum hardware²⁹.

- *Error Sensitivity:* The presence of gate errors and decoherence affects the reliability of QPE results, making implementation on noisy intermediate-scale quantum (NISQ) devices difficult.
- *Initial State Preparation:* The accuracy of QPE depends on how well the initial quantum state approximates an eigenstate of the Hamiltonian. Poor state preparation can lead to convergence issues.

While Quantum Phase Estimation (QPE) is a critical algorithm for fault-tolerant quantum computing, its implementation on near-term noisy intermediate-scale quantum (NISQ) devices remains impractical due to several limitations. QPE requires deep quantum circuits with numerous controlled-unitary operations, leading to high gate depth that exceeds the coherence times of current quantum hardware. Additionally, the accumulation of gate errors and the lack of fully error-corrected qubits further degrade its accuracy. The high number of ancilla qubits required for precise phase estimation also surpasses the capabilities of present-day quantum processors.³²

1.3.1.2 Variational Quantum Eigensolver (VQE)

The Variational Quantum Eigensolver (VQE) is a hybrid quantum-classical algorithm designed specifically for near-term quantum devices^{33,34}. Unlike QPE, which requires deep circuits, VQE leverages a variational principle that allows for approximate solutions to the ground-state energy using shorter quantum circuits. The algorithm follows these steps:

1. *Ansatz Preparation:* A parameterized quantum circuit (ansatz) is chosen to represent the wavefunction. Common ansätze include hardware-efficient circuits and chemically motivated ansätze like the unitary coupled-cluster (UCC) approach³⁵.
2. *Energy Expectation Evaluation:* The quantum computer prepares the trial wavefunction and measures the expectation value of the Hamiltonian.
3. *Classical Optimization:* A classical optimizer (e.g., gradient-based or gradient-free methods) adjusts the ansatz parameters to minimize the energy expectation value iteratively.

VQE has several advantages for NISQ devices:

- *Shallow Circuit Depth:* The algorithm requires significantly shallower circuits than QPE, reducing susceptibility to decoherence and gate errors.
- *Hybrid Approach:* The reliance on classical optimization helps mitigate some quantum hardware limitations.
- *Flexibility in Ansatz Choice:* Different ansätze can be chosen to improve accuracy for specific chemical systems.

Despite its advantages, VQE faces significant scalability challenges, particularly due to the hardware limitations of near-term quantum devices. The number of qubits and constraints on gate fidelity restrict the size of molecular systems that can be accurately simulated. While VQE has demonstrated success for small molecules, its applicability to larger systems is hindered by increasing circuit depth, measurement overhead, and noise accumulation^{34,36}. Additionally, the computational cost grows as more qubits are required to represent complex electronic states, making large-scale applications infeasible on current quantum hardware³⁷.

1.3.1.3 Sample-Based Quantum Diagonalization (SQD)

Sample-Based Quantum Diagonalization (SQD) is a recently developed hybrid quantum-classical method designed to efficiently compute the electronic structure of molecules³⁸. Unlike conventional quantum algorithms, which rely on direct Hamiltonian evolution, SQD exploits quantum sampling techniques to generate electronic configurations and reconstruct the wavefunction using classical diagonalization methods. The SQD algorithm consists of the following key steps:

1. *Quantum State Preparation:* A quantum circuit is used to prepare an approximate molecular wavefunction, Ψ , encoding electronic configurations relevant to the molecule.
2. *Quantum Sampling:* The wavefunction is sampled multiple times to generate bit strings corresponding to electronic configurations.

3. *Configuration Recovery*: The sampled configurations are processed to recover an approximate eigenstate representation, grouping configurations into batches for further refinement.
4. *Self-Consistent Iteration*: The sampled states undergo classical diagonalization using selected configuration interaction (SCI) methods, iteratively improving until convergence.

Beyond these core algorithms, emerging techniques such as Quantum Embedding Methods provide new avenues for tackling large-scale electronic structure problems. One of the most promising approaches is Density Matrix Embedding Theory (DMET), which partitions a large molecular system into smaller subsystems that can be solved using quantum methods³⁹. This hybrid approach allows for scalable quantum simulations by embedding quantum computations within a classical framework, significantly expanding the range of solvable problems.

As quantum hardware continues to advance, these quantum algorithms will become increasingly viable for large-scale electronic structure calculations. While current limitations such as noise, error rates, and circuit depth remain challenges, ongoing research in quantum error correction, improved ansatz design, and hybrid quantum-classical algorithms will drive further progress in quantum chemistry applications.

1.3.2 Challenges and Current Limitations

Although quantum computing holds great promise for revolutionizing fields such as quantum chemistry and materials science, it is still in its early developmental stages. As such, several significant challenges remain that hinder the full realization of its potential for electronic structure calculations and other complex simulations.

- *Noise and Hardware Limitations*: Current quantum processors, often referred to as Noisy Intermediate-Scale Quantum (NISQ) devices, are susceptible to a variety of errors that stem from both environmental noise and intrinsic imperfections in the hardware⁴⁰. These errors manifest as qubit decoherence, gate inaccuracies, and crosstalk between qubits, all of which limit the depth and accuracy of calculations that can be performed. As a result, the number of quantum operations that can be executed before error accumulation becomes significant

is constrained, limiting the complexity of practical calculations on these devices. Efforts such as error mitigation strategies^{41,42} and quantum error correction⁴³ are being developed to overcome these issues.

- *Scaling Challenges:* Many of the quantum algorithms used in electronic structure calculations, such as those based on quantum diagonalization or quantum phase estimation, require large numbers of qubits to encode the quantum state of a system^{29,30}. However, the number of high-quality qubits available in current quantum processors remains relatively small. Furthermore, as the number of qubits increases, maintaining high fidelity and reducing noise become even more challenging⁴⁴. Achieving the required qubit count with the necessary quality to simulate large molecular systems remains a major hurdle for quantum computing to be broadly applicable to real-world chemistry problems.
- *Quantum-Classical Integration:* Many practical quantum chemistry methods, including the hybrid approach combining Density Matrix Embedding Theory (DMET) with Sample-Based Quantum Diagonalization (SQD) explored in this work, continue to rely on quantum-classical workflows^{34,45}. These methods harness the advantages of both computing paradigms: quantum circuits prepare accurate molecular ground states with a balance between accuracy and circuit depth, while classical computation processes quantum measurement samples to extract electronic properties. The classical post-processing step, rooted in selected configuration interaction, performs diagonalization within subspaces defined by computational basis states, enhancing accuracy and mitigating quantum noise. However, the seamless integration of quantum and classical resources remains a significant challenge, necessitating efficient synchronization, optimized data exchange, and minimal communication overhead^{33,35}.

These challenges underscore the current limitations in the field and highlight the need for continued advances in quantum hardware, error correction techniques, and algorithmic innovation. While significant progress is being made, it is clear that overcoming these obstacles will be

essential for quantum computing to realize its full potential in electronic structure simulations and other domains.

BIBLIOGRAPHY

- [1] A. Szabo and N.S. Ostlund. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Dover Books on Chemistry. Dover Publications, 1996.
- [2] T. Helgaker, P. Jorgensen, and J. Olsen. *Molecular Electronic-Structure Theory*. Wiley, 2014.
- [3] F. Jensen. *Introduction to Computational Chemistry*. Wiley, 2017.
- [4] Richard A. Friesner. *ab initio* quantum chemistry: Methodology and applications. *Proceedings of the National Academy of Sciences*, 102(19):6648–6653, 2005.
- [5] R.M. Martin. *Electronic Structure: Basic Theory and Practical Methods*. Cambridge University Press, 2004.
- [6] Jürgen Hafner. Ab-initio simulations of materials using vasp: Density-functional theory and beyond. *Journal of Computational Chemistry*, 29(13):2044–2078, 2008.
- [7] William L. Jorgensen. Efficient drug lead discovery and optimization. *Accounts of Chemical Research*, 42(6):724–733, 2009. PMID: 19317443.
- [8] I. Shavitt and R.J. Bartlett. *Many-Body Methods in Chemistry and Physics: MBPT and Coupled-Cluster Theory*. Cambridge Molecular Science. Cambridge University Press, 2009.
- [9] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal. Quantum monte carlo simulations of solids. *Rev. Mod. Phys.*, 73:33–83, Jan 2001.
- [10] George H Booth, Andreas Grüneis, Georg Kresse, and Ali Alavi. Towards an exact description of electronic wavefunctions in real solids. *Nature*, 493(7432):365—370, January 2013.
- [11] Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements h-pu. *The Journal of Chemical Physics*, 132(15):154104, 04 2010.
- [12] Gerald Knizia and Garnet Chan. Density matrix embedding: A simple alternative to dynamical mean-field theory. *Physical review letters*, 109:186404, 11 2012.
- [13] M. Born and R. Oppenheimer. Zur quantentheorie der molekeln. *Annalen der Physik*, 389(20):457–484, 1927. PDF of an english translation by S.M. Blinder is available.
- [14] Sahil Gulania and James Daniel Whitfield. Limitations of hartree–fock with quantum resources. *The Journal of Chemical Physics*, 154(4):044112, 01 2021.
- [15] Hong Gao, Satoshi Imamura, Akihiko Kasagi, and Eiji Yoshida. Distributed implementation of full configuration interaction for one trillion determinants. *Journal of Chemical Theory*

- and Computation*, 20(3):1185–1192, 2024. PMID: 38314701.
- [16] Rodney J. Bartlett and Monika Musiał. Coupled-cluster theory in quantum chemistry. *Rev. Mod. Phys.*, 79:291–352, Feb 2007.
 - [17] John A. Pople, Martin Head-Gordon, and Krishnan Raghavachari. Quadratic configuration interaction. a general technique for determining electron correlation energies. *Journal of Chemical Physics*, 87:5968–5975, 1987.
 - [18] Jun Shen Ilias Magoulas and Piotr Piecuch. Addressing strong correlation by approximate coupled-pair methods with active-space and full treatments of three-body clusters. *Molecular Physics*, 120(19-20):e2057365, 2022.
 - [19] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864–B871, Nov 1964.
 - [20] Robert G. Parr and Weitao Yang. *Density-Functional Theory of Atoms and Molecules*, volume 16 of *International Series of Monographs on Chemistry*. Oxford University Press, 1994.
 - [21] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133–A1138, Nov 1965.
 - [22] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865–3868, Oct 1996.
 - [23] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, 38:3098–3100, Sep 1988.
 - [24] Axel D. Becke. Density-functional thermochemistry. iii. the role of exact exchange. *The Journal of Chemical Physics*, 98(7):5648–5652, 04 1993.
 - [25] Carlo Adamo and Vincenzo Barone. Toward reliable density functional methods without adjustable parameters: The pbe0 model. *The Journal of Chemical Physics*, 110(13):6158–6170, 04 1999.
 - [26] John P. Perdew, Stefan Kurth, Ale š Zupan, and Peter Blaha. Accurate density functional with correct formal properties: A step beyond the generalized gradient approximation. *Phys. Rev. Lett.*, 82:2544–2547, Mar 1999.
 - [27] Aron J. Cohen, Paula Mori-Sánchez, and Weitao Yang. Insights into current limitations of density functional theory. *Science*, 321(5890):792–794, 2008.
 - [28] Paula Mori-Sánchez, Aron J. Cohen, and Weitao Yang. Many-electron self-interaction error in approximate density functionals. *The Journal of Chemical Physics*, 125(20):201102, 2006.

- [29] Alán Aspuru-Guzik, Anthony D. Dutoi, Peter J. Love, and Martin Head-Gordon. Simulated quantum computation of molecular energies. *Science*, 309(5741):1704–1707, 2005.
- [30] Yudong Cao, Jonathan Romero, Jonathan P. Olson, Matthias Degroote, Peter D. Johnson, Mária Kieferová, Ian D. Kivlichan, Tim Menke, Borja Peropadre, Nicolas P. D. Sawaya, Sukin Sim, Libor Veis, and Alán Aspuru-Guzik. Quantum chemistry in the age of quantum computing. *Chemical Reviews*, 119(19):10856–10915, 2019. PMID: 31469277.
- [31] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010.
- [32] Kishor Bharti, Alba Cervera-Lierta, Thi Ha Kyaw, Tobias Haug, Sumner Alperin-Lea, Abhinav Anand, Matthias Degroote, Hermann Heimonen, Jakob S. Kottmann, Tim Menke, Wai-Keong Mok, Sukin Sim, Leong-Chuan Kwek, and Alán Aspuru-Guzik. Noisy intermediate-scale quantum algorithms. *Rev. Mod. Phys.*, 94:015004, Feb 2022.
- [33] 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(1):4213, 2014.
- [34] Jarrod R. McClean, Jonathan Romero, Ryan Babbush, and Alán Aspuru-Guzik. The theory of variational hybrid quantum-classical algorithms. *New Journal of Physics*, 18:023023, 2016.
- [35] Jonathan Romero, Ryan Babbush, Jarrod R. McClean, Cornelius Hempel, Peter J. Love, and Alán Aspuru-Guzik. Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz. *Quantum Sci. Technol.*, 4(1):014008, 2018.
- [36] 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(7671):242–246, 2017.
- [37] J. Tilly, H. Chen, S. Cao, D. Picozzi, K. Setia, Y. Li, E. Grant, L. Wossnig, I. Rungger, and G. H. Booth. The variational quantum eigensolver: a review of methods and best practices. *Physics Reports*, 986:1–128, 2022.
- [38] J. Robledo-Moreno, M. Motta, H. Haas, A. Javadi-Abhari, P. Jurcevic, W. Kirby, S. Martiel, K. Sharma, S. Sharma, and T. Shirakawa. Chemistry beyond exact solutions on a quantum-centric supercomputer. *arXiv preprint arXiv:2405.05068*, 2024.
- [39] Gerald Knizia and Garnet Kin-Lic Chan. Density matrix embedding: A simple alternative to dynamical mean-field theory. *Phys. Rev. Lett.*, 109(18):186404, 2012.
- [40] John Preskill. Quantum Computing in the NISQ era and beyond. *Quantum*, 2:79, August 2018.

- [41] Kristan Temme, Sergey Bravyi, and Jay M. Gambetta. Error mitigation for short-depth quantum circuits. *Phys. Rev. Lett.*, 119:180509, Nov 2017.
- [42] Abhinav Kandala, Kristan Temme, Antonio D Córcoles, Antonio Mezzacapo, Jerry M Chow, and Jay M Gambetta. Error mitigation extends the computational reach of a noisy quantum processor. *Nature*, 567(7749):491–495, 2019.
- [43] Daniel A. Lidar and Todd A. Brun, editors. *Quantum Error Correction*. Cambridge University Press, 2013.
- [44] Sergio Boixo, Sergei V. Isakov, Vadim N. Smelyanskiy, Ryan Babbush, Nan Ding, Zhang Jiang, Michael J. Bremner, John M. Martinis, and Hartmut Neven. Characterizing quantum supremacy in near-term devices. *Nature Physics*, 14:595–600, 2018.
- [45] Nicholas C. Rubin. A hybrid classical/quantum approach for large-scale studies of quantum systems with density matrix embedding theory. *arXiv preprint*, 2016.

CHAPTER 2

GEOMETRY OPTIMIZATION: A COMPARISON OF DIFFERENT OPEN-SOURCE GEOMETRY OPTIMIZERS

This chapter is reprinted with permissions from *J. Chem. Theory Comput.* **2023**, *19*, 21, 7533-7541

Based on a series of energy minimizations with starting structures obtained from the Baker test set of 30 organic molecules, a comparison is made between various open-source geometry optimization codes that are interfaced with the open-source QUantum Interaction Computational Kernel (QUICK) program for gradient and energy calculations. The findings demonstrate how the choice of the coordinate system influences the optimization process to reach an equilibrium structure. With fewer steps, internal coordinates outperform Cartesian coordinates while the choice of the initial Hessian and Hessian update method in quasi-Newton approaches made by different optimization algorithms also contributes to the rate of convergence. Furthermore, an available open-source machine learning method based on Gaussian Process Regression (GPR) was evaluated for energy minimizations over surrogate potential energy surfaces with both Cartesian and internal coordinates, with internal coordinates outperforming Cartesian. Overall, geomeTRIC and DL-FIND with their default optimization method as well as with GPR-based model using Hartree–Fock theory with the 6-31G** basis set, needed a comparable number of geometry optimization steps to the approach of Baker using a unit matrix as the initial Hessian to reach the optimized geometry. On the other hand, the Berny and Sella offerings in ASE outperformed the other algorithms. Based on this we recommend using the file-based approaches, ASE/Berny and ASE/Sella, for large-scale optimization efforts, while if using a single executable is preferable, we now distribute QUICK integrated with DL-FIND.

2.1 Introduction

The importance of intermolecular interactions in atomistic simulations of materials, chemical reactions, and biological processes is well known. Optimizing the molecular geometry to find the stationary points on the potential energy surface, followed by calculation of molecular properties

and characterization of the interactions at a certain geometry, is a core technique used in theoretical chemistry to analyze molecular structure and intermolecular interactions¹⁻³. By taking into account the energy, $E(x_0)$, at a point x_0 on a potential energy surface, we can use the Taylor series to form a quadratic approximation to describe the energy at a nearby point, $x = x_0 + \Delta x$:

$$E(x) = E(x_0) + g_0^T \Delta x + 1/2 \Delta x^T H_0 \Delta x \quad (2.1)$$

where g_0 is the gradient vector (dE/dx) at x_0 , H_0 is the Hessian matrix (d^2E/dx^2) at x_0 , and $\Delta x = x - x_0$. Most nonlinear optimization algorithms are based on this local quadratic approximation of the potential energy surface.^{4,5} By differentiating with respect to the coordinates, one could form an approximation for the gradient, which is given by:

$$g(x) = g_0 + H_0 \Delta x \quad (2.2)$$

On the potential energy surface, the gradient, $g(x)$, vanishes at a stationary point, $g(x) \equiv \nabla E = 0$. Hence, in the local quadratic approximation of the potential energy surface, the displacement to the minimum at the stationary point is given by:

$$\Delta x = -H_0^{-1} g_0 \quad (2.3)$$

This is known as the Newton-Raphson step. It is an integral part of almost all quantum chemistry geometry optimization approaches. The gradient here can be obtained by differentiation of the energy with respect to the coordinates, and the Hessian can be obtained using numerical or analytical methods of the second derivative of the energy with respect to the coordinates.

Atomistic studies are of interest for a variety of stationary points, such as minima, transition states, and conical intersections⁶⁻¹⁰. The most effective and extensively used techniques for identifying these stationary points are Newton and quasi-Newton techniques, which typically employ a sequential optimization cycle, in which a guess optimal geometry is steadily enhanced by employing the gradient and either an exact or an approximate Hessian. For minimization, the Hessian must contain exclusively positive eigenvalues (i.e., positive definite). If one or more eigenvalues

are negative, the step will be in the direction of a first-order or higher-order saddle point, which can be useful when locating a transition state.

The choice of the coordinate system, the optimization algorithm, the quality of the Hessian, and the algorithm used to determine the step, Δx , are four factors that influence the efficiency of a geometry optimization. Strong coupling between coordinates, narrow gullies, and curved valleys provide significant hurdles to even the best optimization algorithms, making the choice of a suitable coordinate system essential to the efficiency of any optimization. Since the Cartesian coordinate system is the easiest approach to define the molecular geometry, it is usually used for the evaluation of the energy and gradient. However, the potential energy surface is strongly nonlinear and coordinates are coupled, so only small steps can be taken downhill. Cartesian coordinates perform significantly worse for flexible, acyclic systems than they do for constrained, cyclic molecules. In Cartesian coordinates, it is also much harder to impose constraints^{11,12}. Internal coordinates on the other hand are better at reflecting the overall atomic motions. Internal coordinates can describe displacements along curved pathways and decouple various types of molecular displacements. As a result, the optimization algorithm can proceed with greater efficiency^{13,14}.

It is possible to calculate the entire Hessian matrix analytically¹⁵ for each step at the current point on the potential energy surface. Only one step is needed to get to the minimum of the energy if the local potential energy surface is quadratic. Since actual potential energy surfaces are rarely quadratic, one must take several Newton-Raphson steps to get at a stationary point. It is thus often impractical or undesirable to directly calculate the full Hessian due to the amount of computational work this typically involves. Instead one can resort to quasi-Newton techniques, whose foundation is a computationally cheap approximation of the Hessian. At each stage of a quasi-Newton optimization, the Hessian is improved using the difference between the calculated gradient change and the change anticipated by the approximate Hessian.

$$H_{\text{new}} = H_{\text{old}} + \Delta H \quad (2.4)$$

This updated Hessian can be used to determine the step, Δx . The Hessian can be updated in a number of ways, and several algorithmic schemes can be used to determine the step and the new

coordinate system along the minimum to get at the optimized geometry^{16–19}.

Additionally, the initial estimate of the Hessian or second derivative matrix does have an effect on the performance of geometry optimization with quasi-Newton methods; the more precise the initial estimate, the faster the convergence^{2,20,21}. One can use an exact Hessian, which is expensive. However, often the estimate does not have to be extremely precise since the Hessian matrix gradually improves in quality as it is updated using gradient information during the search for the minimum. The unit matrix is the most basic and popular approximation, in which case the first optimization step corresponds to a gradient descent. The nature of the atoms, the bonds connecting them, and other pertinent structural information about the molecule are all ignored in this case despite the fact that this is an unbiased choice. All connection between coordinates is initially ignored, and flexible coordinates (such as torsion and ring deformation) are not separated from stiff modes (such as bond stretching). At the price of additional optimization steps, this data must be gathered during the optimization process. For cyclic compounds, whose coordinates are by nature strongly coupled, this is a poor approximation. Simple schemes to approximate the initial Hessian based on parameterized values for bond, angle and dihedral force constants have thus been proposed to reduce the number of required geometry optimization steps^{2,22,23}.

Recently, the integration of machine learning techniques into traditional methods for calculating potential energy surfaces (PES) has led to even more efficient optimization approaches. In particular, machine learning can accelerate the location of minima^{24,25} and transition states^{26,27}. One of the most popular approaches in this field is to use surrogate potential energy surfaces. Surrogates are constructed by fitting a model, such as a Gaussian process regression (GPR) model, to the data points that have already been evaluated. This model is then used to suggest new geometries for evaluation, without the need for expensive *ab initio* calculations. As more data points become available, the surrogate PES is continuously updated and used to identify the location of the next guess minimum. This guess point is then subject to single-point energy and gradient evaluations on the true PES and then added to the dataset. The accuracy of the surrogate PES predicted by a GPR model heavily relies on the appropriate selection of a kernel function. The kernel function determines

the covariance between the input space points and directly impacts the accuracy of the predicted PES. Compared to classical second-order methods, such as the local quadratic approximation in the quasi-Newton method discussed above, surrogate-based approaches offer a more accurate representation of the PES, particularly in regions where the true PES is highly nonlinear or difficult to sample. The careful selection of the appropriate kernel function and coordinate system is essential in constructing an accurate GPR model that can precisely predict the PES and ultimately reduce the number of required geometry optimization steps.^{28,29} Further discussion on this topic will be presented later in the paper.

These algorithmic schemes can be found in most electronic structure software packages. However, many of these packages are either not open source or their optimized geometry optimization algorithms are not generally available for inclusion into other open-source projects. Here, we discuss the performance of various open-source geometry optimization codes including both conventional methods and a GPR-based machine learning method for energy minimization, coupled with the free and open-source QUantum Interaction Computational Kernel (QUICK) program³⁰ for energy and gradient calculations. Several of the new GPR models are not available in open-source software, restricting our exploration to the method that is available in the development version of DL-FIND. While this model is effective, other more efficient models have been reported in the literature which, however, are not yet publicly available neither in free and open-source nor commercial software. We compare these results to the legacy optimizer in QUICK, basing results on the number of iterations it takes to converge geometries of an established test set of representative molecules.

2.2 Overview of Open-Source Geometry Optimization Software

2.2.1 QUICK–Legacy Optimizer

The QUantum Interaction Computational Kernel (QUICK) program³⁰ is an open-source, GPU enabled^{31,32}, *ab initio* and density functional theory program³³, which has been developed for QM and QM/MM calculations with Gaussian basis functions^{34,35}. It contains a limited-memory Broyden-Fletcher-Goldfarb-Shanno (L–BFGS) optimization algorithm which uses a Cartesian co-

ordinate system as input. L-BFGS is an optimization technique from the family of quasi-Newton methods that limits its memory usage while approximating the Broyden-Fletcher-Goldfarb-Shanno algorithm (BFGS)^{36,37}.

In general, the BFGS update of the inverse Hessian is³:

$$\Delta H^{-1} = \frac{\Delta x \Delta x^T}{\Delta x^T \Delta g} - \frac{H_{old}^{-1} \Delta g \Delta g^T H_{old}^{-1}}{\Delta g^T H_{old}^{-1} \Delta g} \quad (2.5)$$

where the updated inverse Hessian is used to form the Newton step, $\Delta x = H^{-1} \Delta g$. The L-BFGS method avoids the complete Hessian or its inverse from being stored because doing so would need $O(n^2)$ memory for n variables. Since L-BFGS starts with a diagonal inverse Hessian and only stores the x and g vectors from a few prior iterations, the storage requirement is only $O(n)$. The inverse Hessian is expressed as a diagonal Hessian plus the updates using the saved vectors. The new coordinates, $x_{new} = x_{old} - H^{-1} g_{old}$, may be expressed in terms of dot products between vectors, therefore the product of the updated inverse Hessian and the gradient only requires $O(n)$ work.

2.2.2 DL-FIND

DL-FIND is an open-source geometry optimization library that provides methods for local minimization, conical intersection optimization, population-based (global) optimization, reaction path optimization, and transition state search making it a versatile choice for molecular codes.³⁸

Cartesian coordinates, redundant internal coordinates, and hybrid delocalized internal coordinates (HDLC) are available options for DL-FIND geometry optimizations. To compare the optimizers with the GPR model that uses the Matérn kernel³⁹ to build the surrogate potential energy surface, the development version of DL-FIND was utilized alongside the stable version utilizing the L-BFGS method. The stable version of DL-FIND with the L-BFGS method has been the default geometry optimizer in QUICK since version QUICK-22.03.³⁰ The common interface between both versions and the QUICK optimizer library is as shown in Figure 2.1. The main DL-FIND geometry optimization driver routine is called from QUICK and does not exit until the optimization is finished. During the optimization, DL-FIND calls another interface routine when it needs to exchange information with QUICK during an optimization step. The input parameters are

obtained from QUICK using the initial interface call. These are the initial coordinates and the user-specified DL-FIND settings that control the optimization algorithm. The iterative optimization process then starts, and a new call is placed each time DL-FIND requires an energy or gradient. DL-FIND provides a set of Cartesian coordinates to QUICK and receives back the required information. During each optimization step, relevant information on the progress of the optimization is printed, and at the end of the geometry optimization, the routine reports the optimized set of Cartesian coordinates back to QUICK. To employ the GPR model in geometry optimization, a new surrogate PES is constructed incrementally using a specified coordinate system transformed from Cartesian coordinates after each *ab initio* calculation. The conventional L-BFGS algorithm is used to iteratively locate the minimum on this surrogate PES (micro-iterations), and then another *ab initio* calculation is performed in the Cartesian coordinates obtained from back-transformation. This entire cycle (macro-iterations) is repeated until the *ab initio* gradient is below the preset convergence threshold.⁴⁰

2.2.3 GeomeTRIC

GeomeTRIC is an open-source geometry optimization program that takes input in form of Cartesian coordinates and executes external electronic structure codes through wrapper functions with file-based data-exchange to obtain energy and gradients. Translation-rotation-internal coordinates (TRIC), delocalized internal coordinates (DLC), hybrid delocalized internal coordinates (HDLC), and redundant internal coordinates are all implemented by GeomeTRIC⁴¹. The code supports optimizations with constraints and transition state searches.

Translation-rotation-internal coordinates (TRIC) are created by including translations and rotations in the primitive internal coordinates set coupled with existing delocalization techniques. The initial Hessian in the space of primitive internal coordinates is a diagonal matrix with a few minor adjustments adopted from Schlegel’s proposed values⁴². The translations and rotations, and atomic Cartesian coordinates are given force constants of 0.05, bonds and angles involving non-covalent distances are given force constants of 0.1, and dihedral angles have their force constants set to 0.023.

GeomeTRIC uses a wrapper around QUICK, which is unlike DL-FIND as shown in Figure 2.1. Cartesian coordinates are used as input by GeomeTRIC, which transforms them to internal coordinates for optimization. After each cycle of optimization, it updates its own Hessian estimate using the BFGS algorithm. In order to generate an input file for QUICK to obtain the energy and Cartesian gradient, internal coordinates are converted to Cartesian coordinates at the end of each cycle. The Cartesian gradient from QUICK is used to calculate the internal coordinates gradient followed by Hessian update to estimate the search direction and step size.

2.2.4 Atomic Simulation Environment (ASE)

The Python-based Atomic Simulation Environment (ASE) was designed with the goal of setting up, directing, and analyzing atomistic simulations. Numerous features of ASE include molecular dynamics with various controls, including thermostats, structure improvement utilizing atomic forces, saddle point searches on potential energy surfaces, genetic algorithms for structure or chemical composition optimization, basin hopping or minima hopping algorithms for global structure optimization, analysis of phonon modes for solids or molecular vibrational modes⁴³.

Here, we focus exclusively on the geometry optimization methods provided by ASE, in particular the ASE internal L-BFGS optimizer, and the Sella and Berny algorithms. We will emphasize the Sella and Berny methods more as the L-BFGS approach has previously been summarized. The Berny geometry optimization algorithm is based on an earlier program written by H. B. Schlegel.⁴⁴ Sella is an open-source tool used for automating the process of finding saddle points and minimizing molecular structures.⁴⁵ It works by converting Cartesian coordinates into internal coordinates, automatically replacing pathological linear angles with improper dihedrals, and introducing necessary modifications like dummy atoms and constraints to ensure that the dummy atom does not drift unnecessarily over the course of optimization.⁴⁶ The algorithm utilizes Hessian diagonalization based on internal coordinates to determine the direction of minima, leading to a partially-exact Hessian matrix. This matrix guides the optimization process using a state-of-the-art constrained partitioned rational function approach, effectively steering the system towards local energy minima. The Berny method utilizes a valence force field to construct an estimate of a Hessian at the start

of the optimization. This approximation is then updated using the energies and first derivatives computed along the optimization pathway. The update is typically carried out using an iterative BFGS algorithm for minimization and a modified version of the original Schlegel update process for internal coordinate optimizations. It is ideally suited for optimizing covalently bound compounds since it is based on a redundant set of internal coordinates.

ASE uses the L-BFGS method as its internal optimization method of choice, while wrappers are used for the Berny optimizer from the PyBerny implementation and for the Sella optimizer. The actual geometry optimization method as well as the delocalized coordinates, screening functions, etc. are generated either directly or through PyBerny or Sella by the ASE program, which functions as a wrapper around the QUICK program with a file-based interface in the same manner as GeomeTRIC. The ASE program generates all of the QUICK input files and then runs QUICK to obtain the energy and gradient. The working scheme of ASE with QUICK is depicted in Figure 2.1.

2.3 Methodology

We developed calling interfaces for GeomeTRIC and ASE to execute QUICK by passing Cartesian coordinates, writing a QUICK input file, and retrieving energies and Cartesian gradients from the QUICK output file. While ASE uses the Cartesian coordinate system for the L-BFGS method with the unit matrix as the initial Hessian and the internal coordinate system for the Berny algorithm where the Hessian is estimated using Schlegel’s proposed values⁴² and the internal coordinate system for the Sella algorithm utilizes the scheme of Fischer and Almlof⁴⁷ to initialize the Hessian matrix; GeomeTRIC uses Translation-rotation-internal coordinates (TRIC) along with its own Hessian estimate with a few minor adjustments of Schlegel’s proposed values for geometry optimization using the BFGS algorithm. DL-FIND has been implemented internally within QUICK and makes use of a non-redundant hybrid-delocalized internal coordinate system and a unit matrix as the initial Hessian for optimization with the L-BFGS method. The development version of DL-FIND was used for GPR-based geometry optimizations in either Cartesian or hybrid-delocalized internal coordinate space. The built-in QUICK-legacy minimizer employs a Cartesian coordinate system

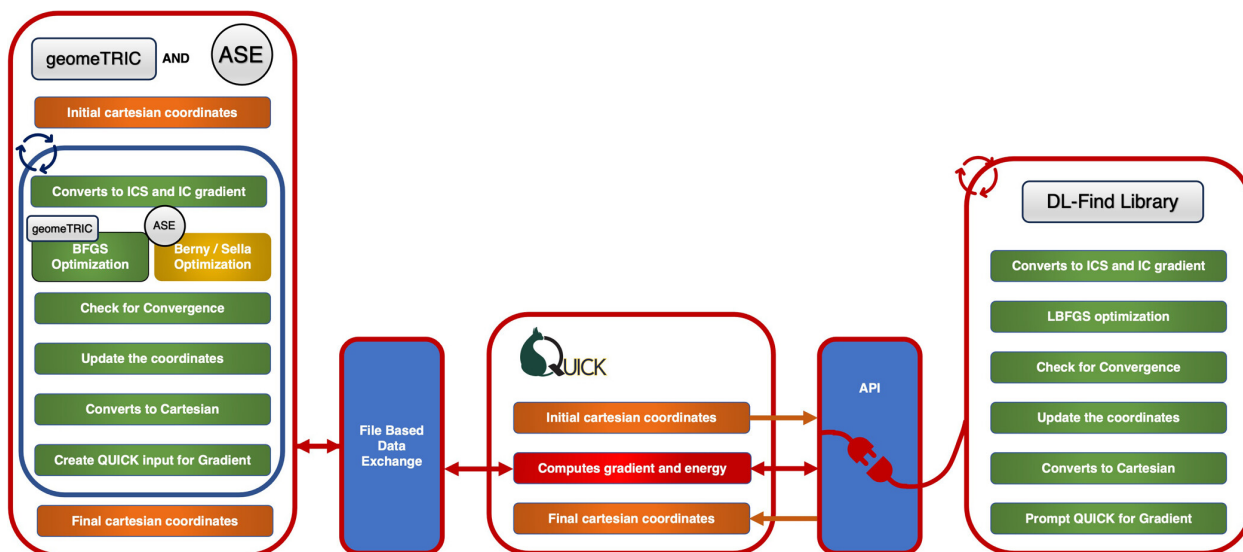


Figure 2.1 QUICK interface with the different open-source optimizers. The optimization steps of various optimizers are indicated by green color boxes where the initial Cartesian coordinates and gradient are transformed to an internal coordinate system (ICS) and internal gradients respectively for further optimization. The red box denotes the gradient computation performed by QUICK for various optimizers. The final coordinates obtained, and the initial coordinates used for calculation are represented by orange boxes. The yellow box denotes the PyBerny or Sella wrapper used by ASE for optimization. QUICK utilizes the DL-FIND library for optimization by passing gradient as required at each iteration. geomeTRIC and ASE carry out optimization independently utilizing QUICK for gradient calculation at each iteration.

with a unit matrix as the initial Hessian for L-BFGS based optimizations. The tests performed with these optimizers give us a clear understanding of the differences in performance between various methods that use the Cartesian coordinate system or any other internal coordinate system along with various Hessian estimations.

All Hartree-Fock (HF) calculations were performed with QUICK using the restricted Hartree-Fock method with the 6-31G** basis set (HF/6-31G**) for the energy and gradient. Results obtained with DL-FIND using different basis sets (HF/STO-3G, HF/6-31G, HF-6-31G*, HF/def2-SVP) as well as a generalized gradient approximation (GGA) and hybrid-GGA density functional method (BP86/def2-SVP, B3LYP/def2-SVP) are summarized in the Supporting Information. The convergence criteria for the geometry optimizations were as follows: a maximum gradient component of less than 0.00045 au, a change in energy from the previous step of less than 10^{-6} Hartree, and a maximum predicted displacement of less than 0.0018 Å per coordinate. These geometry con-

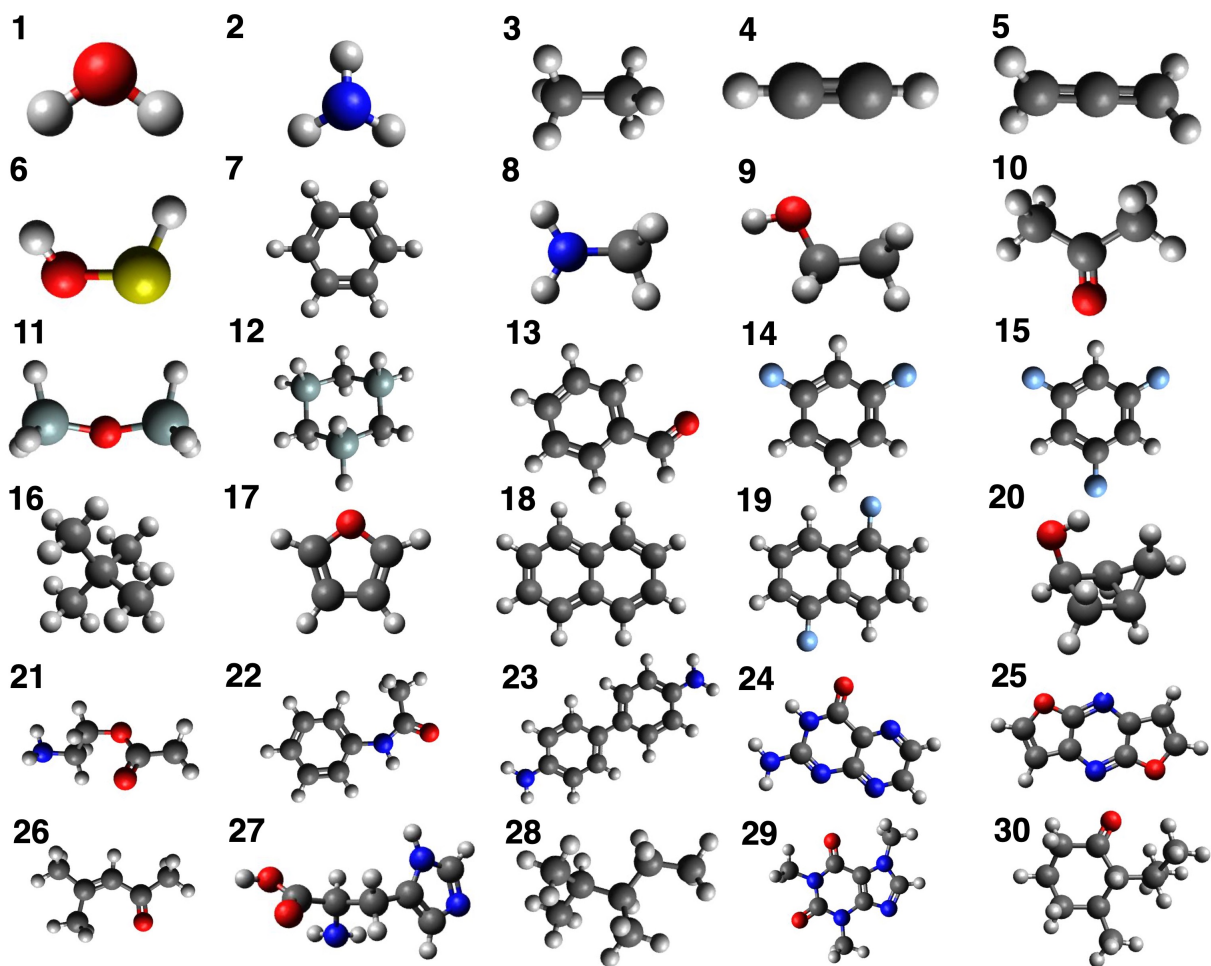


Figure 2.2 Initial geometries for the molecules in the Baker set. Oxygen is depicted in red, nitrogen in blue, carbon in gray, sulfur in yellow, fluorine in light blue, silicon in dark cyan, and hydrogen in white.

vergence criteria have been commonly utilized by the majority of past research that has examined the benchmark being studied, thus enabling direct comparisons of results. In order to guarantee accurate gradients, we used the TIGHTINT keyword in QUICK, which requests tight numerical cutoffs and SCF convergence. With these settings, the root-mean-square (RMS) change in the density matrix is less than 10^{-7} au and the maximum change in the density matrix is less than 10^{-5} au.

2.4 Results and discussion

In order to compare different geometry optimizers, a test suite of 30 molecules as depicted in Figure 2.2, originally suggested by Baker²⁰, was used as the initial geometry inputs for the

calculations. The test suite contains a variety of compounds, including fused bi- and tri-cyclics like caffeine and difluoropyrazine as well as smaller systems like water and ammonia. Input files containing initial geometries for optimization are available in the Supporting Information. Baker used the Eigenvector Following (EF) algorithm⁴⁸ at the restricted Hartree–Fock (RHF) level with the STO-3G basis set (HF/STO-3G) for optimizations in both Cartesian and internal coordinates with a convergence criterion for the gradient of 3.0×10^{-4} au. Table 1 lists the number of steps required to reach the minima for each of the molecules for a number of different geometry optimization algorithms in comparison to Baker’s results. The sum of optimization steps for all molecules in the test suite and the average number of optimization steps per molecule is also reported. Given the vagaries of examining the data on a molecule-to-molecule basis we feel these averages give the best overall assessment of the performance of the geometry optimization algorithms investigated in this work.

With quasi-Newton techniques Baker reported requiring a total of 765 geometry cycles (on average 26 steps per molecule) with a Cartesian coordinate system while requiring only 371 geometry cycles (on average 12 steps per molecule) with an internal coordinate system. In both cases Baker employed the unit matrix as the initial Hessian. Using an initial Hessian from a molecular mechanics model, Baker managed to further reduce the number of steps to 240 (on average 8 steps per molecule)²⁰. This data is not shown in Table 1 but the numbers should be kept in mind for the following discussion.

The total number of geometry cycles required to optimize the same set of molecules using the optimizers tested here ranges from 728 for DL-FIND/GPR-Cartesian, 683 for QUICK-Legacy, 613 for ASE/L-BFGS, 356 for DL-FIND/L-BFGS, 326 for DL-FIND/GPR-Internal, 287 for geo-meTRIC, 190 for ASE/Berny, to as low as 187 for ASE/Sella. Geometry optimization in Cartesian coordinates without sophisticated initial Hessian guess (ASE/L-BFGS, DL-FIND/GPR-Cartesian and QUICK-Legacy) thereby required between 20 and 24 steps per molecule on average. Optimization in internal coordinate space is much more efficient, requiring only 11 to 12 optimization steps (DL-FIND/LBFGS, DL-FIND/GPR-Internal). Using both internal coordinates and a better

initial Hessian reduces this further to 6 to 10 optimization steps (ASE/Berny, ASE/Sella, GeomeTRIC). Here ASE/Berny and ASE/Sella use redundant internal coordinates, while geomeTRIC uses translation-rotational-internal coordinates, and DL-FIND uses delocalized internal coordinates. Except for ASE/Berny and geomeTRIC, which use a modified version of the original Schlegel update process, as well as ASE/Sella, all methods initialize the Hessian to be the unit matrix. These results reconfirm the two well-known observations that i) internal coordinates outperform the Cartesian coordinate system for molecular geometry optimizations, and ii) good approximations to the initial Hessian are also important to reduce the number of required steps.

We finally point out that while Baker used the HF/STO-3G method and we are using HF/6-31G**, the observed overall trends are comparable (see Table 1). To provide a more comprehensive understanding of the test suite, we used the DL-FIND/L-BFGS implementation to compare the effect of different levels of theory and basis sets in addition to HF/6-31G**. The results of this comparison are summarized in Table S1 of the Supporting Information. While there are differences for individual molecules, the total number of steps increases only slightly with increasing flexibility in the basis set, from 331 for HF/STO-3G to 356 as mentioned above for HF/6-31G**. The total number of steps remains very similar between HF and representative generalized gradient approximation (GGA) and hybrid-GGA density functional methods. This comparison allows for a clearer and more complete picture of the observed overall trends, which are consistent with those presented in Table 1.

As mentioned above, the number of geometry cycles required to converge to the optimized geometry can be reduced with a concomitant increase in computational time depending on the Hessian used. Bakken and Helgaker, for instance, already addressed this problem, where it is noted that utilizing precise Hessians at each step reduces the number of required iterations while compromising on total computation time. Only 125 and 111 steps are required for the Baker test set with HF/STO-3G using Cartesian and redundant internal coordinates, respectively, when using exact Hessians at each step.²¹

Baker, in Ref. 20, used natural internal coordinates and the initial Hessian provided by the CVFF

force field in combination with an eigenvector following (EF) approach to reduce the number of iterations required for HF/STO-3G geometry optimization to 240 steps. Bakken and Helgaker managed to optimize the Baker test set in 185 steps using a combination of extra-redundant internal coordinates, a good approximate initial Hessian with BFGS updates, and optimized step size²¹. A slightly larger number of 198 steps was reported when using the more flexible 6-31G* basis set. Swart and Bickelhaupt managed to further reduce the number of geometry optimization steps using the delocalized coordinates setup of Baker, a combination of quasi-Newton steps with GDIIS⁴⁹, and a modification of Lindh’s force constant model to generate initial Hessians⁵⁰, resulting in an impressive 173 steps for the Baker test set using the PW91 density functional and a large triple zeta Slater type basis set with polarization functions (TZP).² It therefore stands to reason that replacing the unit matrix with a better initial Hessian approximation should also reduce the number of geometry optimization steps required by the L-BFGS optimization employed by DL-FIND.

The use of machine learning models such as GPR with efficient kernel functions²⁴ in conjunction with different internal coordinate systems can also reduce the number of required optimization cycles. For instance, Meyer and Hauser achieved 225 steps for the Baker test set by using reduced redundancy Z-matrix-derived internal coordinates with HF/STO-3G.²⁸ Raggi *et al.* introduced a restricted-variance optimization (RVO) scheme that utilizes a Hessian model function²³ to generate a non-redundant set of internal coordinates for molecular geometries, making the surrogate model invariant to translations and rotations. With the HF/6-31G basis set, they achieved 225 steps using RVO, while with DFT(B3LYP)/def2-SVP, they obtained 241 steps.²⁵ Teng, Huang, and Bao demonstrated that combining the gradient-enhanced universal kriging (GEUK) algorithm with an adaptive *ab initio* prior mean function, which incorporates prior physical knowledge into surrogate-based optimization, and incomplete internal/incomplete Cartesian methods can reduce the number of steps to 165 as shown for the BP86/def2-SVP level of theory.²⁹

Herein we explored several different open-source quasi-Newton approaches (L-BFGS and some modified BFGS algorithms) and available machine learning (GPR) models, utilizing various internal coordinate systems with either the unit matrix as the starting Hessian (as in the L-BFGS

approaches) or a more sophisticated first estimate (geomeTRIC, ASE/Berny, ASE/Sella). In all cases, not unexpectedly, internal coordinates outperformed methods using Cartesian coordinates, but the numerical details of the implementation of the internal coordinate methods make a significant difference (*e.g.*, ASE/Berny *versus* geomeTRIC). Furthermore, the use of more sophisticated initial Hessians and geometry update algorithms resulted in a significant reduction of required steps. Importantly, the open-source implementations that are currently available require approximately the same number of steps as the best traditional or machine-learning based approaches that have been reported in the literature offering the community state-of-the-art optimizers in an open-source offering.

2.5 Conclusion

We have presented a comparison of different open-source geometry optimization codes available to the community. The performance of the geometry optimization algorithms implemented in these codes has been evaluated for a test set of 30 molecules that was originally proposed by Baker. We find results in line with expectations for the various optimization algorithms. While there are differences for individual molecules, when Cartesian coordinates are used, the optimization algorithms take at least 20 and in the worst case 24 steps on average per molecule of this test set. For internal coordinate optimization algorithms, the number of steps is greatly reduced requiring about 12 steps per molecule when using a unit matrix as initial Hessian and as little as 6 steps per molecule when using more sophisticated initial Hessians and geometry update algorithms. Overall, the best performing open-source software in terms of the number of iterations was Sella and the Berny algorithm as implemented in PyBerny, both offered through ASE interfaces, closely followed by the BFGS method as implemented in geomeTRIC. Additionally, the release version of DL-FIND, utilizing a conventional L-BFGS algorithm along with delocalized internal coordinates, also demonstrated commendable performance.

In terms of computational efficiency, when running QUICK on a single NVIDIA V100 GPU using the HF/6-31G** level of theory, our assessment revealed notable differences in computation times among the various open-source software packages. Specifically, DL-FIND was the most

Table 2.1 Number of steps required to optimize the geometry of molecules from the Baker test set using open-source QUICK with various open-source geometry optimization routines. Results for Baker/Cartesian and Baker/Internal are for HF/STO-3G and taken from Baker’s work²⁰. All other results are obtained with HF/6-31G**. QUICK-Legacy and ASE/L-BFGS both use L-BFGS in Cartesian coordinate space. ASE/Berny and geomeTRIC use a modified version of the Schlegel update process, while other methods initialize the Hessian as the unit matrix. DL-FIND, ASE/Berny and geomeTRIC use different forms of internal coordinate systems(see discussion).

Molecules	Baker		QUICK-Legacy		ASE		geomeTRIC		DL-FIND	
	EF		L-BFGS		L-BFGS		BFGS		L-BFGS	
	Cartesian	Internal	Cartesian	Internal	Cartesian	Internal	Internal	Sella	Internal	GPR
Water	5	7	7	7	5	4	4	4	6	5
Ammonia	7	7	9	4	6	4	4	4	6	4
Ethane	7	6	10	4	8	4	4	4	5	5
Acetylene	7	5	11	4	6	5	5	4	7	7
Allene	10	8	11	4	5	5	6	4	6	6
Hydroxysulfane	21	17	17	7	16	6	7	7	15	18
Benzene	6	5	16	3	4	3	4	3	4	5
Methylamine	10	7	14	5	8	4	6	5	7	7
Ethanol	18	8	16	6	15	5	5	6	10	9
Acetone	22	7	20	5	18	5	6	5	10	10
Disilyl-ether	27	13	21	9	30	9	14	10	17	10
1,3,5-Trisilacyclohexane	36	14	22	4	25	4	10	5	20	16
Benzaldehyde	17	9	17	7	17	7	11	7	8	7
1,3-Difluorobenzene	8	9	9	4	10	4	9	5	6	7
1,3,5-Trifluorobenzene	7	6	9	4	6	4	11	4	6	8
Neopentane	10	7	11	4	8	4	5	3	6	5
Furan	10	11	13	5	10	5	6	6	7	7
Naphthalene	11	9	12	5	9	5	7	5	7	7
1,5-Difluoronaphthalene	16	11	18	6	13	5	10	6	7	8
2-Hydroxybicyclopentane	45	23	38	11	39	9	12	11	28	23
ACHTAR10	51	19	41	7	19	10	11	7	15	13
ACANIL01	36	10	38	7	33	7	12	7	9	12
Benzidine	26	25	20	7	41	7	13	7	17	16
Pterin	23	12	28	11	22	11	12	9	12	10
Difuroprazine	21	11	21	7	16	7	10	8	9	8
Mesityl-oxide	38	9	30	6	23	7	10	6	11	11
Histidine	102	44	113	13	66	14	22	13	34	33
Dimethylpentane	29	15	10	4	22	6	21	4	18	13
Caffeine	39	11	32	8	31	10	12	8	11	10
Menthone	100	26	68	10	82	10	18	10	32	27
Sum	765	371	683	187	613	190	287	187	356	326
Average per Molecule	26	12	23	6	20	6	10	6	12	11

efficient, averaging 34 seconds per molecule. In contrast, Sella took an average of 81 seconds, and PyBerny took 198 seconds per molecule. It is important to note that the differences in computation times were observed to vary with the size of the molecule, with a more pronounced gap for smaller molecules that gradually reduces for larger molecules. Hence, molecules larger than those in the Baker set will likely show less pronounced differences in computation time or tip the scale in favor of Sella as the QM portion of the calculation will dominate over time spent by the geometry optimizer. The results of this comparison are summarized in Table 2.A.2 of the Supporting Information. DL-FIND's advantage stems from the fact that it is written in Fortran which generates faster code than Python and that it is built into QUICK as a library, eliminating file-based data exchange and repeated program startup time compared to the ASE libraries, which require wrappers for data transfer. Sella and PyBerny also employ more sophisticated Hessian update strategies, which further increases the computation time.

Given this analysis, we are using the release version of DL-FIND as the default optimizer in QUICK for simplified distribution within a single binary executable and usage without requiring Python wrappers. This integration enhances the inherent optimization framework and introduces supplementary features like conical intersection optimization, reaction path optimization, and transition state search. Additionally, it is worth noting that the development version of DL-FIND contains GPR optimization for both local minima and transition state optimization. This upgrade will be incorporated into QUICK once it becomes the release version. While DL-FIND combined with QUICK offers substantial capabilities, we also do recommend utilizing the Berny algorithm as implemented in PyBerny or the Sella geometry optimization software for large-scale production optimization efforts due to their performance (*e.g.*, in the creation of large synthetic data sets for ML/AI efforts). These optimizers are accessible for usage with QUICK through the ASE interface as mentioned in the Supporting Information. The powerful combination of QUICK with the latest version of DL-FIND, ASE/Berny or ASE/Sella provides a reliable and robust open-source solution for efficiently optimizing molecular geometry based on *ab initio* and density functional theory methods, combined with the computational capabilities of graphics processing units (GPUs).

BIBLIOGRAPHY

- [1] H. Bernhard Schlegel. Exploring potential energy surfaces for chemical reactions: An overview of some practical methods. *Journal of Computational Chemistry*, 24(12):1514–1527, 2003.
- [2] Marcel Swart and F. Matthias Bickelhaupt. Optimization of strong and weak coordinates. *International Journal of Quantum Chemistry*, 106(12):2536–2544, 2006.
- [3] H. Bernhard Schlegel. Geometry optimization. *WIREs Computational Molecular Science*, 1(5):790–809, 2011.
- [4] Robert E. Kass, John E. Dennis, and Robert B. Schnabel. Numerical methods for unconstrained optimization and nonlinear equations. *Journal of the American Statistical Association*, 80:247–248, 1985.
- [5] L. E. Scales. *Introduction to Non-Linear Optimization*. Springer-Verlag, 1985.
- [6] Satoshi Maeda, Koichi Ohno, and Keiji Morokuma. Automated global mapping of minimal energy points on seams of crossing by the anharmonic downward distortion following method: A case study of h_2co . *Journal of Physical Chemistry A*, 113:1704–1710, 2009.
- [7] Søren Madsen and Frank Jensen. Locating seam minima for macromolecular systems. *Theoretical Chemistry Accounts*, 123:477–485, 2009.
- [8] Bernhard Dick. Gradient projection method for constraint optimization and relaxed energy paths on conical intersection spaces and potential energy surfaces. *Journal of Chemical Theory and Computation*, 5:116–125, 2009.
- [9] Weinan E and Eric Vanden-Eijnden. Transition-path theory and path-finding algorithms for the study of rare events. *Annual Review of Physical Chemistry*, 61:391–420, 2010.
- [10] H. Bernhard Schlegel. Optimization of equilibrium geometries and transition structures. *Journal of Computational Chemistry*, 3:214–218, 1982.
- [11] Jon Baker. Geometry optimization in cartesian coordinates: Constrained optimization. *Journal of Computational Chemistry*, 13:240–253, 1992.
- [12] John D. Head. Partial optimization of large molecules and clusters. *Journal of Computational Chemistry*, 11:67–75, 1990.
- [13] P. Pulay and G. Fogarasi. Geometry optimization in redundant internal coordinates. *The Journal of Chemical Physics*, 96:2856–2860, 1992.
- [14] Jon Baker, Don Kinghorn, and Peter Pulay. Geometry optimization in delocalized internal

- coordinates: An efficient quadratically scaling algorithm for large molecules. *Journal of Chemical Physics*, 110:4986–4991, 1999.
- [15] J. A. Pople, R. Krishnan, H. B. Schlegel, and J. S. Binkley. Derivative studies in hartree-fock and møller-plesset theories. *International Journal of Quantum Chemistry*, 16:225–241, 1979.
 - [16] B. A. Murtagh and R. W. H. Sargent. Computational experience with quadratically convergent minimisation methods. *The Computer Journal*, 13(2):185–194, 01 1970.
 - [17] Donald Goldfarb. A family of variable-metric methods derived by variational means. *Mathematics of Computation*, 24(109):23–26, 1970.
 - [18] C. G. Broyden. The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations. *IMA Journal of Applied Mathematics*, 6(1):76–90, 03 1970.
 - [19] R. Fletcher. A new approach to variable metric algorithms. *The Computer Journal*, 13(3):317–322, 01 1970.
 - [20] Jon Baker. Techniques for geometry optimization: A comparison of cartesian and natural internal coordinates. *Journal of Computational Chemistry*, 14(9):1085–1100, 1993.
 - [21] Vebjørn Bakken and Trygve Helgaker. The efficient optimization of molecular geometries using redundant internal coordinates. *Journal of Chemical Physics*, 117:9160–9174, 2002.
 - [22] Jon Baker, Alain Kessi, and Bernard Delley. The generation and use of delocalized internal coordinates in geometry optimization. *The Journal of Chemical Physics*, 105(1):192–212, 1996.
 - [23] Roland Lindh, Anders Bernhardsson, Gunnar Karlström, and Per Åke Malmqvist. On the use of a hessian model function in molecular geometry optimizations. *Chemical Physics Letters*, 241(4):423–428, 1995.
 - [24] Alexander Denzel and Johannes Kästner. Gaussian process regression for geometry optimization. *The Journal of Chemical Physics*, 148(9):094114, 2018.
 - [25] Gerardo Raggi, Ignacio Fdez. Galván, Christian L. Ritterhoff, Morgane Vacher, and Roland Lindh. Restricted-variance molecular geometry optimization based on gradient-enhanced kriging. *Journal of Chemical Theory and Computation*, 16(6):3989–4001, 2020.
 - [26] Alexander Denzel and Johannes Kästner. Gaussian process regression for transition state search. *Journal of Chemical Theory and Computation*, 14(11):5777–5786, 2018.
 - [27] Alexander Denzel, Bernard Haasdonk, and Johannes Kästner. Gaussian process regression for minimum energy path optimization and transition state search. *The Journal of Physical Chemistry A*, 123(44):9600–9611, 2019.

- [28] Ralf Meyer and Andreas W. Hauser. Geometry optimization using gaussian process regression in internal coordinate systems. *The Journal of Chemical Physics*, 152(8):084112, 2020.
- [29] Chong Teng, Daniel Huang, and Junwei Lucas Bao. A spur to molecular geometry optimization: Gradient-enhanced universal kriging with on-the-fly adaptive ab initio prior mean functions in curvilinear coordinates. *The Journal of Chemical Physics*, 158(2):024112, 2023.
- [30] M. Manathunga, A. Shajan, T. J. Giese, V. W. D. Cruzeiro, J. Smith, Y. Miao, X. He, K. Ayers, E. Brothers, A. W. Götz, and K. M. Merz. Quick-22.03, university of california san diego, ca and michigan state university, east lansing, mi, 2022.
- [31] Yipu Miao and Kenneth M. Merz, Jr. Acceleration of high angular momentum electron repulsion integrals and integral derivatives on graphics processing units. *Journal of Chemical Theory and Computation*, 11(4):1449–1462, 2015.
- [32] Madushanka Manathunga, Chi Jin, Vinícius Wilian D. Cruzeiro, Yipu Miao, Dawei Mu, Kamesh Arumugam, Kristopher Keipert, Hasan Metin Aktulga, Kenneth M. Merz, Jr., and Andreas W. Götz. Harnessing the power of multi-gpu acceleration into the quantum interaction computational kernel program. *Journal of Chemical Theory and Computation*, 17(7):3955–3966, 2021.
- [33] Madushanka Manathunga, Yipu Miao, Dawei Mu, Andreas W. Götz, and Kenneth M. Merz, Jr. Parallel implementation of density functional theory methods in the quantum interaction computational kernel program. *Journal of Chemical Theory and Computation*, 16(7):4315–4326, 2020.
- [34] Vinícius Wilian D. Cruzeiro, Madushanka Manathunga, Kenneth M. Merz, Jr., and Andreas W. Götz. Open-source multi-gpu-accelerated qm/mm simulations with amber and quick. *Journal of Chemical Information and Modeling*, 61(5):2109–2115, 2021.
- [35] Madushanka Manathunga, Hasan Metin Aktulga, Andreas W. Götz, and Kenneth M. Merz, Jr. Quantum mechanics/molecular mechanics simulations on NVIDIA and AMD graphics processing units. *Journal of Chemical Information and Modeling*, 63(3):711–717, 2023.
- [36] Jorge Nocedal. Updating quasi-newton matrices with limited storage. *Mathematics of Computation*, 35(151):773–782, 1980.
- [37] Dong C. Liu and Jorge Nocedal. On the limited memory bfgs method for large scale optimization. *Mathematical Programming*, 45:503–528, 1989.
- [38] Johannes Kästner, Joanne M. Carr, Thomas W. Keal, Walter Thiel, Adrian Wander, and Paul Sherwood. Dl-find: An open-source geometry optimizer for atomistic simulations. *Journal of Physical Chemistry A*, 113:11856–11865, 2009.
- [39] Bertil Matérn. *Spatial variation*, volume 36. Springer Science & Business Media, 2013.

- [40] Daniel Born and Johannes Kästner. Geometry optimization in internal coordinates based on gaussian process regression: Comparison of two approaches. *Journal of Chemical Theory and Computation*, 17(9):5955–5967, 2021.
- [41] Lee-Ping Wang and Chenchen Song. Geometry optimization made simple with translation and rotation coordinates. *The Journal of Chemical Physics*, 144(21):214108, 2016.
- [42] H. Bernhard Schlegel. Estimating the hessian for gradient-type geometry optimizations. *Theoretica Chimica Acta*, 66:333–340, 1984.
- [43] Ask Hjorth Larsen, Jens JØrgen Mortensen, Jakob Blomqvist, Ivano E. Castelli, Rune Christensen, Marcin Dułak, Jesper Friis, Michael N. Groves, BjØrk Hammer, Cory Hargus, Eric D. Hermes, Paul C. Jennings, Peter Bjerre Jensen, James Kermode, John R. Kitchin, Esben Leonhard Kolsbjerg, Joseph Kubal, Kristen Kaasbjerg, Steen Lysgaard, Jón Bergmann Maronsson, Tristan Maxson, Thomas Olsen, Lars Pastewka, Andrew Peterson, Carsten Rostgaard, Jakob SchiØtz, Ole Schütt, Mikkel Strange, Kristian S. Thygesen, Tejs Vegge, Lasse Vilhelmsen, Michael Walter, Zhenhua Zeng, and Karsten W. Jacobsen. The atomic simulation environment - a python library for working with atoms. *Journal of Physics Condensed Matter*, 29:273002, 2017.
- [44] Adam B. Birkholz and H. Bernhard Schlegel. Exploration of some refinements to geometry optimization methods. *Theoretical Chemistry Accounts*, 135(4):84, Mar 2016.
- [45] Eric D. Hermes, Khachik Sargsyan, Habib N. Najm, and Judit Zádor. Sella, an open-source automation-friendly molecular saddle point optimizer. *Journal of Chemical Theory and Computation*, 18(11):6974–6988, 2022.
- [46] Eric D. Hermes, Khachik Sargsyan, Habib N. Najm, and Judit Zádor. Geometry optimization speedup through a geodesic approach to internal coordinates. *The Journal of Chemical Physics*, 155(9):094105, 09 2021.
- [47] Thomas H. Fischer and Jan Almlof. General methods for geometry and wave function optimization. *The Journal of Physical Chemistry*, 96(24):9768–9774, 1992.
- [48] Jon Baker. An algorithm for the location of transition states. *Journal of Computational Chemistry*, 7(4):385–395, 1986.
- [49] Pál Császár and Péter Pulay. Geometry optimization by direct inversion in the iterative subspace. *Journal of Molecular Structure*, 114:31–34, 1984.
- [50] Roland Lindh, Anders Bernhardsson, and Martin Schütz. Force-constant weighted redundant coordinates in molecular geometry optimizations. *Chemical physics letters*, 303(5-6):567–575, 1999.

APPENDIX

SUPPORTING INFORMATION

2A.1 Geometry optimization with different basis sets and methods

Table 2.A.1 lists the number of steps required to optimize the geometry of the molecules from the Baker test set with QUICK 23.08 and DL-FIND using the L-BFGS algorithm and delocalized internal coordinates. Different levels of theory with single- and double-zeta basis sets with and without polarization functions were used. The same SCF and geometry optimization thresholds as described in the Methodology section of the manuscript were applied. Geometries were considered converged with a maximum gradient component of less than 0.00045 au, a change in energy from the previous step of less than 10^{-6} Hartree, and a maximum predicted displacement of less than 0.0018 Å. The TIGHTINT keyword was used, which for SCF convergence requests that the RMS change in the density matrix is less than 10^{-7} au and the maximum change in the density matrix is less than 10^{-5} au.

2A.2 Geometry optimization wall clock time

Table 2.A.2 lists the computation time for optimizing molecular geometry using QUICK 23.08 and the Baker test set with various optimization approaches by Sella, PyBerny, and DL-FIND using a single NVIDIA V100 GPU, specifically the NVIDIA V100S-PCIe 32GB GPU chipset, and a 2.4 GHz Intel Xeon Platinum 8260 CPU. The quantum mechanical calculations were performed at the Hartree-Fock (HF) level of theory, employing the 6-31G** basis set. The software stack utilized for these calculations includes ASE 3.23.0b1, PyBerny 0.6.3, and Sella 2.3.2. Sella 2.3.2 employed the SciPy 1.10.1, NumPy 1.22.4, JAX and JAXlib 0.4.13 libraries, and the calculations were executed using Python version 3.8.8. In each optimization approach, the converged electron density from each step was employed as the initial guess for the subsequent step's self-consistent field (SCF) calculation. File input and output operations were carried out on a dedicated fast scratch file system within the data center.

Table 2A.1 Number of steps required to optimize the geometry of molecules from the Baker test set using different levels of theory and basis sets with the open-source QUICK and DL-FIND geometry optimizer with use of the L-BFGS optimizer and delocalized internal coordinate systems.

Molecules	HF/STO-3G	HF/6-31G	HF/6-31G*	HF/6-31G**	HF/def2-SVP	B3LYP/def2-SVP	BP86/def2-SVP
Water	6	6	6	6	6	6	6
Ammonia	5	7	6	6	6	6	6
Ethane	5	5	5	5	6	6	6
Acetylene	9	7	8	7	7	9	9
Allene	6	6	6	6	6	5	6
Hydroxysulfane	15	18	15	15	16	15	16
Benzene	4	4	4	4	4	4	5
Methylamine	6	7	7	7	7	6	6
Ethanol	9	9	10	10	10	10	10
Acetone	9	9	10	10	12	10	10
Disilyl-ether	9	14	18	17	18	13	13
1,3,5-Trisilacyclohexane	19	21	19	20	19	22	21
Benzaldehyde	6	8	8	8	8	7	8
1,3-Difluorobenzene	6	6	6	6	6	6	7
1,3,5-Trifluorobenzene	6	6	6	6	6	8	6
Neopentane	6	6	6	6	6	7	9
Furan	7	7	7	7	7	7	7
Naphthalene	7	7	7	7	7	8	7
1,5-Difluoronaphthalene	9	7	7	7	7	7	9
2-Hydroxybicyclopentane	24	28	28	28	29	28	23
ACHTAR 10	19	19	16	15	16	14	15
ACANIL 01	9	9	9	9	9	9	11
Benzidine	18	17	17	17	16	16	11
Pterin	11	11	12	12	12	11	17
Difuroprazine	9	9	9	9	9	12	9
Mesityl-oxide	10	10	11	11	13	12	11
Histidine	35	34	33	34	33	42	42
Dimethylpentane	9	9	15	18	15	17	16
Caffeine	10	11	11	11	11	12	14
Menthone	28	30	30	32	34	34	36
Sum	331	348	352	356	361	369	372
Average	11	12	12	12	12	12	12

Table 2A.2 Wall clock time (seconds) required for geometry optimization of molecules from the Baker test set using HF theory with the 6-31G** basis set and different optimization methods implemented by Sella, PyBernv, and DL-FIND, running QUICK on a single NVIDIA V100 GPU.

Molecules	Sella	PyBernv	DL-Find
Water	26	47	5
Ammonia	26	62	5
Ethane	30	71	6
Acetylene	27	92	6
Allene	29	91	7
Hydroxysulfhane	56	115	13
Benzene	22	52	9
Methylamine	36	72	7
Ethanol	62	94	9
Acetone	38	93	11
Disilyl-ether	102	234	21
1,3,5-Trisilacyclohexane	212	597	62
Benzaldehyde	79	196	21
1,3-Difluorobenzene	49	87	15
1,3,5-Trifluorobenzene	39	91	17
Neopentane	24	85	11
Furan	54	106	10
Naphthalene	64	148	26
1,5-Difluoronaphthalene	90	164	32
2-Hydroxybicyclopentane	133	227	42
ACHTAR 10	74	279	29
ACANIL 01	99	232	33
Benzidine	142	311	103
Pterin	148	354	51
Difuropyrazine	132	243	39
Mesityl-oxide	65	183	24
Histidine	219	518	127
Dimethylpentane	43	160	42
Caffeine	119	426	71
Menthone	205	510	174
Sum	2444	5940	1028
Average	81	198	34

CHAPTER 3

ACCURATE QUANTUM-CENTRIC SIMULATIONS OF SUPRAMOLECULAR INTERACTIONS

We present the first quantum-centric simulations of noncovalent interactions using a supramolecular approach. We simulate the potential energy surfaces (PES) of the water and methane dimers, featuring hydrophilic and hydrophobic interactions, respectively, with a sample-based quantum diagonalization (SQD) approach. Our simulations on quantum processors, using 27- and 36-qubit circuits, are in remarkable agreement with classical methods, deviating from complete active space configuration interaction (CASSI) and coupled-cluster singles, doubles, and perturbative triples (CCSD(T)) within 1 kcal/mol in the equilibrium regions of the PES. Finally, we test the capacity limits of the quantum methods for capturing hydrophobic interactions with an experiment on 54 qubits. These results mark significant progress in the application of quantum computing to chemical problems, paving the way for more accurate modeling of noncovalent interactions in complex systems critical to the biological, chemical and pharmaceutical sciences.

3.1 Introduction

The accurate treatment of noncovalent interactions^{1,2} is extremely important in the biological, chemical, and pharmaceutical sciences^{3,4}. Specifically, non-covalent interactions between hydrophobic species and hydrogen-bonded pairs play pivotal roles in a myriad of biological processes, ranging from protein folding^{5–9} membrane assembly¹⁰, cell signaling¹¹ and drug discovery^{12–16}. The correct modeling of these interactions along with solvation plays a key role in understanding many chemical and biological processes¹⁷.

Traditionally, quantum mechanical^{18,19} methods have been used to study these systems to a high level of accuracy - so-called chemical accuracy (± 1 kcal/mol from experiment). However, these calculations are quite expensive and approaches to accelerate these calculations continue to be explored using classical hardware.^{20–29} Using the results from these calculations, force fields have been fine-tuned for wide use in molecular simulation studies of chemical and biological processes^{30–34}. More recently, machine learning³⁵ methods built using accurate quantum chemical

calculations on thousands of systems have appeared to study largely covalent interactions, but can be extended to non-covalent interactions at a good level of accuracy, but at a reduced computational cost. However, these latter methods build models that can struggle to study diverse systems outside of the training set and can be subject to overfitting^{36,37}.

Quantum computing based studies of these interactions, to a high level of accuracy and speed, would revolutionize our ability to understand complex processes like drug binding, but would also allow for the development of large synthetic datasets that could be used to build even better force fields and quantum machine learning models. However, to date, quantum hardware has struggled to address these problems. In this work we demonstrate that quantum-centric supercomputing (QCSC)³⁸ combined with the sample-based quantum diagonalization (SQD) approach³⁹ allows for the study of intermolecular interactions.

QCSC is a new computational paradigm, in which a quantum computer operates in concert with classical high-performance computing (HPC) resources. Classical processing carried out before, during, and after quantum computations allows for the introduction of quantum subroutines in the workflow of classical HPC algorithms, to extract and amplify signal from noisy quantum devices, and to leverage quantum processors to execute a limited number of large quantum circuits.

The QCSC architecture enables scaling of computational capabilities, as exemplified by methods that use classical diagonalization in subspaces determined by quantum samples such as SQD³⁹ and QSCI⁴⁰. The SQD method is developed based on QSCI. The SQD method use a quantum device to sample electronic configurations from a quantum circuit approximating the ground state of a molecular Hamiltonian, and use classical distributed HPC resources to post-process quantum measurements against known symmetries to obtain recovered configurations³⁹, as well as to solve the Schrödinger equation in the subspace spanned by the recovered configurations. The SQD method recently allowed us to address instances of the electronic structure problem with up to 36 spatial orbitals using up to 77 qubits³⁹. The QCSC workflows produced significant improvements over simulations using quantum computers in isolation – which have in the last decade, used up to a handful of qubits with limited accuracies^{41–78}. The QCSC paradigm coupled with SQD enables

the study of problems heretofore out of reach of quantum computers including static correlation in iron-sulfur complexes³⁹ and well as dynamical correlation as exemplified in the intermolecular interactions studied herein.

Past studies have reported the simulation of noncovalent interactions^{79,80} using symmetry-adapted perturbation theory (SAPT). This method expresses the interaction energy through a perturbative treatment of the intermolecular potential^{81–83}, and requires the simulation of electronic structure of individual monomers on a quantum computer. In addition Anderson et al. demonstrated the possibility of simulations of coarse-grained intermolecular interactions on quantum computer as well.⁸⁴ However, to date, predicting binding energies between monomers using the supramolecular approach, where the electronic structure of entire dimer need to be simulated on quantum hardware, has been an elusive target for quantum simulations, due to lack of accuracy and scale of conventional quantum approaches.

Herein, we present the first quantum-centric simulation for the modeling of noncovalent hydrophilic and hydrophobic interactions with a supramolecular approach. We simulate the potential energy surfaces (PES) of the water dimer and the methane dimer. Our water dimer simulations use 27-qubit circuits, while the methane dimer simulations use 36- and 54-qubit circuits. To assess the accuracy of our quantum solutions, we compare them against heat-bath configuration interaction (HCI)^{85–88} in the case of (16e,24o) calculations, complete active space configuration interaction for the (16e,12o) and (16e,16o) calculations, as well as coupled-cluster singles, doubles and perturbative triples (CCSD(T))⁸⁹ performed for all of the studied instances. The latter is widely recognized as the gold standard for computing intermolecular interactions⁹⁰ to chemical accuracy. For the 27-qubit water dimer and the 36-qubit methane dimer simulations, we demonstrate that SQD energies agree with CASCI nearly exactly, while deviating from CCSD(T) within 1 kcal/mol in the equilibrium region of the PES. For the 54-qubit simulations of the methane dimer, we observe how the accuracy of the quantum solution can be systematically improved by increasing the number of sampled configurations.

3.2 Methods and Computational Details

3.2.1 Classical benchmark

In the supramolecular approach binding energies between two monomers in a dimer is most often expressed as

$$E_{\text{binding}} = E_{AB} - E_A - E_B . \quad (3.1)$$

In Eq. (3.1) E_{AB} , E_A , and E_B denote the ground-state energies of the dimer AB , monomer A , and monomer B , respectively. For calculations utilizing active spaces the highest accuracy obtainable with the supramolecular approach can be achieved if Eq. (3.1) is instead expressed in terms of the energy of bound and unbound dimers ($E_{AB\text{-bound}}$ and $E_{AB\text{-unbound}}$). Better accuracy is achieved within this approximation due to the fact that it allows for a consistent active space in all of the calculations. Hence, in all of our calculations we express the binding energy as

$$E_{\text{binding}} = E_{AB\text{-bound}} - E_{AB\text{-unbound}} . \quad (3.2)$$

Here, the $E_{AB\text{-unbound}}$ term of Eq. (3.2) is approximated as two monomers separated by a 48.000 Å distance, where the chosen distance guarantees the absence of interactions between the monomers.

Table 3.1 Active spaces used in the present work.

species	active space	AOs	Figure
water dimer	(16e,12o)	O[2s,2p], H[1s]	3.1a
methane dimer	(16e,16o)	C[2s,2p], H[1s]	3.1b
methane dimer	(16e,24o)	C[2s,2p,3s,3p], H[1s,2s]	3.1c

Metz et al.⁹¹ and Li et al.⁹² demonstrated that CCSD(T)/aug-cc-pVQZ calculations closely reproduce the results of the CCSD(T)/complete basis set (CBS) limit for the methane dimer. Metz et al.⁹¹ also demonstrated this for water dimer. All of our simulations are therefore done with the aug-cc-pVQZ basis set. We simulate the water and methane dimers with the active spaces listed in Table I.

We construct these active spaces using the atomic valence active space (AVAS) method⁹³ as implemented in the PySCF 2.6.2 software package^{94–96}, and select active-space orbitals that overlap

with the atomic orbitals (AOs) listed in column 3 of the table. The active-space orbitals of the water and methane dimers are shown in Fig. 3.1. Orbital visualization is performed with Pegamoid.⁹⁷

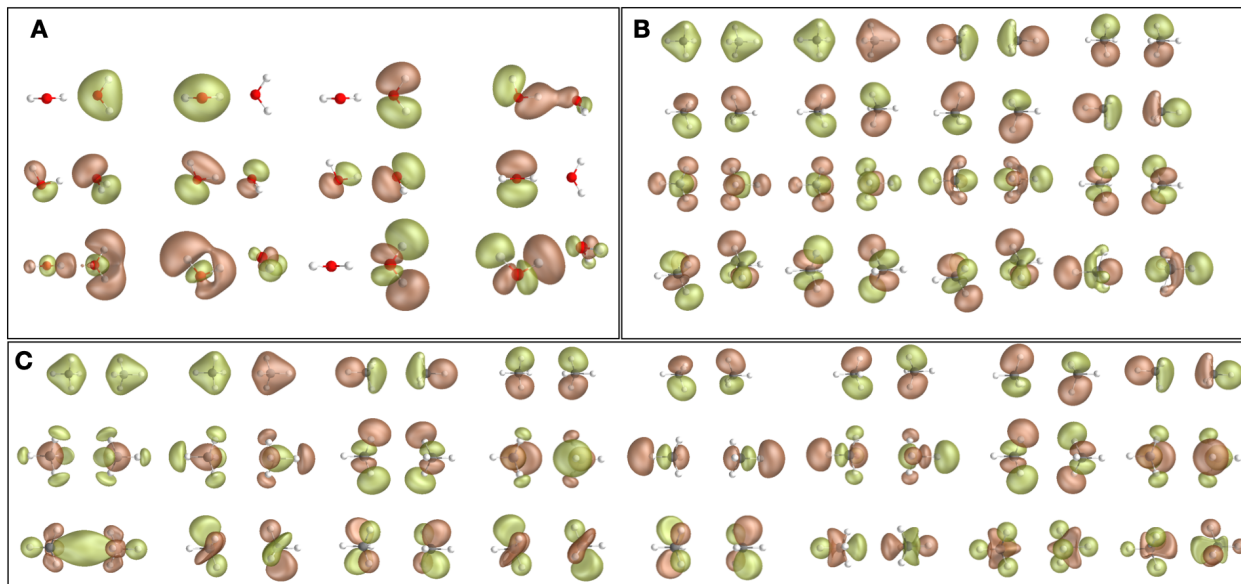


Figure 3.1 Active spaces used in this work: (A) (16e,12o) of the water dimer, (B) (16e,16o) of the methane dimer, (C) (16e,24o) of the methane dimer.

In each active space, we perform CCSD and CCSD(T) calculations with PySCF 2.6.2. For water and methane dimers we also perform CASCI(16e,12o) and CASCI(16e,16o) simulations, respectively, using PySCF 2.6.2. For the (16e,24o) active space of the methane dimer we perform HCI calculations with the SHCI-SCF 0.1 interface between PySCF 2.6.2 and DICE 1.0^{85,87,88}. Further details of HCI calculations can be found in the Supplementary Information. Along with active-space simulations, we perform complete CCSD and CCSD(T) calculations with ORCA 5.0.4⁹⁸. The geometries of equilibrium structures of the water and methane dimer originate from works by Temelso et al⁹⁹ and by Rezac and Hobza¹⁰⁰, sourced through the BEGDB database¹⁰¹. We describe the generation of PES geometries for water and methane dimers in the Supplementary Information.

3.2.2 Quantum Computing

We start from the active-space Hamiltonian, written in second quantization as

$$\hat{H} = E_0 + \sum_{\substack{pr \\ \sigma}} h_{pr} \hat{a}_{p\sigma}^\dagger \hat{a}_{r\sigma} + \sum_{\substack{prqs \\ \sigma\tau}} \frac{(pr|qs)}{2} \hat{a}_{p\sigma}^\dagger \hat{a}_{q\tau}^\dagger \hat{a}_{s\tau} \hat{a}_{r\sigma}, \quad (3.3)$$

where $\hat{a}^\dagger$ ($\hat{a}$) are creation (annihilation) operators, p, r, s , and $q = 1 \dots M$ denote basis set element, σ and τ denote spin- z polarizations, h_{pr} and $(pr|qs)$ are the one- and two-body electronic integrals, and E_0 is a constant accounting for the electrostatic interactions between nuclei and electrons in occupied inactive orbitals. We obtain the quantities E_0 , h_{pr} , and $(pr|qs)$ for the selected active spaces using PySCF.

We prepare our wavefunction guesses $|\Psi\rangle$, used to approximate the ground state of Eq. (3.3), from a truncated version of the local unitary cluster Jastrow (LUCJ) ansatz¹⁰²

$$|\Psi\rangle = \prod_{\mu=0}^{L-1} e^{\hat{K}_\mu} e^{i\hat{J}_\mu} e^{-\hat{K}_\mu} |\mathbf{x}_{\text{RHF}}\rangle, \quad (3.4)$$

where $\hat{K}_\mu = \sum_{pr,\sigma} K_{pr}^\mu \hat{a}_{p\sigma}^\dagger \hat{a}_{r\sigma}$ are one-body operators, $\hat{J}_\mu = \sum_{pr,\sigma\tau} J_{p\sigma,r\tau}^\mu \hat{n}_{p\sigma} \hat{n}_{r\tau}$ are suitable (vide infra) density-density operators, and $|\mathbf{x}_{\text{RHF}}\rangle$ is the restricted closed-shell Hartree-Fock (RHF) state. We use the Jordan-Wigner (JW) transformation¹⁰³ to map the fermionic wavefunction Eq. (3.4) onto a qubit wavefunction that can be prepared executing a quantum circuit. The JW transformation maps the Fock space of fermions in M spatial orbitals onto the Hilbert space of $2M$ qubits, where the basis state $|\mathbf{x}\rangle$ is parametrized by a bitstring $\mathbf{x} \in \{0, 1\}^{2M}$ and represents an electronic configuration where the spin-orbital $p\sigma$ is occupied (empty) if $x_{p\sigma} = 1$ ($x_{p\sigma} = 0$). We prepare the wavefunction Eq. (3.4) by executing the following quantum circuit: a single layer of Pauli-X gates prepares the basis state $|\mathbf{x}_{\text{RHF}}\rangle$, a Bogoliubov circuit¹⁰⁴ (with linear depth, quadratic number of gates, and a 1D qubit connectivity) encodes each orbital rotation $e^{\pm\hat{K}_\mu}$, and a circuit of Pauli-ZZ rotations encodes each density-density interaction $e^{i\hat{J}_\mu}$. When J^μ is a dense matrix, Pauli-ZZ rotations are applied across all pair of qubits, requiring all-to-all qubit connectivity or a substantial overhead of swap gates. To mitigate these quantum hardware requirements LUCJ imposes a “locality” approximation, i.e., it assumes $J_{p\sigma,r\tau}^\mu = 0$ for all pairs of spin-orbitals that are not mapped onto adjacent qubits under JW¹⁰² (as a consequence, a circuit with constant depth and linear number of gates encodes each $e^{i\hat{J}_\mu}$ operator). Hence, the number of layers $(L - 1)$ in Eq. (3.4) is formally equal to 1.5. As the result the specific form of $|\Psi\rangle$ used in this work is expressed as $|\Psi\rangle = e^{-\hat{K}_2} e^{\hat{K}_1} e^{i\hat{J}_1} e^{-\hat{K}_1} |\mathbf{x}_{\text{RHF}}\rangle$. We parametrize the LUCJ circuit based on amplitudes computed from classical restricted closed-shell

CCSD within the given active space³⁹, yet a further quantum-classical parameter optimization could further improve the quality of the ground-state approximation. We produce the LUCJ circuits using the `ffsim` library¹⁰⁵ interfaced with Qiskit 1.1.1^{104,106}.

The qubit layouts of the LUCJ circuits used for (16e,12o) water dimer, (16e,16o) methane dimer, and (16e,24o) methane dimer simulations are shown in Fig. 3.2a, 3.2b, and 3.2c, respectively. We execute these circuits on IBM’s 127-qubit Eagle devices `ibm_cleveland` and `ibm_kyiv`. In all our quantum computing experiments, we used gate (not measurement) twirling over random 2-qubit Clifford gates¹⁰⁷ and dynamical decoupling^{108–111} – available through the `SamplerV2` primitive of Qiskit’s runtime library – to mitigate quantum errors.

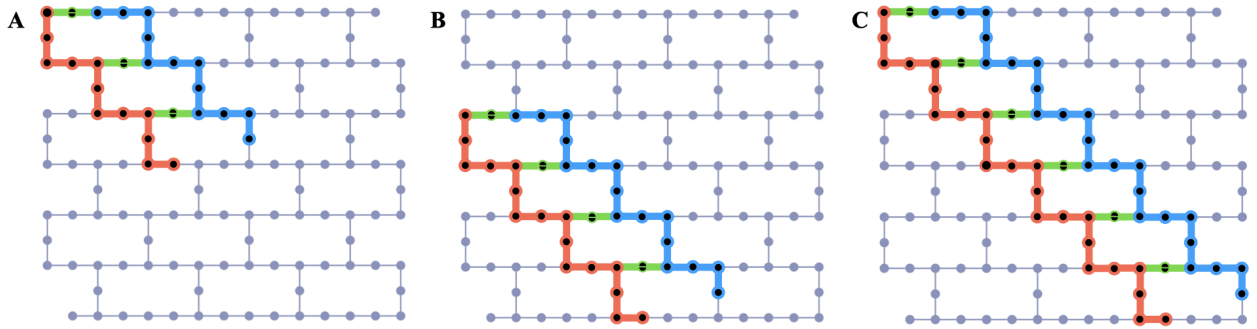


Figure 3.2 Qubit layouts of LUCJ circuits executed in this work: (A) (16e,12o) water dimer simulations using 27 qubits on `ibm_cleveland`, (B) (16e,16o) methane dimer simulations using 36 qubits on `ibm_cleveland`, and (C) (16e,24o) methane dimer simulations using 54 qubits on `ibm_kyiv`. Qubits used to encode occupation numbers of α (β) spin-orbitals are shown in red (blue). Auxiliary qubits used to execute density-density interactions between α and β spin-orbitals are marked in green.

Upon executing the LUCJ circuits, we measure $|\Psi\rangle$ in the computational basis. Repeating this produces a set of measurement outcomes (or “shots”)

$$\tilde{\chi} = \{\mathbf{x} | \mathbf{x} \sim \tilde{p}(\mathbf{x})\} \quad (3.5)$$

in the form of bitstrings $\mathbf{x} \in \{0, 1\}^{2M}$, each representing an electronic configuration (Slater determinants) distributed according to $\tilde{p}(\mathbf{x})$. While on a noiseless device configurations are distributed according to $|\langle \mathbf{x} | \Psi \rangle|^2$, on a noisy device they follow a distribution $\tilde{p}(\mathbf{x}) \neq |\langle \mathbf{x} | \Psi \rangle|^2$. In particular, $\tilde{p}(\mathbf{x})$ breaks particle-number conservation and returns configurations with incorrect particle

Table 3.2 Details of SQD calculations.

species	active space	$ \tilde{\chi} [10^3]$	K	$ \tilde{\chi}_b [10^3]$	d	CPUs, code	steps
water dimer	(16e,12o)	200	10	10	$24.5 \cdot 10^4$	10, PySCF	10
methane dimer	(16e,16o)	200	10	20	$12.6 \cdot 10^7$	10, PySCF	10
methane dimer	(16e,24o)	300	4	8.5	$24.9 \cdot 10^7$	16, DICE	5

number. We use a technique called self-consistent configuration recovery³⁹, executed on a classical computer, to restore particle-number conservation. The associated code is publicly available in the GitHub repository.¹¹² Within each step of self-consistent recovery, we sample K subsets (or batches) of $\tilde{\chi}$ labeled $\tilde{\chi}_b$ with $b = 1 \dots K$. Each batch defines – through a transformation³⁹ informed by an approximation to the ground-state occupation numbers $n_{p\sigma}$ – a subspace $S^{(b)}$ of dimension d , in which we project the many-electron Hamiltonian as^{39,40,113}

$$\hat{H}_{S^{(b)}} = \hat{P}_{S^{(b)}} \hat{H} \hat{P}_{S^{(b)}} , \quad (3.6)$$

where the projector $\hat{P}_{S^{(b)}}$ is

$$\hat{P}_{S^{(b)}} = \sum_{\mathbf{x} \in S^{(b)}} |\mathbf{x}\rangle \langle \mathbf{x}| . \quad (3.7)$$

We compute the ground states and energies of the Hamiltonians in Eq. (3.6), $|\psi^{(b)}\rangle$ and $E^{(b)}$ respectively, and use the lowest energy across the batches, $\min_b E^{(b)}$, as the best approximation to the ground-state energy at the current iteration of the configuration recovery. We use the ground states $|\psi^{(b)}\rangle$ to obtain an updated set of occupation numbers,

$$n_{p\sigma} = \frac{1}{K} \sum_{1 \leq b \leq K} \langle \psi^{(b)} | \hat{n}_{p\sigma} | \psi^{(b)} \rangle , \quad (3.8)$$

that we use in the next iteration of configuration recovery to produce the subspaces $S^{(b)}$. We repeat the iterations of self-consistent configuration recovery until convergence of the energy $\min_b E^{(b)}$. At the first iteration of self-consistent configuration recovery, we initialize $n_{p\sigma}$ from the measurement outcomes in $\tilde{\chi}$ with the correct particle number. We summarize the details of our SQD calculations in Table II.

We demonstrate that for SQD (16e,16o) simulations of the methane dimer at 3.638 Å a $|\tilde{\chi}_b| = 20.0 \cdot 10^3$ is necessary to reach agreement within 0.010 kcal/mol when compared against CASCI

(16e,16o). We show how the predicted total energies in these simulations improve with an increase of $|\tilde{\chi}_b|$ from $5.0 \cdot 10^3$ to $20.0 \cdot 10^3$ in Figure S3. We also demonstrate that in SQD(16e,16o) simulations the linear energy-variance relation allows for utilization of energy extrapolation which reproduces similar binding energies as simulation with $|\tilde{\chi}_b| = 20.0 \cdot 10^3$ while using substantially lower values of $|\tilde{\chi}_b|$. The extrapolation is done for the total energy of the dimer as the function of the Hamiltonian variance divided by the square of the variational energy, where Hamiltonian variance (ΔH) is calculated as $\Delta H = \langle \psi^{(k)} | \hat{H}^2 | \psi^{(k)} \rangle - \langle \psi^{(k)} | \hat{H} | \psi^{(k)} \rangle^2$. The extrapolation is done based on three points with $|\tilde{\chi}_b|$ of $9.0 \cdot 10^3$, $11.0 \cdot 10^3$, and $14.0 \cdot 10^3$, which allows for the reduction of the maximum $|\tilde{\chi}_b|$ by $6.0 \cdot 10^3$. This choice of values for $|\tilde{\chi}_b|$ allows for an even distribution of ΔH values used in extrapolation. The extrapolated energies are compared against CASCI(16e,16o) simulations and SQD(16e,16o) simulations with $|\tilde{\chi}_b| = 20.0 \cdot 10^3$.

We compute the ground-state eigenpairs of the Hamiltonians Eq. (3.6) using the iterative Davidson method on 10 CPUs with PySCF's selected configuration interaction (SCI) solver for SQD (16e,12o) simulations of the water dimer and SQD (16e,16o) simulations of the methane dimer. We achieve parallelization across 10 CPUs with Ray 2.33.0¹⁴ where the eigenstate solver within each of the 10 batches is using 1 CPU. For SQD (16e,24o) simulations of the methane dimer, we utilize the SCI solver of DICE and 16 CPUs, where the eigenstate solver within each of the 4 batches is using 4 CPUs. Further parallelization is possible with the SCI solver of DICE, as was demonstrated previously³⁹. The SQD (16e,12o) simulations of the water dimer and SQD (16e,16o) simulations of the methane dimer are done for the distances that are described in the Supplemental Information while SQD (16e,24o) simulations of the methane dimer are only done for 3.638 Å.

3.3 Results

Figure 3.3 shows the binding energy of the water dimer as a function of the oxygen-oxygen distance using SQD and CASCI. The SQD and CASCI potential energy surfaces closely align, deviating from each by less than 0.001 kcal/mol. This close alignment is an indication that both methods have accurately solved the Schrodinger equation in the active space. The active-space SQD and CASCI calculations cannot capture dynamical correlation from inactive orbitals. To

quantify the extent of the active-space approximation, we also compute the potential energy surface using CCSD and CCSD(T) in the full aug-cc-pVQZ basis. The perturbative triples do not have a drastic effect on the binding energy between water monomers and the close agreement between CCSD and CCSD(T) calculations is shown in Figure 3b. The excellent agreement between CCSD and CCSD(T) in the full basis and between SQD and CASCI in the active space indicates that the differences between SQD and CCSD(T) are due to the active-space approximation underlying the former. The CCSD(T) and SQD potential energy curves are in reasonable agreement with each other, the highest deviation being observed at 1.400 Å and corresponding to 2.263 kcal/mol. Despite this reasonable agreement and the ability of SQD to capture hydrogen bonding, there are quantitative differences in the predicted binding energies, -5.129 kcal/mol and -4.366 kcal/mol for CCSD(T) and SQD respectively, and the lowest-energy distances, 1.962 Å and 2.000 Å CCSD(T) and SQD respectively. The quantitative differences between SQD and CCSD and CCSD(T) shown in Fig. 3 are a consequence of SQD not being carried out in the full basis set.

Figure 3.4 shows the binding energy of the methane dimer – with (16e,16o) active space for the dimer and monomer respectively – as a function of the carbon-carbon distance using SQD and HCI. Figure 3.4 focuses on the attractive region, whereas the full curve is shown in Figure

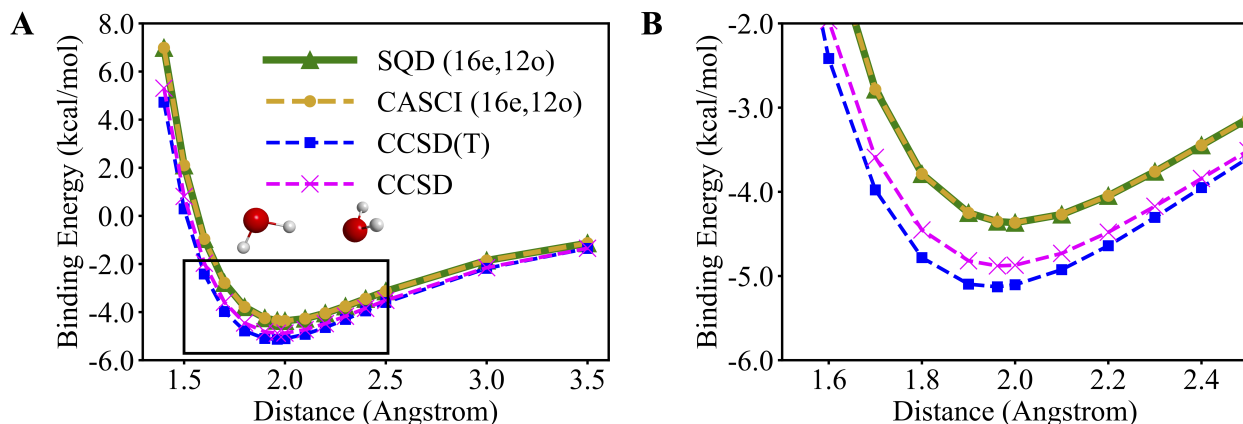


Figure 3.3 Binding energies of the water dimer along the PES, where distances between oxygen atoms range between 1.400 and 3.500 Å. (A) the entire range of bondlengths, and (B) a magnified region near equilibrium, highlighted in panel (A) as a black box. Light brown, blue, and magenta dashed lines with circle, square, and cross markers depict the PES calculated with the CASCI (16e,12o), CCSD(T), and CCSD methods, respectively. The solid green line with triangular markers depicts the PES calculated with the SQD (16e,12o).

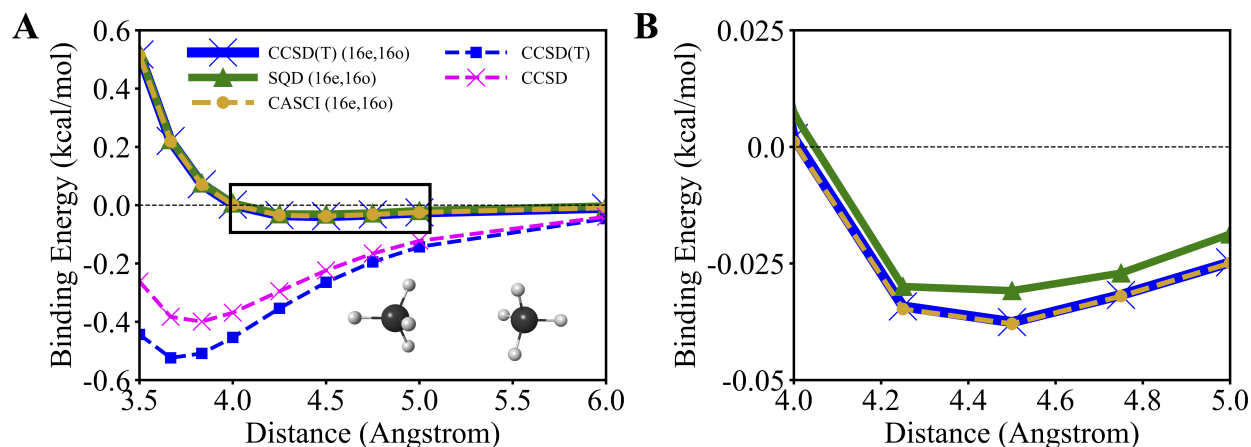


Figure 3.4 Binding energies of the methane dimer along the PES, where the distances between the carbon atoms range between 3.500 and 6.000 Å. Active-space simulations use (16e,16o). (A) the entire range of bondlengths, and (B) a magnified region near equilibrium, highlighted in panel (A) as a black box. Light brown, blue, and magenta dashed lines with circle, square, and cross markers depict the PES calculated with the CASCI (16e,16o), CCSD(T), and CCSD methods, respectively. The solid green line with triangular markers depicts the PES calculated with the SQD (16e,16o). The solid blue line represents CCSD(T) (16e,16o) calculations. The black horizontal dashed line indicates the zero value of the binding energy.

S5 of the SI. The SQD (16e,16o) and CASCI (16e,16o) data are closely aligned, with deviations below 0.005 kcal/mol. We interpret the excellent agreement between SQD (16e,16o) and CASCI (16e,16o) as an indication that the active-space Schrodinger equation is solved accurately. SQD (16e,16o) predicts the interaction between the monomers to be only marginally attractive, with a binding energy of -0.038 kcal/mol and a lowest-energy distance around 4.500 Å. On the other hand, full-basis CCSD and CCSD(T) calculations predict binding energies of -0.399 kcal/mol and -0.524 kcal/mol, respectively, at distances 3.834 Å and 3.667 Å, respectively. Despite some differences quantifying the importance of perturbative triple corrections, both full-basis calculations predict a substantially more pronounced tendency to binding than SQD (16e,16o) and CASCI (16e,16o). This is because, although SQD (16e,16o) and CASCI (16e,16o) calculations can accurately capture the active-space electronic correlation, they cannot account for the residual dynamical electron correlation, unlike full-basis CCSD and CCSD(T).

Before proceeding with the expansion of the active space we first demonstrate that accurate SQD (16e,16o) calculations can be achieved with a reduced number of samples through the extrapolation

of the total energies. The exact SQD (16e,16o) calculations require $|\tilde{\chi}_b| = 20.0 \cdot 10^3$ while the extrapolation is done based on three points with $|\tilde{\chi}_b|$ of $9.0 \cdot 10^3$, $11.0 \cdot 10^3$, and $14.0 \cdot 10^3$. Hence, the extrapolation allows for the reduction of the maximum required $|\tilde{\chi}_b|$ by $6.0 \cdot 10^3$. We show the SQD (16e,16o) total energy extrapolations for 4.000, 4.250, 4.500, 4.750, 5.000, and 6.000 Å distances in Figure 5a-f, while the extrapolation for the 48.000 Å distance is shown in Figure

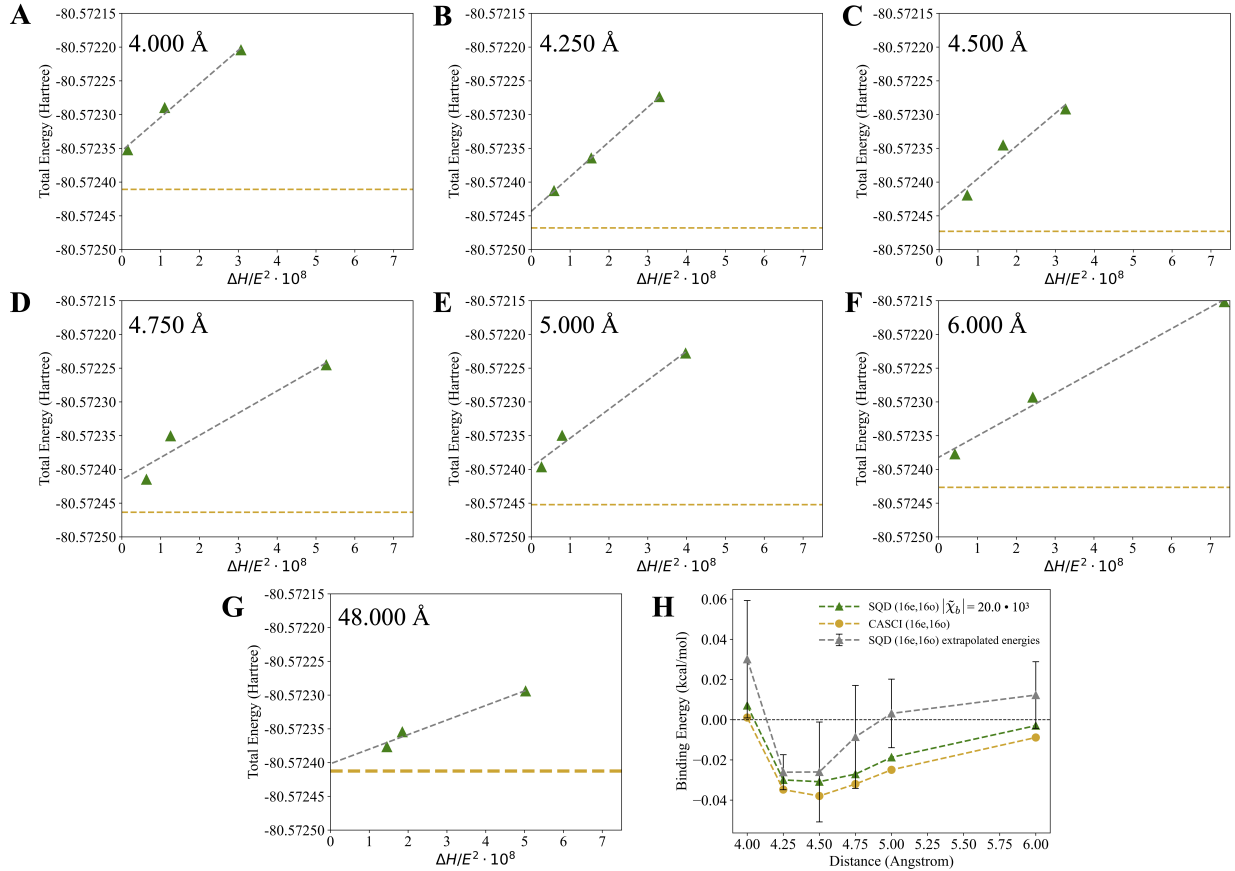


Figure 3.5 Extrapolated SQD (16e,16o) energies of the methane dimer along the PES, for the 4.000, 4.250, 4.500, 4.750, 5.000, and 6.000 Å distances between the carbon atoms. Extrapolations are done using three points with $|\tilde{\chi}_b|$ of $9.0 \cdot 10^3$, $11.0 \cdot 10^3$, and $14.0 \cdot 10^3$. Hamiltonian variance (ΔH) is calculated as $\Delta H = \langle \psi^{(k)} | \hat{H}^2 | \psi^{(k)} \rangle - \langle \psi^{(k)} | \hat{H} | \psi^{(k)} \rangle^2$. (A) - (F) total energy extrapolations for methane dimer 4.000, 4.250, 4.500, 4.750, 5.000, and 6.000 Å distances, (G) total energy extrapolations at 48.000 Å distance, and (H) binding energy in methane dimer calculated with extrapolated SQD (16e,16o) total energies compared against CASCI (16e,16o) simulations and SQD (16e,16o) simulations with $|\tilde{\chi}_b| = 20.0 \cdot 10^3$. Green triangles and green dashed lines indicate SQD (16e,16o) energies calculated with $|\tilde{\chi}_b| = 20.0 \cdot 10^3$. Grey triangles and grey dashed lines indicate extrapolated SQD (16e,16o) energies. Light brown circles and dashed lines indicate CASCI (16e,16o) energies. Black horizontal dashed line indicates the zero value of the binding energy. Error bars indicate magnitude of error estimate in extrapolation.

5g. The resulting binding energies of the methane dimer are shown in Figure 5h and compared against the CASCI (16e,16o) simulations and SQD (16e,16o) simulations with $|\tilde{\chi}_b| = 20.0 \cdot 10^3$. Figure 5g shows that the extrapolated SQD (16e,16o) energies predict a binding energy in good qualitative agreement with exact SQD (16e,16o) simulations and CASCI (16e,16o). This result is promising for future simulations with large active spaces, where classical post-processing of SQD data becomes computationally expensive.

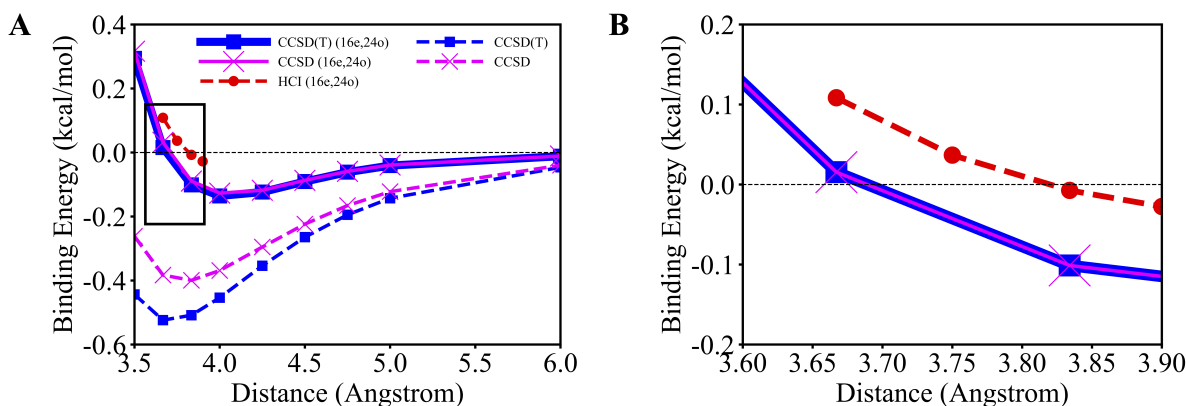


Figure 3.6 Binding energies of the methane dimer along the PES, where the distances between the carbon atoms range between 3.500 and 6.000 Å. Active-space simulations use (16e,24o) and are performed over (A) the entire range of bondlengths, and (B) a magnified region near equilibrium, highlighted in panel (A) as a black box. Blue and magenta dashed lines depict PES calculated with CCSD(T) and CCSD methods, respectively. The dashed red line depicts the HCI (16e,24o) results. The solid blue and magenta lines represents CCSD(T) (16e,24o) and CCSD (16e,24o) calculations, respectively. Black horizontal dashed line indicates the zero value of the binding energy.

Next we analyze the effect of extending the active space on the predicted binding energy via the inclusion of virtual orbitals with carbon 3s and 3p character. First, in Fig. 3.6, we explore the performance of HCI in this extended (16e,24o) active space. Here, HCI is used in place of CASCI due to the fact that the (16e,24o) active space is prohibitively expensive in conventional CASCI simulations. Figure 3.6a shows active-space calculations with HCI, CCSD, and CCSD(T). The CCSD(T) (16e,24o) curve, in good agreement with CCSD (16e,24o), is substantially more attractive than in the (16e,16o) active space, predicting a binding energy of -0.136 kcal/mol at 4.000 Å. The size of the (16e,24o) active space prevents us from significantly lowering the parameter ε_1 which results in the underestimation of the total energy in HCI (16e,24o) calculations. In

particular, at 3.834 Å distance the HCI (16e,24o) calculations underestimate the binding energy by 0.094 kcal/mol comparing to CCSD(T) (16e,24o), as visible in Fig. 3.6b. Note that HCI (16e,24o) calculations were carried out over four distances (3.667, 3.750, 3.834, and 3.900 Å). Within the target accuracy of the present work the prediction of HCI (16e,24o) binding energies for other geometries around the CCSD(T) (16e,24o) minimum is dramatically more computationally expensive.

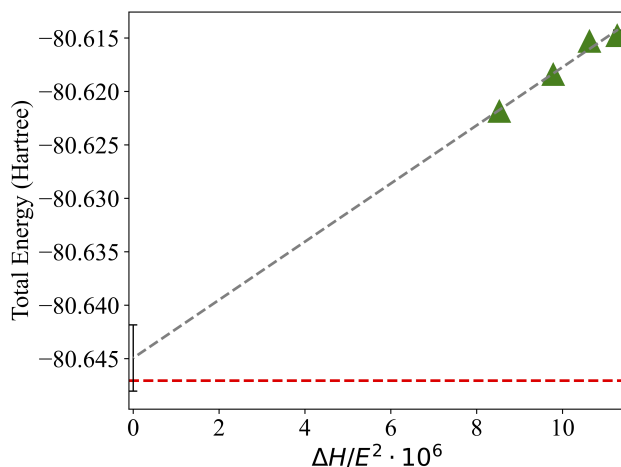


Figure 3.7 Extrapolation of SQD (16e,24o) total energies for the methane dimer at 3.638 Å distance between the carbon atoms. Extrapolations are done using four points with $|\tilde{\chi}_b|$ of $5.5 \cdot 10^3$, $6.5 \cdot 10^3$, $7.5 \cdot 10^3$, and $8.5 \cdot 10^3$. Hamiltonian variance (ΔH) is calculated as $\Delta H = \langle \psi^{(k)} | \hat{H}^2 | \psi^{(k)} \rangle - \langle \psi^{(k)} | \hat{H} | \psi^{(k)} \rangle^2$. Green triangles indicate SQD (16e,24o) energies. Grey dashed lines indicate extrapolated SQD (16e,24o) energies. Dashed red line indicates HCI (16e,24o) energies. Error bars indicate magnitude of error estimate in extrapolation.

We show the decrease in the SQD (16e,24o) total energy for the methane dimer at 3.638 Å with the increase of $|\tilde{\chi}_b|$ from $5.5 \cdot 10^3$ to $8.5 \cdot 10^3$ in Fig. 3.7. The differences between the total energies predicted with SQD (16e,24o) and HCI (16e,24o) reduce from 31.8 milliHartree to 25.2 milliHartree when the $|\tilde{\chi}_s|$ is increased from $5.5 \cdot 10^3$ to $8.5 \cdot 10^3$. The extrapolated total energy based on SQD (16e,24o) simulations with $|\tilde{\chi}_b|$ of $5.5 \cdot 10^3$, $6.5 \cdot 10^3$, $7.5 \cdot 10^3$, and $8.5 \cdot 10^3$ agrees with HCI (16e,24o) results within 2.12 milliHartree. Magnitude of the error estimate in extrapolation is ± 3.10 milliHartree. We believe that a further increase in the number of samples will allow us to advance the accuracy of SQD (16e,24o) calculations. To make SQD (16e,24o) calculations of the methane dimer more computationally feasible we are currently exploring parallelization options

for calculations on this system as well as the analysis of the configurations with low contributions to the total energies.

3.4 Conclusions and Outlook

We have presented quantum-centric simulations of the water and methane dimers using a sample-based quantum diagonalization method on IBM’s Eagle quantum processors. This demonstration is a first simulation of noncovalent supramolecular interactions on quantum processors. The accuracy of SQD and HCI predictions of noncovalent interactions can be systematically improved by the addition of extended shells of virtual orbitals. We anticipate that further expansion of the active space through the inclusion of the virtual orbitals corresponding to the 3d shell of the heavy atoms will allow for an even more accurate description of non-covalent interactions with SQD and HCI, which will be the subject of future studies on quantum processors. Importantly, the present study lays out a framework for electronic structure calculations of noncovalent interactions on quantum hardware.

Our findings demonstrate that SQD is capable of capturing noncovalent interactions between molecules at the level of theory chosen, with potential energy surfaces that closely align with those obtained through classical computational methods. We examine the binding energies of the water and methane dimers by comparing SQD with an analogous classical method, namely HCI. We also compare SQD against the CCSD(T) method, which is considered the gold standard for calculations of binding energies⁹⁰. This comparison aims to evaluate the accuracy of SQD and to understand how the nature of the PES changes with different active-space selections. The ability of HCI and SQD to recover the dispersion interaction is highly dependent on the size and nature of the active space, which is especially critical for predicting the binding energy of the methane dimer. In fact, a previous study by Hapka et al.¹¹⁵ demonstrated that the ability of supramolecular multiconfigurational interaction calculations to recover the dispersion energy depends on the size of the active space and can be improved with systematic expansion of the active space.

The results obtained here demonstrate the improvements both in terms of accuracy and scale of quantum computations on chemical problems, enabling, on current quantum processors, use

cases previously thought to belong to the fault-tolerant domain, such as the largest active space considered here for methane, which has 1.3M Pauli operators. Further examples of problems that could be enabled by our approach include quantum computing simulations of chemical reactivity of CO_2 -fixating ruthenium catalyst proposed by Burg et al.¹¹⁶, Ibrutinib drug simulations proposed by Blunt et al.¹¹⁷, and the drug-discovery workflows proposed by Pyrkov et al.¹¹⁸ and Kumar et al.¹¹⁹, as well as multiple stages of drug optimization as described by Bonde et al.¹²⁰ The SQD method allows for simulations of systems with qubit counts that is essential for projection-based embedding algorithm proposed by Ralli et al.¹²¹ and can enhance the viability of fragment-based quantum computing simulations. Previously, VQE-based fragment molecular orbital (FMO),¹²² divide and conquer (DC),¹²³ and density matrix embedding theory (DMET)^{65,124} simulations were limited to very simple illustrative systems. Shang et al. proposed a DMET-based massively parallel quantum computing approach based on VQE, but execution of their methodology was only possible on a quantum simulator rather than actual hardware.¹²⁵

Such limitations in fragment-based VQE simulations are due to the fact that the number of orbitals that could be described with reasonable accuracy on actual hardware in the VQE formalism within each fragment is very limited. Fragment-based simulations with SQD would allow for substantially higher number of orbitals in each individual fragment, making the quantum computing simulations of proteins and drug molecules possible.

In conclusion, combining quantum and classical computational resources in workflows like SQD opens the way for the use of current and near-future quantum technology to tackle computational challenges in small-molecule conformational search, drug-protein interactions and drug discovery.

BIBLIOGRAPHY

- [1] K. Müller-Dethlefs and P. Hobza. Noncovalent interactions: a challenge for experiment and theory. *Chemical Reviews*, 100(1):143–168, 2000.
- [2] C. Puzzarini, L. Spada, S. Alessandrini, and V. Barone. The challenge of non-covalent interactions: Theory meets experiment for reconciling accuracy and interpretation. *Journal of Physics: Condensed Matter*, 32(34):343002, 2020.
- [3] W. J. Pichler. The important role of non-covalent drug-protein interactions in drug hypersensitivity reactions. *Allergy*, 77(2):404–415, 2022.
- [4] A. Aljoundi, I. Bjij, A. El Rashedy, and M. E. Soliman. Covalent versus non-covalent enzyme inhibition: which route should we take? a justification of the good and bad from molecular modelling perspective. *The Protein Journal*, 39(2):97–105, 2020.
- [5] H. Jane Dyson, Peter E. Wright, and Harold A. Scheraga. The role of hydrophobic interactions in initiation and propagation of protein folding. *Proceedings of the National Academy of Sciences*, 103(35):13057–13061, 2006.
- [6] Neal J. Zondlo. Fold globally, bond locally. *Nature Chemical Biology*, 6(8):567–568, 2010.
- [7] C. N. Pace, H. Fu, K. L. Fryar, J. Landua, S. R. Trevino, B. A. Shirley, M. M. Hendricks, S. Iimura, K. Gajiwala, J. M. Scholtz, and G. R. Grimsley. Contribution of hydrophobic interactions to protein stability. *J Mol Biol*, 408(3):514–28, 2011.
- [8] Carlo Camilloni, Daniela Bonetti, Angela Morrone, Rajanish Giri, Christopher M. Dobson, Maurizio Brunori, Stefano Gianni, and Michele Vendruscolo. Towards a structural biology of the hydrophobic effect in protein folding. *Scientific Reports*, 6(1):28285, 2016.
- [9] S. R. Durell and A. Ben-Naim. Hydrophobic-hydrophilic forces in protein folding. *Biopolymers*, 107(8), 2017.
- [10] R. G. Hanshaw, R. V. Stahelin, and B. D. Smith. Noncovalent keystone interactions controlling biomembrane structure. *Chemistry*, 14(6):1690–7, 2008.
- [11] Vishal Annasaheb Adhav and Kayarat Saikrishnan. The realm of unconventional noncovalent interactions in proteins: Their significance in structure and function. *ACS Omega*, 8(25):22268–22284, 2023.
- [12] Harren Jhoti and Andrew R. Leach. *Structure-Based Drug Discovery*. Springer Netherlands, 2007.
- [13] Edwige Gros, Sebastien Deshayes, May C. Morris, Gudrun Aldrian-Herrada, Julien Depollier, Frederic Heitz, and Gilles Divita. A non-covalent peptide-based strategy for protein and

- peptide nucleic acid transduction. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1758(3):384–393, 2006.
- [14] Y. Wei, L. Ma, L. Zhang, and X. Xu. Noncovalent interaction-assisted drug delivery system with highly efficient uptake and release of paclitaxel for anticancer therapy. *Int J Nanomedicine*, 12:7039–7051, 2017.
 - [15] Y. Yang, H. Zhang, Y. Wanyan, K. Liu, T. Lv, M. Li, and Y. Chen. Effect of hydrophobicity on the anticancer activity of fatty-acyl-conjugated cm4 in breast cancer cells. *ACS Omega*, 5(34):21513–21523, 2020.
 - [16] K. M. Wright, S. R. DiNapoli, M. S. Miller, P. Aitana Azurmendi, X. Zhao, Z. Yu, M. Chakrabarti, W. Shi, J. Douglass, M. S. Hwang, E. H. Hsiue, B. J. Mog, A. H. Pearlman, S. Paul, M. F. Konig, D. M. Pardoll, C. Bettgowda, N. Papadopoulos, K. W. Kinzler, B. Vogelstein, S. Zhou, and S. B. Gabelli. Hydrophobic interactions dominate the recognition of a kras g12v neoantigen. *Nat Commun*, 14(1):5063, 2023.
 - [17] Kevin E. Riley and Pavel Hobza. Noncovalent interactions in biochemistry. *WIREs Computational Molecular Science*, 1(1):3–17, 2011.
 - [18] Gino A. DiLabio and Alberto Otero de-la Roza. Noncovalent interactions in density-functional theory. *arXiv preprint arXiv:1405.1771*, 2014.
 - [19] Lori A. Burns, Álvaro Vázquez Mayagoitia, Bobby G. Sumpter, and C. David Sherrill. Density-functional approaches to noncovalent interactions: A comparison of dispersion corrections (DFT-D), exchange-hole dipole moment (XDM) theory, and specialized functionals. *The Journal of Chemical Physics*, 134(8):084107, 02 2011.
 - [20] Hong Gao, Satoshi Imamura, Akihiko Kasagi, and Eiji Yoshida. Distributed implementation of full configuration interaction for one trillion determinants. *Journal of Chemical Theory and Computation*, 20(3):1185–1192, 2024. PMID: 38314701.
 - [21] Konstantinos D. Vogiatzis, Dongxia Ma, Jeppe Olsen, Laura Gagliardi, and Wibe A. de Jong. Pushing configuration-interaction to the limit: Towards massively parallel MCSCF calculations. *The Journal of Chemical Physics*, 147(18):184111, 11 2017.
 - [22] Andor Menczer, Maarten van Damme, Alan Rask, Lee Huntington, Jeff Hammond, Sotiris S. Xantheas, Martin Ganahl, and Ors Legeza. Parallel implementation of the density matrix renormalization group method achieving a quarter petaflops performance on a single dgx-h100 gpu node. *Journal of Chemical Theory and Computation*, 0(0):null, 2024. PMID: 39297788.
 - [23] Xiaochuan Ren, Jingxiang Zou, Haodong Zhang, Wei Li, and Shuhua Li. Block-correlated coupled cluster theory with up to four-pair correlation for accurate static correlation of strongly correlated systems. *The Journal of Physical Chemistry Letters*, 15(3):693–700,

2024. PMID: 38207241.

- [24] Runfeng Jin, Wenhao Liang, Haoyuan Zhang, Yinxuan Song, Zhen Luo, Haibo Ma, Yingjin Ma, and Zhong Jin. Pasci : A scalable framework for heterogeneous parallel calculation of dynamical electron correlation. In *Proceedings of the 53rd International Conference on Parallel Processing, ICPP '24*, page 1103–1113, New York, NY, USA, 2024. Association for Computing Machinery.
- [25] David W. Schwenke. Introducing MPEC: Massively parallel electron correlation. *The Journal of Chemical Physics*, 158(8):084801, 02 2023.
- [26] Mickaël G. Delcey. Multipsi: A python-driven mscf program for photochemistry and spectroscopy simulations on modern hpc environments. *WIREs Computational Molecular Science*, 13(6):e1675, 2023.
- [27] Hector H. Corzo, Andreas Erbs Hillers-Bendtsen, Ashleigh Barnes, Abdulrahman Y. Zamani, Filip Pawłowski, Jeppe Olsen, Poul Jørgensen, Kurt V. Mikkelsen, and Dmytro Bykov. Coupled cluster theory on modern heterogeneous supercomputers. *Frontiers in Chemistry*, 11, 2023.
- [28] Dipayan Datta and Mark S. Gordon. Accelerating coupled-cluster calculations with gpus: An implementation of the density-fitted ccSD(t) approach for heterogeneous computing architectures using openmp directives. *Journal of Chemical Theory and Computation*, 19(21):7640–7657, 2023.
- [29] Huanchen Zhai and Garnet Kin-Lic Chan. Low communication high performance ab initio density matrix renormalization group algorithms. *The Journal of Chemical Physics*, 154(22):224116, 06 2021.
- [30] X. He, B. Walker, V. H. Man, P. Ren, and J. Wang. Recent progress in general force fields of small molecules. *Curr Opin Struct Biol*, 72:187–193, 2022.
- [31] Paul S Nerenberg and Teresa Head-Gordon. New developments in force fields for biomolecular simulations. *Current Opinion in Structural Biology*, 49:129–138, 2018.
- [32] J. W. Ponder and D. A. Case. Force fields for protein simulations. *Adv. Protein Chem.*, 66:27–85, 2003.
- [33] Jr. Mackerell, A. D. Empirical force fields for biological macromolecules: overview and issues. *J Comput Chem*, 25(13):1584–604, 2004.
- [34] W. L. Jorgensen. Monte carlo simulations for free energies of hydration: Past to present. *J Chem Phys*, 161(6), 2024.
- [35] Oliver T. Unke, Stefan Chmiela, Huziel E. Sauceda, Michael Gastegger, Igor Poltavsky,

- Kristof T. Schütt, Alexandre Tkatchenko, and Klaus-Robert Müller. Machine learning force fields. *Chemical Reviews*, 121(16):10142–10186, 2021.
- [36] J. S. Smith, O. Isayev, and A. E. Roitberg. Ani-1: an extensible neural network potential with dft accuracy at force field computational cost. *Chem. Sci.*, 8:3192–3203, 2017.
- [37] Mozafar Rezaee, Saeid Ekrami, and Seyed Majid Hashemianzadeh. Comparing ani-2x, ani-1ccx neural networks, force field, and dft methods for predicting conformational potential energy of organic molecules. *Scientific Reports*, 14(1):11791, 2024.
- [38] Y. Alexeev, M. Amsler, M. A. Barroca, S. Bassini, T. Battelle, D. Camps, D. Casanova, Y. J. Choi, F. T. Chong, and C. Chung. Quantum-centric supercomputing for materials science: A perspective on challenges and future directions. *Future Generation Computer Systems*, 160:666–710, 2024.
- [39] J. Robledo-Moreno, M. Motta, H. Haas, A. Javadi-Abhari, P. Jurcevic, W. Kirby, S. Martiel, K. Sharma, S. Sharma, and T. Shirakawa. Chemistry beyond exact solutions on a quantum-centric supercomputer. *arXiv preprint arXiv:2405.05068*, 2024.
- [40] K. Kanno, M. Kohda, R. Imai, S. Koh, K. Mitarai, W. Mizukami, and Y. O. Nakagawa. Quantum-selected configuration interaction: Classical diagonalization of hamiltonians in subspaces selected by quantum computers. *arXiv preprint arXiv:2302.11320*, 2023.
- [41] 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(7671):242–246, 2017.
- [42] Abhinav Kandala, Kristan Temme, Antonio D Córcoles, Antonio Mezzacapo, Jerry M Chow, and Jay M Gambetta. Error mitigation extends the computational reach of a noisy quantum processor. *Nature*, 567(7749):491–495, 2019.
- [43] Google AI Quantum and Collaborators. Hartree-fock on a superconducting qubit quantum computer. *Science*, 369(6507):1084–1089, 2020.
- [44] Qian Gao, Hiroki Nakamura, Tanmay P Gujarati, Gregory O Jones, Jeanne E Rice, Sebastian P Wood, Marco Pistoia, Juan M Garcia, and Naoki Yamamoto. Computational investigations of the lithium superoxide dimer rearrangement on noisy quantum devices. *The Journal of Physical Chemistry A*, 125(9):1827–1836, 2021.
- [45] Qian Gao, Gregory O Jones, Mario Motta, Makoto Sugawara, Hiroshi C Watanabe, Takahiro Kobayashi, Eisuke Watanabe, Yoshiyuki Ohnishi, Hiroki Nakamura, and Naoki Yamamoto. Applications of quantum computing for investigations of electronic transitions in phenylsulfonyl-carbazole tadf emitters. *npj Computational Materials*, 7(1):70, 2021.
- [46] S. Gocho, H. Nakamura, S. Kanno, Q. Gao, T. Kobayashi, T. Inagaki, and M. Hatanaka.

- Excited state calculations using variational quantum eigensolver with spin-restricted ansätze and automatically-adjusted constraints. *npj Computational Materials*, 9(1):13, 2023.
- [47] S. Shirai, H. Iwakiri, K. Kanno, T. Horiba, K. Omiya, H. Hirai, and S. Koh. Computational analysis of chemical reactions using a variational quantum eigensolver algorithm without specifying spin multiplicity. *ACS Omega*, 8(22):19917–19925, 2023.
 - [48] Y. Nam, J.-S. Chen, N. C. Pisenti, K. Wright, C. Delaney, D. Maslov, K. R. Brown, S. Allen, J. M. Amini, and J. Apisdorf. Ground-state energy estimation of the water molecule on a trapped-ion quantum computer. *npj Quantum Information*, 6(1):33, 2020.
 - [49] I. T. Khan, M. Tudorovskaya, J. J. M. Kirsopp, D. Muñoz Ramo, P. Warrier, D. K. Papanastasiou, and R. Singh. Chemically aware unitary coupled cluster with ab initio calculations on an ion trap quantum computer: A refrigerant chemicals’ application. *The Journal of Chemical Physics*, 158(21):214114, 2023.
 - [50] T. E. O’Brien, G. Anselmetti, F. Gkritis, V. E. Elfving, S. Polla, W. J. Huggins, O. Oumarou, K. Kechedzhi, D. Abanin, R. Acharya, et al. Purification-based quantum error mitigation of pair-correlated electron simulations. *Nature Physics*, 19(12):1787–1792, 2023.
 - [51] L. Zhao, J. Goings, K. Shin, W. Kyoung, J. I. Fuks, J.-K. Kevin Rhee, Y. M. Rhee, K. Wright, J. Nguyen, J. Kim, and S. Johri. Orbital-optimized pair-correlated electron simulations on trapped-ion quantum computers. *npj Quantum Information*, 9(1):60, 2023.
 - [52] 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(1):3007, 2019.
 - [53] G. Christopoulou, C. Di Paola, F. E. Elzinga, A. Jallat, D. M. Ramo, and M. Krompiec. Quantum hardware calculations of the activation and dissociation of nitrogen on iron clusters and surfaces. *Physical Chemistry Chemical Physics*, 26(7):5895–5906, 2024.
 - [54] A. Eddins, M. Motta, T. Gujarati, S. Bravyi, A. Mezzacapo, C. Hadfield, and S. Sheldon. Doubling the size of quantum simulators by entanglement forging. *PRX Quantum*, 3:010309, 2022.
 - [55] M. Motta, G. O. Jones, J. E. Rice, T. P. Gujarati, R. Sakuma, I. Liepuoniute, J. M. Garcia, and Y.-y. Ohnishi. Quantum chemistry simulation of ground-and excited-state properties of the sulfonium cation on a superconducting quantum processor. *Chemical Science*, 14(11):2915–2927, 2023.
 - [56] M. A. Castellanos, M. Motta, and J. E. Rice. Quantum computation of $\pi \rightarrow \pi^*$ and $n \rightarrow \pi^*$ excited states of aromatic heterocycles. *Molecular Physics*, 122(7-8):e2282736, 2024.
 - [57] I. Liepuoniute, M. Motta, T. Pellegrini, J. E. Rice, T. P. Gujarati, S. Gil, and G. O. Jones. Sim-

- ulation of a diels-alder reaction on a quantum computer. *arXiv preprint arXiv:2403.08107*, 2024.
- [58] T. Weaving, A. Ralli, P. J. Love, S. Succi, and P. V. Coveney. Contextual subspace variational quantum eigensolver calculation of the dissociation curve of molecular nitrogen on a superconducting quantum computer. *arXiv preprint arXiv:2312.04392*, 2024.
 - [59] J. I. Colless, V. V. Ramasesh, D. Dahlen, M. S. Blok, M. E. Kimchi-Schwartz, J. R. McClean, J. Carter, W. A. de Jong, and I. Siddiqi. Computation of molecular spectra on a quantum processor with an error-resilient algorithm. *Physical Review X*, 8(1):011021, 2018.
 - [60] Emil Dimitrov, Goar Sanchez-Sanz, James Nelson, Lee O’Riordan, Myles Doyle, Sean Courtney, Venkatesh Kannan, Hassan Naseri, Alberto Garcia Garcia, James Tricker, et al. Pushing the limits of quantum computing for simulating pfas chemistry. *arXiv preprint arXiv:2311.01242*, 2023.
 - [61] S. Guo, J. Sun, H. Qian, M. Gong, Y. Zhang, F. Chen, Y. Ye, Y. Wu, S. Cao, K. Liu, et al. Experimental quantum computational chemistry with optimized unitary coupled cluster ansatz. *Nature Physics*, 2024.
 - [62] C. Hempel, C. Maier, J. Romero, J. McClean, T. Monz, H. Shen, P. Jurcevic, B. P. Lanyon, P. Love, R. Babbush, et al. Quantum chemistry calculations on a trapped-ion quantum simulator. *Physical Review X*, 8(3):031022, 2018.
 - [63] K. Huang, X. Cai, H. Li, Z.-Y. Ge, R. Hou, H. Li, T. Liu, Y. Shi, C. Chen, D. Zheng, et al. Variational quantum computation of molecular linear response properties on a superconducting quantum processor. *The Journal of Physical Chemistry Letters*, 13(39):9114–9121, 2022.
 - [64] M. A. Jones, H. J. Vallury, and L. C. L. Hollenberg. Ground-state-energy calculation for the water molecule on a superconducting quantum processor. *Physical Review Applied*, 21(6):064017, 2024.
 - [65] Y. Kawashima, E. Lloyd, M. P. Coons, Y. Nam, S. Matsuura, A. J. Garza, S. Johri, L. Huntington, V. Senicourt, A. O. Maksymov, et al. Optimizing electronic structure simulations on a trapped-ion quantum computer using problem decomposition. *Communications Physics*, 4(1):245, 2021.
 - [66] J. J. M. Kirsopp, C. Di Paola, D. Z. Manrique, M. Krompiec, G. Greene-Diniz, W. Guba, A. Meyder, D. Wolf, M. Strahm, and D. Muñoz Ramo. Quantum computational quantification of protein–ligand interactions. *International Journal of Quantum Chemistry*, 122(22):e26975, 2022.
 - [67] V. Leyton-Ortega, S. Majumder, and R. C. Pooser. Quantum error mitigation by hidden inverses protocol in superconducting quantum devices. *Quantum Science and Technology*,

8(1):014008, 2023.

- [68] Z. Liang, J. Cheng, H. Ren, H. Wang, F. Hua, Z. Song, Y. Ding, F. T. Chong, S. Han, X. Qian, and Y. Shi. Napa: Intermediate-level variational native-pulse ansatz for variational quantum algorithms. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 43(6):1834–1847, 2024.
- [69] P. Liu, R. Wang, J.-N. Zhang, Y. Zhang, X. Cai, H. Xu, Z. Li, J. Han, X. Li, G. Xue, et al. Performing $su(d)$ operations and rudimentary algorithms in a superconducting transmon qudit for $d=3$ and $d=4$. *Physical Review X*, 13(2):021028, 2023.
- [70] P. Lolur, M. Skogh, W. Dobrutz, C. Warren, J. Biznárová, A. Osman, G. Tancredi, G. Wendin, J. Bylander, and M. Rahm. Reference-state error mitigation: A strategy for high accuracy quantum computation of chemistry. *Journal of Chemical Theory and Computation*, 19(3):783–789, 2023.
- [71] A. J. McCaskey, Z. P. Parks, J. Jakowski, S. V. Moore, T. D. Morris, T. S. Humble, and R. C. Pooser. Quantum chemistry as a benchmark for near-term quantum computers. *npj Quantum Information*, 5(1):99, 2019.
- [72] J. E. Rice, T. P. Gujarati, M. Motta, T. Y. Takeshita, E. Lee, J. A. Latone, and J. M. Garcia. Quantum computation of dominant products in lithium–sulfur batteries. *The Journal of Chemical Physics*, 154(13):134115, 2021.
- [73] R. Santagati, J. Wang, A. A. Gentile, S. Paesani, N. Wiebe, J. R. McClean, S. Morley-Short, P. J. Shadbolt, D. Bonneau, J. W. Silverstone, et al. Witnessing eigenstates for quantum simulation of hamiltonian spectra. *Science Advances*, 4(1):eaap9646, 2018.
- [74] Y. Shen, X. Zhang, S. Zhang, J.-N. Zhang, M.-H. Yung, and K. Kim. Quantum implementation of the unitary coupled cluster for simulating molecular electronic structure. *Physical Review A*, 95(2):020501, 2017.
- [75] S. E. Smart and D. A. Mazziotti. Quantum-classical hybrid algorithm using an error-mitigating n -representability condition to compute the mott metal-insulator transition. *Physical Review A*, 100(2):022517, 2019.
- [76] K. Yamamoto, D. Z. Manrique, I. T. Khan, H. Sawada, and D. M. Ramo. Quantum hardware calculations of periodic systems with partition-measurement symmetry verification: Simplified models of hydrogen chain and iron crystals. *Physical Review Research*, 4(3):033110, 2022.
- [77] 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(1):4213, 2014.

- [78] P. J. O’Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, and N. Ding. Scalable quantum simulation of molecular energies. *Physical Review X*, 6(3):031007, 2016.
- [79] M. Loipersberger, F. D. Malone, A. R. Welden, R. M. Parrish, T. Fox, M. Degroote, E. Koyseva, N. Moll, R. Santagati, and M. Streif. Accurate non-covalent interaction energies on noisy intermediate-scale quantum computers via second-order symmetry-adapted perturbation theory. *Chemical Science*, 14(13):3587–3599, 2023.
- [80] P. J. Ollitrault, M. Loipersberger, R. M. Parrish, A. Erhard, C. Maier, C. Sommer, J. Ullmanis, T. Monz, C. Gogolin, and C. S. Tautermann. Estimation of electrostatic interaction energies on a trapped-ion quantum computer. *ACS Central Science*, 10(4):882–889, 2024.
- [81] B. Jeziorski, R. Moszynski, and K. Szalewicz. Perturbation theory approach to intermolecular potential energy surfaces of van der waals complexes. *Chemical Reviews*, 94(7):1887–1930, 1994.
- [82] B. Jeziorski, M. Bulski, and L. Piela. First-order perturbation treatment of the short-range repulsion in a system of many closed-shell atoms or molecules. *International Journal of Quantum Chemistry*, 10(2):281–297, 1976.
- [83] K. Patkowski. Recent developments in symmetry-adapted perturbation theory. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 10(3):e1452, 2020.
- [84] Lewis W Anderson, Martin Kiffner, Panagiotis Kl Barkoutsos, Ivano Tavernelli, Jason Crain, and Dieter Jaksch. Coarse-grained intermolecular interactions on quantum processors. *Physical Review A*, 105(6):062409, 2022.
- [85] A. A. Holmes, N. M. Tubman, and C. Umrigar. Heat-bath configuration interaction: An efficient selected configuration interaction algorithm inspired by heat-bath sampling. *Journal of Chemical Theory and Computation*, 12(8):3674–3680, 2016.
- [86] A. A. Holmes, H. J. Changlani, and C. Umrigar. Efficient heat-bath sampling in fock space. *Journal of Chemical Theory and Computation*, 12(4):1561–1571, 2016.
- [87] J. E. Smith, B. Mussard, A. A. Holmes, and S. Sharma. Cheap and near exact casscf with large active spaces. *Journal of Chemical Theory and Computation*, 13(11):5468–5478, 2017.
- [88] S. Sharma, A. A. Holmes, G. Jeanmairet, A. Alavi, and C. J. Umrigar. Semistochastic heat-bath configuration interaction method: Selected configuration interaction with semistochastic perturbation theory. *Journal of Chemical Theory and Computation*, 13(4):1595–1604, 2017.
- [89] R. J. Bartlett. Perspective on coupled-cluster theory. the evolution toward simplicity in

- quantum chemistry. *Physical Chemistry Chemical Physics*, 26(10):8013–8037, 2024.
- [90] P. Cársky, J. Paldus, and J. Pittner. *Recent progress in coupled cluster methods: Theory and applications*, volume 10 of *Challenges and Advances in Computational Chemistry and Physics*. Springer Science & Business Media, Dordrecht, 2010.
 - [91] M. P. Metz, K. Szalewicz, J. Sarka, R. Tóbiás, A. G. Császár, and E. Mátyus. Molecular dimers of methane clathrates: ab initio potential energy surfaces and variational vibrational states. *Physical Chemistry Chemical Physics*, 21(25):13504–13525, 2019.
 - [92] Arvin Huang-Te Li and Sheng D. Chao. Interaction energies of dispersion-bound methane dimer from coupled cluster method at complete basis set limit. *Journal of Molecular Structure: THEOCHEM*, 897(1):90–94, 2009.
 - [93] Elvira R. Sayfutyarova, Qiming Sun, Garnet Kin-Lic Chan, and Gerald Knizia. Automated construction of molecular active spaces from atomic valence orbitals. *Journal of Chemical Theory and Computation*, 13(9):4063–4078, 2017.
 - [94] Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt, N. A. Bogdanov, G. H. Booth, J. Chen, and Z.-H. Cui. Recent developments in the pyscf program package. *The Journal of Chemical Physics*, 153(2):024109, 2020.
 - [95] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, and S. Sharma. Pyscf: the python-based simulations of chemistry framework. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 8(1):e1340, 2018.
 - [96] Q. Sun. Libcint: An efficient general integral library for gaussian basis functions. *Journal of Computational Chemistry*, 36(22):1664–1671, 2015.
 - [97] I. F. Galván. Pegamoid, 2024. Accessed: 2024-09-01.
 - [98] F. Neese. Software update: The orca program system - version 5.0. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 12(5):e1606, 2022.
 - [99] B. Temelso, K. A. Archer, and G. C. Shields. Benchmark structures and binding energies of small water clusters with anharmonicity corrections. *The Journal of Physical Chemistry A*, 115(43):12034–12046, 2011.
 - [100] J. Rezac and P. Hobza. Describing noncovalent interactions beyond the common approximations: how accurate is the “gold standard,” ccSD(T) at the complete basis set limit? *Journal of Chemical Theory and Computation*, 9(5):2151–2155, 2013.
 - [101] J. Řezáč, P. Jurečka, K. E. Riley, J. Černý, H. Valdes, K. Pluháčková, K. Berka, T. Řezáč, M. Pitoňák, and J. Vondrášek. Quantum chemical benchmark energy and geometry database for molecular clusters and complex molecular systems (www.begdb.com): a user’s manual

- and examples. *Collection of Czechoslovak Chemical Communications*, 73(10):1261–1270, 2008.
- [102] 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. *Chemical Science*, 14(40):11213–11227, 2023.
- [103] P. Jordan and E. P. Wigner. *Über das paulische Äquivalenzverbot*. Springer, 1993.
- [104] Gadi Aleksandrowicz, Thomas Alexander, Panagiotis Kl. Barkoutsos, Luciano Bello, Yael Ben-Haim, David Bucher, Francisco Jose Cabrera-Hernández, Jorge Carballo-Franquis, Adrian Chen, Chun-Fu Chen, Jerry M. Chow, Antonio D. Córcoles-Gonzales, Abigail J. Cross, Andrew W. Cross, Juan Cruz-Benito, Chris Culver, Salvador De La Puente González, Enrique De La Torre, Delton Ding, Eugene F. Dumitrescu, Ivan Duran, Pieter T. Eendebak, Mark Everitt, Ismael Faro Sertage, Albert Frisch, Andreas Fuhrer, Jay M. Gambetta, Borja Godoy Gago, Juan Gomez-Mosquera, Donny Greenberg, Ikko Hamamura, Vojtech Havlicek, Joe Hellmers, Łukasz Herok, Hiroshi Horii, Shaohan Hu, Takashi Imamichi, Toshi-nari Itoko, Ali Javadi-Abhari, Naoki Kanazawa, Anton Karazeev, Kevin Krsulich, Peng Liu, Yang Luh, Yunho Maeng, Manoel Marques, Francisco Martín-Fernández, Douglas McClure, David McKay, Srujan Meesala, Antonio Mezzacapo, Nikolaj Moll, Diego Moreda Rodríguez, Giacomo Nannicini, P. D. Nation, Pauline J. Ollitrault, Lee James O’Riordan, Hanhee Paik, Jesús Pérez, Anna Phan, Marco Pistoia, Viktor Prutyanov, Max Reuter, Julia E. Rice, Abdón Rodríguez Davila, Raymond Harry Rudy, Mingi Ryu, Ninad Sathaye, Chris Schnabel, Eddie Schoute, Kanav Setia, Yunong Shi, Adenilton Silva, Yukio Siraichi, Seyon Sivarajah, John A. Smolin, Mathias Soeken, Hitomi Takahashi, Ivano Tavernelli, Charles Taylor, Pete Taylour, Kenso Trabing, Matthew Treinish, Wes Turner, Desiree Vogt-Lee, Christophe Vuillot, Jonathan A. Wildstrom, Jessica Wilson, Erick Winston, Christopher J. Wood, Stephen P. Wood, Stefan Wörner, Ismail Yunus Akhalwaya, and Christa Zoufal. Qiskit: An open-source framework for quantum computing, 2019.
- [105] ffsims developers. ffsim: Faster simulations of fermionic quantum circuits, 2024. Accessed: 2024-09-01.
- [106] A. Javadi-Abhari, M. Treinish, K. Krsulich, C. J. Wood, J. Lishman, J. Gacon, S. Martiel, P. D. Nation, L. S. Bishop, and A. W. Cross. Quantum computing with qiskit. *arXiv preprint arXiv:2405.08810*, 2024.
- [107] J. J. Wallman and J. Emerson. Noise tailoring for scalable quantum computation via randomized compiling. *Physical Review A*, 94(5):052325, 2016.
- [108] L. Viola and S. Lloyd. Dynamical suppression of decoherence in two-state quantum systems. *Physical Review A*, 58(4):2733, 1998.
- [109] A. Kofman and G. Kurizki. Universal dynamical control of quantum mechanical decay: modulation of the coupling to the continuum. *Physical Review Letters*, 87(27):270405,

2001.

- [110] M. J. Biercuk, H. Uys, A. P. VanDevender, N. Shiga, W. M. Itano, and J. J. Bollinger. Optimized dynamical decoupling in a model quantum memory. *Nature*, 458(7241):996–1000, 2009.
- [111] S. Niu and A. Todri-Sanial. Effects of dynamical decoupling and pulse-level optimizations on ibm quantum computers. *IEEE Transactions on Quantum Engineering*, 3:1–10, 2022.
- [112] Abdullah Ash Saki, Stefano Barison, Bryce Fuller, James R. Garrison, Jennifer R. Glick, Caleb Johnson, Antonio Mezzacapo, Javier Robledo-Moreno, Max Rossmannek, Paul Schweigert, Iskandar Sitdikov, and Kevin J. Sung. Qiskit addon: sample-based quantum diagonalization, 2024.
- [113] Y. O. Nakagawa, M. Kamoshita, W. Mizukami, S. Sudo, and Y.-y. Ohnishi. Adapt-qsci: Adaptive construction of input state for quantum-selected configuration interaction. *arXiv preprint arXiv:2311.01105*, 2023.
- [114] P. Moritz, R. Nishihara, S. Wang, A. Tumanov, R. Liaw, E. Liang, M. Elibol, Z. Yang, W. Paul, and M. I. Jordan. Ray: A distributed framework for emerging ai applications. *arXiv preprint arXiv:1712.05889*, 2017.
- [115] M. Hapka, A. Krzemińska, and K. Pernal. How much dispersion energy is included in the multiconfigurational interaction energy? *Journal of Chemical Theory and Computation*, 16(10):6280–6293, 2020.
- [116] Vera von Burg, Guang Hao Low, Thomas Häner, Damian S Steiger, Markus Reiher, Martin Roetteler, and Matthias Troyer. Quantum computing enhanced computational catalysis. *Physical Review Research*, 3(3):033055, 2021.
- [117] Nick S. Blunt, Joan Camps, Ophelia Crawford, Róbert Izsák, Sebastian Leontica, Arjun Mirani, Alexandra E. Moylett, Sam A. Scivier, Christoph Sünderhauf, Patrick Schopf, Jacob M. Taylor, and Nicole Holzmann. Perspective on the current state-of-the-art of quantum computing for drug discovery applications. *Journal of Chemical Theory and Computation*, 18(12):7001–7023, 2022. PMID: 36355616.
- [118] Alexey Pyrkov, Alex Aliper, Dmitry Bezrukov, Yen-Chu Lin, Daniil Polykovskiy, Petrina Kamy, Feng Ren, and Alex Zhavoronkov. Quantum computing for near-term applications in generative chemistry and drug discovery. *Drug Discovery Today*, 28(8):103675, 2023.
- [119] Gautam Kumar, Sahil Yadav, Aniruddha Mukherjee, Vikas Hassija, and Mohsen Guizani. Recent advances in quantum computing for drug discovery and development. *IEEE Access*, 12:64491–64509, 2024.
- [120] Bonde Bhushan, Patil Pratik, and Choubey Bhaskar. Chapter 7: The future of drug develop-

- ment with quantum computing. In Alexander Heifetz, editor, *High Performance Computing for Drug Discovery and Biomedicine*, pages 153–179. Springer, 2023.
- [121] Alexis Ralli, Michael Williams de la Bastida, and Peter V. Coveney. Scalable approach to quantum simulation via projection-based embedding. *Phys. Rev. A*, 109:022418, Feb 2024.
 - [122] Hocheol Lim, Doo Hyung Kang, Jeonghoon Kim, Aidan Pellow-Jarman, Shane McFarthing, Rowan Pellow-Jarman, Hyeon-Nae Jeon, Byungdu Oh, June-Koo Kevin Rhee, and Kyoung Tai No. Fragment molecular orbital-based variational quantum eigensolver for quantum chemistry in the age of quantum computing. *Scientific Reports*, 14(1):2422, Jan 2024.
 - [123] Takeshi Yoshikawa, Tomoya Takanashi, and Hiromi Nakai. Quantum algorithm of the divide-and-conquer unitary coupled cluster method with a variational quantum eigensolver. *Journal of Chemical Theory and Computation*, 18(9):5360–5373, 2022. PMID: 35926142.
 - [124] Naoki Iijima, Satoshi Imamura, Mikio Morita, Sho Takemori, Akihiko Kasagi, Yuhei Umeda, and Eiji Yoshida. Towards accurate quantum chemical calculations on noisy quantum computers. *arXiv preprint arXiv:2311.09634*, 2023. This preprint is related to the press release of Fujitsu Limited.
 - [125] Honghui Shang, Yi Fan, Li Shen, Chu Guo, Jie Liu, Xiaohui Duan, Fang Li, and Zhenyu Li. Towards practical and massively parallel quantum computing emulation for quantum chemistry. *npj Quantum Information*, 9(1):33, Apr 2023.

APPENDIX

SUPPLEMENTARY INFORMATION

3A.1 Geometries of potential energy surfaces

The PES for the water dimer is calculated for distances between two oxygen atoms ranging between 1.400 and 3.500 Å. We distribute the points for water dimer PES as 1.400, 1.500, 1.600, 1.700, 1.800, 1.900, 1.962, 2.000, 2.100, 2.200, 2.300, 2.400, 2.500, 3.000, and 3.500 Å. All of water dimer simulations are done for all of these distances. The PES for the methane dimer is calculated for the distances between two carbon atoms ranging between 2.500 and 6.000 Å. We distribute the points for methane dimer PES as 2.500, 2.750, 3.000, 3.167, 3.334, 3.500, 3.667, 3.834, 4.000, 4.250, 4.500, 4.750, 5.000, and 6.000 Å. To calculate the total energy of unbound dimer we utilize the distance of 48.000 Å for both water and methane dimers. All of CASCI (16e,16o), CCSD, CCSD(T), CCSD (16e,16o), and CCSD(T) (16e,16o) simulations of methane dimer as well as SQD (16e,16o) simulations with $|\tilde{\chi}_b| = 20.0 \cdot 10^3$ are done for all of the distances described earlier. The methane dimer CASCI (16e,16o) simulations and SQD (16e,16o) simulations with $|\tilde{\chi}_b| = 20.0 \cdot 10^3$ are also performed for an additional distance of 3.638 Å. The SQD (16e,16o) energy extrapolations using $|\tilde{\chi}_b|$ of $9.0 \cdot 10^3$, $11.0 \cdot 10^3$, and $14.0 \cdot 10^3$ are done for 4.000, 4.250, 4.500, 4.750, 5.000, 6.000, and 48.000 Å distances. In the case of HCI (16e,24o) simulations of the methane dimer, we use only distances of 3.638, 3.667, 3.750, 3.834, and 3.900 Å. In case of SQD (16e,24o) simulations we only use the distance of 3.638 Å. All calculations are done as single-point energy calculations with no geometry optimizations. To produce the geometries studies in this work, we start from the equilibrium geometries and change the distance between the centers of the monomers, with the geometries of the individual monomers fixed.

3A.2 Details of HCI calculations

In HCI simulations instead of generating all of the single and double excitations one generates only those single and double excitations that correspond to Hamiltonian matrix elements exceeding a threshold ε . In HCI ε_1 controls which determinants will be included in the variational wave function. In our HCI calculations, we use values of ε_1 equal to $5 \cdot 10^{-6}$ (during initial variational

steps) and $1 \cdot 10^{-6}$ (during later variational steps). We do not use the non-variational perturbative correction in our HCI calculations. Hence, our HCI calculations are fully variational, which allows for more appropriate comparison between the SQD and HCI results.

3A.3 Effect of number of samples on total energy in SQD(16e,16o)

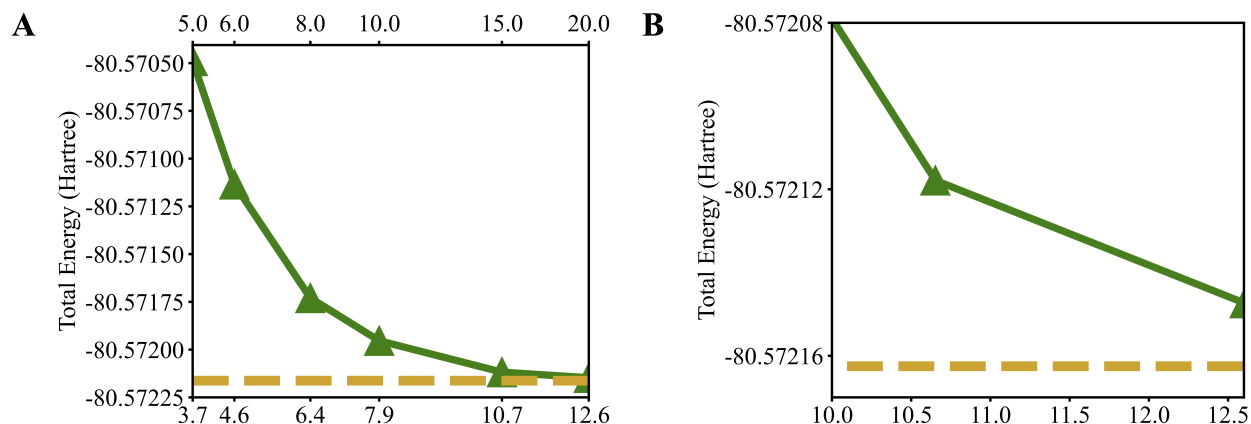


Figure 3A.1 Total energy of methane dimer predicted with SQD (16e,16o) at 3.638 Å distance between the monomers as the function of $d \cdot 10^7$. (A) the entire range of d , and (B) a magnified region with largest values of d , highlighted in panel (A) as a black box. The secondary x-axis demonstrates the value of $|\tilde{\chi}_b| \cdot 10^3$ producing the given value of $d \cdot 10^7$. Solid green line with triangular markers shows SQD (16e,16o) results. Horizontal dashed light brown line indicates the total energy from CASCI (16e,16o) calculation.

3A.4 PES of methane dimer including the repulsive region

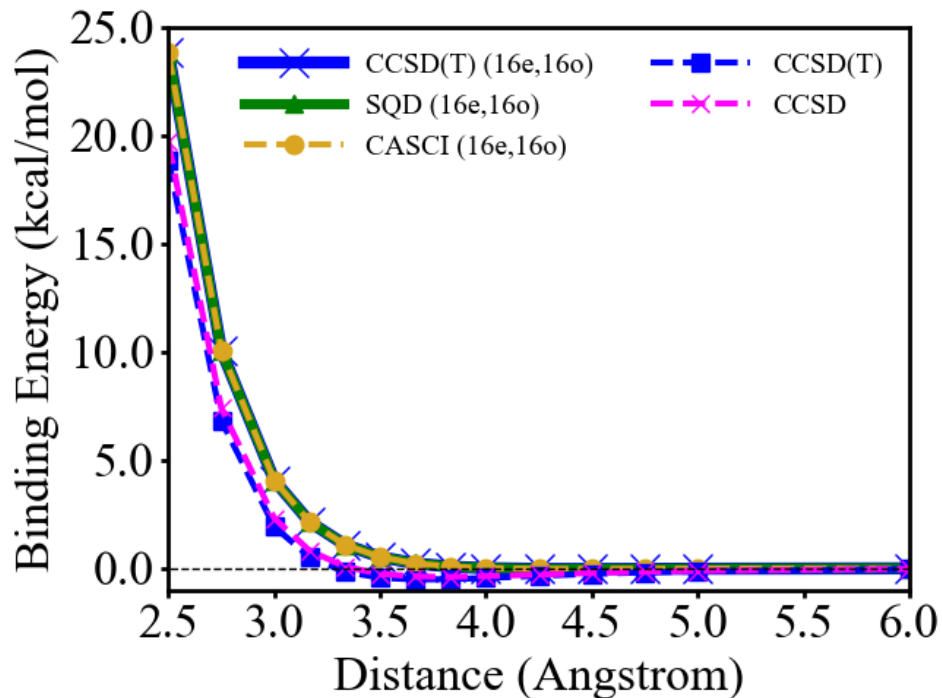


Figure 3A.2 Binding energies of the methane dimer along its PES, where the distances between the centers of methane molecules range between 2.500 and 6.000 Å. Active space simulations are performed with (16e,16o). Light brown, orange, and blue dashed lines with circle markers depict PES calculated with CASCI, CCSD, and CCSD(T) methods, respectively. Solid yellow line with triangular markers depicts the PES calculated with the SQR method. Solid blue line represents CCSD(T) calculations using an active space. Black horizontal dashed line indicates the zero value of the binding energy.

CHAPTER 4

TOWARDS QUANTUM-CENTRIC SIMULATIONS OF EXTENDED MOLECULES: SAMPLE-BASED QUANTUM DIAGONALIZATION ENHANCED WITH DENSITY MATRIX EMBEDDING THEORY

Computing ground-state properties of molecules is a promising application for quantum computers operating in concert with classical high-performance computing resources. Quantum embedding methods are a family of algorithms particularly suited to these computational platforms: they combine high-level calculations on active regions of a molecule with low-level calculations on the surrounding environment, thereby avoiding expensive high-level full-molecule calculations and allowing to distribute computational cost across multiple and heterogeneous computing units. Here, we present the first density matrix embedding theory (DMET) simulations performed in combination with the sample-based quantum diagonalization (SQD) method. We employ the DMET-SQD formalism to compute the ground-state energy of a ring of 18 hydrogen atoms, and the relative energies of the chair, half-chair, twist-boat, and boat conformers of cyclohexane. The full-molecule 41- and 89-qubit simulations are decomposed into 27- and 32-qubit active-region simulations, that we carry out on the `ibm_cleveland` device, obtaining results in agreement with reference classical methods. Our DMET-SQD calculations mark a tangible progress in the size of active regions that can be accurately tackled by near-term quantum computers, and are an early demonstration of the potential for quantum-centric simulations to accurately treat the electronic structure of large molecules, with the ultimate goal of tackling systems such as peptides and proteins.

4.1 Introduction

The accurate treatment of interacting many-electron systems by first-principle computational methods is a grand challenge of contemporary science. Progress in addressing this challenge will affect diverse fields, including pharmaceutical research and in particular computer-aided drug design, where it is important to characterize the energy landscape of the conformations of large molecule and their complexes at room temperature^{1,2}. Existing computational resources, however, are insufficient to carry out first-principle electronic structure simulations on many large molecules, e.g. proteins, by methods beyond the mean-field approximation. For example, the simulation of

insulin³ using minimal and correlation-consistent cc-pVxZ (x=D,T,Q) basis sets requires 2418 and 7561, 17420, 33559 molecular orbitals (MOs) respectively, at the boundaries of conventional electronic structure methods^{4,5}.

Some approximations, therefore, are needed to realize quantum mechanical calculations on such systems. An example is offered by the large family of fragment-based electronic structure methods⁶⁻⁸. These algorithms are based on the decomposition of an intractably large system into tractable subsystems, and typically employ the quantum embedding scheme, that combines the treatment of subsystems by a higher-level method (i.e. more accurate and expensive) with a lower-level computation on the entire system⁹. They thus offer an important path to leverage advances in first-principle electronic structure methods in the modeling of otherwise intractably large molecules.

Quantum embedding methods are also natural and compelling targets for quantum-centric supercomputing (QCSC) architectures¹⁰, where a quantum computer operates in concert with classical high-performance computing (HPC) resources. In this framework, high-level quantum computations can be carried out on subsystems, with HPC resources providing feedback between the subsystems and their environment. Various authors have recognized this possibility, and have proposed to integrate quantum computing subroutines in the workflow of quantum embedding methods¹¹⁻¹³ where the feedback between subsystems and their environment is based on the electronic density, the Green's function, or the one-particle reduced density matrix (depending on the specific embedding theory)⁹. Before the availability of large-scale fault-tolerant quantum computers, a critical aspect of the application of QCSC to quantum embedding is the optimization of the resources of near-term quantum computers, limited in both qubit number and quantum circuit size. Motivated by this consideration, several authors have proposed to use the variational quantum eigensolver (VQE) method¹⁴⁻¹⁶ in combination with the fragment molecular orbital (FMO)¹⁷, divide and conquer (DC)¹⁸, many-body expansion (MBE)¹⁹, density matrix embedding theory (DMET)¹⁹⁻²⁵, and other wavefunction-based embedding²⁶⁻²⁸ methods.

However, there are some drawbacks that are inevitable upon introducing fragmentation ap-

proximations, that are exacerbated by the constrained size of subsystems accurately tractable by near-term quantum algorithms and computers: for example, the dispersion and induction of many-body interactions and the cooperativity of hydrogen bonds are considerably impacted. Therefore, extending the reach of quantum computers in terms of tractable subsystem sizes and accuracy of results stands to mitigate the impact of the fragmentation approximation, allowing more meaningful QCSC applications in the framework of quantum embedding.

The recently introduced sample-based quantum diagonalization (SQD) method²⁹ has allowed electronic structure simulations for active sites of metallo-proteins using up to 77 qubits (corresponding to 36 MOs), a substantial advancement in application of quantum computers to quantum chemistry^{30–33}, and for molecular dimers bound by non-covalent interactions³⁴, as well as the large active space simulations of methylene³⁵. In view of these results, it is timely and compelling to investigate the use of SQD as a subsystem solver in the framework of the well-established DMET quantum embedding method. Herein, we present the first DMET-SQD simulations of molecular systems. We use DMET-SQD to compute the ground-state potential energy curve of a ring of 18 hydrogen atoms, a well-established benchmark test^{36–39}. We then compute the relative energies of the chair, half-chair, twist-boat, and boat conformers of cyclohexane, the structure and dynamics of which are important prototypes of a wide range of organic compounds^{40–43}. Our calculations use 27 and 32 qubits for the hydrogen ring and cyclohexane respectively, a substantial reduction from the 41 and 89 qubits required for unfragmented calculations, and are carried out on the `ibm_cleveland` quantum computer from IBM’s Eagle processor family. To assess the accuracy of our results, we compare them against DMET calculations using exact diagonalization (FCI) as a subsystem solver – to estimate the accuracy of SQD as a subsystem solver – and unfragmented calculations with heat-bath configuration interaction (HCI) and coupled cluster singles and doubles (CCSD) – to quantify the accuracy of the fragmentation approximation.

4.2 Methods

4.2.1 Density matrix embedding theory

DMET^{44–48} is a wave-function-based quantum embedding method, that allows high-level treatments for multiple subsystems using accurate and expensive beyond mean-field methods in conjunction with a mean-field calculation on the entire system. Formally, DMET divides the D -dimensional Hilbert space of a quantum system into a fragment, i.e. a subspace of dimension $d_F \ll D$, and an environment, i.e. a subspace of dimension $D_E = D - d_F \gg d_F$. The exact and unknown ground-state of the system's Hamiltonian $\hat{H}$, $|\Psi\rangle = \sum_{f=1}^{d_F} \sum_{e=1}^{D_E} \Psi_{fe} |f\rangle \otimes |e\rangle$, is written through a Schmidt decomposition of the matrix $\Psi_{fe} = \sum_{\mu=1}^{d_F} U_{f\mu} \lambda_{\mu} V_{\mu e}$ as $|\Psi\rangle = \sum_{\mu=1}^{d_F} \lambda_{\mu} |u_{\mu}\rangle \otimes |b_{\mu}\rangle$. The states $|b_{\mu}\rangle = \sum_e V_{\mu e} |e\rangle$ are called bath states and, although they lie in a D_E -dimensional subspace, they span a d_F -dimensional subspace. The ground state of $\hat{H}$ is also the ground state of the subsystem Hamiltonian $\hat{H}_{emb} = \hat{P} \hat{H} \hat{P}$, where $\hat{P} = \sum_{\mu\nu} |u_{\mu}\rangle \langle u_{\nu}| \otimes |b_{\nu}\rangle \langle b_{\mu}|$ is the projection on a $2d_F$ -dimensional subspace, drastically smaller than the original D -dimensional space. While this construction is formally exact and shows that the electronic structure of the entire system can be described exactly by that of a fragment and its surrounding bath, it is impractical because it assumes knowledge of the exact ground-state wavefunction and its Schmidt decomposition.

In practical implementations, DMET employs a mean-field approximation to the ground-state wavefunction⁴⁶. For a closed-shell system of N_{elec} electrons in N_{orb} spatial orbitals, this is a Slater determinant of the form $|\Phi\rangle = \prod_{i\sigma} \hat{a}_{ci\sigma}^{\dagger} |\emptyset\rangle$, where $|\emptyset\rangle$ is the vacuum state and $\hat{a}_{ci\sigma}^{\dagger}$ creates a spin- σ electron in the spatial orbital $|c_i\rangle = \sum_p C_{pi} |p\rangle$, with $i = 1 \dots N_{elec}/2$ and $|p\rangle$ a finite basis set of N_{orb} orthonormal and spatially localized single-electron orbitals. Basis set elements are divided in fragment ($p \in F$) and environment ($e \notin F$) orbitals. Diagonalizing the environment-environment block of the one-particle reduced density matrix, $\rho_{ee'} = \langle \Phi | \hat{E}_{e'}^e | \Phi \rangle$ with $\hat{E}_{e'}^e = \sum_{\sigma} \hat{a}_{e\sigma}^{\dagger} \hat{a}_{e'\sigma}$, yields a set of eigenpairs, $\sum_{e'} \rho_{ee'} W_{e'\alpha} = W_{e\alpha} \delta_{\alpha}$. The eigenvalues $\delta_{\alpha} = 0$, $0 < \delta_{\alpha} < 2$, and $\delta_{\alpha} = 2$ respectively correspond to virtual environment orbitals, so-called “bath orbitals”, and occupied environment orbitals. The state $|\Phi\rangle$ can be factored in two terms, respectively containing occupied orbitals without overlap on F (occupied environment) and with overlap on F (linear

combinations of fragment and bath orbitals). Since bath orbitals and local fragment orbitals are in general partially occupied in the DMET high-level wavefunction, the subsystem Hamiltonian has the form of a quantum chemistry active-space Hamiltonian^{45,46} where occupied and virtual environment orbitals are inactive orbitals. Denoting the electronic Hamiltonian of the full system as $\hat{H} = \sum_{pr} t_{pr} \hat{E}_r^p + \sum_{prqs} \frac{(pr|qs)}{2} \hat{E}_{rs}^{pq}$ with $\hat{E}_{rs}^{pq} = \sum_{\sigma\tau} \hat{a}_{p\sigma}^\dagger \hat{a}_{q\tau}^\dagger \hat{a}_{s\tau} \hat{a}_{r\sigma}$, the subsystem Hamiltonian takes the form

$$\begin{aligned} \hat{H}_{emb} &= \sum_{tu} \tilde{t}_{tu} \hat{E}_u^t + \sum_{prqs} \frac{(tu|vw)}{2} \hat{E}_{uw}^{tv}, \\ \tilde{t}_{tu} &= t_{tu} + \sum_{ee'} \left[(tu|ee') - (te'|eu) \right] \tilde{\rho}_{ee'}, \end{aligned} \quad (4.1)$$

where tu range over fragment and bath orbitals (defining the subsystem, also called “impurity” in the DMET literature), and $\tilde{\rho}_{ee'} = \sum_{\alpha: \delta_\alpha=2} W_{e\alpha} W_{e'\alpha}^*$ is the density matrix of the N_{occ}^{env} occupied environment orbitals. The lowest-energy eigenvector (with $N_{elec} - 2N_{occ}^{env}$ electrons) of the subsystem Hamiltonian Eq. (4.1) can be computed with a high-level method like full configuration interaction (FCI)^{44,45}, density matrix renormalization theory (DMRG)⁴⁹, complete active space self-consistent field (CASSCF)^{48,50}, or CCSD⁵¹.

Typical DMET calculations employ multiple fragments, in which case multiple subsystem Hamiltonians are produced and diagonalized, yielding multiple one- and two-body reduced density matrices that are used to evaluate properties of the full system, e.g. the total energy and electron number^{45,46}. The partitioning into multiple overlapping fragments can lead to non-variationality of the DMET energy, and discrepancies between the target (N_{elec}) and DMET electron number. To remedy this deficiency, we follow the strategy of “one-shot” DMET calculations⁴⁶, where the subsystem Hamiltonians Eq. (4.1) are modified adding a global chemical potential (i.e. independent of fragment and orbital indices) for the local fragment orbitals, $\hat{H}_{emb} \rightarrow \hat{H}_{emb} - \mu_{\text{glob}} \sum_{f \in F} \hat{a}_f^\dagger \hat{a}_f$, optimized to ensure that DMET predicts the correct particle number. Optimization of chemical potential is done iteratively until the threshold is reached.

4.2.2 Sample-based quantum diagonalization

SQD^{29,52,53} is a quantum subspace method⁵⁴ that uses a quantum circuit $|\Phi_{\text{qc}}\rangle$ to sample a set of computational basis states $\chi = \{\mathbf{x}_1 \dots \mathbf{x}_d\}$ from the probability distribution $p(\mathbf{x}) = |\langle \mathbf{x} | \Phi_{\text{qc}} \rangle|^2$ and a classical computer to solve the Schrödinger equation in the subspace spanned by such computational basis states. Since in the standard Jordan-Wigner (JW) mapping of fermions to qubits^{55–57} a computational basis state parametrizes a Slater determinant, the Slater-Condon rules allow to efficiently compute the matrix elements $\langle \mathbf{x}_i | \hat{H} | \mathbf{x}_j \rangle$ within Davidson diagonalization.

In this work, we use SQD to perform conventional unfragmented active-space calculations^{29,34,58} and to approximate the ground state of the DMET subspace Hamiltonians (4.1). We sample computational basis states from the following truncated version of the local unitary cluster Jastrow (LUCJ) ansatz⁵⁹

$$|\Phi_{\text{qc}}\rangle = e^{-\hat{K}_2} e^{\hat{K}_1} e^{i\hat{J}_1} e^{-\hat{K}_1} |\mathbf{x}_{\text{RHF}}\rangle, \quad (4.2)$$

where $\hat{K}_1$ and $\hat{K}_2$ are one-body operators, $\hat{J}_1$ is density-density operator, and $|\mathbf{x}_{\text{RHF}}\rangle$ is the restricted closed-shell Hartree-Fock (RHF) state. The parameters defining the LUCJ wavefunction (4.2) are derived from a classical restricted closed-shell CCSD, as detailed in Ref. 29.

On a noisy quantum device, computational basis states are sampled from a probability distribution $\tilde{p}(\mathbf{x})$ that unavoidably differs from $p(\mathbf{x})$ due to quantum noise. As a result, (i) particle number and total spin- z may not be conserved, i.e. the sampled computational basis states may correspond to Slater determinants with incorrect number of spin-up and/or spin-down electrons, and (ii) the subspace spanned by the sampled computational basis states may not allow to produce eigenstates of the total spin operator $\hat{S}^2$. We emphasize that situations (i) and (ii) may occur also on noiseless quantum devices, respectively because the quantum circuit Ψ may break particle-number and spin- z symmetries (although this is not the case for LUCJ) and Slater determinants are generally not eigenstates of total spin. To restore the broken particle-number and spin- z symmetries, Ref. 29 proposed an iterative self-consistent configuration recovery (S-CORE) procedure. Each iteration of S-CORE has two inputs: a fixed set of computational basis states $\tilde{\chi}$ sampled from a quantum computer, and an approximation to the ground-state occupation number distribution $n_{p\sigma} = \langle \Psi | \hat{a}_{p\sigma}^\dagger \hat{a}_{p\sigma} | \Psi \rangle$. In

each iteration, we randomly flip the entries of the computational basis states in $\tilde{\chi}$ based on $n_{p\sigma}$ until particle number and spin- z assume target values, thereby producing a new set $\tilde{\chi}_R$. We then sample K subsets (batches) from $\tilde{\chi}_R$, that we label $\tilde{\chi}_b$ with $b = 1 \dots K$. Each batch yields a subspace $S^{(b)}$ of dimension d^{29} , in which we project the many-electron Hamiltonian as

$$\hat{H}_{S^{(b)}} = \hat{P}_{S^{(b)}} \hat{H} \hat{P}_{S^{(b)}} , \quad (4.3)$$

where the projector $\hat{P}_{S^{(b)}}$ is

$$\hat{P}_{S^{(b)}} = \sum_{\mathbf{x} \in S^{(b)}} |\mathbf{x}\rangle \langle \mathbf{x}| . \quad (4.4)$$

We compute the ground-state wavefunctions and energies of Eq. (4.3), respectively $|\psi^{(b)}\rangle$ and $E^{(b)}$, and use the lowest energy across the batches, $\min_b E^{(b)}$, as the best approximation to the ground-state energy. We use the wavefunctions $|\psi^{(b)}\rangle$ to update the occupation number distribution,

$$n_{p\sigma} = \frac{1}{K} \sum_{1 \leq b \leq K} \langle \psi^{(b)} | \hat{a}_{p\sigma}^\dagger \hat{a}_{p\sigma} | \psi^{(b)} \rangle , \quad (4.5)$$

and use it as an input in the next S-CORE iteration. At the first S-CORE iteration, we initialize $n_{p\sigma}$ by post-selecting⁶⁰ measurement outcomes in $\tilde{\chi}$ with the correct particle number.

While S-CORE restores particle number and spin- z conservation (and can be immediately generalized to any other symmetry operator having Slater determinants as eigenstates, e.g., molecular point-group symmetries in a basis of symmetry-adapted MOs), it does not ensure conservation of total spin $\hat{S}^2$. To mitigate this limitation, in the construction of the subspaces $S^{(b)}$, we extend the set of configurations $\tilde{\chi}_b$ to ensure its closure under spin inversion symmetry as detailed in Ref. 29. For this reason, d can be larger than $|\tilde{\chi}_b|$.

4.2.3 Computational details

Target applications We define the geometries of H_{18} as $\mathbf{R}_k = (\rho \cos(k\Delta\theta), \rho \sin(k\Delta\theta), 0)$ with $k = 1 \dots 18$ and $\Delta\theta = 2\pi/18$, i.e. along a circle of radius ρ , chosen so that the distance between adjacent atoms is $R = 0.70, 0.85, 1.00, 1.10$, and $1.30 \text{ \AA}$ (see Fig. 4.1). We compute the geometries of the chair, half-chair, twist-boat, and boat conformers of cyclohexane by performing a geometry optimization at the MP2/aug-cc-pVDZ level of theory using the ORCA software package⁶¹, and

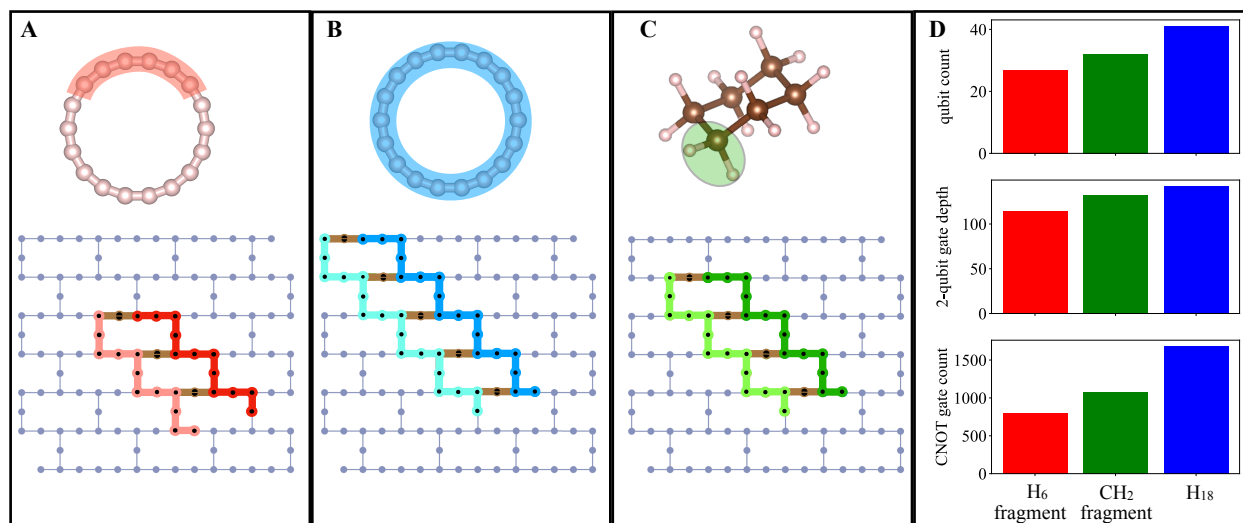


Figure 4.1 Overview of quantum circuits. (A-C) Qubit layouts of LUCJ circuits for: (A) DMET-SQD simulations of H₁₈ with H₆ fragments and (12e,12o) subsystems using 27 qubits of `ibm_cleveland`, (B) unfragmented H₁₈ simulations using 41 qubits, and (C) DMET-SQD simulations of cyclohexane with CH₂ fragments and (14e,14o) subsystems using 32 qubits. Colored highlights over the molecular structures indicate DMET fragments (red, blue, green for systems A, B, C respectively). The layout of `ibm_cleveland` is shown in gray, qubits used to encode occupation numbers of spin-up/down electrons are marked in light/dark colors, and ancilla qubits in brown. (D) Number of qubits, 2-qubit gate depth, and CNOT gate count of the LUCJ circuits of systems A, B, C (red, blue, green columns respectively).

confirming the local minimum and transition-state character of each geometry by a vibrational frequency analysis.

DMET We perform DMET calculations as implemented in Tangelo 0.4.3⁶², using PySCF 2.6.2^{4,63,64}. In our DMET calculations, carried out at the STO-3G level of theory, we localize orbitals^{46,48} with the meta-Lowdin scheme⁶⁵, and optimize the chemical potential μ_{glob} with a convergence threshold of $1 \cdot 10^{-5}$.

LUCJ We generate the LUCJ circuits in Eq. (4.2) using the `ffsim` library⁶⁶ interfaced with Qiskit 1.1.1^{67,68}. We execute them on the `ibm_cleveland` quantum computer, using the qubit layouts shown in Fig. 4.1A, 4.1B, and 4.1C for DMET calculations on H₁₈ with an H₆ fragment, unfragmented SQD calculations on H₁₈, and DMET calculations on cyclohexane with a CH₂ fragment, respectively. To mitigate quantum errors, we use gate twirling (but not measurement

Table 4.1 Details of SQD and DMET-SQD calculations. Fragments are defined by MOs localized spatially in the regions highlighted in Fig. 4.1. We use the following abbreviations: “unfrag.” for unfragmented (i.e. conventional SQD), “AS” for active space (a subsystem, i.e. fragment+bath, in DMET-SQD calculations) and D_{AS} for Hilbert-space dimension as determined from the active-space orbitals and electrons in column 3. $|\tilde{\chi}_b|$ and d is defined as in the main text, and d' is the number of configurations in an SQD wavefunction $|\psi^{(b)}\rangle = \sum_{m=1}^d c_m |\mathbf{x}_m\rangle$ with amplitudes $|c_m|^2 \geq 1 \cdot 10^{-8}$. The values of d and d' are computed for H_{18} at bondlength 1.0 Å and for the chair conformation of cyclohexane, and more extensive studies are reported in the Appendix.

system	fragment	AS	method	$ \tilde{\chi}_b [10^3]$	$d [10^5]$	$d' [10^5]$	$D_{\text{AS}} [10^5]$
H_{18}	H_6	(12e,12o)	DMET-SQD	1	5.0	1.190	8.5
H_{18}	H_6	(12e,12o)	DMET-SQD	2	6.9	1.723	8.5
H_{18}	H_6	(12e,12o)	DMET-SQD	3	8.0	2.050	8.5
H_{18}	H_6	(12e,12o)	DMET-SQD	4	8.3	2.174	8.5
H_{18}	H_6	(12e,12o)	DMET-SQD	5	8.5	2.224	8.5
H_{18}	unfrag.	(18e,18o)	SQD	7	1163	0.731	23639
H_{18}	unfrag.	(18e,18o)	SQD	8	1408	0.825	23639
H_{18}	unfrag.	(18e,18o)	SQD	9	1712	0.876	23639
H_{18}	unfrag.	(18e,18o)	SQD	10	2034	0.980	23639
H_{18}	unfrag.	(18e,18o)	SQD	11	2346	1.048	23639
H_{18}	unfrag.	(18e,18o)	SQD	12	2656	1.075	23639
C_6H_{12}	CH_2	(14e,14o)	DMET-SQD	6	86.7	0.858	118
C_6H_{12}	CH_2	(14e,14o)	DMET-SQD	7	88.7	0.886	118
C_6H_{12}	CH_2	(14e,14o)	DMET-SQD	8	97.2	0.889	118
C_6H_{12}	CH_2	(14e,14o)	DMET-SQD	9	99.9	0.906	118
C_6H_{12}	CH_2	(14e,14o)	DMET-SQD	10	103.8	0.907	118

twirling) over random 2-qubit Clifford gates⁶⁹ and dynamical decoupling^{70–73} as available through the SamplerV2 primitive of Qiskit’s runtime library. The number of qubits, 2-qubit gate depth, and number of CNOT gates are listed, for the circuits executed in this work, in Fig. 4.1D. Since we elected to perform “one-shot” DMET calculations, and to optimize μ_{glob} using the configurations sampled at $\mu_{\text{glob}} = 0$ and simply repeating the S-CORE procedure for several values of μ_{glob} , the total number of circuits executed in this work is 3, 1, and 6 for the systems in Fig. 4.1A, 4.1B, and 4.1C respectively.

SQD We perform SQD calculations using the implementation in the GitHub repository Ref. 74. In DMET-SQD simulations, we perform 10 and 3 iterations of S-CORE, respectively prior and during the optimization of the chemical potential. In unfragmented SQD calculations, we perform

10 iterations of S-CORE. The details of SQD and DMET-SQD calculations performed in this work are listed in Table 4.1.

Classical benchmarks To assess the accuracy of DMET-SQD calculations, we perform unfragmented calculations using: CCSD, CCSD with perturbative triples (CCSD(T))⁷⁵ as implemented in PySCF 2.6.2, and heat-bath configuration interaction (HCI)^{76–79} as implemented in the SHCI-SCF 0.1 interface between PySCF 2.6.2 and DICE 1.0. In HCI simulations, we fix the parameter ε_1 (controlling which determinants are included in the variational wavefunction⁷⁶) to $5 \cdot 10^{-6}$ during initial variational steps and $1 \cdot 10^{-6}$ during later variational steps.

4.3 Results

We now present results for a ring of 18 hydrogen atoms, closely related to the hydrogen chain described in Ref. 36 and studied at finite lengths and finite basis sets by several groups^{37,80–84} both as a benchmark for numerical methods in presence of electronic correlation of varying nature and strength³⁸ and a model system to investigate a variety of physical phenomena³⁹.

In Fig. 4.2, we show the potential energy curve of the ring from unfragmented restricted closed-shell HF, CCSD, CCSD(T), SQD, and HCI methods (top) and DMET with CCSD, SQD, and FCI as subsystem solver (bottom). Correlated methods agree qualitatively for $R \leq 1.0 \text{ \AA}$ and their potential energy curves visibly depart from each other for $R \geq 1.1 \text{ \AA}$, as the simultaneous breaking of multiple bonds leads to stronger electronic correlation effects. The different behavior of correlated methods is analyzed in more detail in the bottom panel of Fig. 4.2, where we show deviations between computed energies and HCI, taken as reference. As R increases restricted CCSD and CCSD(T) underestimate the ground-state energy by $\sim 2 \text{ kcal/mol}$ per atom, a phenomenon well-known^{38,85} to occur in the presence of static electronic correlation. On the other hand SQD overestimates it by $\sim 2.5 \text{ kcal/mol}$ per atom: as R increases, the ground-state wavefunction acquires multireference character, requiring more configurations for accurate energy estimates, and highlighting the sensitivity of SQD to several algorithmic elements including the LUCJ circuit, its parametrization via CCSD amplitudes, and device noise (affecting e.g. the quality of the occupation number distribution used

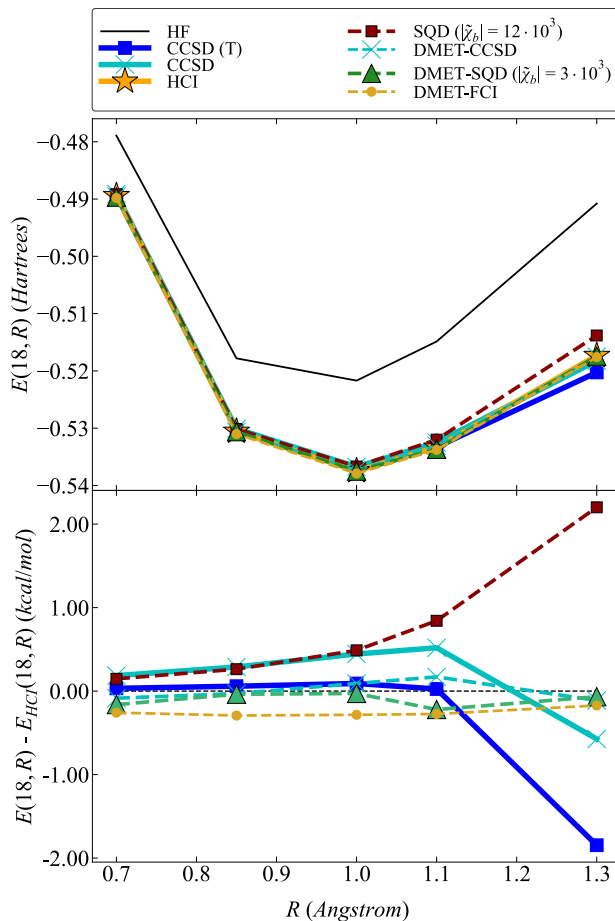


Figure 4.2 Potential energy curve of a hydrogen ring. Top: ground-state energy per atom along the symmetric expansion of a ring of 18 hydrogen atoms using various unfragmented (RHF, CCSD, CCSD(T), SQD, HCI) methods and DMET with CCSD, SQD, and FCI as subsystem solver. Bottom: deviation between HCI and other potential energy curves. Energies per atom are shown.

in the first iteration of S-CORE).

DMET-CCSD and DMET-SQD yield potential energy curves in better agreement with HCI than equivalent unfragmented calculations and with considerably lower non-parallelity error (beneficial in the calculation of energy differences and derivatives) and sub-kcal/mol non-variationality biases. The observed improvements are accounted for by the smaller size of DMET subsystems as compared to the entire hydrogen ring, which alleviates the breakdown of CCSD in DMET-CCSD calculations, and the impact of inefficient configuration sampling in DMET-SQD calculations (the ratio between the number d' of configurations with coefficients above a given threshold, and the total number d of configurations in the SQD wavefunction is higher for DMET-SQD than for unfragmented SQD,

see Table 4.1 and the Appendix).

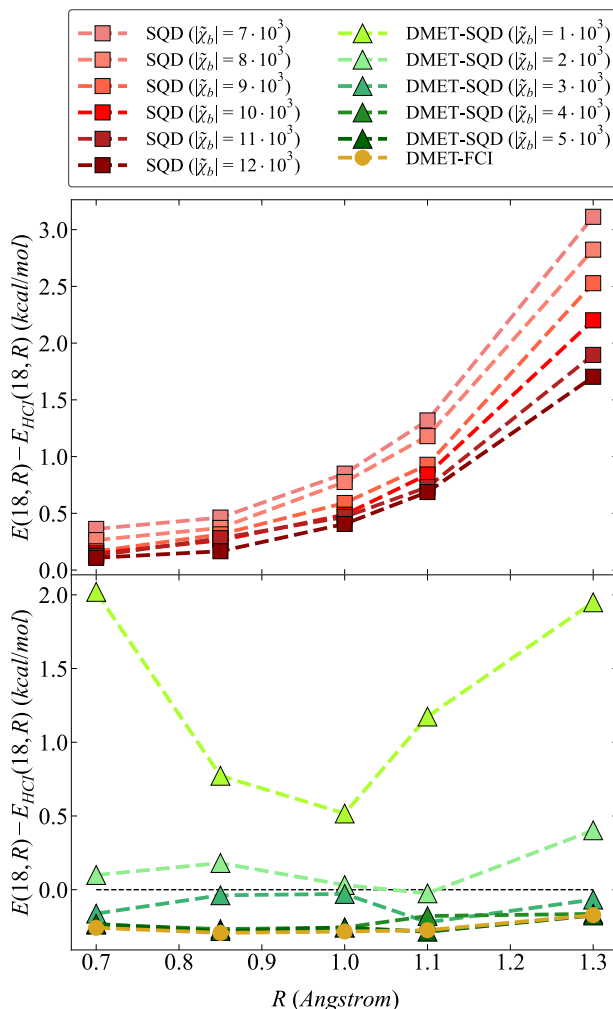


Figure 4.3 SQR performance versus number of configurations. Deviations from HCl for unfragmented SQR (top) and DMET-SQR energies (bottom), along the symmetric expansion of a ring of 18 hydrogen atoms, for several values of $|\tilde{\chi}_b|$.

In Fig. 4.3, we explore the behavior of SQR and DMET-SQR calculations as a function of the number of batch configurations $|\tilde{\chi}_b|$. As naturally expected, increasing $|\tilde{\chi}_b|$ leads to increasing d (see also Table 4.1) and thus to a closer agreement with HCl and DMET-FCI, respectively. Increasing $|\tilde{\chi}_b|$ also reduces fluctuations in SQR energies caused by the randomness of sampled configurations, which are particularly visible in the bottom panel of Fig. 4.3 (light green curve, $|\tilde{\chi}_b| = 1 \cdot 10^3$). The rate at which SQR and DMET-SQR results approach HCl and DMET-FCI counterparts is determined by $|\tilde{\chi}_b|$ through the ratio between d and the dimension of the active

space and through the ratio between d' and d , which is lower for unfragmented H_{18} calculations (see Table 4.1 and the appendix), suggesting inefficient sampling.

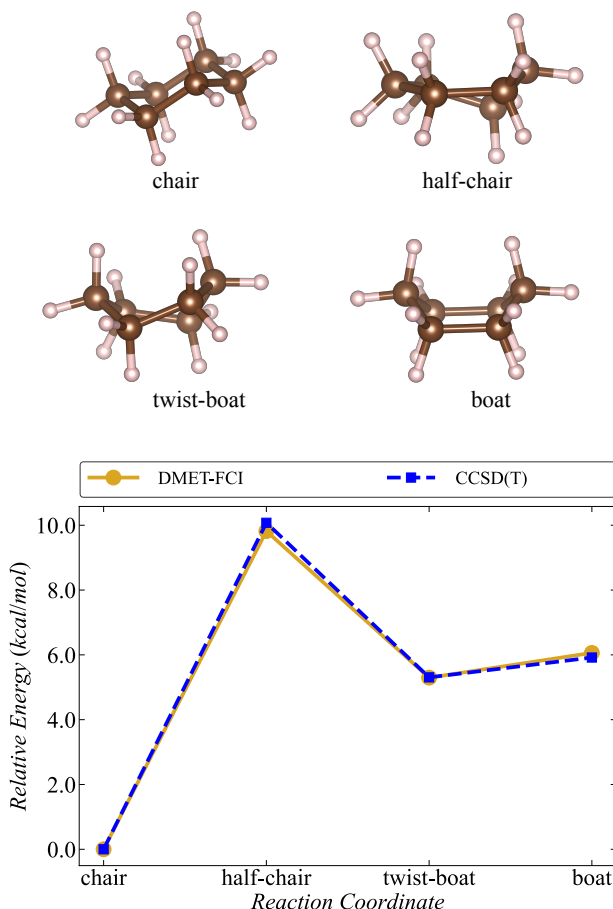


Figure 4.4 Conformations of cyclohexane. Schematic representations of the chair, half-chair, twist-boat, and boat conformations of cyclohexane (top) and their energies relative to the chair using DMET-FCI and unfragmented CCSD(T) (bottom).

We now turn our attention to cyclohexane. This compound can adopt several conformations, illustrated in Fig. 4.4, owing to the flexibility of its carbon-carbon single bonds: (i) in the lowest-energy “chair” conformation, C atoms are positioned so that C-C ring bonds assume energetically favorable angles (thereby minimizing angle strain) and C-H ring bonds are staggered (eliminating torsional strain), (ii) in the less stable “boat” conformation, two C atoms are at the same height, creating steric strain (and hindrance) due to the proximity of two H atoms bound to them (above the boat) and creating torsional strain because eight H atoms (below the boat) are forced into eclipsed positions, (iii) in the “twist-boat” conformation, a slight twisting deformation from the “boat”

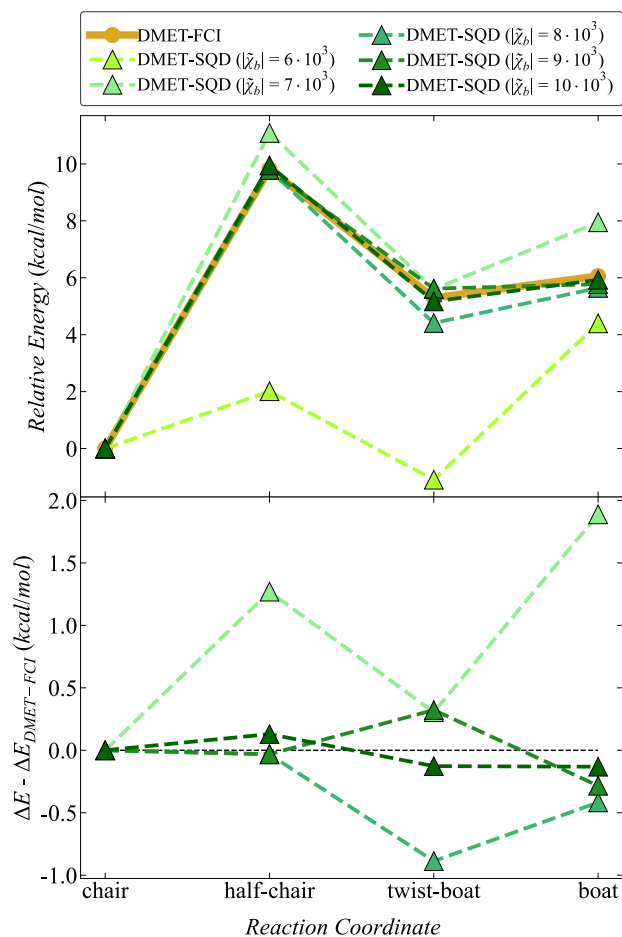


Figure 4.5 DMET-SQD calculations on cyclohexane. Top: comparison between DMET-FCI and DMET-SQD for several values of $|\tilde{\chi}_b|$. Bottom: deviations of DMET-SQD from DMET-FCI.

reduces steric strain by moving H atoms “above the boat” far apart, and eclipsing interactions by moving H atoms “below the boat” into largely (but not completely) staggered positions, (iv) in the “half-chair” conformation, cyclohexane assumes a partially planar shape at the cost of significant ring strain, as in the planar portion the C-C bond angles are forced to energetically unfavorable angles and the corresponding C-H bonds are fully eclipsed.

Computing energy differences between cyclohexane conformations in a qualitatively (i.e. correctly ranking) and quantitatively accurate way is an important methodological test, particularly because these energy differences are of the order of a few kcal/mol, and a necessary step towards richer and more realistic studies. In Fig. 4.4, we compare DMET-FCI with unfragmented CCSD(T) energies of cyclohexane conformations, relative to the chair. As seen, the fragmentation approx-

imation underlying DMET affects energy differences rather modestly, well below 1 kcal/mol. In Fig. 4.5 (top), we compare DMET-FCI and DMET-SQD energy differences for several values of $|\tilde{\chi}_b|$, showing deviations between DMET-FCI and DMET-SQD for each conformation in Fig. 4.4 (bottom). For $|\tilde{\chi}_b| \geq 8 \cdot 10^3$, deviations from DMET-FCI are mostly within 1 kcal/mol and conformations are ordered correctly by DMET-SQD. For $|\tilde{\chi}_b| = 6 \cdot 10^3$, on the other hand, the ordering of conformations is incorrect, with twist-boat predicted to be the most stable geometry and half-chair predicted to be more stable than boat.

4.4 Conclusion

In this study, we presented the first DMET calculations using the quantum computing SQD method as a subsystem solver. We tested the DMET+SQD combination by computing the the ground-state potential energy curve of a ring of 18 hydrogen atoms, a standard benchmark of first-principle electronic structure methods, and the relative energies of the conformers of cyclohexane, a use case of more practical significance to organic chemistry. The relative energies of these conformations do not results from the breaking and formation of chemical bonds, but from the delicate balance of electrostatic interactions between molecular moieties in different geometries. In this aspect, the calculations performed in this work can be considered a severe test of DMET-SQD.

The use of SQD as a subsystem solver allowed the quantum computation of subsystems with more electrons and orbitals than previously possible, by executing fewer and larger quantum circuits (as documented in Fig. 4.1) on a quantum processor assisted by classical HPC resources carrying out pre-, peri-, and post-processing operations (respectively preliminary mean-field calculations and subsystem Hamiltonian construction, S-CORE and the associate subspace diagonalization, and computation of system properties such as particle number and ground-state energy by the collation of results from individual subsystems). In addition, the use of SQD as a subsystem solver allows for higher accuracy and precision than previously possible in studies involving present-day quantum computers, due to the noise-resilience properties of the algorithm²⁹.

It is important to continue to expand the application of DMET-SQD to more complex instances of the electronic structure problem. Continued development in error rates of quantum comput-

ers, error mitigation techniques, and construction and optimization of quantum circuits for SQD applications have the potential to improve the efficiency of sampling computational basis states, leading to more compact eigenvalues problems. This would in turn improve the accuracy and time to solution of DMET-SQD, further extending the range of accessible applications and/or the reliance on computationally expensive HPC peri-processing (especially matrix diagonalization). Further studies building on the work done here, involving chemical species of greater relevance to organic and biological chemistry, and non-minimal basis sets necessary for qualitatively correct and quantitatively accurate results, would also be very desirable.

More generally, our study serves as a proof-of-concept for the concerted use of quantum and classical computers, as complementary elements of QCSC algorithms and architectures, as an innovative and promising mode of attack of correlated many-electron systems by first-principle electronic structure methods, with the ultimate goal to enhance molecular design platforms.

BIBLIOGRAPHY

- [1] David A Bardwell, Claire S Adjiman, Yelena A Arnautova, Ekaterina Bartashevich, Stephan XM Boerrigter, Doris E Braun, Aurora J Cruz-Cabeza, Graeme M Day, Raffaele G Della Valle, Gautam R Desiraju, et al. Towards crystal structure prediction of complex organic compounds—a report on the fifth blind test. *Acta Crystallogr. B*, 67(6):535–551, 2011.
- [2] Jun Yang, Weifeng Hu, Denis Usvyat, Devin Matthews, Martin Schütz, and Garnet Kin-Lic Chan. Ab initio determination of the crystalline benzene lattice energy to sub-kilojoule/mole accuracy. *Science*, 345(6197):640–643, 2014.
- [3] VI Timofeev, RN Chuprov-Netochin, VR Samigina, VV Bezuglov, KA Miroshnikov, and IP Kuranova. X-ray investigation of gene-engineered human insulin crystallized from a solution containing polysialic acid. *Acta Crystallogr. F*, 66(3):259–263, 2010.
- [4] Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt, N. A. Bogdanov, G. H. Booth, J. Chen, and Z.-H. Cui. Recent developments in the pyscf program package. *The Journal of Chemical Physics*, 153(2):024109, 2020.
- [5] Marat Valiev, Eric J Bylaska, Niranjana Govind, Karol Kowalski, Tjerk P Straatsma, Hubertus Johannes Jacobus Van Dam, Dunyou Wang, Jarek Nieplocha, Edoardo Aprà, Theresa L Windus, et al. Nwchem: A comprehensive and scalable open-source solution for large scale molecular simulations. *Comp. Phys. Comm*, 181(9):1477–1489, 2010.
- [6] Kazuo Kitaura, Eiji Ikeo, Toshio Asada, Tatsuya Nakano, and Masami Uebayasi. Fragment molecular orbital method: an approximate computational method for large molecules. *Chem. Phys. Lett*, 313(3-4):701–706, 1999.
- [7] Krishnan Raghavachari and Arjun Saha. Accurate composite and fragment-based quantum chemical models for large molecules. *Chem. Rev*, 115(12):5643–5677, 2015.
- [8] Daniel A Erlanson, Ben J Davis, and Wolfgang Jahnke. Fragment-based drug discovery: advancing fragments in the absence of crystal structures. *Cell Chem. Bio*, 26(1):9–15, 2019.
- [9] Qiming Sun and Garnet Kin-Lic Chan. Quantum embedding theories. *Acc. Chem. Res*, 49(12):2705–2712, 2016.
- [10] Yuri Alexeev, Maximilian Amsler, Marco Antonio Barroca, Sanzio Bassini, Torey Battelle, Daan Camps, David Casanova, Young Jay Choi, Frederic T Chong, Charles Chung, et al. Quantum-centric supercomputing for materials science: A perspective on challenges and future directions. *Future Gener. Comp. Sys*, 160:666–710, 2024.
- [11] Bela Bauer, Dave Wecker, Andrew J Millis, Matthew B Hastings, and Matthias Troyer. Hybrid quantum-classical approach to correlated materials. *Phys. Rev. X*, 6(3):031045, 2016.

- [12] Juha M Kreula, Laura García-Álvarez, Lucas Lamata, Stephen R Clark, Enrique Solano, and Dieter Jaksch. Few-qubit quantum-classical simulation of strongly correlated lattice fermions. *EPJ Quant. Tech*, 3(1):1–19, 2016.
- [13] Sergey Bravyi and David Gosset. Complexity of quantum impurity problems. *Comm. Math. Phys*, 356:451–500, 2017.
- [14] 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(1):4213, 2014.
- [15] Mario Motta and Julia E Rice. Emerging quantum computing algorithms for quantum chemistry. *WIREs Comput. Mol. Sci*, 12(3):e1580, 2022.
- [16] Bela Bauer, Sergey Bravyi, Mario Motta, and Garnet Kin-Lic Chan. Quantum algorithms for quantum chemistry and quantum materials science. *Chem. Rev*, 120(22):12685–12717, 2020.
- [17] Hocheol Lim, Doo Hyung Kang, Jeonghoon Kim, Aidan Pellow-Jarman, Shane McFarthing, Rowan Pellow-Jarman, Hyeon-Nae Jeon, Byungdu Oh, June-Koo Kevin Rhee, and Kyoung Tai No. Fragment molecular orbital-based variational quantum eigensolver for quantum chemistry in the age of quantum computing. *Sci. Rep*, 14(1):2422, 2024.
- [18] Takeshi Yoshikawa, Tomoya Takanashi, and Hiromi Nakai. Quantum algorithm of the divide-and-conquer unitary coupled cluster method with a variational quantum eigensolver. *J. Chem. Theory Comput*, 18(9):5360–5373, 2022.
- [19] Huan Ma, Jie Liu, Honghui Shang, Yi Fan, Zhenyu Li, and Jinlong Yang. Multiscale quantum algorithms for quantum chemistry. *Chem. Sci*, 14(12):3190–3205, 2023.
- [20] Nicholas C. Rubin. A hybrid classical/quantum approach for large-scale studies of quantum systems with density matrix embedding theory. *arXiv preprint*, 2016.
- [21] Y. Kawashima, E. Lloyd, M. P. Coons, Y. Nam, S. Matsuura, A. J. Garza, S. Johri, L. Huntington, V. Senicourt, A. O. Maksymov, et al. Optimizing electronic structure simulations on a trapped-ion quantum computer using problem decomposition. *Communications Physics*, 4(1):245, 2021.
- [22] Naoki Iijima, Satoshi Imamura, Mikio Morita, Sho Takemori, Akihiko Kasagi, Yuhei Umeda, and Eiji Yoshida. Towards accurate quantum chemical calculations on noisy quantum computers. *arXiv:2311.09634*, 2023.
- [23] Weitang Li, Zigeng Huang, Changsu Cao, Yifei Huang, Zhigang Shuai, Xiaoming Sun, Jinzhao Sun, Xiao Yuan, and Dingshun Lv. Toward practical quantum embedding simulation of realistic chemical systems on near-term quantum computers. *Chem. Sci*, 13(31):8953–8962, 2022.

- [24] J. J. M. Kirsopp, C. Di Paola, D. Z. Manrique, M. Krompiec, G. Greene-Diniz, W. Guba, A. Meyder, D. Wolf, M. Strahm, and D. Muñoz Ramo. Quantum computational quantification of protein–ligand interactions. *International Journal of Quantum Chemistry*, 122(22):e26975, 2022.
- [25] Changsu Cao, Jinzhao Sun, Xiao Yuan, Han-Shi Hu, Hung Q Pham, and Dingshun Lv. Ab initio quantum simulation of strongly correlated materials with quantum embedding. *npj Comput. Mater*, 9(1):78, 2023.
- [26] Christian Vorwerk, Nan Sheng, Marco Govoni, Bencheng Huang, and Giulia Galli. Quantum embedding theories to simulate condensed systems on quantum computers. *Nat. Comp. Sci*, 2(7):424–432, 2022.
- [27] Max Rossmannek, Fabijan Pavosevic, Angel Rubio, and Ivano Tavernelli. Quantum embedding method for the simulation of strongly correlated systems on quantum computers. *J. Phys. Chem. Lett*, 14(14):3491–3497, 2023.
- [28] Tanvi P Gujarati, Mario Motta, Triet Nguyen Friedhoff, Julia E Rice, Nam Nguyen, Panagiotis Kl Barkoutsos, Richard J Thompson, Tyler Smith, Marna Kagele, Mark Brei, et al. Quantum computation of reactions on surfaces using local embedding. *npj Quantum Inf*, 9(1):88, 2023.
- [29] J. Robledo-Moreno, M. Motta, H. Haas, A. Javadi-Abhari, P. Jurcevic, W. Kirby, S. Martiel, K. Sharma, S. Sharma, and T. Shirakawa. Chemistry beyond exact solutions on a quantum-centric supercomputer. *arXiv preprint arXiv:2405.05068*, 2024.
- [30] S. Guo, J. Sun, H. Qian, M. Gong, Y. Zhang, F. Chen, Y. Ye, Y. Wu, S. Cao, K. Liu, et al. Experimental quantum computational chemistry with optimized unitary coupled cluster ansatz. *Nature Physics*, 2024.
- [31] L. Zhao, J. Goings, K. Shin, W. Kyoung, J. I. Fuks, J.-K. Kevin Rhee, Y. M. Rhee, K. Wright, J. Nguyen, J. Kim, and S. Johri. Orbital-optimized pair-correlated electron simulations on trapped-ion quantum computers. *npj Quantum Information*, 9(1):60, 2023.
- [32] Google AI Quantum and Collaborators. Hartree-fock on a superconducting qubit quantum computer. *Science*, 369(6507):1084–1089, 2020.
- [33] Renke Huang, Chenyang Li, and Francesco A Evangelista. Leveraging small-scale quantum computers with unitarily downfolded hamiltonians. *PRX Quantum*, 4(2):020313, 2023.
- [34] Danil Kaliakin, Akhil Shajan, Javier Robledo Moreno, Zhen Li, Abhishek Mitra, Mario Motta, Caleb Johnson, Abdullah Ash Saki, Susanta Das, Iskandar Sitdikov, et al. Accurate quantum-centric simulations of supramolecular interactions. *arXiv:2410.09209*, 2024.
- [35] Ieva Liepuoniute, Kirstin D. Doney, Javier Robledo, Joshua A. Job, Will S. Friend, and

- Gavin O. Jones. Quantum-centric study of methylene singlet and triplet states. *arXiv preprint arXiv:2411.04827*, 2024.
- [36] Johannes Hachmann, Wim Cardoen, and Garnet Kin Chan. Multireference correlation in long molecules with the quadratic scaling density matrix renormalization group. *J. Chem. Phys.*, 125(14), 2006.
 - [37] Lorenzo Stella, Claudio Attaccalite, Sandro Sorella, and Angel Rubio. Strong electronic correlation in the hydrogen chain: A variational monte carlo study. *Phys. Rev. B*, 84(24):245117, 2011.
 - [38] Mario Motta, David M Ceperley, Garnet Kin-Lic Chan, John A Gomez, Emanuel Gull, Sheng Guo, Carlos A Jiménez-Hoyos, Tran Nguyen Lan, Jia Li, Fengjie Ma, et al. Towards the solution of the many-electron problem in real materials: Equation of state of the hydrogen chain with state-of-the-art many-body methods. *Phys. Rev. X*, 7(3):031059, 2017.
 - [39] Mario Motta, Claudio Genovese, Fengjie Ma, Zhi-Hao Cui, Randy Sawaya, Garnet Kin-Lic Chan, Natalia Chepiga, Phillip Helms, Carlos Jiménez-Hoyos, Andrew J Millis, et al. Ground-state properties of the hydrogen chain: dimerization, insulator-to-metal transition, and magnetic phases. *Phys. Rev. X*, 10(3):031058, 2020.
 - [40] DHR Barton and RC Cookson. The principles of conformational analysis. *Chem. Soc. Rev.*, 10(1):44–82, 1956.
 - [41] Herbert L Strauss and Herbert M Pickett. Conformational structure, energy, and inversion rates of cyclohexane and some related oxanes. *J. Am. Chem. Soc.*, 92(25):7281–7290, 1970.
 - [42] David A Dixon and Andrew Komornicki. Ab initio conformational analysis of cyclohexane. *J. Phys. Chem.*, 94(14):5630–5636, 1990.
 - [43] Ernest L Eliel. Conformational analysis in heterocyclic systems: recent results and applications. *Angew. Chem., Int. Ed. Engl.*, 11(9):739–750, 1972.
 - [44] Gerald Knizia and Garnet Kin-Lic Chan. Density matrix embedding: A simple alternative to dynamical mean-field theory. *Phys. Rev. Lett.*, 109(18):186404, 2012.
 - [45] Gerald Knizia and Garnet Kin-Lic Chan. Density matrix embedding: A strong-coupling quantum embedding theory. *J. Chem. Theory Comput.*, 9(3):1428–1432, 2013.
 - [46] Sebastian Wouters, Carlos A Jiménez-Hoyos, Qiming Sun, and Garnet K-L Chan. A practical guide to density matrix embedding theory in quantum chemistry. *Journal of chemical theory and computation*, 12(6):2706–2719, 2016.
 - [47] Sebastian Wouters, Carlos A. Jiménez-Hoyos, and Garnet KL Chan. Five years of density matrix embedding theory. *Fragmentation: toward accurate calculations on complex molecular*

systems, pages 227–243, 2017.

- [48] Hung Q Pham, Varinia Bernales, and Laura Gagliardi. Can density matrix embedding theory with the complete activate space self-consistent field solver describe single and double bond breaking in molecular systems? *J. Chem. Theory Comput*, 14(4):1960–1968, 2018.
- [49] Bo-Xiao Zheng and Garnet Kin-Lic Chan. Ground-state phase diagram of the square lattice hubbard model from density matrix embedding theory. *Phys. Rev. B*, 93(3):035126, 2016.
- [50] Qiaoni Chen, George H Booth, Sandeep Sharma, Gerald Knizia, and Garnet Kin-Lic Chan. Intermediate and spin-liquid phase of the half-filled honeycomb hubbard model. *Phys. Rev. B*, 89(16):165134, 2014.
- [51] Ireneusz W Bulik, Weibing Chen, and Gustavo E Scuseria. Electron correlation in solids via density embedding theory. *J. Chem. Phys*, 141(5), 2014.
- [52] K. Kanno, M. Kohda, R. Imai, S. Koh, K. Mitarai, W. Mizukami, and Y. O. Nakagawa. Quantum-selected configuration interaction: Classical diagonalization of hamiltonians in subspaces selected by quantum computers. *arXiv preprint arXiv:2302.11320*, 2023.
- [53] Y. O. Nakagawa, M. Kamoshita, W. Mizukami, S. Sudo, and Y.-y. Ohnishi. Adapt-qsci: Adaptive construction of input state for quantum-selected configuration interaction. *arXiv preprint arXiv:2311.01105*, 2023.
- [54] Mario Motta, William Kirby, Ieva Liepuoniute, Kevin J Sung, Jeffrey Cohn, Antonio Mezzacapo, Katherine Klymko, Nam Nguyen, Nobuyuki Yoshioka, and Julia E Rice. Subspace methods for electronic structure simulations on quantum computers. *Electron. Struct*, 6(1):013001, 2024.
- [55] G Ortiz, JE Gubernatis, E Knill, and R Laflamme. Simulating fermions on a quantum computer. *Comp. Phys. Comm*, 146(3):302–316, 2002.
- [56] Rolando Somma, Gerardo Ortiz, James E Gubernatis, Emanuel Knill, and Raymond Laflamme. Simulating physical phenomena by quantum networks. *Phys. Rev. A*, 65(4):042323, 2002.
- [57] Rolando D Somma. *Quantum computation, complexity, and many-body physics*. PhD thesis, Instituto Balseiro, S.C. de Bariloche, Argentina and Los Alamos National Laboratory, Los Alamos, USA, 2005. arXiv:quant-ph/0512209.
- [58] Stefano Barison, Javier Robledo Moreno, and Mario Motta. Quantum-centric computation of molecular excited states with extended sample-based quantum diagonalization, 2024.
- [59] 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. *Chemical Science*, 14(40):11213–11227, 2023.
- [60] William J Huggins, Jarrod R McClean, Nicholas C Rubin, Zhang Jiang, Nathan Wiebe, K Birgitta Whaley, and Ryan Babbush. Efficient and noise resilient measurements for quantum chemistry on near-term quantum computers. *npj Quantum Inf*, 7(1):23, 2021.
- [61] F. Neese. Software update: The orca program system - version 5.0. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 12(5):e1606, 2022.
- [62] Valentin Senicourt, James Brown, Alexandre Fleury, Ryan Day, Erika Lloyd, Marc P. Coons, Krzysztof Bieniasz, Lee Huntington, Alejandro J. Garza, Shunji Matsuura, Rudi Plesch, Takeshi Yamazaki, and Arman Zaribafiyani. Tangelo: An open-source python package for end-to-end chemistry workflows on quantum computers. *arXiv:2206.12424*, 2022.
- [63] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, and S. Sharma. Pyscf: the python-based simulations of chemistry framework. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 8(1):e1340, 2018.
- [64] Q. Sun. Libcint: An efficient general integral library for gaussian basis functions. *Journal of Computational Chemistry*, 36(22):1664–1671, 2015.
- [65] Peter Fulde and Hermann Stoll. Dealing with the exponential wall in electronic structure calculations. *J. Chem. Phys.*, 146(19), 2017.
- [66] ffsims developers. ffsim: Faster simulations of fermionic quantum circuits, 2024. Accessed: 2024-09-01.
- [67] Gadi Aleksandrowicz, Thomas Alexander, Panagiotis Kl. Barkoutsos, Luciano Bello, Yael Ben-Haim, David Bucher, Francisco Jose Cabrera-Hernández, Jorge Carballo-Franquis, Adrian Chen, Chun-Fu Chen, Jerry M. Chow, Antonio D. Córcoles-Gonzales, Abigail J. Cross, Andrew W. Cross, Juan Cruz-Benito, Chris Culver, Salvador De La Puente González, Enrique De La Torre, Delton Ding, Eugene F. Dumitrescu, Ivan Duran, Pieter T. Eendebak, Mark Everitt, Ismael Faro Sertage, Albert Frisch, Andreas Fuhrer, Jay M. Gambetta, Borja Godoy Gago, Juan Gomez-Mosquera, Donny Greenberg, Ikko Hamamura, Vojtech Havlicek, Joe Hellmers, Łukasz Herok, Hiroshi Horii, Shaohan Hu, Takashi Imamichi, Toshinari Itoko, Ali Javadi-Abhari, Naoki Kanazawa, Anton Karazeev, Kevin Krsulich, Peng Liu, Yang Luh, Yunho Maeng, Manoel Marques, Francisco Martín-Fernández, Douglas McClure, David McKay, Srujan Meesala, Antonio Mezzacapo, Nikolaj Moll, Diego Moreda Rodríguez, Giacomo Nannicini, P. D. Nation, Pauline J. Ollitrault, Lee James O’Riordan, Hanhee Paik, Jesús Pérez, Anna Phan, Marco Pistoia, Viktor Prutyantov, Max Reuter, Julia E. Rice, Abdón Rodríguez Davila, Raymond Harry Rudy, Mingi Ryu, Ninad Sathaye, Chris Schnabel, Eddie Schoute, Kanav Setia, Yunong Shi, Adenilton Silva, Yukio Siraichi, Seyon Sivarajah, John A. Smolin, Mathias Soeken, Hitomi Takahashi, Ivano Tavernelli, Charles Taylor, Pete Taylour, Kenso Trabing, Matthew Treinish, Wes Turner, Desiree Vogt-Lee, Christophe Vuillot, Jonathan A. Wildstrom, Jessica Wilson, Erick Winston, Christopher J. Wood, Stephen P.

- Wood, Stefan Wörner, Ismail Yunus Akhalwaya, and Christa Zoufal. Qiskit: An open-source framework for quantum computing, 2019.
- [68] A. Javadi-Abhari, M. Treinish, K. Krsulich, C. J. Wood, J. Lishman, J. Gacon, S. Martiel, P. D. Nation, L. S. Bishop, and A. W. Cross. Quantum computing with qiskit. *arXiv preprint arXiv:2405.08810*, 2024.
 - [69] J. J. Wallman and J. Emerson. Noise tailoring for scalable quantum computation via randomized compiling. *Physical Review A*, 94(5):052325, 2016.
 - [70] L. Viola and S. Lloyd. Dynamical suppression of decoherence in two-state quantum systems. *Physical Review A*, 58(4):2733, 1998.
 - [71] A. Kofman and G. Kurizki. Universal dynamical control of quantum mechanical decay: modulation of the coupling to the continuum. *Physical Review Letters*, 87(27):270405, 2001.
 - [72] M. J. Biercuk, H. Uys, A. P. VanDevender, N. Shiga, W. M. Itano, and J. J. Bollinger. Optimized dynamical decoupling in a model quantum memory. *Nature*, 458(7241):996–1000, 2009.
 - [73] S. Niu and A. Todri-Sanial. Effects of dynamical decoupling and pulse-level optimizations on ibm quantum computers. *IEEE Transactions on Quantum Engineering*, 3:1–10, 2022.
 - [74] Abdullah Ash Saki, Stefano Barison, Bryce Fuller, James R. Garrison, Jennifer R. Glick, Caleb Johnson, Antonio Mezzacapo, Javier Robledo-Moreno, Max Rossmannek, Paul Schweigert, Iskandar Sitdikov, and Kevin J. Sung. Qiskit addon: sample-based quantum diagonalization, 2024.
 - [75] R. J. Bartlett. Perspective on coupled-cluster theory. the evolution toward simplicity in quantum chemistry. *Physical Chemistry Chemical Physics*, 26(10):8013–8037, 2024.
 - [76] A. A. Holmes, N. M. Tubman, and C. Umrigar. Heat-bath configuration interaction: An efficient selected configuration interaction algorithm inspired by heat-bath sampling. *Journal of Chemical Theory and Computation*, 12(8):3674–3680, 2016.
 - [77] A. A. Holmes, H. J. Changlani, and C. Umrigar. Efficient heat-bath sampling in fock space. *Journal of Chemical Theory and Computation*, 12(4):1561–1571, 2016.
 - [78] J. E. Smith, B. Mussard, A. A. Holmes, and S. Sharma. Cheap and near exact casscf with large active spaces. *Journal of Chemical Theory and Computation*, 13(11):5468–5478, 2017.
 - [79] S. Sharma, A. A. Holmes, G. Jeanmairet, A. Alavi, and C. J. Umrigar. Semistochastic heat-bath configuration interaction method: Selected configuration interaction with semistochastic perturbation theory. *Journal of Chemical Theory and Computation*, 13(4):1595–1604, 2017.
 - [80] Wissam A Al-Saidi, Shiwei Zhang, and Henry Krakauer. Bond breaking with auxiliary-field

- quantum monte carlo. *J. Chem. Phys*, 127(14), 2007.
- [81] Takashi Tsuchimochi and Gustavo E Scuseria. Strong correlations via constrained-pairing mean-field theory. *J. Chem. Phys*, 131(12), 2009.
- [82] Anton V Sinitskiy, Loren Greenman, and David A Mazziotti. Strong correlation in hydrogen chains and lattices using the variational two-electron reduced density matrix method. *J. Chem. Phys*, 133(1), 2010.
- [83] Nan Lin, CA Marianetti, Andrew J Millis, and David R Reichman. Dynamical mean-field theory for quantum chemistry. *Phys. Rev. Lett*, 106(9):096402, 2011.
- [84] David A Mazziotti. Large-scale semidefinite programming for many-electron quantum mechanics. *Phys. Rev. Lett*, 106(8):083001, 2011.
- [85] Ireneusz W Bulik, Thomas M Henderson, and Gustavo E Scuseria. Can single-reference coupled cluster theory describe static correlation? *J. Chem. Theory Comput*, 11(7):3171–3179, 2015.

APPENDIX

ADDITIONAL DATA

In Figs. 4A.1 and 4A.2, we show the number of configurations in an SQD wavefunction $|\psi^{(b)}\rangle = \sum_{m=1}^d c_m |\mathbf{x}_m\rangle$, labeled d , the number of such configurations with $|c_m|^2 \geq 1 \cdot 10^{-5}$ labeled d' , and the Hilbert space dimension, for H_{18} as a function of bondlength using $|\tilde{\chi}_b| = 12 \cdot 10^3$ and for cyclohexane as a function of reaction coordinate using $|\tilde{\chi}_b| = 8 \cdot 10^3$ respectively.

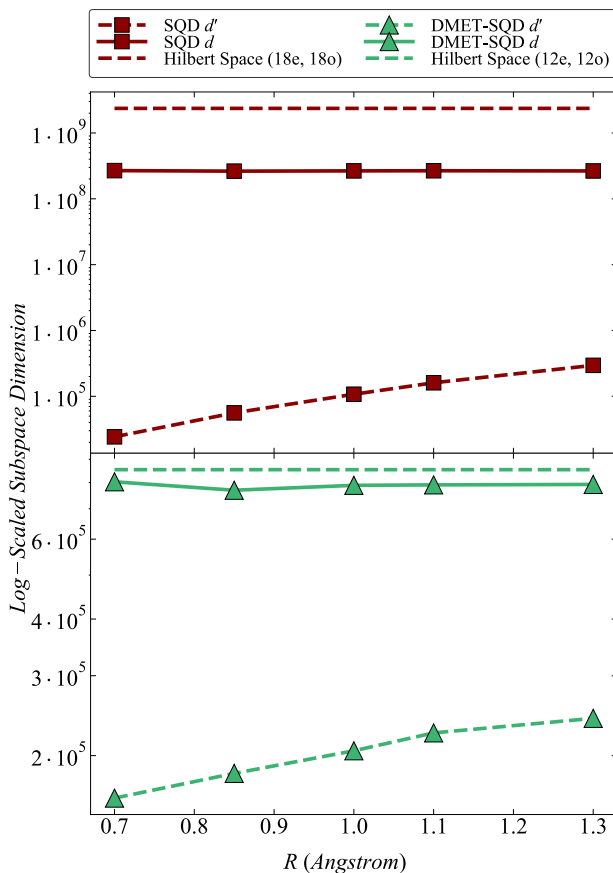


Figure 4A.1 Sparsity of SQD wavefunctions for a hydrogen ring. Hilbert space dimension (dashed line), number d of configurations in the SQD wavefunction (symbols connected by a solid line) and number d' of such configurations with coefficients above $1 \cdot 10^{-8}$ in absolute value squared (symbols connected by a dashed line), for a ring of 18 hydrogen atoms as a function of bondlength, studied with SQD (top, red squares) and DMET-SQD (bottom, green triangles).

In Fig. 4A.1, note how d is relatively close to the Hilbert space dimension and roughly independent of R , whereas d' increases with R , albeit at different rates for SQD and DMET-SQD.

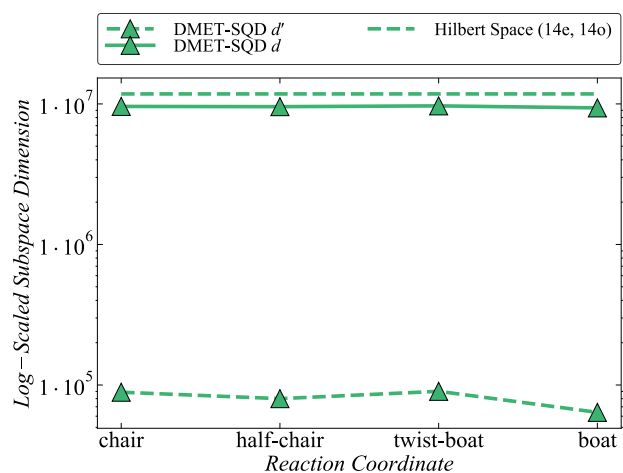


Figure 4A.2 Sparsity of SQD wavefunctions for cyclohexane. Hilbert space dimension (dashed line), number d of configurations in the SQD wavefunction (triangles connected by a solid line) and number d' of such configurations with coefficients above $1.0 \cdot 10^{-8}$ in absolute value squared (triangles connected by a dashed line), for cyclohexane conformations studied with DMET-SQD.