Quantum Simulation of Physical and Chemical Systems Using Noisy Quantum Computers and Machine Learning

Dissertation
zur Erlangung des Grades
des Doktors der Naturwissenschaften
der Naturwissenschaftlich-Technischen Fakultät
der Universität des Saarlandes

von

Tomislav Piskor

Saarbrücken

2023

UNIVERSITÄT
DES
SAARLANDES

Tag des Kolloquiums: 03. Mai 2024

Dekan: Prof. Dr. Michael Vielhaber

Berichterstatter: Prof. Dr. Frank Wilhelm-Mauch
Prof. Dr. Peter P. Orth

Vorsitz: Prof. Dr. Christoph Becher

Akad. Mitarbeiter: Dr. Adam Wysocki

ABSTRACT

The development of quantum computers has ushered in a new epoch of computing. Currently we are in an era, where the hardware is not yet sophisticated enough to produce reliable results. However, quantum hardware is already being used to compute simple problems. In addition, machine learning algorithms are gaining popularity and are being used in various areas. This thesis provides concepts for efficient simulations using quantum algorithms and machine learning methods.

The first part of this thesis gives a brief overview on basic concepts of quantum computing and computational quantum chemistry.

The next part is divided into two chapters and deals with methods for more efficient simulations when using quantum hardware or simulators. The first work describes the differences between numerical and analytical gradients on simple models. The second chapter of this part presents a post-processing method to improve results from quantum simulations.

In the third and last part of this thesis, a method is presented that allows a machine learning model to be extracted from only a small number of data points from two different chemical reactions. It can be seen that with this model, for example, the interpolation of the potential energy surface can be achieved with an error below chemical accuracy.

ZUSAMMENFASSUNG

Die Entwicklung von Quantencomputern hat eine neue Epoche der Datenverarbeitung eingeläutet. Derzeit befinden wir uns noch in einer Ära, in der die Hardware noch nicht ausgereift genug, um zuverlässige Ergebnisse zu erzielen. Allerdings wird Quantenhardware bereits zur Berechnung einfacher Probleme eingesetzt. Darüber hinaus werden Algorithmen des maschinellen Lernens immer beliebter und in verschiedenen Bereichen eingesetzt. In dieser Arbeit werden Konzepte für effiziente Simulationen mit Quantenalgorithmen und Methoden des maschinellen Lernens vorgestellt.

Der erste Teil dieser Arbeit gibt einen kurzen Überblick über die grundlegenden Konzepte des Quantencomputings und der Quantenchemie.

Der nächste Teil ist in zwei Hauptkapitel unterteilt und befasst sich mit Methoden für effizientere Simulationen beim Einsatz von Quantenhardware oder -simulatoren. Die erste Arbeit beschreibt die Unterschiede zwischen numerischen und analytischen Gradienten an einfachen Modellen. Das zweite Kapitel von diesem Teil stellt eine Methode vor, um die Ergebnisse von Quantensimulationen zu verbessern.

Im dritten und letzten Teil dieser Arbeit wird eine Methode vorgestellt, die es ermöglicht, ein Modell für maschinelles Lernen aus einer geringen Anzahl von Datenpunkten von zwei verschiedenen chemischen Reaktionen zu extrahieren. Es zeigt sich, dass mit diesem Modell zum Beispiel die Interpolation der potenziellen Energiefläche mit einem Fehler unterhalb der chemischen Genauigkeit erreicht werden kann.

PUBLICATION LIST

Published

- T. Piskor, J.-M. Reiner, S. Zanker, N. Vogt, M. Marthaler, F. K. Wilhelm, and F. G. Eich
  Using gradient-based algorithms to determine ground-state energies on a quantum computer
  Phys. Rev. A **105**, 062415 (2022)

Submitted

- T. Piskor, F. G. Eich, M. Marthaler, F. K. Wilhelm, and J.-M. Reiner
  Post-processing noisy quantum computations utilising N-representability constraints
  arXiv:2304.13401 (2023)
- T. Piskor, P. Pinski, T. Mast, V. Rybkin
  Multi-level protocol for mechanistic reaction studies using semi-local fitted potential energy surfaces
  arXiv:2304.00942 (2023)

Previous Publications

- I.A. Golovchanskiy, N.N. Abramov, M. Pfirrmann, T. Piskor, J.N. Voss, D.S. Baranov, R.A. Hovhannisyan, V.S. Stolyarov, C. Dubs, A.A. Golubov, V.V. Ryazanov, A.V. Ustinov, and M. Weides
  Interplay of Magnetization Dynamics with a Microwave Waveguide at Cryogenic Temperatures
  Phys. Rev. Applied **11**, 044076 (2019)

ACKNOWLEDGEMENTS

First of all, I would like to thank Michael Marthaler from HQS Quantum Simulations, who made it possible for me to do this thesis within the company and at Saarland University. I would also like to thank Prof. Frank Wilhelm-Mauch, who was my supervisor during this time. It was an honour to be part of the group of such an outstanding researcher.

Most of all, I would like to thank my parents, who have always supported me in good and difficult times. Without their help it would never have been possible to complete this thesis.

Finally, I would like to thank the entire team from HQS Quantum Simulations and especially Jan Reiner for supporting me during my studies.

CONTENT

I Introduction 1

1 Quantum Computing 3

1.1 Quantum Bits 4
1.2 Product and Entangled States 5
1.3 Quantum Gates 6
1.4 Quantum Measurements 8
1.5 The Variational Quantum Eigensolver 9
1.6 Density Matrix 10
1.7 Decoherence 11
1.8 Fermion-Qubit Encoding 14

2 Computational Quantum Chemistry 15

2.1 Fermionic Anticommutation Relations 15
2.2 Chemical Basis Sets 16
2.3 Molecular Hamiltonian and Reduced Density Matrices 18
2.4 Ensemble N-Representability Conditions 19
2.5 Hartree-Fock 20
2.6 Density Functional Theory 21
2.7 Complete Active Space Self Consistent Field 22
2.8 Configuration Interaction 22
2.9 Coupled Cluster 23
2.10 Data Sets within Machine Learning 24

II Improving Simulations on Noisy Quantum Computers 27

3 Analyzing Gradients in Variational Algorithms 29

3.1 The Variational Hamiltonian Ansatz and 1D Hubbard Model 31

3.2 Gradient Descent Optimizer 32
3.3 Circuit Evaluation 35
3.4 Results 36
3.4.1 Simple Circuit 36
3.4.2 Results for the 2-site Hubbard Model 39
3.4.3 Results for the 6-site Hubbard Model 41
3.5 Conclusion 42

Appendices 45
3.A Derivation of the Parameter-Shift Rule 45

4 Imposing N-Representability Constraints on Reduced Density Matrices 47
4.1 N-Representability Conditions for Hole and Particle-Hole Reduced Density Matrices 48
4.2 Application of Noise to Perfect Measurements 49
4.2.1 Shot Noise 50
4.2.2 Influence of Noise on N-Representability 50
4.3 Procedure 51
4.4 Numerical Analysis 54
4.4.1 Simulations without Shot Noise 54
4.4.2 Simulations with Shot Noise 58
4.5 Conclusion 60

Appendices 63
4.A Dephasing and N-Representability Conditions 63

III Data Interpolation for Quantum Chemistry Methods with Machine Learning 67

5 Mechanistic Reaction Studies 69
5.1 Methodology 71
5.1.1 Simulation Protocol 71
5.1.2 Electronic Structure Calculations 72
5.1.3 Data Set Generation 72
5.1.4 Machine Learning Fit of the Semi-Local Reactive PES 72
5.1.5 Property Calculation 73
5.2 Results 73
5.2.1 Bergman Cyclization of Enediyne 74
5.2.2 S_N2 Reaction of Chloromethane with Bromide 77

5.3 Discussion 80
5.4 Conclusions and Outlook 81

Appendices **83**
5.A Reaction Rate Coefficients 83
5.B Harmonic Vibrations 84
  5.B.1 Vibrational Modes for the Bergman Cyclization of Enediyne 84
  5.B.2 Vibrational Modes for the S_N2 Reaction 85
5.C Reaction Coordinate (NEB) for the S_N2 reaction 86

Conclusion **87**

Bibliography **90**

List of Figures

Figure 1.1 Graphical representation of a qubit 5
Figure 1.2 Representation of the Hadamard gate in terms of rotation gates. . 7
Figure 1.3 Graphical illustration of a noisy quantum circuit 13

Figure 3.1 Optimization runs for the simple circuit and $N = 50000$ 38
Figure 3.2 Same plot as Fig. 3.1, but with a depolarization rate of $\gamma = 10^{-4}$. 38
Figure 3.3 Same plot as Fig. 3.2, but with a depolarization rate of $\gamma = 10^{-3}$. . 39
Figure 3.4 Same plot as Fig. 3.2, but with a depolarization rate of $\gamma = 10^{-2}$. . 39
Figure 3.5 Optimization runs for the 2-site Hubbard model and $N = 50000$. 41
Figure 3.6 Same as Fig. 3.5, but with a depolarization rate of $\gamma = 10^{-4}$ 41
Figure 3.7 Optimization runs for the 6-site Hubbard model 43

Figure 4.4.1 Energy errors and fidelities for damping noise 55
Figure 4.4.2 Energy errors and fidelities for depolarizing noise 56
Figure 4.4.3 Energy errors and measurement variances for damping and shot noise 58
Figure 4.4.4 Energy errors and measurement variances for depolarizing and shot noise 59

Figure 5.2.1 Bergman cyclization of enediyne. 74
Figure 5.2.2 Potential energy profile for the Bergman cyclization of enediyne. . 75
Figure 5.2.3 Accuracy of geometric structures for the Bergman cyclization. . . . 75
Figure 5.2.4 IRC for the Bergman cyclization of enediyne. 76
Figure 5.2.5 S_N2 reaction of chloromethane with bromide. 77
Figure 5.2.6 Potential energy profile for the S_N2 reaction. 78
Figure 5.2.7 Accuracy of geometric structures for the S_N2 reaction. 78
Figure 5.2.8 IRC for the S_N2 reaction of chloromethane with bromide. 79
Figure 5.C.1PES comparison between CCSD and DFT (PBE). 86

To my parents.

Part I

Introduction

CHAPTER 1

QUANTUM COMPUTING

Classical computers as we know them today are powerful tools to determine and simulate a large number of problems. One of the pioneers in this field was Konrad Zuse, who developed the world's first programmable computer in 1941 called the Z3. The Z3, which used vacuum tubes to perform calculations, was the first computer to use binary arithmetic and floating-point numbers, as well as electromechanical relays for processing. Since its development in 1947, transistors replaced vacuum tubes and made computers smaller, more reliable and less expensive to build. This invention led to the development of the first commercial computers in the 1950s, such as the IBM 701 and the UNIVAC I. With the birth of the integrated circuit in the 1960s and 1970s, even smaller and more powerful computers came to life. In 1965 Gordon Moore predicted that the number of transistors on a chip would double every two years, a prediction that came to be known as Moore's law. This prediction has largely held true, leading to an increase in computing power and a decrease in cost.

In today's computers we already observe chips containing billions of transistors using very sophisticated fabrication methods. However, it gets more difficult to fabricate chips that follow Moore's law and thus the computing power might reach a saturating point in a few years. In the 1980s, Richard Feynman proposed the idea of using quantum computers to simulate quantum systems, which are difficult or impossible to simulate on classical devices [1]. With the establishment of this idea important quantum algorithms such as Shor's algorithm for factoring large numbers [2] and the Deutsch-Jozsa algorithm for finding entries in large databases [3] have been developed.

Recent advances in quantum computing hardware [4–6] have led to an increasing interest in developing quantum algorithms for noisy intermediate scale quantum (NISQ) computers. One of the most promising applications for these near-term quantum computers is the simulation of fermionic systems, such as lattice systems or small molecules. The determination of the ground state and its energy is one of the key aspects to gain information about the system. Hybrid quantum-classical algorithms, such as the Variational Quantum Eigensolver (VQE) [7] turn out to have lower resource requirements – compared to for instance Shor's algorithm – making them suitable for NISQ computers. These algorithms determine the ground state energy of a system via preparing a parametrized trial state on a quantum computer and optimizing the parameter set with a classical optimization routine.

Even though we are still in the era of noisy intermediate scale quantum devices [8], a number of experiments showed highly promising results regarding possible advantages of quantum computing over conventional computing [4, 5, 9–11], with one of the most promising field of future application being in the realm of quantum simulations, e.g., of materials or chemical systems [12, 13], with impressive recent demonstrations [14–21].

Yet, at the current stage, quantum resources are fairly limited. Despite tremendous achievements concerning quantum error correction [22, 23], current qubit systems are too small for the application of quantum error correction on a sufficient scale [24–26], leading to erroneous results due to decoherence. Furthermore, error-prone measurements are another source of inaccuracy for algorithms discussed for NISQ devices; particularly for variational algorithms [7, 27–29] where a large number of measurements needs to be taken during the challenging optimization procedure of the algorithm's variational parameters [30–34].

Despite these challenges, Part II of this thesis will provide solutions on how to deal with gradient calculations and post-processing on a noisy quantum computer. This chapter gives a brief introduction to some important topics within quantum computing, such as quantum bits, entanglement, quantum operations and decoherence.

1.1 Quantum Bits

Quantum bits, or qubits, are the basic units of a quantum computer. Unlike the classical bit, which can only exist in one of two states, 0 or 1, qubits can exist in a superposition of both states.

Qubits can be implemented in a variety of physical systems such as superconducting circuits [4, 35, 36] or trapped ions [25, 37]. Each of these types of qubits has its own strengths and weaknesses, such as fast gate and moderate coherence times for superconducting qubits and vice versa for trapped ions. However, improvements are constantly being made to make these qubits more stable, easier to control and less sensitive to environmental noise.

The state of a qubit can be generally written as:

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle\,, \tag{1.1}$$

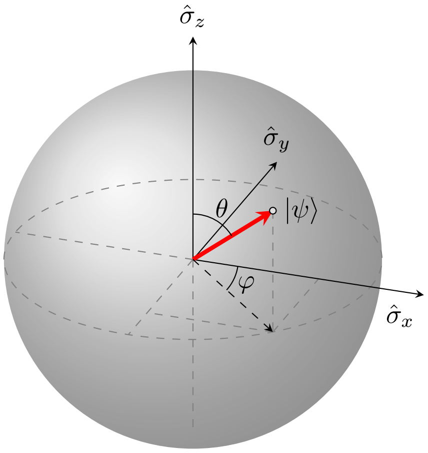


FIGURE 1.1: Graphical representation of a qubit in an arbitrary pure state $|\psi\rangle$, which can be completely described by the two angles θ and φ.

where α and β are complex values. Here, we consider computational basis states with $|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $|1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. As we will see later, these parameters describe the probability amplitude of a qubit being either in the $|0\rangle$ or $|1\rangle$ state.

If we consider the Bloch sphere representation as in Fig. 1.1, the state of a qubit can be described in terms of its polar angle θ and azimuthal angle φ by the following equation

$$|\psi\left(\theta, \varphi\right)\rangle = \cos\left(\frac{\theta}{2}\right) |0\rangle + e^{i\varphi} \sin\left(\frac{\theta}{2}\right) |1\rangle \, . \tag{1.2}$$

1.2 Product and Entangled States

If two or more qubits interact with each other we can distinguish between product and entangled states. Product states can be written as a tensor product of single qubit states, such as the following example:

$$\frac{1}{\sqrt{2}} \left(|00\rangle + |10\rangle\right) = \frac{1}{\sqrt{2}} \left(|0\rangle + |1\rangle\right) \otimes |0\rangle \, . \tag{1.3}$$

This is not possible for entangled states, where the famous Bell states [38] are an example for two qubits:

$$\frac{1}{\sqrt{2}} \left(|01\rangle + |10\rangle\right), \quad \frac{1}{\sqrt{2}} \left(|00\rangle + |11\rangle\right). \tag{1.4}$$

While product states could be represented classically in an efficient way for a small number of qubits this would not be possible for entangled states. Consequently, in order to create an entangled state between two qubits, one would require a two-qubit entangling gate (called controlled NOT or `CNOT` which will be discussed in the next section).

It is simple to construct the states in Eq. 1.4 by solely using the so-called `CNOT` and Hadamard gate. Another prominent example for entangled states is the three-qubit GHZ state [39], which is similar to the aforementioned Bell state:

$$\frac{1}{\sqrt{2}}\left(|000\rangle + |111\rangle\right). \tag{1.5}$$

The entanglement of qubits leads to an exponential increase in the Hilbert space, which can be used as a computational resource. An entanglement of 50 qubits would already lead to a Hilbert space of dimension 2^{50}, which would require a memory of several Petabyte on a classical supercomputer to simulate this number of qubits, while adding only one more qubit would double the amount of memory. Consequently, this feature allows quantum computers to be much faster than their classical counterparts for certain types of algorithms and even solve problems that would not be possible on a modern supercomputer.

1.3 Quantum Gates

Quantum gates are the basic building blocks of quantum circuits, much as logic gates are the building blocks of classical circuits in today's computers. A quantum gate is a unitary transformation that can usually operate on one or two qubits to manipulate their states, such as flipping a qubit's state or adding a phase to it and acts on the state of a qubit register. These operations can be represented as matrices within the computational basis with $|0\rangle = \begin{pmatrix}1\\0\end{pmatrix}$ and $|1\rangle = \begin{pmatrix}0\\1\end{pmatrix}$ for the single-qubit case, and in the two-qubit case the tensor-product basis $|0\rangle \otimes |0\rangle, |0\rangle \otimes |1\rangle, |1\rangle \otimes |0\rangle$, and $|1\rangle \otimes |1\rangle$. Assuming we have a universal gate set, each unitary operation acting on n qubits can be divided into single- and two-qubit gates. A typical example for a universal gate set is an entangling gate, such as `CNOT` or `CZ`, and single-qubit rotations (as defined in Eq. (1.7)). Note that these rotations can have arbitrary angles, making them not a member of the Clifford group. If we would solely consider Clifford gates, we could efficiently simulate our systems classically (according to the Gottesman-Knill theorem [40]) and no quantum computer would be required.

In the following, a few examples of one- or two-qubit gates will be presented:

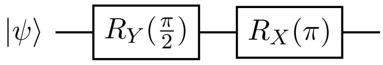


FIGURE 1.2: Representation of the Hadamard gate in terms of rotation gates.

- `Pauli` - For instance the `X` gate that flips the qubit state or the `Z` gate that adds a phase to the qubit. The matrix representations are as follows

$$\mathtt{X} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \mathtt{Y} = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \mathtt{Z} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{1.6}$$

- `Single-qubit rotation gates` - Moreover, we can use rotation gates which are especially important when it comes to compiling quantum circuits. Rotation gates have the following structure:

$$R_j(\theta) = e^{-\frac{i\theta}{2}\sigma_j}, \tag{1.7}$$

  where θ is a tunable parameter and σ_j is one of the Pauli matrices.

- `Hadamard` - The Hadamard gate is also known as the entangling gate and puts the qubit in a superposition of the ground and excited state.

$$\mathtt{H} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{1.8}$$

  Depending on the initial state of the qubit, we find

$$\mathtt{H}\,|0\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \quad \mathtt{H}\,|1\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle). \tag{1.9}$$

  If we for instance want to transform the Hadamard gate in terms of rotation gates up to a global phase, we could apply a 90° rotation around the σ_y axis followed by a 180° rotation around the σ_x axis. Note, that this applies a basis change, where the eigenvectors of Z are transformed into the eigenvectors of X, and vice versa. The corresponding circuit is illustrated in Fig. 1.2.

- `CNOT` - The `CNOT` or `CX` gate flips the qubit state of the target qubit when the control qubit is in the $|1\rangle$ state.

$$\mathtt{CNOT} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{1.10}$$

- `CZ` - The `CZ` gate adds a phase to the entire state if the control and target qubit are both in the $|1\rangle$ state.

$$\mathtt{CZ} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \tag{1.11}$$

1.4 Quantum Measurements

One of the most important paradigms of quantum mechanics is the quantum measurement postulate. These are described by a collection $\{M_m\}$ of measurement operators and are operators acting on the state space of the system being measured [38]. The index m indicates the possible measurement outcome within the experiment. If we consider that the state of the investigated quantum system immediately before the measurement is $|\psi\rangle$ then the probability that the result m occurs is given by

$$p(m) = \langle\psi| M_m^\dagger M_m |\psi\rangle . \tag{1.12}$$

The state of the system after the measurement is

$$\frac{M_m |\psi\rangle}{\langle\psi| M_m^\dagger M_m |\psi\rangle}. \tag{1.13}$$

The measurement operators satisfy the completeness equation,

$$\sum_m M_m^\dagger M_m = I, \tag{1.14}$$

which also states that probabilities sum to one:

$$1 = \sum_m p(m) = \sum_m \langle\psi| M_m^\dagger M_m |\psi\rangle . \tag{1.15}$$

To illustrate this postulate better, we will perform a measurement of one qubit in the computational basis with two outcomes defined by the two measurement operators $M_0 = |0\rangle\langle 0|$ and $M_1 = |1\rangle\langle 1|$. Each measurement operator is Hermitian, and we find $M_0^2 = M_0$ and $M_1^2 = M_1$. In addition, the completeness relation is fulfilled, $I = M_0^\dagger M_0 + M_1^\dagger M_1 = M_0 + M_1$. The probability of obtaining measurement outcome 0 measuring the state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ is

$$p(0) = \langle\psi| M_0^\dagger M_0 |\psi\rangle = \langle\psi| M_0 |\psi\rangle = |\alpha|^2 . \tag{1.16}$$

Equivalently, the probability of measuring 1 is $p(1) = |\beta|^2$. The states after the measurement for the two cases is given by

$$\begin{aligned} \frac{M_0 |\psi\rangle}{|\alpha|} &= \frac{\alpha}{|\alpha|} |0\rangle , \\ \frac{M_1 |\psi\rangle}{|\beta|} &= \frac{\beta}{|\beta|} |1\rangle . \end{aligned} \tag{1.17}$$

Multipliers, such as $\frac{\alpha}{|\alpha|}$ can be ignored since they are simply a global phase factor which can be ignored. Hence, the two post-measurement states are effectively $|0\rangle$ and $|1\rangle$.

It should be emphasized that while we stay in the computational basis the measurement of other operators than the ones described above or the Pauli Z matrix requires a decomposition into sums of Pauli operators, where the single terms of these operators have to be determined via basis rotations, e.g., Pauli X can be measured by applying a Hadamard and then measuring in the Z basis, since the Hadamard gate transform between the X and Z basis as discussed above. This is why in Part II of this work we will see that a lot of measurements have to be performed due to these basis rotations, since we have a large number of Pauli products.

1.5 The Variational Quantum Eigensolver

One important type of algorithms within the NISQ era are so-called variational algorithms such as the VQE. They start from a trial state given by

$$|\psi(\boldsymbol{\theta})\rangle = \hat{U}(\boldsymbol{\theta}) |\psi_0\rangle , \tag{1.18}$$

with $\boldsymbol{\theta}$ defining the parameter set, $\hat{U}(\boldsymbol{\theta})$ being a parametrized unitary operator, which is implemented by a quantum circuit and $|\psi_0\rangle$ an initial state usually chosen to be a single Slater determinant of orbitals obtained from the mean-field solution. The energy of this trial state for a system with Hamiltonian $\hat{\mathcal{H}}$ can then be written as

$$E(\boldsymbol{\theta}) = \langle\psi(\boldsymbol{\theta})| \hat{\mathcal{H}} |\psi(\boldsymbol{\theta})\rangle . \tag{1.19}$$

Minimizing the energy with respect to the parameter set leads to the optimal energy attainable with the Ansatz and the corresponding parameter set $\boldsymbol{\theta}_{\text{opt}}$ satisfying

$$E(\boldsymbol{\theta}_{\text{opt}}) = \min_{\boldsymbol{\theta}} E(\boldsymbol{\theta}). \tag{1.20}$$

According to the Rayleigh-Ritz variational principle [41, 42], the exact ground state of the Hamiltonian E_{exact} sets a lower bound for the variationally determined energy

$$E_{\text{exact}} \leq E(\boldsymbol{\theta}_{\text{opt}}). \tag{1.21}$$

1.6 Density Matrix

The density matrix is defined as

$$\rho = \sum_i p_i \left|\psi_i\right\rangle \left\langle\psi_i\right|, \tag{1.22}$$

with p_i describing the probabilities for a system being in the corresponding state ψ_i. The sum of all probabilities has to be equal to 1, i.e. $\sum_i p_i = 1$. A system whose state can be represented by one single vector within the Hilbert space is defined as a pure state [38] and can be written as:

$$\rho = \left|\psi\right\rangle \left\langle\psi\right|. \tag{1.23}$$

The properties of a density matrix are as follows:

- The trace of a density matrix is equal to one: $\text{Tr}[\rho] = 1$.
- Density matrices are positive semidefinite, meaning that all eigenvalues are non-negative: $\rho \succcurlyeq 0$.
- They are Hermitian, i.e. $\rho^\dagger = \rho$.

A further property occurs for pure states which is idempotency, meaning that the squared density matrix is equal to the density matrix (i.e. $\rho^2 = \rho$). We can represent the state defined in Eq. (1.1) in terms of a density matrix:

$$\rho = \begin{pmatrix} |\alpha|^2 & \alpha^*\beta \\ \beta^*\alpha & |\beta|^2 \end{pmatrix}. \tag{1.24}$$

If we now recall the trace property of a density matrix, we find that $|\alpha|^2 + |\beta|^2 = 1$. These are the probability amplitudes for finding the qubit either in state $|0\rangle$ with probability $|\alpha|^2$ or in state $|1\rangle$ with probability $|\beta|^2$.

In case of mixed states, we find that the density matrix is not idempotent anymore which is due to the "tracing out" a part of the system. If we want to describe the effect of decoherence (see next section) we use density matrices, where the environment causing the noise has been traced out. For this we need mixed states and hence density matrices.

The expectation value $\langle M \rangle$ of an operator M can be calculated via

$$\langle M \rangle = \mathrm{Tr}\left(\rho M\right), \tag{1.25}$$

which is how one would describe the outcome of a quantum measurement. In the case of a pure state $\rho = |\psi\rangle \langle\psi|$, this simplifies to

$$\mathrm{Tr}\left(\rho M\right) = \langle\psi| M |\psi\rangle, \tag{1.26}$$

analogously to Eq. (1.16).

1.7 Decoherence

NISQ computers face with decoherence issues, which occur due to the interaction of the qubits with the environment. The following sections present two different formalisms that describe the change of a one-qubit density matrix under influence of different decoherence types.

To describe the effect of decoherence noise we make use of the superoperator formalism after vectorizing the density matrix [43, 44]. For each noise type there exists a corresponding superoperator, which will be defined in the following. Assuming a general one qubit density matrix, we can transform it to a vectorized form

$$\rho_1 = \begin{pmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{pmatrix} \rightarrow \vec{\rho}_1 = \begin{pmatrix} \rho_{00} \\ \rho_{10} \\ \rho_{01} \\ \rho_{11} \end{pmatrix}. \tag{1.27}$$

As a next step, we apply the one-qubit superoperator $\mathcal{L}$ to the vectorized density matrix,

$$\vec{\rho}_1' = \mathcal{L}\vec{\rho}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1-2p & 0 & 0 \\ 0 & 0 & 1-2p & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \rho_{00} \\ \rho_{10} \\ \rho_{01} \\ \rho_{11} \end{pmatrix} = \begin{pmatrix} \rho_{00} \\ (1-2p)\rho_{10} \\ (1-2p)\rho_{01} \\ \rho_{11} \end{pmatrix}, \tag{1.28}$$

where here we used the dephasing superoperator as an example, with dephasing probability $p = \frac{1}{2}\left(1 - e^{-2\Gamma t}\right)$, where Γ denotes the dephasing rate and t the time the noise acted on the system. Transforming Eq. (1.28) back to matrix form yields

$$\rho_1' = \begin{pmatrix} \rho_{00} & (1-2p)\rho_{01} \\ (1-2p)\rho_{10} & \rho_{11} \end{pmatrix}. \tag{1.29}$$

The superoperator formalism can be extended to multiple qubits. Here, one vectorizes the multi-qubit density matrix and expands the superoperator such that it acts on the subspace of the respective qubit. Throughout this work, we assume that each noise type affects all qubits equally. Hence, we sequentially apply the superoperators acting each individual qubits, with equal noise rates for all qubits.

Dephasing was just used as an example with the superoperator given in Eq. (1.28). It can be understood as random phase errors, i.e., Pauli Z applications on qubits with a certain rate. Averaging over many random instances yields a density matrix equivalent to using the superoperator formalism.

The effect of *damping noise* can be seen as a de-excitation of a qubit with a certain probability, i.e., the qubit – initially being in the excited state $|1\rangle$ – decays after a certain amount of time to the ground state $|0\rangle$. The superoperator for the damping channel is given as:

$$\mathcal{L} = \begin{pmatrix} 1 & 0 & 0 & p \\ 0 & \sqrt{1-p} & 0 & 0 \\ 0 & 0 & \sqrt{1-p} & 0 \\ 0 & 0 & 0 & 1-p \end{pmatrix}, \tag{1.30}$$

with probability p being defined as $p = 1 - e^{-\Gamma t}$. Here, Γ denotes the damping rate.

Depolarization can be seen as bit and phase flip errors acting on the qubits. The superoperator representation of this noise gate is defined as:

$$\mathcal{L} = \begin{pmatrix} 1-\frac{2}{3}p & 0 & 0 & \frac{2}{3}p \\ 0 & 1-\frac{4}{3}p & 0 & 0 \\ 0 & 0 & 1-\frac{4}{3}p & 0 \\ \frac{2}{3}p & 0 & 0 & 1-\frac{2}{3}p \end{pmatrix}, \tag{1.31}$$

with the depolarizing probability given as $p = \frac{3}{4}\left(1 - e^{-\Gamma t}\right)$, and the depolarizing rate Γ. Note, that for our numerical simulations in the following we set the evolution time to $t = 1$ and scale the rates Γ accordingly in dimensionless units.

Another method to describe (for instance depolarizing) noise is via the Kraus operator formalism. Since a quantum device can effectively be considered as an open quantum system, i.e., the qubits on the quantum chip interacting with the environment, we simulate the execution of the circuits as follows: the quantum circuit is arranged in such a way, that as many gates as possible are executed in parallel. After that the noise gate is

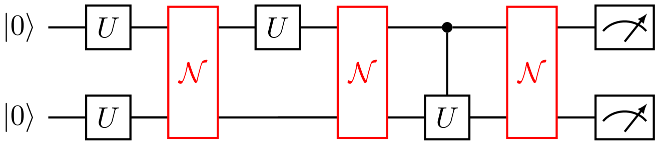


FIGURE 1.3: Graphical illustration of a simple two-qubit circuit with one- and two-qubit unitary operations U and three noise gates, labeled by $\mathcal{N}$, describing the effect of depolarization.

applied to all qubits in the circuit, which can effectively be described as the application of Kraus operators to all qubits. This procedure is repeated until the measurement of observables is performed and is illustrated in Fig. 1.3. The noise channel $\hat{\mathcal{N}}$, acting on a single qubit, can be described as [45]

$$\hat{\mathcal{N}}(\hat{\rho}) = \sum_{i=0}^{3} \hat{K}_i \hat{\rho} \hat{K}_i^{\dagger}, \tag{1.32}$$

with $\hat{\rho}$ being the density matrix representing the quantum system. The Kraus operators $\hat{K}_i$ are defined as follows:

$$\hat{K}_0 = \sqrt{1 - \frac{3}{4}\Gamma} \hat{I}, \tag{1.33a}$$

$$\hat{K}_i = \frac{\sqrt{\Gamma}}{2} \hat{\sigma}_i, i \in \{1, 2, 3\}, \tag{1.33b}$$

with $\Gamma = 1 - e^{-\gamma}$ defining the depolarizing term and $\hat{\sigma}_i$ being the Pauli operators. It can be easily seen that in the noiseless case, i.e. $\gamma = 0$, the depolarizing term Γ vanishes and only the Kraus operator with the identity operator $\hat{I}$ is left, leading to a noiseless quantum operation $\hat{\mathcal{N}}(\hat{\rho}) = \hat{\rho}$. In the noisy case with $\gamma > 0$, not only $\hat{K}_0$, but also the three other Kraus operators $\hat{K}_i$ including the Pauli operators are taken into account, so that the quantum operation $\hat{\mathcal{N}}$ induces depolarization in the system. The parameter γ, determining the strength of the noise, is estimated as the ratio of the gate time T_g and the coherence time T_2 of a qubit.

The gate time T_g represents the time it takes to perform a given operation, for instance the duration of a microwave pulse for controlling superconducting qubits. In Fig. 1.3 the gate time can be viewed as the width of a single gate column, if we interpret the horizontal direction of the algorithm as time axis. An example for a gate operation is a qubit flip, where a qubit prepared in the $|0\rangle$ state is flipped to the $|1\rangle$ state after applying an X gate on that qubit. For superconducting architectures, usual single and two-qubit gate times range from $5 - 500\,\text{ns}$ [35, 46], whereas trapped-ion designs show single-qubit gate times in the microseconds and two-qubit gate times in the $10 - 100\,\mu\text{s}$ regime [37, 47]. The depolarization time T_1 determines the time it takes for a *undesired* qubit flip to occur, due to the coupling of the qubit to its environment. Hence, if we would like to perform N operations on a qubit, we require $NT_g \ll T_1$ or equivalently $\gamma \ll \frac{1}{N}$. In passing we note that the dephasing time T_2 is limited by the relaxation time T_1 of the qubit, i.e., $T_2 \leq 2 \cdot T_1$. While typical coherence times have values in the order of $100\,\mu$s for superconducting qubits [35, 36, 48, 49], these values are several orders of magnitude higher for ion trap devices, which can be several seconds [37]. Considering these numbers, the ratio γ, which characterizes the noise strength, in our simulations is chosen within

$$\gamma = \frac{T_g}{T_2} \in \left[10^{-4}, 10^{-2}\right]. \tag{1.34}$$

1.8 Fermion-Qubit Encoding

The Jordan-Wigner (JW) transformation is one way to encode a fermionic problem into qubits. This is necessary to faithfully express fermionic operators, with anti-commutation relations (which are defined in the next chapter in Eq. (2.1)), in terms of qubit operators, that are spin degrees of freedom and intrinsically obey bosonic commutation relations which are given by

$$\begin{aligned} [a_p, a_q^\dagger] &= a_p a_q^\dagger - a_q^\dagger a_p = \delta_{pq}, \\ [a_p, a_q] &= [a_p^\dagger, a_q^\dagger] = 0. \end{aligned} \tag{1.35}$$

with bosonic annihilation and creation operators a_p, $a_p^\dagger$. The transformations between the fermionic annihilation and creation operator and the qubit representation are given as

$$c_k = Z_0 \otimes \cdots \otimes Z_{k-1} \otimes \sigma_k^+, \qquad \sigma_k^+ = \frac{1}{2}\left(X_k + \mathrm{i}Y_k\right), \tag{1.36a}$$

$$c_k^\dagger = Z_0 \otimes \cdots \otimes Z_{k-1} \otimes \sigma_k^-, \qquad \sigma_k^- = \frac{1}{2}\left(X_k - \mathrm{i}Y_k\right). \tag{1.36b}$$

In these equations, X_k, Y_k, and Z_k are the Pauli matrices of qubit k. The σ_k^+ and σ_k^- operators change the occupation number of the kth orbital, while the string of Z operators recovers the exchange phase factor [42].

Due to its simplicity the JW encoding is a prominent example for the mapping of fermions to qubits. It would only require M qubits to store a wavefunction with 2^M computational basis states. One downside of this encoding scheme lies in the parity. It is stored non-locally, and it would take $O(M)$ (with M being the number of orbitals or qubits) operations to apply a fermionic operator.

In the next section, we will introduce fermionic anticommutation relations which are fulfilled for this encoding scheme.

Chapter 2

Computational Quantum Chemistry

While quantum computers offer a wide range of applications, quantum chemistry is one prominent field providing numerous promising results even on today's devices, raising hope that this research area will offer meaningful results in the near-term future and also in the quantum error correcting era.

This chapter provides a brief overview of important quantum chemistry methods, such as the concept of reduced density matrices and mean-field methods like Hartree-Fock.

2.1 Fermionic Anticommutation Relations

With the creation operator $c_p^\dagger$ electrons are excited into the single electron orbitals and annihilated with the annihilation operator c_p. These operators obey fermionic anticommutations relations [42]:

$$\begin{aligned} \{c_p, c_q^\dagger\} &= c_p c_q^\dagger + c_q^\dagger c_p = \delta_{pq}, \\ \{c_p, c_q\} &= \{c_p^\dagger, c_q^\dagger\} = 0. \end{aligned} \tag{2.1}$$

In the following we assume that a wavefunction of a system can be represented as

$$\psi(\mathbf{x}_0 \dots \mathbf{x}_{\mathbf{N}-1}) = \left| f_{M-1}, \dots, f_p, \dots, f_0 \right\rangle, \tag{2.2}$$

where $\mathbf{x_i} = (\mathbf{r_i}, s_i)$ is the spatial and spin coordinate of the i^{th} electron, N is the number of electrons and M the number of orbitals in the system. $f_p = 1$ when the p^{th} orbital ϕ_p is occupied, and $f_p = 0$ when ϕ_p is unoccupied. The vector $|f\rangle$ is known as an occupation number vector and can be referred to as a Slater determinant.

The actions of the fermionic operators on the determinants $|f\rangle$ are given by

$$\begin{aligned} c_p \left|f_{M-1}, f_{M-2}, \ldots, f_0\right\rangle &= \delta_{f_p,1}(-1)^{\sum_{i=0}^{p-1} f_i} \left|f_{M-1}, f_{M-2}, \ldots, f_p \oplus 1, \ldots, f_0\right\rangle, \\ c_p^\dagger \left|f_{M-1}, f_{M-2}, \ldots, f_0\right\rangle &= \delta_{f_p,0}(-1)^{\sum_{i=0}^{p-1} f_i} \left|f_{M-1}, f_{M-2}, \ldots, f_p \oplus 1, \ldots, f_0\right\rangle, \end{aligned} \tag{2.3}$$

where $\oplus$ is the addition modulo 2 and the phase term $(-1)^{\sum_{i=0}^{p-1} f_i}$ enforces the exchange anti-symmetry of fermions. From the creation and annihilation operator, we can define the orbital occupation operator, which counts the number of electrons in a given orbital:

$$\begin{aligned} n_i &= c_i^\dagger c_i, \\ n_i \left|f_{M-1}, \ldots, f_i, \ldots, f_0\right\rangle &= f_i \left|f_{M-1}, \ldots, f_i, \ldots, f_0\right\rangle. \end{aligned} \tag{2.4}$$

Using these operators, it is possible to write down any Hamiltonian within the so-called second quantization. An example for such a Hamiltonian can be found in Eq. (2.9).

2.2 Chemical Basis Sets

Depending on the size of the system within quantum chemistry calculations there is a variety of basis sets that can be used to determine important properties. Since a basis could theoretically be infinitely large, it would be a computationally impossible task to obtain solutions. Furthermore, in general the choice of a basis is not unique due to the fact that it can be transformed as follows:

$$\tilde{c}_i = \sum_j U_{ij} c_j, \tag{2.5}$$

where U_{ij} are the elements of a unitary matrix.

Therefore, several basis sets can be used which will be presented in the following. These basis sets usually differ from different approximations, for instance with respect to the radial distribution of an atom or the number of different basis functions within orbitals.

Single particle atomic orbitals are best understood by revisiting the most simple atomic system, namely the non-relativistic hydrogen atom. The Hamiltonian of that system has solutions of the form [42]

$$\psi_{nlm} = R_{nl}(r) Y_{lm}(\theta, \phi). \tag{2.6}$$

Here, n denotes the energy level of the orbital, l and m represent the angular momentum. $R_{nl}(r)$ are products of Laguerre polynomials and a term decaying exponentially with the distance r, and $Y_{lm}(\theta, \phi)$ are spherical harmonics. Even though these solutions are exact for one electron atoms, they perform badly for many-electron systems, since these solutions become rapidly diffuse and hence can not describe the behavior of the core electrons precisely. Consequently, different solutions have to be used as basis states to describe the behavior of many-electron atoms.

One option is to retain only the term in $R_{nl}(r)$ with the highest power of r and adding a parameter ζ. These functions are called Slater-type orbitals (STO) and are written as

$$R_n^{\mathrm{STO}}(r) \propto (\zeta r)^{n-1}\, e^{-\zeta r}, \tag{2.7}$$

with n being the energy level and ζ being a fitting parameter, where different values are being used for each orbital to create a good basis function. Furthermore, linear combinations of STO's are required to approximate the true orbitals, as these functions do not represent oscillatory behavior of an atom. A single basis function can be considered for each orbital in the molecule, where for each basis function a different value for ζ is given, known as the single-zeta representation. It is also possible to introduce n basis functions, each with a different ζ value for each orbital, which is known as the n-zeta representation.

While STO functions provide many desirable features, they are not well suited for two-electron integral calculations, as this would be computationally expensive, and are therefore not used as basis functions in practice. Consequently, Gaussian basis functions are used to simplify two-electron integrals. These functions stem from considering the Schrödinger equation with a three-dimensional Harmonic oscillator potential and form of the Gaussian type orbitals (GTOs) is given by

$$R_{nl}^{\mathrm{GTO}}(r) \propto (\sqrt{\alpha_{nl}}r)^{l}\, e^{-\alpha_{nl} r^2}, \tag{2.8}$$

where α_{nl} is a fitting parameter. As can be seen easily, due to the dependence on r^2 in the exponent, GTOs are more localized than STOs leading to a weak approximation of the atomic charge distribution. This means that more GTOs would be required to describe a given orbital, however this limitation is compensated by the ease of the two-electron integral evaluation.

To get a good starting point and a qualitative description of the system, so called STO-nG basis sets can be introduced, where nG stands for n-Gaussians. In this basis set, each atomic orbital is considered to be an STO, being approximated using n GTOs. These basis sets are often referred to as minimal basis sets, as they include only orbitals required to write the Hartree-Fock state.

Further common basis sets are the split-valence (or Pople [50]) and the correlation-consistent basis sets, which can be used to obtain more accurate results. Split-valence basis sets, such as the 6-31G basis include only the minimal orbitals as well, but better approximate the true orbitals than the STO-nG bases. This is due to the increased radial flexibility for the valence electrons. The core orbitals are described by one approximate STO, constructed from a linear combination of six GTOs, with each valence shell orbital having a double-zeta representation. Two approximate STOs are introduced for each valence orbital: the more localized STO consists of three GTOs, while the more diffuse one is represented by a single GTO.

To even more increase the accuracy, we can introduce so-called correlation consistent polarized valence n-zeta (or cc-pVnZ) basis sets, which were introduced by Dunning [51]. To recover the correlation energy these basis sets include additional unoccupied (or virtual) orbitals being generated from correlated calculations on atoms. The core atoms have a single-zeta and the valence atoms have an n-zeta representation. The considered virtual orbitals are polarization functions with higher angular momenta than the valence orbitals and these polarization functions are selected by the size of their contribution to the correlation energy.

2.3 Molecular Hamiltonian and Reduced Density Matrices

We consider systems described by a spin-separated molecular Hamiltonian

$$\mathcal{H} = \text{const.} + \sum_{ij} h_{ij} c_i^\dagger c_j + \sum_{ijkl} V_{ijkl} c_i^\dagger c_j^\dagger c_l c_k, \tag{2.9}$$

where const. collects all non-electron effects such as the interaction between the nuclei, $c^{(\dagger)}$ denotes the creation (annihilation) operators of the spin orbitals, and with the one- and two-electron tensors h_{ij} and V_{ijkl}:

$$h_{ij} = \int dr \phi_i^*(r) \left(-\frac{\nabla^2}{2m} + \sum_I \frac{Z_I}{|r - R_I|} \right) \phi_j(r), \tag{2.10a}$$

$$V_{ijkl} = \int dr dr' \frac{\phi_i^*(r)\phi_j^*(r')\phi_k(r')\phi_l(r)}{|r - r'|}. \tag{2.10b}$$

Here, h_{ij} contains all one-electron effects such as the kinetic energy and the Coulomb interaction between the electron and the nuclei, where $\phi(r)$ denotes the spatial basis function. The two-electron integral V_{ijkl} describes the Coulomb interaction between the electron located at position r and the electron located at r'.

The energy of a state $|\psi\rangle$ with respect to this system is given by the expectation value of the Hamiltonian given in Eq. (2.9),

$$E = \langle \mathcal{H} \rangle = \text{const.} + \sum_{ij} h_{ij} \langle c_i^\dagger c_j \rangle + \sum_{ijkl} V_{ijkl} \langle c_i^\dagger c_j^\dagger c_l c_k \rangle = \text{const.} + \sum_{ij} h_{ij}^1 D_j^i + \sum_{ijkl} V_{ijkl}^2 D_{kl}^{ij}, \tag{2.11}$$

where we introduced the shorthand notation $\langle\cdot\rangle = \langle\psi|\cdot|\psi\rangle$, and the one-particle and two-particle reduced density matrices (1-RDM and 2-RDM):

$$ {}^1D^i_j = \langle c_i^\dagger c_j\rangle = \langle\psi|\, c_i^\dagger c_j \,|\psi\rangle\,, \tag{2.12a} $$

$$ {}^2D^{ij}_{kl} = \langle c_i^\dagger c_j^\dagger c_l c_k\rangle = \langle\psi|\, c_i^\dagger c_j^\dagger c_l c_k \,|\psi\rangle\,. \tag{2.12b} $$

When calculating the energy of a state on a quantum computer, one would map the fermionic operators onto qubits, e.g., using the Jordan-Wigner transformation, and measure the elements of the RDM. However, with a NISQ device in particular and finite computational resources one will obtain an erroneous result due to decoherence and shot noise (more on our considered noise types and their respective descriptions in Sec. 4.2).

2.4 Ensemble N-Representability Conditions

As we will analyze in Chapter 4 we can reduce energy errors and measurement variances by post-processing the result where we utilize knowledge about certain constraints that the 1- and 2-RDM need to fulfill. Specifically, we utilize the fact that an RDM needs to obey the so-called N-representability conditions if it is derived from a proper state of N fermions. There is a variety of these conditions, especially when specifying how many particles are distributed over how many orbitals; but this work we focus on just a few constraints which generally hold for any fermionic systems with a well-defined particle number (note that this is the case in chemical electronic structure problems) [52, 53]:

1. **Hermiticity** – It is easy to see from Eqs. (2.12a) and (2.12b) that the 1- and 2-RDM are Hermitian, meaning that:

$$ {}^1D^i_j = ({}^1D^j_i)^*, \tag{2.13a} $$

$$ {}^2D^{ij}_{kl} = ({}^2D^{kl}_{ij})^* \tag{2.13b} $$

2. **Antisymmetry** – Making use of fermionic anticommutation relations, we can rewrite the elements of the 2-RDM:

$$ {}^2D^{ij}_{kl} = -{}^2D^{ji}_{kl} = -{}^2D^{ij}_{lk} = {}^2D^{ji}_{lk}. \tag{2.14} $$

3. **Positive semidefiniteness** – The 1- and 2-RDM has to be positive semidefinite, meaning that all eigenvalues of the matrices have to be non-negative.

4. **Trace integrity** – From their definition one can derive that for a state with a well-defined particle number N the traces of the 1- and 2-RDM are given by:

$$\sum_i {}^1D^i_i = N, \tag{2.15a}$$

$$\sum_{ij} {}^2D^{ij}_{ij} = N(N-1). \tag{2.15b}$$

5. **Contractibility** – Related to the trace relation, for an N particle state, one can find the 1-RDM elements by contraction of the 2-RDM:

$$ {}^1D^i_j = \frac{1}{N-1}\sum_k {}^2D^{ik}_{jk}, \tag{2.16}$$

2.5 Hartree-Fock

The Hartree-Fock (HF) theory is a fundamental method in quantum chemistry that is used to calculate the electronic structure of a molecule [42]. It is a self-consistent field (SCF) method that provides an approximate solution to the Schrödinger equation for a many-electron system and aims to find the dominant Slater determinant in the system wavefunction. HF is based on the idea of a mean-field approximation, which assumes that the behavior of each electron in the system is influenced by an average field generated by all the other electrons. The theory provides a reasonable approximation to the true wave function and is a useful starting point for more sophisticated methods – such as CASSCF or CCSD – that take into account electron correlation effects that are not captured by the simple Slater determinant.

To find the dominant Slater determinant of the system, the spatial shape of the spin-orbitals is optimized in order to minimize the energy of the wavefunction [42]. Usually, a set of M orbitals is considered which is larger than the total number of electrons N in the molecule. This means that N orbitals are occupied, while $M - N$ orbitals remain unoccupied or virtual. Since the electron-electron repulsion term (2.10b) is neglected within the HF method the problem is reduced to N independent electrons. As mentioned above, it is then assumed that each electron moves in an average charge distribution of all other electrons, introducing an effective potential. Following this assumption, the N coupled equations can be solved iteratively: first the positions of all electrons are determined, then the potential is updated, and the process is repeated until the orbitals converge. In the context of second quantization, this procedure is carried out by repeatedly updating the orbitals to construct the Fock operator – which contains the Coulomb and exchange operator to approximate the single-electron energy

– diagonalizing the Fock operator to obtain new orbitals, and repeating until the orbitals converge.

2.6 Density Functional Theory

The aim within density functional theory (DFT) is to solve the Kohn-Sham equation, which is very similar to the time independent Schrödinger equation. The idea behind DFT is to write the energy as a functional of the electron density:

$$E[n(\vec{r})] = \int \epsilon_{\mathrm{xc}}(n) n(\vec{r}) d\vec{r}, \tag{2.17}$$

where $\epsilon_{\mathrm{xc}}(n)$ is the exchange-correlation energy density and $n(\vec{r})$ is the electron density. Eq. (2.17) makes use of the so-called local density approximation (LDA), where the functional only depends on the electron density at the coordinate where the functional is evaluated [54].

The usual procedure within a DFT calculation is to start with an initial guess for the electron density $n(\vec{r})$ and determine the effective potential, which is given as

$$V_{\mathrm{eff}}(\vec{r}) = V_{\mathrm{ext}}(\vec{r}) + V_{\mathrm{H}}[n(\vec{r})] + V_{\mathrm{xc}}[n(\vec{r})]. \tag{2.18}$$

Here, $V_{\mathrm{ext}}(\vec{r})$ is the external potential, $V_{\mathrm{H}}[n(\vec{r})]$ is the Hartree term covering the electron-electron Coulomb repulsion, and the last term $V_{\mathrm{xc}}[n(\vec{r})]$ is the exchange correlation potential, including all many-particle interactions.

Next, the Kohn-Sham equations are solved, which take the following form:

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V_{\mathrm{eff}}(\vec{r})\right)\psi_i(\vec{r}) = \epsilon_i \psi_i(\vec{r}). \tag{2.19}$$

From the solutions of the Kohn-Sham equations, we can derive the electron density:

$$n\left(\vec{r}\right) = \sum_{i \in \mathrm{occ}} \left|\psi_i\left(\vec{r}\right)\right|^2, \tag{2.20}$$

where the index i solely runs over the occupied orbitals. The final step checks for self-consistency; if the procedure did not converge, the steps are repeated exactly as described above, however if convergence is achieved the electron density of the many-body system is found and can be used to determine properties, such as the energy and gradients of a system.

DFT can return good initial guesses using fewer computational resources and requiring less time than its higher level counterparts, such as CASSCF and CCSD (see next sections). Depending on the system and investigated property, it is possible to choose between a variety of exchange and correlation functionals. One prominent example is the `B3LYP` functional [55], which includes three exchange and two correlation functionals. This hybrid functional performs well for basic properties such as the determination of atomization energies, whereas it is struggling with reaction energies and some metals. Consequently, one has to be careful with the specific choice of functionals since not all of them can "cover" certain physical or chemical properties. Additionally, it is very difficult to find the exact functional(s) for one specific problem. Therefore, a heuristic choice of functionals is used which highlights the limits of DFT in terms of accuracy when compared to other methods. For this reason usually one has to recall higher-level methods such as CASSCF or CCSD to get better results.

2.7 Complete Active Space Self Consistent Field

If we now consider strongly correlated systems, the above-mentioned HF method performs only poorly. Methods such as configuration interaction or coupled clusters (see next section) are effective at recovering dynamic correlations, whereas for states where multiple Slater determinants are equally important, static correlation dominates. Examples for static correlation are excited states and systems at the dissociation limit. One method to solve these problems is to use the complete active space self-consistent field method (CASSCF) [56]. This method considers the most important orbitals in the system (the so-called active space) and performs a multiconfigurational self-consistent field (MCSCF) calculation on all the determinants that could be generated from distributing all electrons in the orbitals within this active space. The MCSCF approach includes a wavefunction with several Slater determinants and performs a variational optimization for the molecular orbitals and determinant amplitudes. The exact quantum mechanical description of the active space is computationally expensive, since the Hilbert space scales exponentially with the number of active orbitals.

2.8 Configuration Interaction

The configuration interaction (CI) method generates a correlated wavefunction by considering excitations above a reference state, which is usually chosen to be the HF state. The full configuration interaction (FCI) wavefunction is recovered if all determinants are included. Considering all excitations above the HF wavefunction we find the expression for the FCI wavefunction

$$|\psi_{\mathrm{FCI}}\rangle = \left(I + \sum_{i,\alpha} T_{i\alpha} c_i^\dagger c_\alpha + \sum_{i,j,\alpha,\beta} T_{ij\alpha\beta} c_i^\dagger c_j^\dagger c_\alpha c_\beta + \dots \right) |\psi_{\mathrm{HF}}\rangle \,, \tag{2.21}$$

where T are the parameters to be variationally optimized. This method has to be restricted to a small number of excitations, as the inclusion of all determinants is classically intractable. Usually, single (CIS), double (CISD) and sometimes triple (CISDT) excitations are used in calculations. These restrictions lead to good approximations to the ground state energy because low energy excitations dominate the ground state wavefunction. The disadvantages of this method are the slow convergence to the FCI wavefunction and that it scales exponentially with the number of orbitals.

2.9 Coupled Cluster

The coupled cluster (CC) wavefunction is given by

$$|\psi_{\mathrm{CC}}\rangle = e^T |\psi_{\mathrm{HF}}\rangle \,, \tag{2.22}$$

with $T = \sum_i T_i$ and

$$\begin{aligned} T_1 &= \sum_{i\in\mathrm{virt},\alpha\in\mathrm{occ}} t_{i\alpha} c_i^\dagger c_\alpha, \\ T_2 &= \sum_{i,j\in\mathrm{virt},\alpha,\beta\in\mathrm{occ}} t_{ij\alpha\beta} c_i^\dagger c_j^\dagger c_\alpha c_\beta, \\ &\dots \end{aligned} \tag{2.23}$$

where occ denotes occupied and virt denotes unoccupied or virtual orbitals in the Hartree-Fock state. The excitation amplitudes t are variationally optimized. In this case we will only consider single and double excitations (CCSD), since it is classically intractable to include all excitations. The CCSD method generates a trial wavefunction which includes all possible determinants and because of its product parametrization it provides faster convergence than the CI method. Due to its non-unitary nature the resulting CC wavefunction does not obey the Rayleigh-Ritz variational principle. In contrast to CASSCF, CC is not able to deal with systems underlying strong static correlation.

2.10 Data Sets within Machine Learning

There is a large number of different methods within machine learning ranging from kernel-based methods to neural networks [57, 58]. In this section we will give a short overview of the three main components within a dataset that are used for finding the ideal machine-learned model: the training, validation and test set [59].

A training data set is a set of data points used during the learning process and used to fit parameters (such as weights) of, for example, a classifier. Such a classifier can be kernel-based and defined as

$$(K_\sigma + \lambda I)\,\vec{\alpha} = \vec{O}, \tag{2.24}$$

where K_σ is the kernel containing the input of the dataset, the vector $\vec{\alpha}$ includes the fit parameters (or weights) and $\vec{O}$ is the output vector. σ and λ are hyperparameters that are optimized during the validation step – we will discuss these parameters later.

The aim is to determine the optimal combinations of variables that will generate a good predictive model and to produce a trained (fitted) model that generalizes well to unknown data. The fitted model is evaluated using data points from different datasets (i.e. validation and test set) to estimate the model's accuracy in classifying new data. To reduce problems such as over-fitting, the data points within the validation and test set should not be used to train the model.

A validation data set is used to tune the hyperparameters (such as σ and λ in Eq. (2.24)) of a classifier. It is important to note that all data sets have to follow the same probability distribution. The validation set is another important pillar to avoid over-fitting. If, for instance, we are interested in finding the most suitable classifier for a problem, the training set is used to train the different candidate classifiers, the validation set is used to compare their performances and decide which one to take. The test set is used to obtain the performance characteristics (such as the accuracy) as a final step. In our example Eq. (2.24), we have defined two hyperparameters, where λ is set to a small constant value and σ is the parameter to be optimized, which is used within symmetric gradient domain machine learning (sGDML) [57]. To find the optimal value for σ, sGDML makes use of early stopping, where the candidate models are successive iterations of the same kernel, and training stops when the error on the validation set grows, making the routine to choose the previous model.

A test set is independent of the training set, but follows the same probability distribution as the training set. If a model fit to the training set fits the test set equally well, minimal over-fitting has taken place. Therefore, a test set is only used to assess the performance of a fully specified classifier. To do this, the final trained model is used to predict classifications, such as the accuracy, of data points within the test set. In a case where both validation and test sets are used, the test set is typically used to assess the final model that is selected during the validation process. It is possible to divide the entire data set into two subsets only, namely the training and test set, making the test set assessing the model only once. In other words, the hyperparameter σ is already given a fixed value, so that the validation step is skipped. While this approach is not recommended, one could use a method such as cross-validation, the partitioning into two subsets might be sufficient since the results are averaged after repeated rounds of model training and testing.

Part II

Improving Simulations on Noisy Quantum Computers

CHAPTER 3

ANALYZING GRADIENTS IN VARIATIONAL ALGORITHMS

The definition of the trial state depends on the choice of the Variational Quantum Eigensolver (VQE) Ansatz. One Ansatz that is well suited for lattice systems is the Variational Hamiltonian Ansatz (VHA) [28, 60, 61], stating that for each term of a generic decomposition of the Hamiltonian a separate parameter is defined. This way, the number of parameters, and thus the circuit depth can be reduced compared to other VQE Ansätze, such as the unitary coupled cluster with single and double excitations (uCCSD) [62]. In the uCCSD Ansatz parameters are defined based on single and double excitations of electrons from occupied to unoccupied (molecular) orbitals thus the number of parameters grows polynomially with the system size – with the excitation level determining the highest power of the polynomial scaling. By contrast, the VHA uses the structure of the Hamiltonian to determine the variational form. Therefore, periodic lattice Hamiltonians, which are typically characterized by a sparse Hamiltonian with few independent coupling constants yield an Ansatz where the number of parameters does not scale with the system size.

The VHA is inspired by an adaptation of the so-called "adiabatic connection" from many-body perturbation theory [63, 64]. Starting from the solution (ground state) of non-interacting electrons the state is evolved on the quantum computer using a sequence of unitary propagators constructed using parts of the fully-interacting Hamiltonian. The variational parameters correspond to the propagation times of each unitary operator. Having opted for the VHA to provide the parametrized trial state, the parameter set is optimized such that the expectation value of the interacting Hamiltonian is minimized.

This chapter has been published in "T. Piskor, J.-M. Reiner, S. Zanker, N. Vogt, M. Marthaler, Frank K. Wilhelm, and F. G. Eich, Phys. Rev. A **105**, 062415 (2022)". Copyright (2022) by the American Physical Society. The text was largely authored by T. Piskor. All numerical simulations were performed by T. Piskor.

There are several approaches to determine the optimal parameter set for a given cost function ranging from gradient-free [65, 66] to gradient-based algorithms [67], which usually lead to faster convergence compared to the gradient-free alternatives. In the present work we focus on the analysis of the evaluation of gradients on NISQ computers, which means that we are less concerned about the specific gradient-based optimization algorithm. Hence, we use a simple steepest-descent approach for the optimization, where the parameters are updated using the gradient of the energy directly, employing a fixed learning rate (damping of the gradient).

The main question we address in this work is how measurement statistics and noise affect the optimization of the parameters appearing in the quantum circuit. Since we always have to perform measurements in order to get the expectation value of an observable, shot noise will always be existent – even on error-corrected quantum devices. However, when simulating the quantum device we can obtain the limit of an infinite number of measurements, providing us with the ideal reference result for the algorithm. Furthermore, due to adverse coupling of the qubits to the environment, another source of error has to be considered, namely the intrinsic depolarization of qubits. However, we find that the dominant error for the noise model chosen in our simulations stems from measurement statistics.

For the determination of the cost function's gradient we compare two procedures: 1) the finite-difference approximation to compute the gradients, which we also refer to as "numerical" method, 2) the so-called parameter-shift rule, which we also refer to as "analytical" method [33, 68–71]. While calculating the gradient numerically, the outcome of the result is more susceptible towards noise effects, such as shot noise or depolarization effects occurring in qubits, because there is a competition of improving the numerical gradient by reducing the finite step size to evaluate the gradient versus resolving differences of the cost function evaluated for two nearby values of a given parameter. To bypass this hurdle, gradients can also be determined analytically via the parameter-shift rule, with the hope to show a more resilient behavior [68] to noise effects, because it relies on a *fixed* finite difference. However, there is an important caveat for using the parameter-shift rule for optimizing a VHA: Compared to the numerical case, where the number of additional circuit evaluations only grows with the number of parameters, the number of additional circuit evaluations grows with twice the number of parametrized gates for the parameter-shift rule. While the number of parameters solely depends on the Ansatz, the number of parametrized gates depends on the implementation of the trial state in the quantum circuit. In general, it has to be considered that one parameter might occur in multiple parametrized gates – especially in case of the VHA (see section 1.5) – and thus the number of parameters might be much smaller than the number of parametrized gates.

In the following we compare both methods for determining the gradients, using a simple gradient-based optimization algorithm, by investigating a simple one-qubit circuit and a 2- and 6-site Hubbard model, mapped onto 4 and 12 qubits, respectively. Sec. 1.5 discusses the VHA in more detail and Sec. 3.2 presents the simple gradient-based optimization routine and compares the two approaches for determining the gradient. Sec. 3.3 details how many additional circuits are required to determine numerical and analytical gradients. A brief overview on how noise has been implemented in our simulations is given in Sec. 1.7. Finally, results are shown for the simple quantum circuit in Sec. 3.4.1 and for the one-dimensional Hubbard model in Sec. 3.4.2.

3.1 The Variational Hamiltonian Ansatz and 1D Hubbard Model

Starting with an initial guess, $\boldsymbol{\theta}_0$, the parameter set is updated iteratively, using the procedure described in Sec. 3.2, to approach the optimal solution.

The explicit form of the unitary operator within the variational Hamiltonian Ansatz (VHA), $\hat{U}(\boldsymbol{\theta})$, is constructed by decomposing the Hamiltonian into P separate terms, as stated in Eq. (3.1a). The partial contributions to the Hamiltonian are Hermitian operators, so we can use them as generators for unitary rotations. Since the partial contributions do not commute in general, the order in which the unitaries are applied matters. In practice, we repeat the application of the unitaries R times and allow for different parameters (or rotation angles) in each repetition (Eq. (3.1b)) yielding a total number of $n = RP$ free variational parameters,

$$\hat{\mathcal{H}} = \sum_{\alpha=1}^{P} \hat{\mathcal{H}}_\alpha, \tag{3.1a}$$

$$\hat{U}(\boldsymbol{\theta}) = \prod_{k=1}^{R} \prod_{\alpha=1}^{P} e^{i\theta_{\alpha,k}\hat{\mathcal{H}}_\alpha}. \tag{3.1b}$$

In this work we investigate the 1D Hubbard model which is described by the Hamiltonian

$$\begin{aligned} \hat{\mathcal{H}} &= \hat{\mathcal{T}} + \hat{\mathcal{W}} \\ &= -t \sum_i \sum_\sigma \left(\hat{c}^\dagger_{i,\sigma} \hat{c}_{i+1,\sigma} + \hat{c}^\dagger_{i+1,\sigma} \hat{c}_{i,\sigma} \right) \\ &\quad + U \sum_i \left(\hat{c}^\dagger_{i,\uparrow} \hat{c}_{i,\uparrow} - \frac{1}{2} \right) \left(\hat{c}^\dagger_{i,\downarrow} \hat{c}_{i,\downarrow} - \frac{1}{2} \right), \end{aligned} \tag{3.2}$$

with the first term describing nearest-neighbor hopping between site i and $i + 1$ and spin σ, being either in the spin up $\uparrow$ or spin down $\downarrow$ state (kinetic energy) with hopping amplitude t and the second term describing the repulsion of two electrons on the same site with interaction strength U. We consider a system is at half-filling – an average of one electron per site. In our studies we focus on the case where the kinetic energy and the interaction have the same magnitude and use t as our unit of energy, i.e., $U = t = 1$. Moreover, we employ periodic boundary conditions. The Hubbard model at half filling is a prototypical model describing a Mott insulator, i.e., an insulator where the fundamental gap is due to electron-electron interaction and any (static) mean-field theory would yield a metal (unless the symmetry is artificially broken).

Quite naturally Hamiltonian (3.2) can be split into two contributions, i.e., the kinetic energy $\hat{\mathcal{T}}$ and the interaction energy $\hat{\mathcal{W}}$. For system sizes larger than 2 sites the hopping operator is split further into a so-called "even" and "odd" contribution, $\hat{\mathcal{T}}_e$ and $\hat{\mathcal{T}}_o$. The individual contributions to each of the three parts of the Hamiltonian, $\hat{\mathcal{T}}_e$, $\hat{\mathcal{T}}_o$ and $\hat{\mathcal{W}}$, commute among each other, which implies that each exponential can be split and reordered easily without using the Baker-Campbell-Hausdorff formula.

3.2 Gradient Descent Optimizer

In this work we are interested in investigating how noise affects the synthesis of gradients of the cost function (cf. Eq. (1.19)). While there are several established algorithms, such as the BFGS algorithm [72–75], for gradient-based optimization, we decided to use a very simple approach, namely a (damped) steepest-descent optimizer, in order to reduce the complexity in the optimization process and focus on the gradient. Specifically, we update a parameter θ_i according to

$$\theta_i^{\tau+1} = \theta_i^{\tau} - \eta \partial_{\theta_i} E(\boldsymbol{\theta}^{\tau}), \tag{3.3}$$

where $\partial_{\theta_i} E(\boldsymbol{\theta})$ denotes the derivative of the cost function with respect to θ_i, η is a *fixed* learning rate (damping) controlling the step size towards the minimum of the cost function and θ_i^{τ} is the i^{th} parameter at iteration step τ. The parameter set at iteration $\tau + 1$ is given by the parameter set at step τ modified by the cost function's gradient scaled by the learning rate η. Compared to other gradient-based optimizers [67], the steepest-descent optimizer is very simplistic in its form, since it contains only one fixed hyperparameter. Since no information beyond the gradient, such as an approximate Hessian matrix, is used, algorithm (3.3) may converge slower than, for example, the aforementioned BFGS optimizer.

In order to determine the gradients for the optimization routine, two possibilities will be discussed throughout this work. On the one hand, gradients can be determined numerically with a finite-difference method:

$$\begin{aligned}\partial_{\theta_i} E(\boldsymbol{\theta}) \approx \frac{1}{\epsilon}\Big(&E(\theta_1, \ldots, \theta_i + \epsilon, \ldots, \theta_n) \\ &-E(\theta_1, \ldots, \theta_i, \ldots, \theta_n)\Big),\end{aligned} \tag{3.4}$$

with ϵ defining a small *but finite* step by which the parameter θ_i is shifted. This immediately implies that the number of additional circuit evaluations for obtaining the gradient using the finite-difference method corresponds to the number of parameters. However, this method may be more susceptible towards noise effects, such as statistical or depolarization noise due to the fact that the energy difference (cf. Eq. (3.4)) is getting smaller, and therefore harder to resolve, if we improve the accuracy of the gradient by making the step width ϵ smaller.

On the other hand we determine gradients analytically with the so-called parameter-shift rule [68]. At its core the parameter-shift rule uses the fact that the cost function generally is represented by a quantum circuit composed of (parametrized) single-qubit rotations and *fixed* two-qubit gates. Note, that every unitary can be represented in such a form (e.g., using `CNOT` plus arbitrary single-qubit rotations as a universal gate set [38]) and that we use such decompositions in our simulations to implement the exponentials of the VHA in Eq (3.1b). Note also, that for the decompositions of the VHA evolution a single parameter $\theta_{\alpha,k}$ will in general appear in multiple rotation gates; however, for the application of the parameter-shift rule they will need to be treated individually in order to evaluate the gradient. Focusing on the dependence of the energy on a single parameter θ_1, which we assume to only control *one* single-qubit gate, the energy, Eq. (1.19), can be written as,

$$\begin{aligned}E(\theta_1; \theta_2, \ldots, \theta_n) = A_1 \cos(\omega\theta_1 + \varphi_1) + \ldots \\ \ldots + A_\alpha \cos(\omega\theta_1 + \varphi_\alpha) + \ldots + C,\end{aligned} \tag{3.5}$$

with amplitudes A_α, phase shifts φ_α and constant C, which depend on all other parameters $\theta_2, \ldots, \theta_n$ (see appendix 3.A for a detailed derivation). The distance between the two eigenvalues of the generator for the single-qubit rotation is denoted by ω – for Pauli matrices we have $\omega = 2$. The main idea of the parameter-shift rule is to consider the derivative of a trigonometric function,

$$\begin{aligned}\partial_\theta \cos(\omega\theta + \varphi) &= -\omega \sin(\omega\theta + \varphi) \\ &= \frac{\omega}{2}\left[\cos\left(\omega\left[\theta + \frac{\pi}{2\omega}\right] + \varphi\right)\right. \\ &\qquad \left. - \cos\left(\omega\left[\theta - \frac{\pi}{2\omega}\right] + \varphi\right)\right] \ ,\end{aligned} \tag{3.6}$$

which highlights that the *exact* derivative with respect to the parameter θ is obtained by taking the difference of two quantum circuits, where the parameter is shifted by $\pm\frac{\pi}{2\omega}$, respectively. The parameter-shift rule, as presented above, is based on the assumption that a parameter θ_i *only* controls a single single-qubit gate (cf. App. (3.A)). However, the parameter θ_i can appear in more than one gate for the VHA Ansatz. This is due to the fact that the parameters are defined in Eq. (3.1b) in reference to generators from the electronic Hamiltonian. In transforming the electronic Hamiltonian into the quantum circuit several single-qubit rotations are parametrized, in general, by the same parameter θ_i. Hence, the parameter-shift rule *cannot* be simply applied to the parameters defined in Eq. (3.1b). Use of the parameter-shift rule can be vindicated by defining a new set of parameters, $\boldsymbol{\mu}$, which assigns each parametrized single-qubit gate its own parameter μ_j ($j = 1, 2, \cdots, m-1, m$, where $m \geq n$). Then we can compute the gradient with respect to the new set of parameters as

$$\begin{aligned}\partial_{\mu_j} E(\boldsymbol{\mu}) = r \Big[& E\left(\mu_1, \ldots, \mu_j + \frac{\pi}{4r}, \ldots, \mu_m\right) \\ & - E\left(\mu_1, \ldots, \mu_j - \frac{\pi}{4r}, \ldots, \mu_m\right)\Big] \ .\end{aligned} \tag{3.7}$$

Note that the parameter r, controlling the “step width”, in principle depends on the explicit form of the unitary operator, Eq. (3.1b), but we can always define a linear map from the parameter set $\boldsymbol{\theta}$ to $\boldsymbol{\mu}$, such that r is the same for all parametrized gates ($r = \frac{1}{2}$ for standard Pauli rotations, see e.g. Ref. [68] for details).

We emphasize that the parameter-shift rule yields, in principle, the exact gradient and not an approximation as the finite-difference method described above. Moreover, the parameter-shift rule is potentially more resilient towards noise and other effects, because – in spite of being exact – it is evaluated using a finite difference with a step width on the order of the spectral width of the generator (typically a scaled Pauli matrix). However, the number of circuit evaluations scales with the number of parametrized gates and thus, with increasing system size, requires more circuits than the finite-difference method. In our work we consider a specific Ansatz, the VHA, for generating the variational quantum circuit, which requires discussing the difference between the number of parameters, n (number of components for $\boldsymbol{\theta}$), and the number of parametrized gates, m (number of components for $\boldsymbol{\mu}$). For a generic variational quantum circuit, which aims at optimizing all single-qubit gates independently, the number of parameters coincides trivially with the number of parametrized gates.

3.3 Circuit Evaluation

In this section, we highlight the difference between the number of parametrized gates, m, and the number of variational parameters, n, mentioned in the previous section, to expose the additional overhead for using the parameter-shift rule, i.e., the fact that $m \geq n$. An overview for the one-dimensional Hubbard model is given in Table 3.1, where we compare the number of parameters and the number of parametrized gates for 1D Hubbard rings of various sizes (number of sites). Therefore, a circuit was generated using a two-qubit decomposition [76, 77], consisting of one-qubit rotation gates and a two-qubit controlled-Z gate.

# sites M	Hilbert space size 2^{2M}	# parameters	# parametrized gates
2	16	2	10
4	256	3	28
6	4,096	3	42
8	65,536	3	56
10	1,048,576	3	70
12	16,777,216	3	84
14	268,435,456	3	98
16	4,294,967,296	3	112

Table 3.1: Number of parameters and parametrized gates for 1D Hubbard chains with varying number of sites.

Table 3.1 shows the number of parameters and parametrized circuits for various sizes, determined by the number of sites M in the one-dimensional Hubbard model. In the case of $M = 2$ with one repetition for the Ansatz made in Eq. (3.1b), we only have two parameters, namely one parameter for the hopping operator and one for the interaction operator. Increasing the site number to $M = 4$ or $M = 6$ yields one further parameter per repetition, since we now also split the hopping term (kinetic energy) into two internally commuting contributions (labeled “even” and “odd”). The total number of circuits to evaluate for calculating energy and gradient with the finite-difference method is given by the number of parameters, i.e., one circuit for each parameter shifted by ϵ, plus the circuit with no shift applied to any parameter. Considering the 2-site Hubbard model with one repetition in the VHA, which implies two parameters, a total number of three circuits would be required in order to evaluate the gradient and the energy. Increasing the number of repetitions from one to two, leading to four parameters, 5 circuits have to be evaluated in order to determine the energy and its gradient. In general, the total number of circuit evaluations required for gradient and energy determination is given by $N_{\text{fd}} = RP + 1 = n + 1$ when using the finite-difference method, where P defines the number of parameters and R the number of repetitions in the VHA (cf. Eq. (3.1b)).

For the parameter-shift rule, the additional circuits for evaluating the gradient is proportional to twice the number of parametrized gates, leading to 21 circuits for the 2-site Hubbard model with one repetition. Similarly to the finite-difference case, the number of parametrized gates scales linearly with the number of parameters and therefore also with the number of repetitions. In general, the number of circuits required using the parameter-shift rule can be written as $N_{\text{ps}} = 2C(M)RP + 1 = 2m + 1$, where we introduced the coefficient $C(M)$, which translates the number of parameters to the number of parametrized gates. Considering the data in table 3.1 and system sizes larger than 2 sites, we get $C(M) = \frac{7}{3}M$, which implies that the number of additional circuits scales linearly with the system size using the Jordan-Wigner transformation [78] to map fermionic sites to qubits. Note, that this scaling is still superior to uCCSD which scales quadratic at best [7].

3.4 Results

In this section, we present numerical results for a simple quantum circuit and the one-dimensional Hubbard model with two and six sites. In all cases, 50000 measurements have been performed in order to get the averaged result of the respective observables. Both, the finite-difference and parameter-shift algorithm have been performed five times to get the stochastic effect of the finite number of measurements. The shaded areas in the upcoming plots thus mark the worst (highest value) and the best (lowest value) energy for each iteration step and do not represent the standard deviations of the optimization runs. The solid lines mark runs without any effect of shot noise, representing the case where the number of measurements $N \to \infty$.

3.4.1 Simple Circuit

In order to get a better understanding of the gradient descent optimizer, a simple one-qubit circuit has been investigated, where at first a Hadamard gate is applied to the prepared ground state $|0\rangle$ followed by a rotation Z gate (see Eq. (3.8a)). After this operation, a measurement of the Pauli X operator is performed, leading to a trivial periodic function

$$
\begin{aligned}
|\psi(\theta)\rangle &= \hat{U}(\theta)\,|\psi_0\rangle = \hat{R}_z(\theta)\hat{H}\,|\psi_0\rangle \\
&= \begin{pmatrix} e^{-i\theta/2} & 0 \\ 0 & e^{i\theta/2} \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\
&= \frac{1}{\sqrt{2}} \begin{pmatrix} e^{-i\theta/2} \\ e^{i\theta/2} \end{pmatrix},
\end{aligned} \tag{3.8a}
$$

$$
E(\theta) = \langle \hat{X} \rangle = \langle \psi(\theta)|\,\hat{X}\,|\psi(\theta)\rangle = \cos(\theta). \tag{3.8b}
$$

The minimum of Eq. (3.8b) is attained for $\theta = (1 + 2p)\pi$ ($p \in \mathbb{Z}$) with a minimal value of $E_{\text{exact}} = -1$.

3.4.1.1 Shot Noise

Fig. 3.1 shows optimization runs for different values of ϵ for the finite-difference method, as well as runs performed with the parameter-shift rule with $N = 50000$ measurement shots. Note that in all plots, starting from Fig. 3.1, we are using a logarithmic scale for the energy differences shown on the y-axis. The solid lines indicate simulations without any source of noise ($N \to \infty$) and the shaded areas with the corresponding color are obtained as the envelope of five runs performed with statistical errors ($N = 50000$), i.e., the upper line marks the maximum and the bottom line indicates the minimum of the relative deviation from the optimal value at a given iteration of the optimization, respectively. Focusing on the $N \to \infty$ results, it can be seen that with decreasing ϵ the accuracy gets better, which highlights the error introduced by approximating the derivative using a finite difference. However, if statistical effects are included it can also be seen that the fluctuations around the optimal curve increase with smaller ϵ, clearly exposing that smaller ϵ are more prone to statistical errors. Considering the optimization runs performed with the parameter-shift rule it can be seen that the on the one hand, a more accurate result is achieved for the noiseless run ($N \to \infty$), demonstrating that the parameter-shift rule yields the exact gradient, and on the other hand the fluctuations around the optimal curve are smaller compared to the finite-difference runs with $\epsilon = 0.02$.

3.4.1.2 Shot and Depolarization Noise

Including the effect of depolarization noise, we observe that the deviation from the cost function's minimum increases. This is most pronounced for the parameter-shift rule and small values of ϵ for the finite-difference method. We can attribute this to the fact that

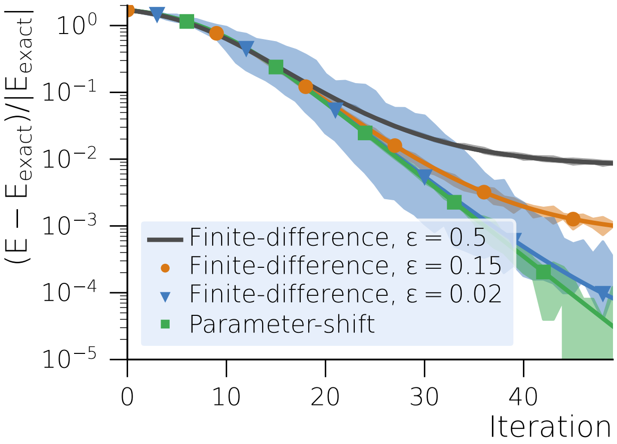


FIGURE 3.1: Optimization runs using a simple gradient descent Ansatz performed with the finite-difference method compared to the parameter-shift rule for the simple circuit and $N = 50000$. The bold line indicates a run without shot noise, whereas the corresponding shaded region marks five optimization runs performed with shot noise. For all optimization runs, a learning rate of $\eta = 0.5$ has been used.

the error introduced by using a larger ϵ in the finite difference dominates the overall error. This is depicted in Fig. 3.2, where for the two mentioned cases a small shift away from the minimum can be seen, compared to the noiseless run in Fig. 3.1. Increasing the depolarization rate further leads to a higher deviation from the minimum for all cases, as can be seen in Figs. 3.3 and 3.4. Especially for the case where $\gamma = 10^{-2}$, the finite-difference curves with the two smallest ϵ and the parameter-shift curve nearly show the same result for the cost function after 50 iteration steps, clearly showing that the depolarization error dominates the overall error.

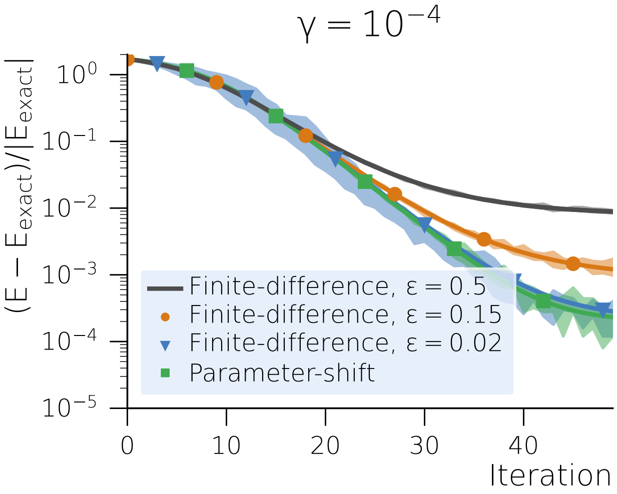


FIGURE 3.2: Same plot as Fig. 3.1, but with depolarization noise, characterized by the rate $\gamma = 10^{-4}$ included in the simulation.

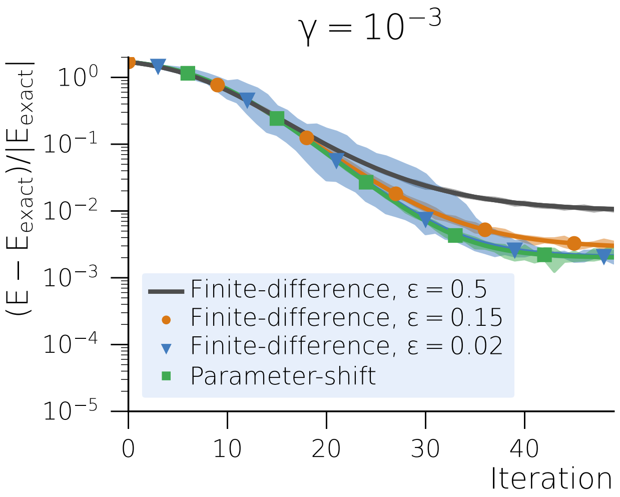


FIGURE 3.3: Same plot as Fig. 3.2, but with a depolarization rate of $\gamma = 10^{-3}$.

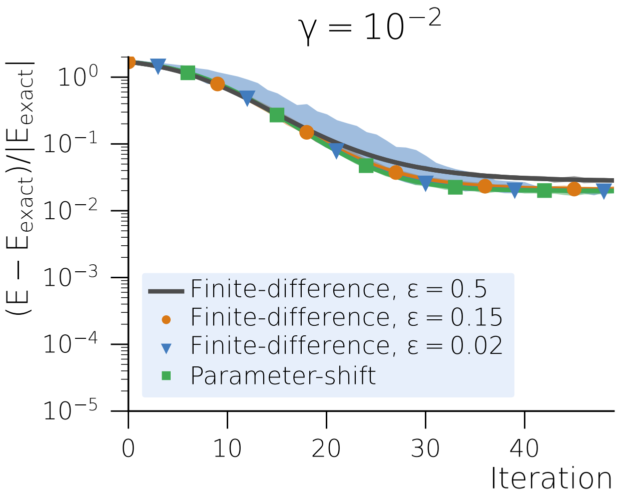


FIGURE 3.4: Same plot as Fig. 3.2, but with a depolarization rate of $\gamma = 10^{-2}$.

3.4.2 Results for the 2-site Hubbard Model

Next, we study the one-dimensional Hubbard model by performing optimization runs with shot noise only and with shot and depolarization noise. The Hubbard Hamiltonian is defined in Eq. (3.2), where the initial state, $|\psi_0\rangle$, is chosen as the ground-state Slater determinant for the non-interacting system.

For the 2-site Hubbard model we decompose the Hamiltonian simply into kinetic energy and the interaction energy. Thus, the trial state for the 2-site Hubbard model is given as

$$|\psi(\boldsymbol{\theta})\rangle = e^{i\theta_2 \hat{\mathcal{T}}} e^{i\theta_1 \hat{\mathcal{W}}} |\psi_0\rangle \,, \tag{3.9}$$

where we only use one repetition of the unitaries, since this parametrization already encompasses the exact ground state for the 2-site Hubbard model. As in the previous example the parameter set $\boldsymbol{\theta}$ has been determined with the gradient descent method explained in Sec. 3.2. The exact analytical result for the 2-site Hubbard model [79] with Eq. (3.9) as the trial state has been taken as the reference value E_{exact}.

3.4.2.1 Shot Noise

First, optimization runs including only shot noise have been performed, where the number of measurement shots has been set to $N = 50000$. As in the previous case the solid lines in Fig. 3.5 show runs without shot noise, corresponding to the limit $N \to \infty$. The shaded areas represent the maximum and minimum of five optimization runs performed with shot noise, respectively. From the graph it can be seen that the shaded region for $\epsilon = 0.5$ is vanishingly small at the scale of the plot. Starting around iteration step 10 it can be observed, at the scale of the plot, that the fluctuations of the shaded regions increase with smaller ϵ. The more accurate the gradients are in principle, the more relevant becomes the fact that in practice we always perform a finite amount of measurements. Fig. 3.5 shows that the energy deviation exhibits a minimum for gradients obtained via finite difference, which can be attributed to the fact that the approximate derivatives do not even in principle correspond to the true derivatives of the cost function. This has to be contrasted with the energy deviation obtained by using the parameter-shift rule to synthesize the gradients for the optimization procedure. Here, the gradients are in principle ($N \to \infty$) exact and the only error stems from the finite accuracy due to a finite number of measurements for the gradient. However, our results show that for a fixed number of measurements ($N = 50000$) the intrinsic error in calculating the gradients via finite difference using $\epsilon = 0.05$ is smaller than the statistical error due to the finite number of measurements. Considering the measurement overhead (cf. Tab. 3.1 and Sec. 3.3) in synthesizing the gradients via the parameter-shift rule, favors the finite-difference gradients.

3.4.2.2 Shot and Depolarization Noise

By adding depolarization noise with a rate of $\gamma = 10^{-4}$ to the previous simulations it can be seen that there are no significant differences for $\epsilon = 0.5$ and $\epsilon = 0.2$, except a minor increase in the deviation from the exact energy, as shown in Fig. 3.6. However, there is a larger offset for the parameter-shift run. For example, around the 20th iteration step, the $N \to \infty$ parameter-shift curve flattens out. As in the previous case, parameter-shift simulations show a similar behaviour as the finite-difference runs with small ϵ.

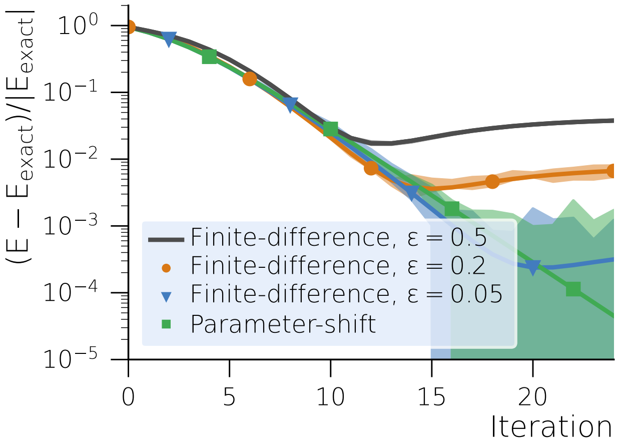


FIGURE 3.5: Optimization runs using a simple gradient descent Ansatz performed with the finite-difference method compared to the parameter-shift rule for the 2-site Hubbard model and $N = 50000$. The solid lines indicate runs without shot noise, whereas the corresponding shaded region marks five optimization runs performed with shot noise. For all optimization runs, a learning rate of $\eta = 0.1$ has been used.

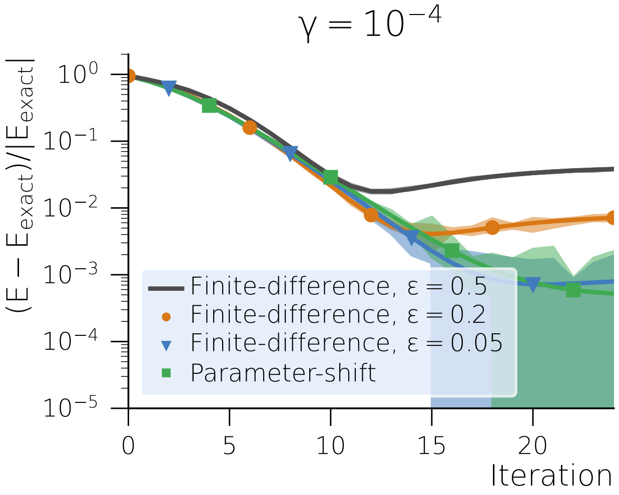


FIGURE 3.6: Same as Fig. 3.5, but now including depolarization noise, characterized by the rate $\gamma = 10^{-4}$, in the simulations.

3.4.3 Results for the 6-site Hubbard Model

In contrast to the 2-site Hubbard model, the number of repetitions has been set to 2 for the 6-site Hubbard model, leading to a trial state given as

$$|\psi(\boldsymbol{\theta})\rangle = e^{i\theta_6 \hat{\mathcal{T}}_o} e^{i\theta_5 \hat{\mathcal{T}}_e} e^{i\theta_4 \hat{\mathcal{W}}} e^{i\theta_3 \hat{\mathcal{T}}_o} e^{i\theta_2 \hat{\mathcal{T}}_e} e^{i\theta_1 \hat{\mathcal{W}}} |\psi_0\rangle , \tag{3.10}$$

with $\hat{\mathcal{T}}_e$ and $\hat{\mathcal{T}}_o$ defined in Sec. 1.5.

By increasing the number of repetitions and thus the number of parameters in the Ansatz, the variationally determined energy is closer to the true ground state energy of the system. Note that we use the optimal energy achievable with the given Ansatz (3.10) as reference energy and denote it as E_{exact}. Due to the system size, runs with depolarization noise were not performed.

3.4.3.1 Shot Noise

For the 6-site Hubbard model we perform optimization runs for three different choices of ϵ and the parameter-shift rule with $N = 50000$. Taking a look at the solid lines in Fig. 3.7, representing noiseless runs, it can be seen that with decreasing ϵ the energy accuracy improves, which could also be observed in the two previous cases. Taking shot noise into account, some minor fluctuations around the optimal curve for $\epsilon = 0.1$ can be spotted. However, these fluctuations increase with decreasing ϵ besides that an offset seems to occur, especially for $\epsilon = 0.01$. While the noiseless run suggests an energy accuracy in the order of 10^{-3} after 50 iterations steps with the gradient-descent optimizer, this value roughly increases by an order of magnitude for the noisy runs.

The noiseless parameter-shift run shows a worse accuracy in energy after the first 50 iterations, however it shows better results compared to the finite-difference method when shot noise is included. Fluctuations occur only around the noiseless run, whereas no offset, as in the case of $\epsilon = 0.01$ performed with the finite-difference method, can be observed and thus showing an overall better performance compared to its numerical counterpart. It has to be emphasized that the total number of circuits is considerably higher for the parameter-shift rule – namely a factor of 24 – for this particular system and consequently requires a lot more time to finish the 50 iterations compared to the finite-difference method. Therefore, a compromise between runtime and accuracy in energy is a plausible conclusion. Having a slightly worse accuracy but therefore requiring a smaller number of additional circuits, one might choose the finite-difference method with $\epsilon = 0.05$.

3.5 Conclusion

In this work, two possibilities for gradient determination have been investigated and tested on a simple toy model and the one-dimensional Hubbard model using the VHA and a simple gradient-descent algorithm for parameter optimization.

On the one hand, gradients have been determined numerically with a simple finite-difference method, where the difference between the obtained and exact minimum of the cost function for noiseless runs gets smaller with decreasing step sizes ϵ. The number of circuit evaluations scales only with the number of parameters defined in the parameter set and thus shows a fast runtime compared to the parameter-shift method.

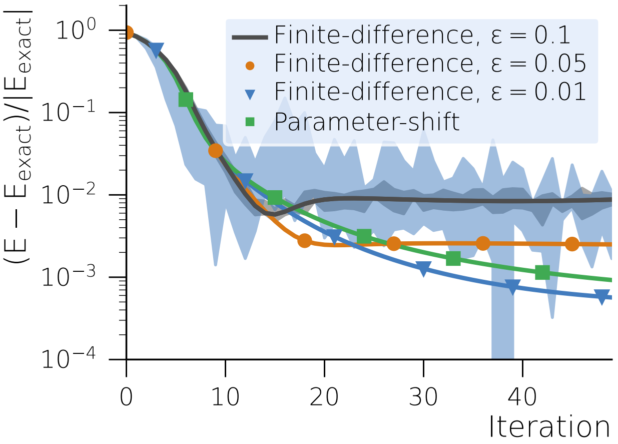


FIGURE 3.7: Optimization runs using a simple gradient descent Ansatz performed with the finite-difference method compared to the parameter-shift rule for the 6-site Hubbard model and $N = 50000$. The solid lines indicate runs without shot noise, whereas the corresponding shaded region marks five optimization runs performed with shot noise. For all optimization runs, a learning rate of $\eta = 0.03$ has been used.

On the other hand, analytical gradients, determined with the parameter-shift rule, lead to more accurate results and a more resilient behavior towards statistical noise. Especially in the case of the 6-site Hubbard model, a better accuracy can be achieved with the parameter-shift rule when simulating the system taking into account shot noise. However, a major drawback of the parameter-shift rule, applied to the optimization of a VHA circuit, is its scaling. Since the VHA does not only scale with the number of parameters but also with the system size (which affects the number of parametrized gates), the number of required measurements increases linearly with the system size as detailed in Sec. 3.3. For example, in the 6-site case the number of circuit evaluations is almost two orders of magnitude higher for the parameter-shift method compared to the numerical counterpart. Thus, a compromise between runtime and accuracy can be made for the finite-difference method, where the step size ϵ is chosen optimally to avoid statistical noise and provide the required accuracy. The obvious downside of this approach that this optimal step width has to be determined and in general depends on the number of measurements and other hyper-parameters of the optimization algorithm. Taking a look at the simple toy model and the 2-site Hubbard model it can be concluded that the finite-difference method is the method of choice when weighing both runtime and accuracy. Even with shot and depolarization noise for both cases, there is no substantial difference between optimization runs performed with the finite-difference method choosing small step sizes ϵ and the parameter-shift rule.

We emphasize that the results presented in this work apply to the generic situation where the trial state is constructed with the goal of reducing the number of parameters optimized in the classical optimization loop of a quantum-classical hybrid optimization procedure, which leads to the important difference of number of parameters and number of parametrized gates when analyzing the measurement costs when considering the parameter-shift rule. One motivation for reducing the number of parameters in the trial state is given by the fact that so-called Barren plateaus [30, 80] hamper the optimization of functions defined in high-dimensional spaces using gradient-based algorithms. However, the trigonometric building blocks of any quantum circuit, highlighted in the discussion of the parameter-shift rule in Sec. 3.2, suggests to use alternative (gradient-free) optimization algorithms [81].

We stress, however, that our conclusions do not apply to *generic* variational quantum circuits, which recently gained considerable attention in the context of quantum machine learning [82–85]. These generic quantum circuits are directly composed using a universal gate set, e.g., `CNOT` and arbitrary single qubit rotations [38], so the distinction of number of parameters and number of parametrized gates does not apply – all parametrized gates are treated as independent, so the number of parametrized gates corresponds trivially to the number of variational parameters. The additional structure imposed by simulating fermions via the fermion-qubit mapping (Jordan-Wigner transformation, etc.) and using parts of the Hamiltonian to generate the variational unitaries, is responsible for the overhead of the parameter-shift rule, compared to the finite-difference method, in our studies.

Finally, this suggests another possibility to generate the trial state, i.e., using the VHA to generate the circuit, but optimizing each parametrized gate individually in the optimization procedure. An interesting question for future studies is to investigate whether treating each parametrized gate independently considerably improves the minimal achievable energy compared to the standard VHA.

Appendices

3.A Derivation of the Parameter-Shift Rule

In this appendix we present an explicit derivation of Eq. (3.5). We start by choosing a specific one-qubit rotation, $\hat{R}_{\theta_i}$, in order to decompose the parametrized unitary quantum circuit into

$$\hat{U}(\boldsymbol{\theta}) = \hat{V}\hat{R}_{\theta_i}\hat{W} \ , \tag{3.11}$$

where $\hat{V}$ and $\hat{W}$ are unitary rotations parametrized by all angles θ_j except for the angle θ_i, which *only* parametrizes the selected single-qubit gate. The single-qubit rotation, $\hat{R}_{\theta_i}$, is explicitly given by

$$\hat{R}_{\theta_i} = e^{-i\theta_i \boldsymbol{n}\cdot\boldsymbol{\sigma}} = \cos(\theta_i)1 - i\sin(\theta_i)\boldsymbol{n}\cdot\boldsymbol{\sigma} \ , \tag{3.12}$$

where $\boldsymbol{n}$ is a unit vector defining the rotation axis (e.g., Cartesian unit vectors would result in rotations around the x-, y-, or z-axis), θ_i provides the rotation angle and $\boldsymbol{\sigma}$ denotes the vector of elementary Pauli matrices. Note that the Pauli matrices are, in principle, also labeled by the qubit index on which they act, which is suppressed here in order to keep the notation concise.

The cost function (1.19) can be rewritten as

$$E(\boldsymbol{\theta}) = \sum_m \langle\psi_0|\, \hat{U}(\boldsymbol{\theta})^\dagger \hat{O}_m \hat{U}(\boldsymbol{\theta})\, |\psi_0\rangle \ , \tag{3.13}$$

highlighting that the energy is synthesized by measuring several Hermitian operators $\hat{O}_m$. Defining $|\Psi\rangle = \hat{W}\,|\psi_0\rangle$ and $\hat{O}'_m = \hat{V}^\dagger\hat{O}_m\hat{V}$ a single term of the sum (3.13) is given by

$$E_m = \langle\Psi|\, \hat{R}^\dagger_{\theta_i}\hat{O}'_m\hat{R}_{\theta_i}\, |\Psi\rangle \ . \tag{3.14}$$

Using Eq. (3.12) this leads to

$$\begin{aligned}
E_m &= \langle\Psi| \big[\cos(\theta_i)1 + i\sin(\theta_i)\boldsymbol{n}\cdot\boldsymbol{\sigma}\big]\hat{O}'_m \\
&\qquad \big[\cos(\theta_i)1 - i\sin(\theta_i)\boldsymbol{n}\cdot\boldsymbol{\sigma}\big]|\Psi\rangle \\
&= \cos^2(\theta_i)\langle\Psi|\hat{O}'_m|\Psi\rangle \\
&\quad + \sin(\theta_i)\cos(\theta_i)\langle\Psi| i\Big[\boldsymbol{n}\cdot\boldsymbol{\sigma}, \hat{O}'_m\Big]|\Psi\rangle \\
&\quad + \sin^2(\theta_i)\langle\Psi|\boldsymbol{n}\cdot\boldsymbol{\sigma}\hat{O}'_m\boldsymbol{n}\cdot\boldsymbol{\sigma}|\Psi\rangle \ ,
\end{aligned}$$

where all three expectation values are given in terms of Hermitian operators. By virtue of standard trigonometric identities this can be expressed as

$$E_m = A_m\cos(2\theta_i + \varphi_m) + C_m \ . \tag{3.15}$$

We stress that A_m, φ_m and C_m depend on all angles θ_j with $j \neq i$, since our initial assumption stated that θ_i does only appear in the explicitly selected single-qubit gate. The generalization to Pauli matrices scaled by a factor s can be obtained by replacing $\theta_i \to s\theta_i$. Defining $\omega = 2s$ and summing all constant contributions C_m leads to Eq. (3.5).

Chapter 4

Imposing N-Representability Constraints on Reduced Density Matrices

In this chapter, we present and analyze a method for improving the accuracy of a quantum chemical simulation on a noisy quantum computer by the means of post-processing. We are interested in calculations where the energy of a system in a given state is evaluated via measuring the expectation values $\langle c_i^\dagger c_j \rangle$ and $\langle c_i^\dagger c_j^\dagger c_l c_k \rangle$, i.e., the elements of the one- and two-particle reduced density matrix (1- and 2-RDM), where $c_i^{(\dagger)}$ are fermionic annihilation (creation) operators of the system in second quantization. On a NISQ computer, we will obtain flawed values for these elements. However, we know that the 1- and 2-RDM need to satisfy certain conditions and, hence, could try to mitigate the error by post-processing the result accordingly, to conform with these constraints.

Specifically, we consider the so-called N-representability constraints that the RDMs need to obey [52, 53]. They originate from the N-representability problem posed when trying to guarantee a 2-RDM remains derivable from a valid density matrix of N fermions, while modifying the 2-RDM in a variational approach to minimize the energy [86, 87]. Using N-representability constraints to improve on a quantum chemistry calculation (on conventional computers) in a post-processing manner was previously suggested [88], and also considered in a quantum computing context [89].

In these two references [88, 89], the concept is to pick a subset of the N-representability constraints and project an erroneous RDM from a calculation to (in some norm) the

---

This chapter is available as a preprint in "T. Piskor, F. G. Eich, M. Marthaler, Frank K. Wilhelm, and J.-M. Reiner, arXiv:2304.13401 (2023)". It has been submitted to Physical Review A. The text was largely authored by T. Piskor. All numerical simulations were performed by T. Piskor.

closest matrix that fulfills the selected constraints to obtain an improved RDM. Furthermore, this projection is not solely performed for the two-particle RDM (with elements $\langle c_i^\dagger c_j^\dagger c_l c_k \rangle$), but also the two-hole RDM ($\langle c_i c_j c_l^\dagger c_k^\dagger \rangle$), and particle-hole RDM ($\langle c_i^\dagger c_j c_l^\dagger c_k \rangle$).

We expand on this work by providing a more thorough analysis for the application of this concept in energy calculations using a quantum computer. We investigate, through analytical consideration and in particular numerical simulation, how different sources of noise in a quantum computation affect the measured RDMs; and how the three options of projecting in the respective particle, hole, or particle-hole sector, enhance the result. We consider as test systems three molecules, H_2, LiH and BeH_2, and we consider three quantum noise channels, dephasing, damping, and depolarization, and furthermore shot noise, stemming from performing a limited amount of measurements to evaluate the expectation values for the RDM elements. Based on our results, we propose a practical approach how to utilize having multiple sectors as options to perform the projection in.

This chapter is structured as follows: In Sec. 4.1 we give a short overview on the basic principles of N-representability as well as present notations that are valid throughout this work. Sec. 4.2 explains how we are simulating noise via the superoperator formalism and also shows the operators for the three investigated decoherence types: dephasing, damping and depolarization. Furthermore, we give a brief description how we were simulating shot noise in combination with decoherence, and comment on how each type of noise leads to states violating N-representability. A thorough description of our simulation and post-processing procedure, followed by the definitions of the considered metrics is given in Sec. 4.3. Finally, Sec. 4.4 gives a presentation of the results of our numerical analysis, before we conclude in Sec. 4.5.

4.1 N-Representability Conditions for Hole and Particle-Hole Reduced Density Matrices

In Sec. 2.4, we have already defined the 1- and 2-particle reduced density matrices with its corresponding ensemble N-representability conditions. These conditions not only hold for the 1- and 2-particle RDMs; similarly they also hold for the 1- and 2-hole, as well as the particle-hole RDMs:

$$
\begin{aligned}
{}^1Q^i_j &= \langle \psi | \, c_i c_j^\dagger \, | \psi \rangle \, , && \text{(4.1a)} \\
{}^2Q^{ij}_{kl} &= \langle \psi | \, c_i c_j c_l^\dagger c_k^\dagger \, | \psi \rangle \, , && \text{(4.1b)} \\
{}^2G^{ij}_{kl} &= \langle \psi | \, c_i^\dagger c_j c_l^\dagger c_k \, | \psi \rangle \, . && \text{(4.1c)}
\end{aligned}
$$

One can obtain these from the one- and two-particle RDMs using the following identities:

$$
\begin{aligned}
{}^1Q_i^j &= \delta_{ij} -^1 D_j^i, && \text{(4.2a)}\\
{}^2Q_{ji}^{lk} &= {}^2D_{kl}^{ij} - \delta_{jl}{}^1D_k^i + \delta_{il}{}^1D_k^j \\
&\qquad + \delta_{jk}{}^1D_l^i - \delta_{ik}{}^1D_l^j + \delta_{jl}\delta_{ik} - \delta_{il}\delta_{jk}, && \text{(4.2b)}\\
{}^2G_{kj}^{il} &= \delta_{jl}{}^1D_k^i - {}^2D_{kl}^{ij}. && \text{(4.2c)}
\end{aligned}
$$

As in the case of the 1- and 2-particle RDM, we find the same constraints for the Q- and G formulations. The only difference is the trace for the different sections. These are defined by

$$
\begin{aligned}
\mathrm{Tr}[{}^1Q] &= \eta, && \text{(4.3a)}\\
\mathrm{Tr}[{}^2Q] &= \eta(\eta-1), && \text{(4.3b)}\\
\mathrm{Tr}[{}^2G] &= N(\eta+1), && \text{(4.3c)}
\end{aligned}
$$

where η is the number of orbitals in the system.

N-representability constraints were long utilized to improve quantum chemical calculations. In fact, they stem from the N-representability problem posed when trying to guarantee a 2-RDM can be represented by, i.e., derived from, a proper state of N fermions, while modifying the 2-RDM in a variational approach to minimize the energy [86, 87]. It was also proposed to look not just at the particle, but also the hole and particle-hole sectors to improve numerical methods [88]. Furthermore, the constraints of the particle sector were applied in the context of quantum computing to improve measurement results of RDM elements [89].

We expand on this work by focusing on a quantum computing application, but providing a method that exploits N-representability conditions in the particle, hole, and particle-hole sector. Furthermore, we perform a more thorough analysis how individual noise types affect the performance of our method. To this end, we continue in the next section to explain which kinds of noise we consider, how we describe them, and how the individual N-representability constraints above are affected though them.

4.2 Application of Noise to Perfect Measurements

In this work we consider three different types of decoherence noise, namely damping, depolarization and dephasing. To include stochastic effects, we will also investigate shot noise. In this section we will discuss the superoperator formalism which we utilize, as well as the single noise types and present how the individual noise types influence the

N-representability conditions listed above. The different decoherence types have been discussed in Sec. 1.7.

4.2.1 Shot Noise

Performing computations on a real quantum device requires multiple projective measurements of qubits in the computational basis in order to extract operator expectation values. In this work we faithfully simulated this measurement process to obtain the expectation values influenced by shot noise using the HQS software package `qoqo` [90].

The software does this by grouping the operators (Pauli products) to be measured into sets that can be measured simultaneously. For each set, the quantum circuit is extended by the respective single-qubit rotations such that the Pauli products can be measured in the computational basis. Each extended circuit is then simulated and the resulting final state vectors are obtained. From there, for each measurement shot a bit string is drawn from a probability distribution based on the prefactors of the according final state vector in the computational basis. This bit string is then used to calculate the simulated result of a projective measurement of a certain Pauli string.

We performed $M = 1000$ measurement shots for every Pauli string that we evaluate and took the average of these to calculate the expectation value of the Hamiltonian for a specific system and geometry. We are interested in statistical effects as well, hence, we repeated these steps $R = 100$ times. Therefore, we find 100 different expectation values for all geometries and thus can determine the measurement variance for a measurement protocol relying on M shots per operator.

4.2.2 Influence of Noise on N-Representability

Now we will discuss how the presented types of noise affect the validity of the five N-representability constrains listed in Sec. 4.1. First, we examine the quantum decoherence channels, i.e., dephasing, damping, and depolarization.

The effect of these channels will depend on how we encode the fermionic problem into qubits. In this work, we rely on the Jordan-Wigner transformation (1.36a) and (1.36b).

The first two listed N-representability conditions, hermiticity and antisymmetry, are not affected by any quantum channel. The application still results in a physical state and therefore the measurement of the fermionic operators will always reveal these fundamental properties. This also means that this is independent of the encoding, as long as the resulting fermionic operators are valid. Likewise, the condition of positive semidefiniteness remains intact, since the regarded quantum noise channels are completely positive maps [38].

The final two listed constraints, the trace and contraction relations, on the other hand both rely on the particle number N staying constant. In the Jordan-Wigner encoding the particle number N is dependent on a sum of Pauli Z matrices, since from Eqs. (1.36a) and (1.36b) follows that $c_k^\dagger c_k = \frac{1}{2}(1 - Z_k)$. Dephasing noise, which – as explained above in Sec. 1.7 – is essentially random Z_k errors, therefore commutes with the particle number, i.e., $[Z_k, N] = 0$ for all k. Hence, using the Jordan-Wigner transformation, none of the N-representability constraints that we consider in this work are violated by dephasing noise. For this reason, in our numerical analysis that follows in Sec. 5.2, we do not study dephasing. Note again, that this is due to our choice of encoding, using instead other fermion to qubit mappings like, e.g., the Bravyi-Kitaev transformation [91], the situation would be different. Appendix 4.A shows an explicit calculation for two qubits under the influence of dephasing noise, where it is also proved that neither the trace nor the positive-semidefiniteness is violated. Staying with the Jordan-Wigner encoding, the case is also different for damping and depolarizing noise, since here, e.g., Pauli X errors may occur, and $[X_k, N] \neq 0$, meaning the last to constraints can in fact be violated. Consequently, these decoherence channels will be investigated in our numerics.

For shot noise the case is simple: It is completely statistical in its nature may violate any of the five N-representability constrains mentioned above.

In the following, not all of the five constraints will be dealt with post-processing. Hermiticity and antisymmetry will be guaranteed by construction – we will simulate measuring only the minimal necessary amount of RDM elements and calculate the rest using the according hermiticity and antisymmetry relations in Eq. (2.13a), (2.13b), and (2.14). Positive semidefiniteness and the corrected trace will be attempted to enforce in a post-processing manner as explained in the subsequent Sec. 4.3. Obtaining ${}^1D$ via the contraction in Eq. (2.16) is also nontrivial if the ${}^2D$ measurement is impaired by noise. In order to perform the energy calculation according to Eq. (2.11), or perform the transformations to the two-hole or particle hole sector as in Eq. (4.2a), (4.2b), and (4.2c), without needing to first correct ${}^2D$, we simply simulate the measurement of ${}^1D$ additionally to the 2-RDM; which is an insignificant overhead as the 2-RDM contains quadratically more elements as the 1-RDM.

4.3 Procedure

In this section we explain how we calculated the data used to produce the plots and results presented in the following Sec. 4.4. There, we have chosen to analyze three molecules as example systems, namely H_2, LiH and BeH_2. The first step was to derive the molecular Hamiltonian (2.9), where we represented H_2 in the STO-3G basis, and LiH as well as BeH_2 in the MinAO basis. Full configuration interaction (FCI) – which is equivalent to exact diagonalization – runs have been performed for all three systems, yielding our reference ground state $|\psi_{\mathrm{FCI}}\rangle$ and its respective energy E_{FCI} for each system. Having obtained the state, using the equations given in Sec. 4.1 gives access to the respective RDMs ${}^1D_{\mathrm{FCI}}$ and ${}^2D_{\mathrm{FCI}}$.

Next we simulate decoherence, where we apply either damping, depolarizing, or dephasing noise to the reference state through the superoperator formalism described above in Sec. 4.2 (each with a rate $\Gamma = 10^{-2}$). Note again, that we use the Jordan-Wigner transformation to translate the fermionic operators to Pauli operators, which also defines the representation of the state and how the respective noise types affect it. This process yields a new state $|\psi\rangle$ which mimics the result of a ground state calculation, for each of the example molecules, on a quantum computer, under the influence of decoherence. With this state one can then derive the RDMs ${}^1D_{\mathrm{QC}}$ and ${}^2D_{QC}$, as well as the respective energy E_{QC}, where the index indicates that this a simulated quantum computation.

Now we tried improving on the energy result by projecting the RDMs to the closest RDM that fulfills the N-representability constraints listed in Sec. 4.1. We guarantee that hermiticity and antisymmetry properties are valid by constructing the RDMs from measuring as little as necessary of the matrix elements and calculating the rest of the matrix via the respective relations. Note that hermiticity and antisymmetry are not violated by applying decoherence channels, but this step will be important when we will consider shot noise below.

We have three options for projecting, fixing the two-particle RDMs, the two-hole RDMs, or the particle-hole RDM; these options we call D-projection, Q-projection, and G-projection, respectively. To perform a D-projection, we take the measured particle 1-RDM (${}^1D_{\mathrm{QC}}$) and the two-particle 2-RDM (${}^2D_{\mathrm{QC}}$) and perform a fixed trace and positivity projection to both RDMs, which follows the algorithm from Ref. [89] and is available as a function in the open source software `OpenFermion` [92]. This function enforces the matrices to have a fixed trace – which in this case is related to the particle number as given by Eqs. (2.15a) and (2.15b) – and to be positive semi-definite. For the Q-projection, we use ${}^1D_{\mathrm{QC}}$ and ${}^2D_{\mathrm{QC}}$ together with Eqs. (4.2a) and (4.2b) to obtain ${}^1Q_{\mathrm{QC}}$ and ${}^2Q_{\mathrm{QC}}$. These are then projected with the same function as in the case of the D-projection (fixing the number of holes instead of the number of particles). After the projection we transform back to the two-particle sector to evaluate the energy using Eq. (2.11). Similarly, we proceeded for the G-projection, transforming from the two-particle to the particle-hole sector using Eq. (4.2c), performing the projection, and then transforming back.

The same procedure is done again for simulations where we assume on top of decoherence also a finite amount of projective measurements, i.e., where the results are affected by shot noise (see Sec. 4.2.1).

Note again, that we measured ${}^1D_{\mathrm{QC}}$ additionally to ${}^2D_{\mathrm{QC}}$ (which is a negligible overhead). Since the contraction from the 2-RDM to the 1-RDM following Eq. (2.16) is violated under the presence of noise, we use the measured ${}^1D$ elements to evaluate the energy E_{QC}, and perform the transformation to the hole and particle-hole sectors according to Eq. (4.2a), (4.2b), and (4.2c) to utilize the Q- and G-projections. For the raw measurement, this yielded lower energy results than using a faulty contraction. Furthermore, we observed that we could achieve lower energies for the individual Q- and G-projections; this is in comparison to using faulty transformations, and, in particular, also compared to otherwise needing to perform a D-projection first in order to allow for a reasonable contraction (which would be necessary for the transformations).

We evaluated in the case where we did not include shot noise the energy difference ΔE between the energy E, either of the raw measurement or after post-processing by on of the individual projections, versus the FCI energy for each geometry,

$$\Delta E = E - E_{\mathrm{FCI}}. \tag{4.4}$$

Furthermore, we investigated how close this procedure brings us to the FCI 2-RDM by looking at the fidelity:

$$\mathcal{F}({}^2D, {}^2D_{\mathrm{FCI}}) = \left(\mathrm{Tr}\sqrt{\sqrt{{}^2D}\,{}^2D_{\mathrm{FCI}}\sqrt{{}^2D}}\right)^2, \tag{4.5}$$

where ${}^2D$ describes the measured or once projected 2-RDM. Note, that the square roots are well-defined as we are dealing only with positive semidefinite matrices in this case.

Including shot noise, we utilize following paradigm: We envision an experiment, where we assume resources to perform a total of $M = 1000$ measurement shots for every operator to measure; that is also for every data point in the plots in the following Sec. 5.2. In order to analyze the statistics of performing such experiments, we repeat the same process $R = 100$ times. For each data point, we therefore obtain 100 different values $\mathcal{E}_i$ with $i \in \{1, \dots, R\}$, where each $\mathcal{E}_i$ itself is the result of averaging over 1000 shots (hence, for every point a total of $R \cdot M = 10^5$ shots are simulated). Finally, we average over the repetitions, yielding

$$\mathcal{E} = \frac{1}{R}\sum_{i=1}^{R}\mathcal{E}_i, \tag{4.6}$$

which is the energy we calculate again the energy difference to the reference energy:

$$\Delta E = \mathcal{E} - E_{\mathrm{FCI}}. \tag{4.7}$$

The reason for this averaging scheme is that, particularly for NISQ hardware, the number of measurements are a scarce resource. Hence, we assume only 1000 shots per operator in a single experiment. Averaging again over multiple repetitions gives not only a more reliability expectation value to compare to the reference energy, but also allows to analyze the variance over the repetitions,

$$\mathrm{Var}(\mathcal{E}) = \frac{1}{R}\sum_{i=1}^{R}\left(\mathcal{E}_i - \mu\right)^2, \quad \mu = \frac{1}{R}\sum_{i=1}^{R}\mathcal{E}_i. \tag{4.8}$$

This quantity gives insight about the expected accuracy of an energy measurement with only a limited number of 1000 shots; specifically, it is interesting whether the proposed post-processing method lowers the variance.

In the following section we will show the results of our calculations. There, we provide an analysis how individual projections improved on the above quantities for the different systems and noise types. Furthermore, we will comment on approaches we tried to concatenate multiple projections to improve the results as much as possible, and how we propose to select the best projection method.

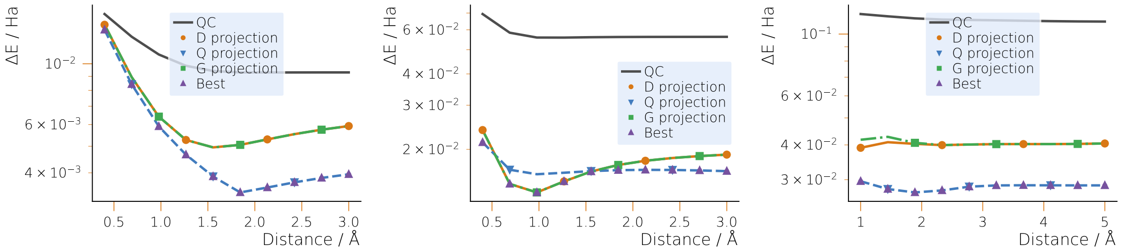
4.4 Numerical Analysis

After discussing the procedure above, here we show our data on how much improvement the D-, Q-, and G-projections grant on our simulated measurement results when dealing with shot and decoherence noise. We examine if there is a preferred projection type for certain noise types, systems or geometries, and furthermore investigate how combining the different projection types alters the result.

The effects of damping and depolarization in addition to shot noise have been investigated on three systems, namely H_2, LiH, and BeH_2. At first, simulations without shot noise will be discussed in Sec. 4.4.1 and afterwards the effect of shot noise will be included in Sec. 4.4.2. Besides the investigation of the energy deviations with respect to the FCI solution, we will furthermore take a look at the corresponding state fidelities for the individual projection methods. The section discussing shot noise will provide figures with energy errors, as well as measurement variances.

4.4.1 Simulations without Shot Noise

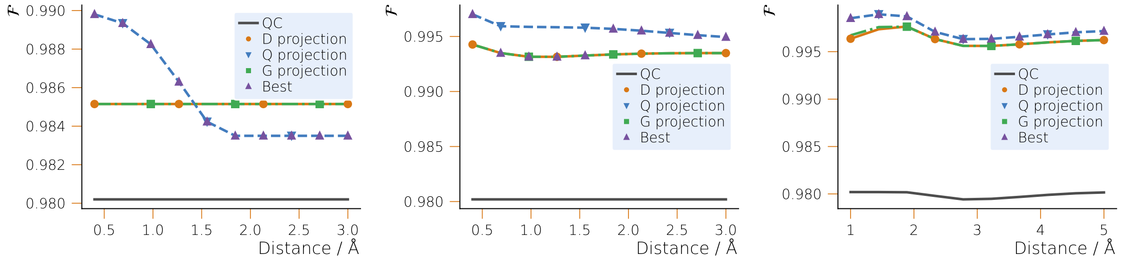
In this section we will investigate the energy errors and fidelities when performing simulations of the three systems with damping or depolarization present, but without shot noise, and compare the simulations without post-processing and with the three projection types to see how these improve the analyzed quantities. Our simulation results are compiled in Fig. 4.4.1 and Fig. 4.4.2.



(a) Energy errors for H_2 under influence of damping noise.

(b) Energy errors for LiH under influence of damping noise.

(c) Energy errors for BeH_2 under influence of damping noise.



(d) Fidelities for H_2 under influence of damping noise.

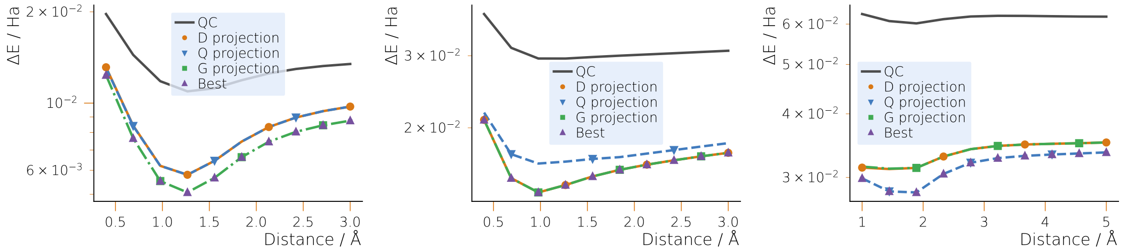
(e) Fidelities for LiH under influence of damping noise.

(f) Fidelities for BeH_2 under influence of damping noise.

FIGURE 4.4.1: Energy errors and fidelities for the three investigated examples H_2, LiH and BeH_2 for damping noise with a rate of $\Gamma = 10^{-2}$. Shown are the results from measurements of the 1- and 2-RDM without post-processing, which are indicated as *QC*, as well as the results from the single projections D, Q and G. *Best* indicates the results yielding the best energies.

At first, we will analyze the data shown in Fig. 4.4.1, where damping noise has been considered. In Fig. 4.4.1a to 4.4.1c we plot the energy errors, and in Fig. 4.4.1d to 4.4.1f the fidelity, each time for H_2, LiH, and BeH_2, respectively, at different inter-atomic distances. For all cases, we plot the quantities derived from measurements of the 1- and 2-RDM without post-processing (labeled as QC), as well as after performing a single D-, Q-, or G-projection (labeled accordingly as D, Q, or G). We also highlight the projection line that resulted in the smallest energy error (labeled as Best), making it easier to track the projection that yields the best energy.

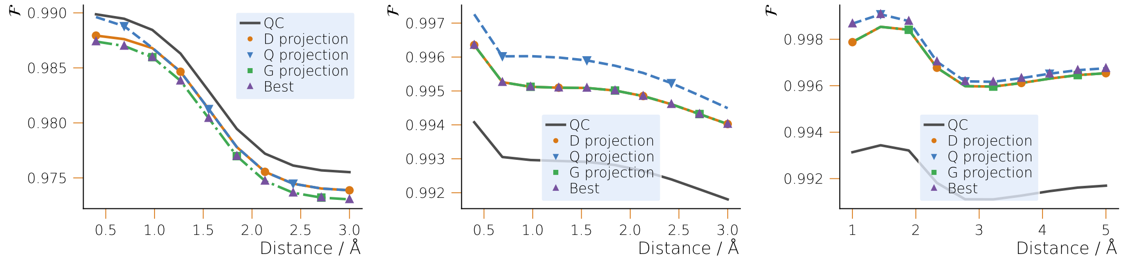
From all sub-figures we can observe a significant improvement in terms of the energy error for all projection types; almost an order of magnitude lower errors are reached. The data also verifies numerically that the projections do not fall below the FCI energy (which would be unphysical). We furthermore observe that certain projections lead to somewhat smaller energy errors. As for instance in Fig. 4.4.1a it can be seen that the Q-projection leads to the best results for all distances between the hydrogen atoms, whereas the D- and G-projection lead to the same, worse result. Interestingly, D- and G-projection lie on top of each other and the Q-projection deviates, which hints at a fundamental way damping affects the respective two-particle, particle-hole, and two-hole



(a) Energy errors for H_2 under influence of depolarizing noise.

(b) Energy errors for LiH under influence of depolarizing noise.

(c) Energy errors for BeH_2 under influence of depolarizing noise.



(d) Fidelities for H_2 under influence of depolarizing noise.

(e) Fidelities for LiH under influence of depolarizing noise.

(f) Fidelities for BeH_2 under influence of depolarizing noise.

FIGURE 4.4.2: Energy errors and fidelities for the three investigated examples H_2, LiH and BeH_2 for depolarizing noise with a rate of $\Gamma = 10^{-2}$. Shown are the results from measurements of the 1- and 2-RDM without post-processing, which are indicated as *QC*, as well as the results from the single projections D, Q and G. *Best* indicates the results yielding the best energies.

sectors. However, it is dependent on the system, and even within a system dependent on the inter-atomic distance (see Fig. 4.4.1c), which projection type yields the best energy.

The fidelity curves in Fig. 4.4.1d to 4.4.1f show general improvement of the fidelity performing post-processing, often times coming much closer to perfect fidelity than the initial QC result. We also observe the same behavior w.r.t. the D- and G-projection yielding the same value and Q deviating. However, we see that the best projection in terms of the energy value does not necessarily yield the highest fidelity. While from a heuristic argument one would expect lowering the energy error leads to approaching the correct 2-RDM as well, yet there is no direct connection of course. This can be easily seen from considering examples with a dense lower spectrum, or even a degenerate ground state. But it is important to realize that optimizing for the best energy will not guarantee all properties of the 2-RDM to be optimized as well. If one is interested in quantities other than the energy, one might try to alternate the approach.

Comparing the above observations with the data in Fig. 4.4.2, where depolarizing noise was considered, we generally see similar results. However, a few differences can be pointed out: Which projection type yielded the best energy in the three systems (and for particular inter-atomic distances) is not equal to the case of damping noise. We also do not observe anymore the strong link between the D- and G-projections (see Fig. 4.4.2a). Hence, the best projection type does depend really on all variables considered; the chemical system, the inter-atomic distance, as well as the noise type, and there is no obvious a priori choice. Another key difference to the previous figure is in Fig. 4.4.2d, where we find that, unfortunately, the fidelity is actually reduced for all projections compared to the raw QC calculation. This highlights on the other hand our advisory of being cautious when looking at properties other than the energy of the system.

Another point we would like to address is that so far we only looked at applying a single projection, either D, Q, or G. In Refs. [88, 89] an iterative approach, where one applies one projection type after the other in an alternating fashion until the result converges, is proposed. This is following the hope that in this way, the end result is as closely N-representable as possible using the projections at hand.

Pursuing this idea, we tried alternating sequences of the D-, Q-, and G-projections as well. Following a projection with another one of a different kind in some cases changed (not necessarily lowered) the energy difference to the FCI reference. However, in our systems we could not observe an improved energy when using a projection series versus the best energy result after only one projection. Looking at the fidelity instead, we also could see quantitative changes, but not find a conclusive improvement in the sense that projection series would lead to higher fidelities.

On top of this analysis of projection series, we also tested if one can reach lower energy values if one performs *partial* projections, possibly avoiding phenomena like local minima. Here, we tried to iteratively post-process RDM's by only changing it towards the fully projected RDM by a small amount,

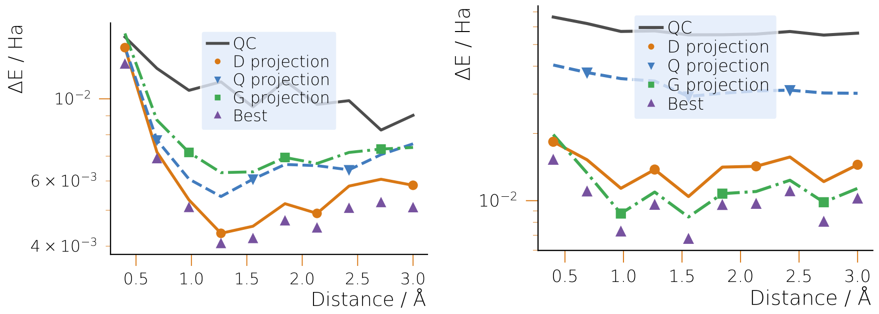
$$^2D_{i+1} = \alpha B(^2D_i) + (1-\alpha)(^2D_i), \tag{4.9}$$

where $B(^2D_i)$ is the two-particle RDM that stems from the energetically best projection (D, Q, or G) of ^2D_i, i is the iteration step, and $\alpha \in [0,1]$ is the projection percentage. However, we have observed that this procedure, even for very small values of $\alpha = 0.001$, converges again to the best result of simply one of the three projections.

Particularly since we are interested mostly in the energy calculation, and that the projections do not yield unphysical energies below the FCI reference, we found the easiest approach for obtaining the best possible energy with the post-processing options available is to apply once the D-, Q, and G-projection and pick the best energy result. This is a very simple to implement strategy that is furthermore very efficient computationally, particularly compared to iterative approaches.

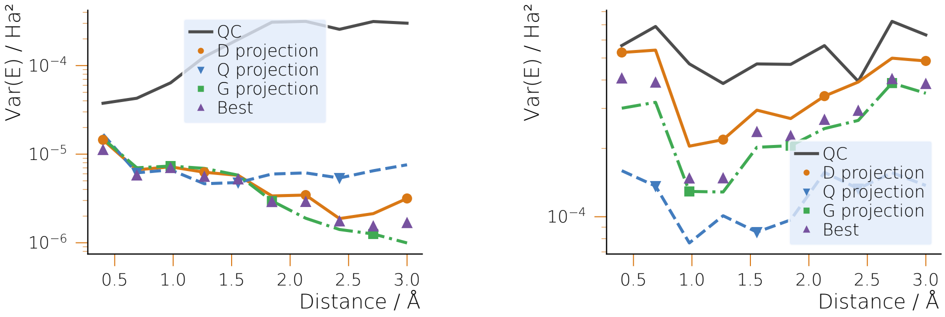
4.4.2 Simulations with Shot Noise

We now include effects of a finite number of measurements in the simulations of our chemical systems. Though, we refrained from simulating the effect of decoherence in combination with shot noise for BeH_2 due to time and resource constraints. We simulated again damping in Fig. 4.4.3 and depolarization in Fig. 4.4.4, now with the addition of shot noise stemming from measuring Pauli strings each with 1000 shots, furthermore averaging over 100 repetitions of such a scenario. Note, that our investigated regime, the effect of decoherence was always strong compared to the statistical fluctuations from limited measurements. This guarantees in our simulations for energy results to remain above the FCI reference, which would not necessarily be the case for pure shot noise. However, if one could drop below the FCI energy, our approach for picking the lowest energy projection as *best* projection would not be meaningful. On the other hand, in



(a) Energy errors for H_2 under influence of damping and shot noise.

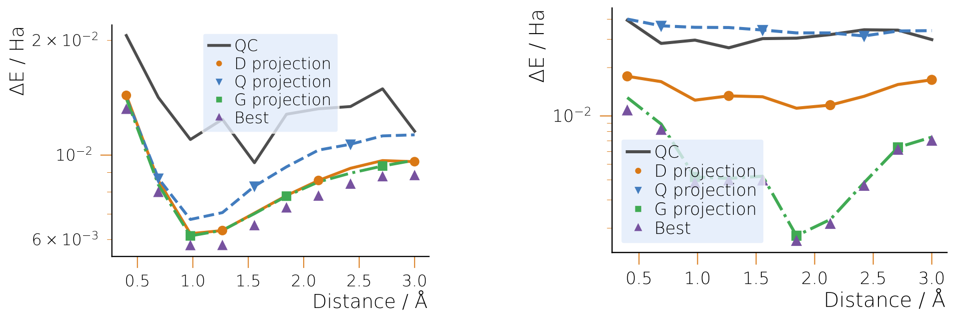
(b) Energy errors for LiH under influence of damping and shot noise.



(c) Variances for H_2 under influence of damping and shot noise.

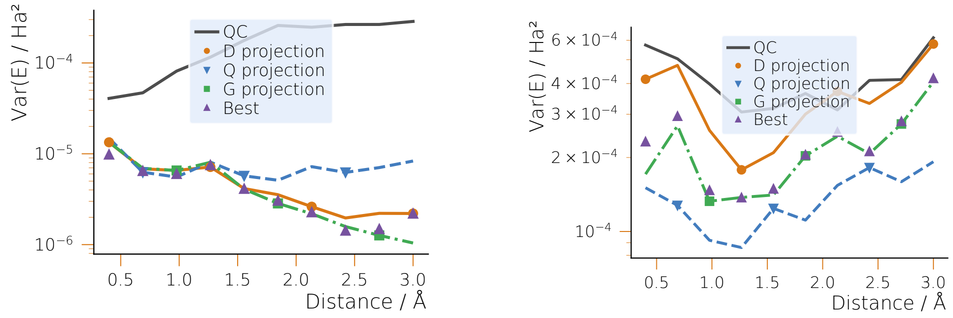
(d) Variances for LiH under influence of damping and shot noise.

FIGURE 4.4.3: Energy errors and measurement variances for the three investigated examples H_2, LiH and BeH_2 for damping and shot noise with 1000 measurement shots, 100 repetitions and a rate of $\Gamma = 10^{-2}$. Shown are the results from measurements of the 1- and 2-RDM without post-processing, which are indicated as *QC*, as well as the results from the single projections D, Q and G. *Best* indicates the results yielding the best energies.



(a) Energy errors for H_2 under influence of depolarizing and shot noise.

(b) Energy errors for LiH under influence of depolarizing and shot noise.



(c) Variances for H_2 under influence of depolarizing and shot noise.

(d) Variances for LiH under influence of depolarizing and shot noise.

FIGURE 4.4.4: Energy errors and measurement variances for the three investigated examples H_2 and LiH for depolarizing and shot noise with 1000 measurement shots, 100 repetitions and a rate of $\Gamma = 10^{-2}$. Shown are the results from measurements of the 1- and 2-RDM without post-processing, which are indicated as *QC*, as well as the results from the single projections D, Q and G. *Best* indicates the results yielding the best energies.

our chosen regime this is not an issue, and we believe that this regime of decoherence dominating shot noise is realistic to assume for NISQ applications.

Looking at Fig. 4.4.3 and Fig. 4.4.4, we observe very similar results to Fig. 4.4.1 and Fig. 4.4.2: We find a similar order of improvement in terms of energy error, and the best projection w.r.t. the energy is again not easily predictable but varies between systems, noise type and inter-atomic distance. Note, that now the Best label does not follow a specific projection type; here, we choose the best projection w.r.t. the energy for every repetition individually. Hence, for every repetition a different projection turns out to be the favored one, averaging over 100 repetitions yields a better value than for every other pure projection type.

Another difference to the plots of Sec. 4.4.1 is that instead of the fidelity we now plot the measurement variance as a relevant quantity in the context of measurement errors due to statistical shot noise. Again, the points labeled Best do not present necessarily the best variance, as the optimization happened according to the energy. But importantly, we see a significant reduction in the variance when post-processing the results by projecting to fulfill our selected N-representability constraints. Specifically, in the case of larger distances in H_2, the reduction spans two orders of magnitude.

This variance reduction is a remarkable feature of the presented projection method. It could potentially enable to measure quantities with a rather low number of shots while still remaining confident about the accuracy of the result. Hence, the method could be a good candidate to reduce the measurement overhead which is a considerable obstacle in quantum computing, particularly in the NISQ era where quantum resources are fairly limited. This sparks interest in studying this property of our, or alike, methods more extensively in further research.

4.5 Conclusion

The aim of this work was to try to reduce the energy error from 1- and 2-RDM calculations on a (simulated) quantum computer limited by decoherence and finite number of measurement shots. We investigated a post-processing method that enforced certain general N-representability constraints by projecting the measured RDMs into the subspace where these conditions were fulfilled. Here, we regarded projecting the RDMs not just in the particle sector, but also the hole and particle-hole sector – where one can switch between the sectors by simple transformations.

Specifically, we guaranteed hermiticity and antisymmetry by construction of the RDM from the measurements, and enforced positive semidefiniteness as well as the correct trace through the post-processing projection. Analyzed were then ground state calculations of H_2, LiH and BeH_2 under the influence of damping and depolarizing channels, as well as shot noise.

We found the post-processing according to the N-representability constraints led to an improvement in terms of the energy error for all investigated example systems and noise types for all projections, i.e., the D-, Q- and G-projection in their respective particle, hole, and particle-hole sectors. Similarly, the state fidelity was generally improved as well. We could not observe an easily explained behavior which of the D-, Q-, or G-projection performed best depending on the system, the inter-atomic distance, or noise type. On the other hand, we also found the best approach to always find the smallest energy error – independent of the system or noise type – is to simply take the smallest value of the three presented projection types. Using approaches with series of alternating projections as previously suggested [88, 89], or an iterative variant relying on partially projecting the RDM in each step, did not lead to better energy results; furthermore, they did not necessarily lead to the lower state fidelities either. We note, that this simple way of finding the lowest energy hinges on the fact that we operate in a regime that is dominated by decoherence (versus shot noise), where we did not observe projecting to energies lower than the FCI reference. However, this regime is reasonable to assume for NISQ devices.

In terms of the measurement variance, another investigated metric for shot noise simulations, we could see that the application of the proposed post-processing method led to a decrease by up to an order of magnitude. The precise reduction in variance was again depending on the sector in which the projection was performed, without a clear choice to be made a priori. While our approach of optimizing for the lowest energy did not necessarily yield the lowest measurement variance, it performed very well in general. This makes it a viable approach for not just improving the energy error but also the variance; which is a particularly compelling feature, as this would enable to significantly increase the confidence in the accuracy of a quantum computation with a restricted number of measurement shots, especially considering the magnitude of the improvement.

In conclusion, we found our presented method of mitigating decoherence and shot noise to be very useful for improving energy calculations, particularly with respect to the highly effective reduction of the measurement variance. In the considered noise regime, our practical approach to utilize the three different sectors to project proved not just to be simple but also fruitful. The post-processing has low computational effort and there is no overhead to the quantum computation itself. Our positive results spark interest for expanding on the method by including more constraints, e.g., further N-representability conditions, or other system-specific conserved symmetries. Investigating these ideas we leave for future work.

APPENDICES

4.A Dephasing and N-Representability Conditions

In the following, we consider a two-qubit system with a given noiseless density matrix ρ underlying the effect of dephasing. Hence, we observe a transformation $\rho \to \rho'$, where the transformed noisy density matrix is given as

$$\rho' = \sum_i M_i \rho M_i^\dagger . \tag{4.10}$$

Here, the matrices M_i describe the different combinations of the Kraus operators for the corresponding noise type. In case of two qubits, we can write $M_0 = K_0 \otimes \mathrm{I} \cdot \mathrm{I} \otimes K_0$, $M_1 = K_1 \otimes \mathrm{I} \cdot \mathrm{I} \otimes K_0$ and so on. Following the assumption that the dephasing rates are equal for both qubits, we find three different Kraus operators:

$$K_0 = \begin{pmatrix} \sqrt{1-p} & 0 \\ 0 & \sqrt{1-p} \end{pmatrix}, K_1 = \begin{pmatrix} \sqrt{p} & 0 \\ 0 & 0 \end{pmatrix}, K_2 = \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{p} \end{pmatrix}, \tag{4.11}$$

with $p = 1 - \mathrm{e}^{-\Gamma}$ being the dephasing term with dephasing rate $\Gamma = \frac{T_g}{T_{\mathrm{coh}}}$. Here, T_g and T_{coh} describe the gate and coherence time, respectively. Now we can write the dephased density matrix ρ' as

$$\rho' = \begin{pmatrix} \rho_{00} & (1-p)\rho_{01} & (1-p)\rho_{02} & (1-p)^2\rho_{03} \\ (1-p)\rho_{10} & \rho_{11} & (1-p)^2\rho_{12} & (1-p)\rho_{13} \\ (1-p)\rho_{20} & (1-p)^2\rho_{21} & \rho_{22} & (1-p)\rho_{23} \\ (1-p)^2\rho_{30} & (1-p)\rho_{31} & (1-p)\rho_{32} & \rho_{33} \end{pmatrix}. \tag{4.12}$$

Since we are interested in obtaining the 1-RDM under influence of dephasing noise, we can now determine the expectation value for each element ${}^1D^i_j$

$$\left({}^1D^i_j\right)' = \langle c_i^\dagger c_j \rangle' = \mathrm{Tr}\left(\rho' c_i^\dagger c_j\right). \tag{4.13}$$

Before we can write the entire dephased 1-RDM for two qubits, we have to perform a fermion-qubit mapping (for instance with a Jordan-Wigner transformation) at first:

$$c_k = Z_0 \otimes \cdots \otimes Z_{k-1} \otimes \sigma_k^+, \tag{4.14a}$$

$$c_k^\dagger = Z_0 \otimes \cdots \otimes Z_{k-1} \otimes \sigma_k^-, \tag{4.14b}$$

with $\sigma^+ = \frac{1}{2}(X + iY)$ and $\sigma^- = \frac{1}{2}(X - iY)$. Finally, the dephased 1-RDM can be evaluated:

$$^1D' = \begin{pmatrix} \rho_{11} + \rho_{33} & \frac{1}{2}(1-p)^2(\rho_{12} + \rho_{21}) \\ \frac{1}{2}(1-p)^2(\rho_{12} + \rho_{21}) & \rho_{22} + \rho_{33} \end{pmatrix}. \tag{4.15}$$

It can be easily seen that the trace of the dephased 1-RDM equals the trace of the noiseless 1-RDM.

To show that the positive semi-definiteness is neither violated, we can use the fact that for a $n \times n$ Hermitian matrix all principal minors and the determinant of the matrix itself are non-negative. In case of a 2×2 matrix there is only one principal minor, which is simply the trace of the matrix. Since the principal minors solely contain the diagonal elements of the 1-RDM ($\rho_{11} + \rho_{33}$ and $\rho_{22} + \rho_{33}$), which per definition are positive, we find that the first criterion fulfills the condition for positive semidefiniteness.

In order to prove the second criterion (i.e. showing that the determinant of the entire dephased 1-RDM $^1D'$ is positive) we rewrite the 1-RDM as

$$^1D' = \begin{pmatrix} \Lambda_0 & \frac{1}{2}(1-p)^2\Lambda_1 \\ \frac{1}{2}(1-p)^2\Lambda_1 & \Lambda_2 \end{pmatrix}, \tag{4.16}$$

with $\Lambda_0 = \rho_{11} + \rho_{33}, \Lambda_1 = \rho_{12} + \rho_{21}, \Lambda_2 = \rho_{22} + \rho_{33}$. If we at first consider the case without dephasing noise (i.e. $p = 0$) the determinant of the dephased 1-RDM is given as:

$$|^1D'| = \Lambda_0\Lambda_2 - \frac{1}{4}\Lambda_1^2 \geq 0. \tag{4.17}$$

We know that Eq. (4.17) is non-negative, since per definition a noiseless 1-RDM has to be positive semi-definite. On the other hand, we can consider $p = 1$. In that case we have

$$|^1D'| = \Lambda_0\Lambda_2 \geq 0, \tag{4.18}$$

which is also non-negative since Eq. (4.18) contains only the diagonal elements of the 1-RDM. Finally, the last case with an arbitrary value for $p \in (0, 1)$ also guarantees that we have a non-negative determinant and thus a positive semi-definite matrix:

$$\Lambda_0\Lambda_2 \geq \Lambda_0\Lambda_2 - \frac{1}{4}(1-p)^4\Lambda_1^2 \geq \Lambda_0\Lambda_2 - \frac{1}{4}\Lambda_1^2 \geq 0. \tag{4.19}$$

Another way to show that the dephased reduced density matrix is positive semi-definite is via the square root of a matrix. If there exists a positive semi-definite, Hermitian matrix B satisfying $M = BB$, then M is positive semi-definite. This unique matrix B is denoted with $B = M^{\frac{1}{2}}$. One way to obtain the square root of a matrix is via a Schur or Cholesky decomposition.

Part III

Data Interpolation for Quantum Chemistry Methods with Machine Learning

Chapter 5

Mechanistic Reaction Studies

Mechanistic studies of chemical reactions are one of the most important and wide-spread applications of computational quantum chemistry [93]. A minimal meaningful workflow consists of locating reactants, products and a corresponding transition state (TS). This allows one to assess the thermodynamics and kinetics of reactions *without* thermal contributions and thereby reveal reaction mechanisms. It is highly desirable to perform vibrational analysis for stationary points of a Potential Energy Surface (PES) to elucidate whether the structures correspond to minima or saddle points. This also allows computing reaction rate coefficients via canonical Transition-State Theory (TST) taking vibrational degrees of freedom into account [94]. To be more rigorous, one should also compute reaction paths connecting the stationary points, one option being following the Intrinsic Reaction Coordinate (IRC) starting from the TS [95].

These objectives are routinely reached by direct dynamics approaches [96], *i.e.* methods evaluating energy and its derivatives on-the-fly. The number of *ab initio* quantum chemistry calls typically reaches hundreds and more. One should not ignore unsuccessful attempts to locate TS and find IRC, which are not uncommon in mechanistic reaction studies. Qualitatively accurate description of reaction paths involving breaking or forming of chemical bonds can only be achieved with advanced electronic structure methods. Often but not always those include static correlation, *i.e.* multi-reference and multi-configurational wave function approaches [97]. As an alternative, Density Functional Theory (DFT) with carefully selected hybrid and double-hydrid functionals can provide accurate results [98]. All these methods are, however, computationally demanding even for moderate-size systems. Therefore, the researchers often apply computationally cheaper methods for the mechanistic reaction studies, employing more accurate electronic structure theories only to stationary points (for an example, see the study of

---

This chapter is available as a preprint in ”T. Piskor, P. Pinski, T. Mast, and V. V. Rybkin, arXiv:2304.00942 (2023)”. It has been submitted to Physical Chemistry Chemical Physics. The text was largely authored by T. Piskor. All numerical simulations were performed by T. Piskor.

the reaction between ferrocenium and trimethylphosphine [99]). Some such heuristic method combinations are known as composite methods, or recipes, *e.g.* the Gaussian-n family [100–102] and CBS-QB3 [103]. Despite many successful applications, these approaches imply properties inconsistent with geometries. They are designed not for PES exploration, but rather for stationary point calculations, and are prone to unpredictable errors (for an example, see [104]).

Ideally, a composite method should provide accurate energies consistent with geometric structures at least in a relevant part of the configuration space, while being computationally feasible, *i. e.* based only on few energy/force evaluations at the high-level of electronic structure theory. This implies obtaining a locally fitted PES: once it is available, one can complete many tasks "free" of charge, including reactive molecular dynamics simulations, canonical and variational TST calculations [94] and so on.

A fitted PES can be efficiently generated using one of the rapidly evolving Machine-Learning (ML) approaches (for the general method overviews see review [105] and perspective [106]). These typically require little, if any, feature design: mainly, the structures with associated energies, and sometimes also energy gradients, are needed as input. They demonstrate remarkable flexibility, successfully describing even non-adiabatic processes [107, 108] and systems without a classical atomistic structural formula [109]. The main type of application for such ML-PES is extensive Molecular Dynamics (MD) sampling (see exemplary applications to organic crystals [110] and liquid water [111]). The resulting PES is often called "global" as it embraces vast regions of the configuration space, although the reactive regions of it are typically not covered. There are only a handful of applications of ML-PES fitting to reactive systems: second-order nucleophilic substitution (S_N2) [112], pericyclic [113], decomposition [114], dissociation [115], Diels-Alder [116] and proton-transfer reactions [117]. In addition, ML-PES have been successfully used for automatic mechanism discovery [118]. All these applications, however, aimed at "global" PES fitting required abundant data: typical sets include at least thousands of data points. Most of them, therefore, used Density Functional Theory (DFT) methods or cheaper many-body correlated wave functions as underlying electronic structure theory.

To facilitate routine mechanistic reaction studies one should be able to construct semi-local reactive PES with only a few hundreds of data points using high-level *ab initio* methods. Such work has been performed by Young *et al.* [119] for several model processes, although the authors restricted themselves to DFT methods, which are typically insufficient to describe chemical reactions due to their single-reference character.

In this work we propose a multi-level protocol for generating semi-local reactive PES for routine mechanistic reaction studies based on ML methods with small data sets and a combination of electronic structure theories. Computationally cheaper DFT is used to generate relevant structures, whereas high-level *ab initio* methods are applied to refine the energetics. The multi-level nature of the protocol makes it suitable for incorporating evolving quantum computing methods which promise quantum advantage for correlated electronic-structure problems [120, 121], while remaining non-routine. The feasibility of the approach is demonstrated with two organic reactions: monomolecular Bergman cyclization of enediyne to para-benzyne [122] and a bimolecular S_N2 reaction of chloromethane with a bromide ion [123]. The former is a textbook example of a chemical reaction involving multi-reference character [124] and is treated with a complete-active-space self-consistent-field method, whereas the latter can be described with a single-reference correlated method [125] and is treated with a coupled-cluster approach.

This paper is organized as follows. In Sec. 5.1, we introduce the simulation protocol. In particular, we focus on obtaining relevant geometries for the investigated reactions, as well as reference methods and the machine learning technique of choice. Next, we present the results: PES sections, energy and force-prediction errors, and performance of the ML-PES for geometry optimization, vibrations and reaction rates. We conclude and give an outlook for this work in Sec. 5.4.

5.1 Methodology

5.1.1 Simulation Protocol

We applied the general protocol for fitting a semi-local reactive PES, which includes the following steps:

1. Optimize reactants and products with a cheaper electronic structure method;
2. Find an approximate reaction path using the Nudged Elastic Band (NEB) [126] method with a cheaper electronic structure method;
3. Select points along the reaction path to form the data set;
4. Calculate energies and forces with a reference correlated method;
5. Split this data set into training, validation and test subsets;
6. Fit the ML-PES, validate and test using the corresponding subsets.

After being generated, the fitted PES was employed to compute the properties: stationary points, harmonic vibrations and intrinsic reaction coordinates in both directions from the TS. In addition, we calculated rate coefficients using transition state theory

(TST). To evaluate the quality of the fitted semi-local PES the properties were compared with those obtained directly by the reference electronic structure method with energies and forces calculated on-the-fly. The details of each step are described below.

5.1.2 Electronic Structure Calculations

We used DFT with the PBE exchange-correlation functional [127] in combination with the double-zeta split-valence def2-SVP basis set [128] as a cheaper electronic structure method for both reactions. These calculations have been performed with `NWChem` [129].

For the Bergman cyclization the complete active space self-consistent field [130] (CASSCF) method with the double-zeta split-valence def2-SVP basis set [128] was used as the reference.

For the initial configuration, we used MP2 natural orbitals to select the active space guess, which consisted of 12 electrons in 12 orbitals (12, 12) as suggested by Lindh and Persson [131]. The CASSCF wave function for atomic configurations in the data set was calculated with the previous guess for the molecular orbitals, so that no discontinuities occurred in the potential energy surface. CASSCF calculations were performed using `PySCF` [132, 133].

The reference correlated method for the $\mathrm{S_N2}$ reaction was coupled-cluster singles and doubles (CCSD) [134] based on the Hartree-Fock reference within the spin-restricted formalism. As in the previous example, CCSD calculations have been performed with `PySCF` [133]. The basis set of choice has been the double-zeta correlation-consistent basis set cc-PVDZ [135–137].

5.1.3 Data Set Generation

`NWChem` [129] has been used to generate the data sets by performing NEB [126] calculations. For both reactions, DFT (the cheaper method) energies and forces were used, with reactants as start points and products as end points. We used 100 beads for the $\mathrm{S_N2}$ reaction and 200 for the Bergman cyclization. The geometries were randomly selected from the optimized elastic band. Thus, they lie on the approximate transition path of the PES obtained from the cheaper method, rather than on the path of the reference method.

5.1.4 Machine Learning Fit of the Semi-Local Reactive PES

We opted for the symmetric gradient domain machine learning (sGDML) method [57, 138, 139] to fit the PES as it efficiently employs the forces and does not require feature design: sGDML uses inverse atomic distances as features and the interatomic forces as the output. Importantly, sGDML has been shown to faithfully reproduce PES from a limited number of structures. Moreover, it has been successfully applied to small molecules using coupled-cluster reference data [140].

To perform one sGDML fit, a certain number of points from the data set is taken to define the training, validation and test set. These numbers are detailed in Sec. 5.2. The first step within one sGDML run is to generate a first model by constructing the kernel matrix out of the training points and setting values for the hyper-parameters σ and λ. The regularization parameter is hereby fixed to a certain value and the length scale σ is varied during the learning process. After the training step, the model is validated against the validation set, where the energy and gradient errors are determined. This procedure is continued until the best σ is found. As a last step the optimized model is validated against the test set, which has no data from neither the training nor validation set and the errors in energy and gradients are compared.

5.1.5 Property Calculation

We optimized the stationary points of the PES using DFT-optimized structures as initial guess and followed the IRC from the TS in both directions with `pysisyphus` [141]. Both, the fitted PES and the reference electronic structure method in `PySCF` [133] were employed as energy functions. Gradients for both PES types were calculated analytically.

We calculated harmonic vibrations for all stationary points on the PES and computed the reaction rate coefficients using conventional TST with a harmonic oscillator-rigid rotor approximation for the partition functions. The working equations and the corresponding rates are given in Appendix 5.A.

5.2 Results

In this section, we present the results for our investigated examples. At first, the PES scans for the reference method and sGDML, as well as the energy differences between both methods will be shown. The next analysis will provide information about the mean absolute errors of the geometries between the reference and machine learning method for the reactant and TS. To perform geometry optimizations for both structures, we used the open-source package `pysisyphus`. In addition, we computed and compared harmonic vibrations for the optimized structures. Besides the intrinsic reaction coordinate we finally compared the reaction rate coefficients for all methods and reactions. The intrinsic

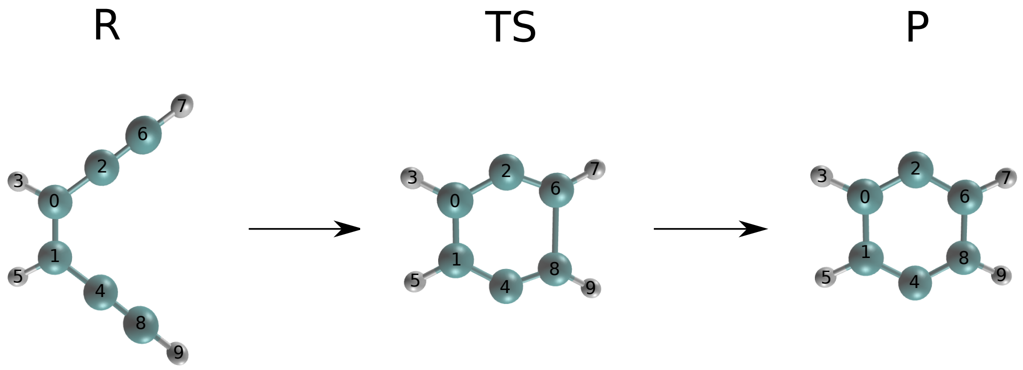


FIGURE 5.2.1: Bergman cyclization of enediyne. The reactant, transition and product states are labeled with R, TS, and P, respectively. Elements are colored as follows: carbon (cyan), hydrogen (white).

reaction coordinate has been determined with `pysisyphus` as well, where the Euler-Predictor-Corrector integrator was used.

5.2.1 Bergman Cyclization of Enediyne

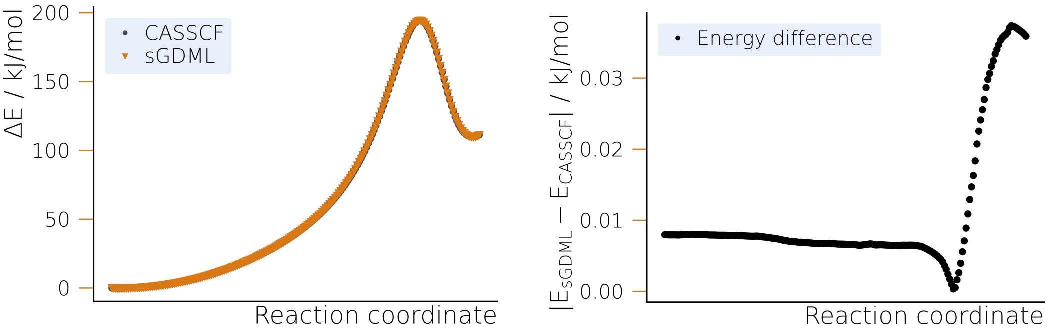
The reaction is shown in Fig. 5.2.1, including the optimized reactant and product.

5.2.1.1 Model Training

We created an sGDML model for the data set consisting of 200 data points: we used 150 training, 30 validation and 20 test points. The regularization parameter was set to $\lambda = 10^{-15}$ and the best length scale was found to be $\sigma = 6$. With these parameters, we obtained a model with a Mean Absolute Error (MAE) of 0.0067 kJ/mol and a Root Mean Square Error (RMSE) of 0.0071 kJ/mol for the energy, whereas the MAE and RMSE for the forces was $0.0033\,\mathrm{kJ/Å \cdot mol}$ and $0.0075\,\mathrm{kJ/Å \cdot mol}$, respectively, on the test data set.

5.2.1.2 Potential Energy Surface Scan

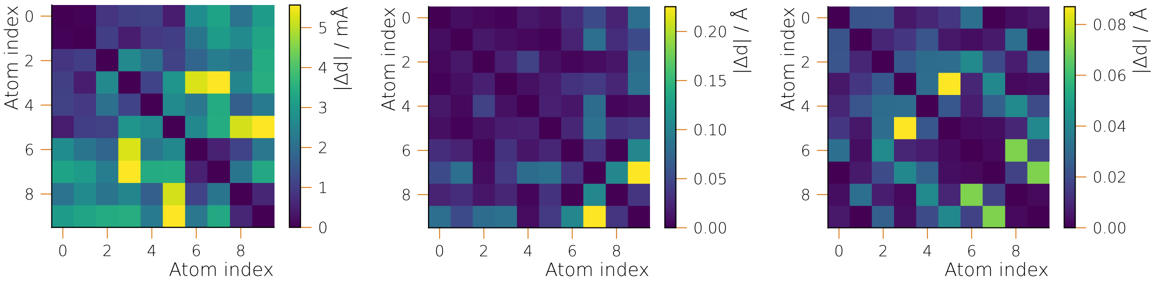
The PES profiles corresponding to the DFT optimized NEB path computed with sGDML and CASSCF are shown in Fig. 5.2.2a. As expected from the small values of MAE and RMSE, the two surfaces agree within chemical accuracy (see Fig. 5.2.2b). Starting from the reactant state and moving towards the TS the energy error is small and practically constant. The smallest error can be found in the vicinity of the TS, increasing significantly towards the product state, although still being two orders of magnitude within chemical accuracy.



(a) Potential energy surface for CASSCF and sGDML.

(b) Energy differences between the learned and CASSCF potential energy surface.

FIGURE 5.2.2: Potential energy profile for the Bergman cyclization of enediyne along the optimized NEB path.



(a) Reactant state.

(b) Transition state.

(c) Product state.

FIGURE 5.2.3: Accuracy of geometric structures from the semi-local fitted PES for the Bergman cyclization of enediyne: position differences (MAE) for sGDML- and CASSCF-optimized structures. Atom numbering is given in Fig. 5.2.1.

5.2.1.3 Geometry Optimization

In Fig. 5.2.3 we compare the geometric structures of the stationary structures as obtained by sGDML and the reference method, CASSCF. Directly comparing the optimized geometries from both methods, we find that the MAE for all inter-atomic distances never reaches the value of 0.01 Å for the reactant state. However, for the TS, the error is considerably larger reaching a deviation of 0.22 Å in the worst case. Although it is a significant value, the corresponding distance is between two non-bonded hydrogen atoms separated by more than 5 Å. For other distances, the error does not exceed 0.1 Å.

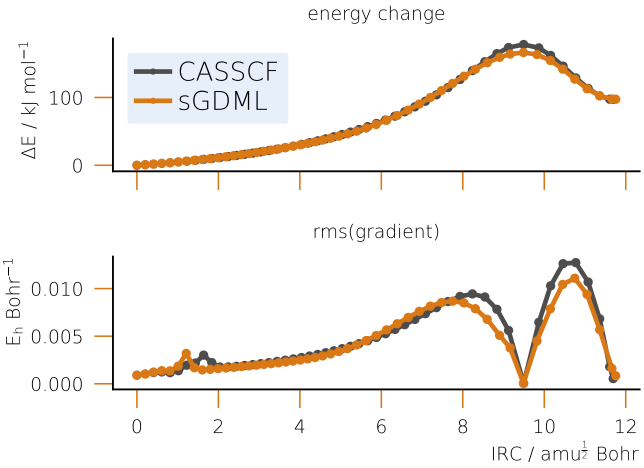


FIGURE 5.2.4: IRC for the Bergman cyclization of enediyne: relative energies and Root Mean Square (RMS) gradients. The TS corresponds to the maximum energy value at approximately 9.5 value of the IRC displacement.

5.2.1.4 Vibrations, Intrinsic Reaction Coordinates and Reaction Rate Coefficients

After obtaining the optimized TS, we analyze the connection between it and the basins of reactants and products by integrating the IRC as shown in Fig. 5.2.4. On the fitted semi-local PES, the TS does connect the reactants and products by a minimum energy path, which is in qualitative agreement with the reference CASSCF calculation along the entire curve as indicated by the maximum and RMS errors in the gradients.

The successful computation of the IRC gives the first positive accuracy assessment of vibrational modes on the semi-local fitted PES: the imaginary frequency at the TS is needed to define the direction of the path. A more detailed look at vibrational frequencies reveals only semi-qualitative agreement between the sGDML and the reference CASSCF PES (see Table 5.B.1 in Appendix 5.B.1): the differences between the frequencies reach up to approximately 200 cm^{-1} for high-frequency nodes, which is still less than 10 %.

Rate coefficients dependent on structures, vibrations and reaction barriers are good integrated indicators of the fitted PES quality. The barrier heights for the Bergman cyclization are $\Delta E^{\ddagger}_{\text{CASSCF}} = 194.59\,\text{kJ}$ for the reference method and $\Delta E^{\ddagger}_{\text{sGDML}} = 194.57\,\text{kJ}$ for the fitted PES, which are in excellent qualitative agreement. Assuming $\text{T} = 300\,\text{K}$, we obtain the following reaction rate coefficients from the conventional TST (as described in Appendix 5.A) for the two methods: $k_{\text{CASSCF}} = 3.4548 \times 10^{-22}\,\text{m}^3\text{s}^{-1}$ for the CAS(12, 12) and $k_{\text{sGDML}} = 4.3856 \times 10^{-22}\,\text{m}^3\text{s}^{-1}$ for sGDML. The agreement is semi-qualitative and stems from the pre-exponential factors defined by the partition functions, which in turn depend on vibrations and structures being less accurate than the barriers.

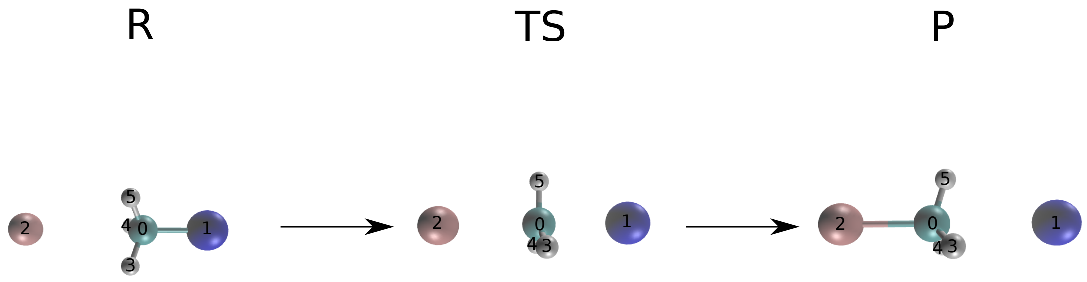


FIGURE 5.2.5: S_N2 reaction of chloromethane with bromide. The reactant, transition and product states are labeled with R, TS, and P, respectively. Elements are colored as follows: carbon (cyan), hydrogen (white), bromine (pink), chlorine (blue).

5.2.2 S_N2 Reaction of Chloromethane with Bromide

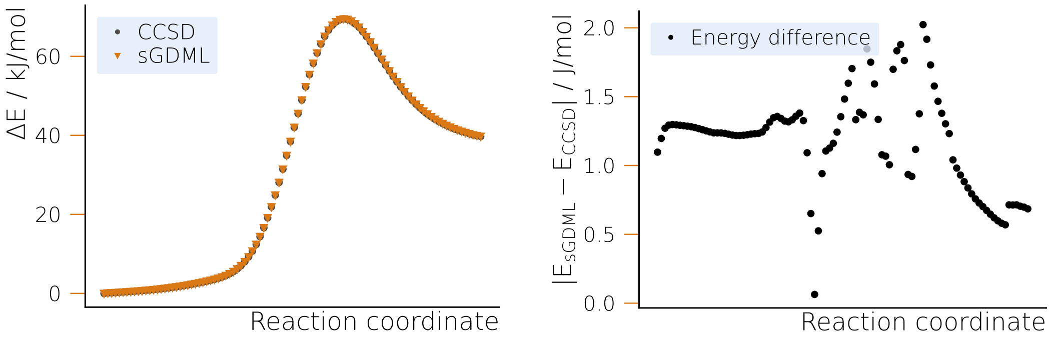
The reaction between CH_3Cl and Br^- is shown in Fig. 5.2.5, including the optimized reactant and product.

5.2.2.1 Model Training

To generate a machine learning model, we created a data set with a total of 100 geometries separated into 50 training, 30 validation and 20 test points. The regularization parameter was set to $\lambda = 10^{-15}$ and the best length scale was found to be $\sigma = 32$. With these parameters, we obtained a model with both the MAE and RMSE being 0.0013 kJ/mol, whereas the MAE and RMSE for the forces were $0.0105\,\mathrm{kJ/Å \cdot mol}$ and $0.0293\,\mathrm{kJ/Å \cdot mol}$, respectively, on the test data set.

5.2.2.2 Potential Energy Surface

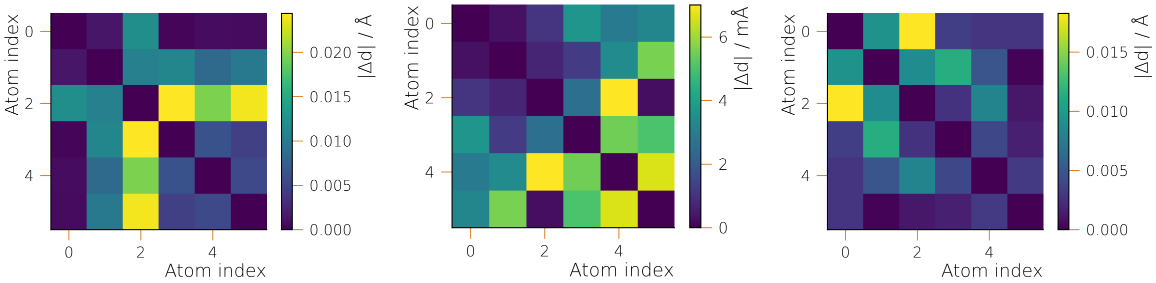
The potential energy profiles of the S_N2 reaction for the reference method, CCSD, and sGDML are shown in Fig. 5.2.6a. The good agreement between CCSD and the sGDML fit along the (DFT) NEB path can be immediately seen in Fig. 5.2.6b. The energy difference is of similar order as for the Bergman cyclization and does not exceed 0.02 kJ/mol. However, we observe the higher errors in the TS region rather than in the vicinity of the product state.



(a) Potential energy surface for CCSD and sGDML.

(b) Energy differences between the learned and CCSD potential energy surface.

FIGURE 5.2.6: Potential energy profile for the S_N2 reaction of chloromethane and bromide along the optimized NEB path.



(a) Reactant state.

(b) Transition state.

(c) Product state.

FIGURE 5.2.7: Accuracy of geometric structures from the semi-local fitted PES for the S_N2 reaction of chloromethane and bromide: position differences (MAE) for sGDML- and CCSD-optimized structures. Atom numbering is given in Fig. 5.2.5.

5.2.2.3 Geometry Optimization

In Fig. 5.2.7 we compare the geometric structures of the stationary structures obtained by sGDML and the reference method. The general agreement between the reference and fitted PES is better than for the Bergman cyclization, where the MAE never exceeds 0.025 Å for a particular interatomic distance. For the S_N2 reaction, the better agreement is reached for the TS with the MAE staying below a value of 0.01 Å. The largest deviation in case of the reactant state can be found for the distances between the non-bonded bromide ion (indicated by index 2) and the hydrogen atoms from the methylene group (indices 3, 4 and 5).

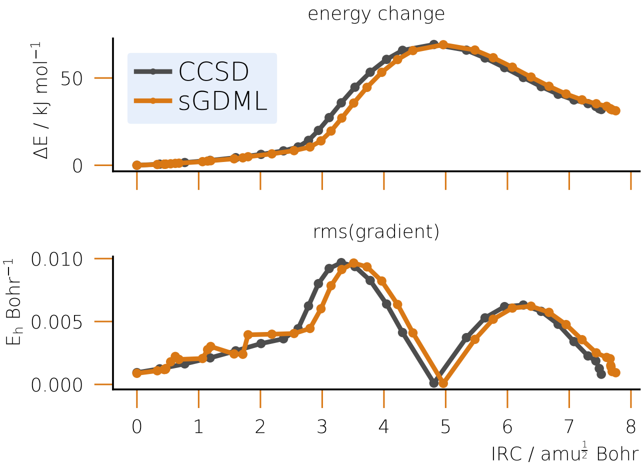


FIGURE 5.2.8: IRC for the S_N2 reaction of chloromethane with bromide: relative energies and Root Mean Square (RMS) gradients. The TS corresponds to the maximum energy value at approximately 4.8 value of the IRC displacement.

5.2.2.4 Vibrations, Intrinsic Reaction Coordinates and Reaction Rate Constants

After obtaining the optimized TS, we analyze the connection between it and the basins of the reactants and products by integrating the IRC as shown in Fig. 5.2.8. On the fitted semi-local PES, the TS does connect the reactants and products by a minimum energy path, which is in qualitative agreement with the reference CCSD calculation along the entire path as indicated by the maximum and RMSE in the gradients.

As in the previous example, successful computation of the IRC indicates the correct Hessian structure at the TS with the only imaginary eigenvalue corresponding to the same eigenvector as in the reference method. A more detailed look at vibrational frequencies reveals only qualitative agreement between the sGDML and the reference CCSD PES (see Table 5.B.2 in Appendix 5.B.2): the differences between the frequencies reach several hundreds of cm^{-1} for both high- and low-frequency modes.

The barrier heights for the S_N2 reaction were identical up to the second digit: $\Delta E^{\ddagger} = 69.47\,\mathrm{kJ}$ for CCSD and the fitted PES. At $T = 300\,\mathrm{K}$, the CCSD reaction rate coefficient is $1.8187 \times 10^{-31}\,\mathrm{m}^3\mathrm{s}^{-1}$, whereas this value is $2.2287 \times 10^{-32}\,\mathrm{m}^3\mathrm{s}^{-1}$ for sGDML, which is an order of magnitude smaller than the reference method. This difference occurs due to the larger deviations in the harmonic vibrations, defining the pre-exponential factors in the TST equations.

5.3 Discussion

The most important property of the fitted semi-local PES surfaces obtained by the proposed protocol is the general stability with respect to PES exploration techniques. Indeed, for both unimolecular and bimolecular reactions geometry optimizations have converged to the stationary points, connected by physically meaningful minimum energy paths. This is achieved by the multi-level nature of our approach.

On the one hand, the NEB driven by the cheaper method (here DFT with a PBE exchange-correlation functional) generates structures relatively close to the reactive region of the system as defined by more accurate reference electronic structure methods (here, CASSCF or CCSD). Indeed, DFT methods are known to predict sensitive geometric structures [142] (even if the energy barriers are not precise). This is illustrated by the S_N2 reaction studied in this work: in Fig. S1 of the SI we see that the highest energy point on the NEB path for both PBE and CCSD is located at the similar values of the reaction coordinate, although the energy barriers differ dramatically.

On the other hand, structures sampled by NEB using the cheaper method should be far enough from the minimum energy path of the reference method. The points on the true minimum energy path (approximated by NEB [126]) have only one non-zero gradient component – the one along the path tangent direction [95]. Keeping in mind that sGDML uses gradients as inputs for fitting, providing only structures on the minimum energy path would provide no information about the nature of the PES along orthogonal directions and make model very sensitive to numerical noise. Consequently, using a cheaper electronic structure method for the NEB simulation is not only computational effort saving, but also essential for the quality of the data set used to train a gradient-based ML model.

Despite qualitative agreement achieved by the semi-local fitted PES in structures, energy barriers and IRC, a more subtle property, vibrational spectrum, is computed less accurately. Although the structure of Hessian is qualitatively correct, some frequencies can differ by several hundreds of cm^{-1}. This is particularly noticeable for the S_N2 example and as a consequence leads to a significant error (order of magnitude) in the reaction rate coefficient as compared to the reference value. This effect must have to do with a smaller number of training points (only 50) used for the model as compared to the Bergman cyclization (150 points). Another reason is the fact that the relevant PES region for the S_N2 is flatter as this for the Bergman cyclization (compare Fig. 5.2.4 and Fig. 5.2.8), one indication of which is the lower energy barrier for the former reaction. This makes finite-difference evaluation of the second derivatives of energy less reliable and requires a better PES sampling for obtaining qualitative results.

5.4 Conclusions and Outlook

In this work, we have proposed a multi-level protocol for reaction mechanism studies aiming to match the accurate electronic structure theory description at the reduced computational cost. The approach involves cheaper electronic structure method to generate structures along the reaction path via the NEB method, evaluating energies and gradients for a restricted set of them with an accurate (reference) theory and fitting a semi-local reactive PES using machine learning method, sGDML. The fitted PES is then used for computing properties.

The protocol has been applied to a unimolecular (Bergman cyclization) and a bimolecular (S_N2) reactions using PBE/CASSCF and PBE/CCSD, respectively, as cheap-/reference method pairs, the results being compared with those obtained on-the-fly by the reference method. Our approach achieves quantitative agreement in structure optimization for the PES stationary points, reaction energy barriers and IRC following using as little as 50-150 training points (corresponding to energy and gradient calculations with the reference method), whereas the agreement in vibrational frequencies and reaction coefficients is only (semi-)qualitative. The key to the performance of the protocol is its multi-level character as the differences between the reference and cheap methods are essential to provide meaningful information about the reactive PES region.

The results are encouraging as important objective of mechanistic reaction study can be performed with high-level electronic structure method without prior knowledge of the TS structure (due to the application of the NEB) requiring only scarce amount of data. The protocol is simple and can be automatized for routine calculations.

At the same time, we have not considered more complex reactions with multiple steps, product branching and shallow minima. These cases would require more subtle approaches to sampling and larger amounts of training data.

Our simple protocol can be further improved by applying smart selection of training points using molecular fingerprints [143] and using methods even computationally cheaper than DFT (such as modern tight-binding, GFN2-XTB [144]) for sampling structures. Moderately accurate reference methods applied here should be substituted by those providing qualitatively correct PES, such as multi-reference approaches, to match the experimental accuracy. Furthermore, the potential of different ML fitting techniques should be explored.

APPENDICES

5.A Reaction Rate Coefficients

We use conventional transition state theory to compute rate coefficients for the two reactions as follows [145]:

$$k = \left(\frac{k_B T}{h}\right)\left(\frac{q_{trans,TS}/V}{q_{trans,R}/V}\right)\left(\frac{q_{rot,TS}/V}{q_{rot,R}/V}\right)\left(\frac{q_{vib,TS}/V}{q_{vib,R}/V}\right)e^{-\Delta E^{\ddagger}/k_B T}, \tag{5.1}$$

where k_B being the Boltzman constant, T - the temperature and h - the Planck constant. $q_{trans,TS}$, $q_{rot,TS}$ and $q_{vib,TS}$ denote the translational, rotational and vibrational partition functions for the transition state, whereas $q_{trans,R}$, $q_{rot,R}$ and $q_{vib,R}$ are their counterparts for the reactants.

The calculation of the rotational partition functions requires the eigenvalues of the inertia tensor:

$$q_{rot}(T) = \frac{\pi^2}{\sigma}\sqrt{(8\pi I_A k_B T)/h^2}\sqrt{(8\pi I_B k_B T)/h^2}\sqrt{(8\pi I_C k_B T)/h^2}, \tag{5.2}$$

where I_A, I_B and I_C are the eigenvalues of the inertia tensor.

The harmonic vibrational partition function is defined as:

$$q_{vib}(T) = \prod_i^N \left(\frac{1}{1 - e^{\frac{-hc\nu_i}{k_B T}}}\right), \tag{5.3}$$

where ν_i are vibrational frequencies and N is the number of normal modes.

This translational partition function is defined as:

$$q_{trans}(V,T) = V\left(\frac{2\pi k_B T m}{h^2}\right)^{3/2}, \tag{5.4}$$

where m is the mass of the molecule.

For the unimolecular Bergman cyclization, the translational partition functions in Eq. 5.1 cancel out, whereas for the bimolecular $\mathrm{S_N2}$ reaction they have to be included.

5.B Harmonic Vibrations

5.B.1 Vibrational Modes for the Bergman Cyclization of Enediyne

Mode number	ν^{R}_{CASSCF}/m^{-1}	ν^{R}_{sGDML}/m^{-1}	ν^{TS}_{CASSCF}/m^{-1}	ν^{TS}_{sGDML}/m^{-1}
0	10.3i	7.9i	799.5i	646.8i
1	7.2i	2.4i	0.9i	0.1i
2	0.7i	0.1i	0.0	0.0
3	0.1i	0.1i	0.4	0.0
4	2.9	0.3	3.5	7.2
5	6.3	3.0	7.7	9.0
6	100.5	102.7	8.1	11.2
7	211.7	206.7	348.1	353.1
8	220.4	225.0	422.5	403.4
9	323.1	324.8	485.2	491.0
10	402.5	422.9	543.8	502.1
11	564.7	548.1	598.4	528.5
12	577.9	567.8	639.9	617.9
13	578.1	574.4	679.2	665.1
14	591.1	581.0	736.2	677.5
15	594.1	589.8	818.2	835.3
16	728.9	725.9	856.7	949.5
17	778.1	742.7	975.7	956.7
18	900.5	883.0	1024.2	1015.0
19	1001.9	966.9	1033.1	1032.2
20	1063.1	1073.7	1148.2	1043.7
21	1291.9	1348.8	1210.1	1120.0
22	1497.3	1500.6	1453.9	1333.9
23	1658.1	1613.1	1479.7	1427.4
24	2238.8	2123.6	1713.0	1685.2
25	2246.3	2230.1	1885.0	1832.8
26	3338.3	3326.4	3365.9	3144.6
27	3354.6	3327.8	3381.2	3275.8
28	3611.0	3572.6	3482.0	3367.0
29	3611.2	3577.6	3487.3	3441.6

Table 5.B.1: Comparison between CASSCF and sGDML frequencies for the reactant and transition state.

5.B.2 Vibrational Modes for the $\mathrm{S_N}2$ Reaction

Mode number	ν^{R}_{CCSD}/m^{-1}	ν^{R}_{sGDML}/m^{-1}	ν^{TS}_{CCSD}/m^{-1}	ν^{TS}_{sGDML}/m^{-1}
0	36.0i	3.2i	447.7i	448.1i
1	0.7i	1.4i	0.1i	0.3i
2	0.4i	0.7i	0.1i	0.2i
3	0.1	0.0	0.1	0.0
4	0.4	0.2	3.0	3.0
5	0.5	19.2	3.7	3.7
6	73.4	100.5	40.9	19.7
7	73.4	405.2	184.1	176.7
8	92.4	411.1	188.9	440.2
9	666.6	652.9	188.9	443.5
10	988.8	1022.0	921.5	1093.0
11	989.2	1030.7	922.1	1102.4
12	1328.1	1639.9	1003.4	1117.8
13	1468.1	1677.2	1409.1	1171.4
14	1468.4	1764.9	1409.2	1220.6
15	3152.2	2911.4	3228.6	2690.0
16	3275.1	3376.1	3440.0	3422.7
17	3275.6	3550.1	3441.8	3584.3

TABLE 5.B.2: Comparison between CCSD and sGDML frequencies for the reactant and transition state.

5.C Reaction Coordinate (NEB) for the S_N2 reaction

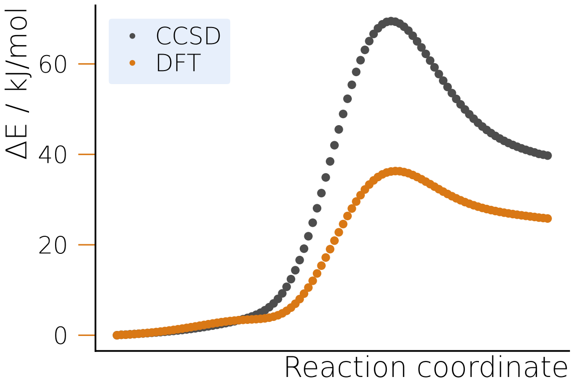


Figure 5.C.1: PES comparison between CCSD and DFT (PBE). Energies have been calculated for the same structures corresponding to the final state of the NEB calculation with DFT(PBE).

CONCLUSION

Quantum computing is a rapidly growing field of research, not only in academia but also for industrial use cases. As hardware performance tends to increase every year, more and more applications can be mapped and simulated on quantum devices. However, as we are still in the NISQ era, noise effects lead to results that may not be reliable enough. Therefore, the right choice of algorithms and post-processing of quantum measurements is key in the current quantum age to obtain representative results. In addition, obtaining a reliable expectation value on a quantum device is relatively expensive and would require at least a million measurement shots, so we would like to make due with determining as little expectation values as possible. In this case, sophisticated classical methods such as machine learning techniques can be used to gain more information about the system under investigation and to build representative models.

Chapters 1 and 2 gave brief introductions to the fields of quantum computing and computational quantum chemistry, respectively.

In Chapter 3 of this thesis we compared two methods to determine gradients – namely the finite-difference or numerical method, and the parameter-shift rule or analytical method – on a quantum computer taking depolarization noise into account. Due to its simplicity, we have used a gradient descent optimizer which just contains one parameter, the so-called learning rate. For a simple toy system we could observe that the parameter-shift rule reaches a smaller error after a certain amount of iteration steps, compared to the finite-difference method with three different choices for the learning rate. In case of more complicated systems, specifically small one-dimensional Hubbard models, an advantage of the analytical method compared to its classical counterpart could not be observed. This might be due to the fact that the circuit has a larger circuit depth and width but more importantly that the same parameters appear in multiple parametrized gates. The question if another gradient method, such as BFGS would provide better results with the parameter-shift rule is left for a future work.

Chapter 4 provided insights into the post-processing of measurements on a noisy quantum device using N-representability conditions. We studied not only the case of decoherence noise, but also the combination of decoherence and shot noise in a regime that is realistic for today's quantum computers. Considering only decoherence noise, we saw that the application of two N-representability conditions – namely fixed trace and positive semidefiniteness – could reduce the energy error compared to runs without post-processing. In particular, the case of H_2 has shown that we could reduce the energy error by almost an order of magnitude. In addition, we considered the 2-RDM fidelity for all cases. Here we partially observed an improvement compared to the raw measurement when applying the above constraints to the reduced density matrices. Unfortunately, with this projection method there is no strong correlation between the energy error and the 2-RDM fidelity. With the inclusion of shot noise in our simulations, the method led to improved results. A further analysis of the measurement variance confirmed the efficiency of this method. While for LiH calculations we observed a reduction of the measurement variance in the order of half to about one order of magnitude, for H_2 we observed an improvement of up to two orders of magnitude. This deviation could be due to the low number of measurements and repetitions in the case of larger molecules, which typically require a higher number of data for representative results. Another reason might be the possibly bad scalability of this method for larger systems, where the investigation of said scalability would be a task for future studies. This post-processing method could be particularly interesting in the NISQ era, where we still have to deal with decoherence noise. When we reach the era where quantum error correction will be possible and we would only have to deal with shot noise, this method might be less advantageous. We have also seen that this method is not helpful to mitigate pure dephasing noise in combination with the Jordan-Wigner encoding. However, this approach shows that we can improve the energy errors by one to two orders of magnitude when using NISQ devices.

Finally, in Chapter 5 we have used a machine learning approach that requires a small number of data points to produce a representative model. The machine learning model of choice was sGDML, which uses the inverse interatomic distances of the molecule as input and the corresponding forces, or energy gradients with respect to positions, as output. By training only 150 (Bergman cyclization) and 50 (S_N2 reaction) points, we were able to obtain good models for further processing. This further processing included, for example, interpolation of the entire potential energy surfaces and geometry optimizations, where all observables were compared with the corresponding reference methods, namely CASSCF for the Bergman cyclization and CCSD for the S_N2 reaction. In all cases good agreement with the reference methods was observed. The idea behind using as few data points as possible is to replace computations on classical devices with computations on real quantum hardware. In the case of CASSCF, the active space would be represented by system (or “good“) qubits, while all other unoccupied states would be represented by bath (or “bad“) qubits. As quantum hardware is expensive today, the aim is to perform as few computations as possible to achieve a relatively low throughput. These points can then be used and fed into an approach as described in this chapter.

The aim of this thesis was to find methods for reducing the amount of resources that are required for computations on today’s quantum computers. Our analysis of calculating gradients in an optimization procedure using the parameter-shift rule to calculate compared to a simple finite-element method for algorithms of different complexity provided insights about the strengths, but also the limitations, of the parameter-shift rule, and sparks ideas for further research in this realm. While our post-processing approach offers positive and promising results for the near-future application on quantum devices, the presented machine learning Ansatz could already be beneficial using today's purely classical computers. It would be interesting to see how these methods perform on real quantum hardware and on larger systems.

BIBLIOGRAPHY

[1] R. P. Feynman, International Journal of Theoretical Physics **21**, 467 (1982).

[2] P. W. Shor, SIAM Journal on Computing **26**, 1484 (1997).

[3] D. Deutsch and R. Jozsa, Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences **439**, 553 (1992).

[4] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, K. Guerin, S. Habegger, M. P. Harrigan, M. J. Hartmann, A. Ho, M. Hoffmann, T. Huang, T. S. Humble, S. V. Isakov, E. Jeffrey, Z. Jiang, D. Kafri, K. Kechedzhi, J. Kelly, P. V. Klimov, S. Knysh, A. Korotkov, F. Kostritsa, D. Landhuis, M. Lindmark, E. Lucero, D. Lyakh, S. Mandrà, J. R. McClean, M. McEwen, A. Megrant, X. Mi, K. Michielsen, M. Mohseni, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Y. Niu, E. Ostby, A. Petukhov, J. C. Platt, C. Quintana, E. G. Rieffel, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, K. J. Sung, M. D. Trevithick, A. Vainsencher, B. Villalonga, T. White, Z. J. Yao, P. Yeh, A. Zalcman, H. Neven, and J. M. Martinis, Nature **574**, 505 (2019).

[5] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu, P. Hu, X.-Y. Yang, W.-J. Zhang, H. Li, Y. Li, X. Jiang, L. Gan, G. Yang, L. You, Z. Wang, L. Li, N.-L. Liu, C.-Y. Lu, and J.-W. Pan, Science **370**, 1460 (2020).

[6] J. M. Pino, J. M. Dreiling, C. Figgatt, J. P. Gaebler, S. A. Moses, M. S. Allman, C. H. Baldwin, M. Foss-Feig, D. Hayes, K. Mayer, C. Ryan-Anderson, and B. Neyenhuis, Nature **592**, 209 (2021).

[7] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, Nature Communications **5**, 4213 (2014).

[8] J. Preskill, Quantum **2**, 79 (2018).

[9] Y. Wu, W.-S. Bao, S. Cao, F. Chen, M.-C. Chen, X. Chen, T.-H. Chung, H. Deng, Y. Du, D. Fan, M. Gong, C. Guo, C. Guo, S. Guo, L. Han, L. Hong, H.-L. Huang, Y.-H. Huo, L. Li, N. Li, S. Li, Y. Li, F. Liang, C. Lin, J. Lin, H. Qian, D. Qiao,

H. Rong, H. Su, L. Sun, L. Wang, S. Wang, D. Wu, Y. Xu, K. Yan, W. Yang, Y. Yang, Y. Ye, J. Yin, C. Ying, J. Yu, C. Zha, C. Zhang, H. Zhang, K. Zhang, Y. Zhang, H. Zhao, Y. Zhao, L. Zhou, Q. Zhu, C.-Y. Lu, C.-Z. Peng, X. Zhu, and J.-W. Pan, Phys. Rev. Lett. **127**, 180501 (2021).

[10] Q. Zhu, S. Cao, F. Chen, M.-C. Chen, X. Chen, T.-H. Chung, H. Deng, Y. Du, D. Fan, M. Gong, C. Guo, C. Guo, S. Guo, L. Han, L. Hong, H.-L. Huang, Y.-H. Huo, L. Li, N. Li, S. Li, Y. Li, F. Liang, C. Lin, J. Lin, H. Qian, D. Qiao, H. Rong, H. Su, L. Sun, L. Wang, S. Wang, D. Wu, Y. Wu, Y. Xu, K. Yan, W. Yang, Y. Yang, Y. Ye, J. Yin, C. Ying, J. Yu, C. Zha, C. Zhang, H. Zhang, K. Zhang, Y. Zhang, H. Zhao, Y. Zhao, L. Zhou, C.-Y. Lu, C.-Z. Peng, X. Zhu, and J.-W. Pan, Quantum computational advantage via 60-qubit 24-cycle random circuit sampling (2021), arXiv:2109.03494 [quant-ph] .

[11] L. S. Madsen, F. Laudenbach, M. F. Askarani, F. Rortais, T. Vincent, J. F. F. Bulmer, F. M. Miatto, L. Neuhaus, L. G. Helt, M. J. Collins, A. E. Lita, T. Gerrits, S. W. Nam, V. D. Vaidya, M. Menotti, I. Dhand, Z. Vernon, N. Quesada, and J. Lavoie, Nature **606**, 75 (2022).

[12] D. S. Abrams and S. Lloyd, Phys. Rev. Lett. **79**, 2586 (1997).

[13] D. S. Abrams and S. Lloyd, Phys. Rev. Lett. **83**, 5162 (1999).

[14] A. Mazurenko, C. S. Chiu, G. Ji, M. F. Parsons, M. Kanász-Nagy, R. Schmidt, F. Grusdt, E. Demler, D. Greif, and M. Greiner, Nature **545**, 462 (2017).

[15] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Nature **549**, 242 (2017).

[16] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, Nature **567**, 491 (2019).

[17] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, A. Bengtsson, S. Boixo, M. Broughton, B. B. Buckley, D. A. Buell, B. Burkett, N. Bushnell, Y. Chen, Z. Chen, Y.-A. Chen, B. Chiaro, R. Collins, S. J. Cotton, W. Courtney, S. Demura, A. Derk, A. Dunsworth, D. Eppens, T. Eckl, C. Erickson, E. Farhi, A. Fowler, B. Foxen, C. Gidney, M. Giustina, R. Graff, J. A. Gross, S. Habegger, M. P. Harrigan, A. Ho, S. Hong, T. Huang, W. Huggins, L. B. Ioffe, S. V. Isakov, E. Jeffrey, Z. Jiang, C. Jones, D. Kafri, K. Kechedzhi, J. Kelly, S. Kim, P. V. Klimov, A. N. Korotkov, F. Kostritsa, D. Landhuis, P. Laptev, M. Lindmark, E. Lucero, M. Marthaler, O. Martin, J. M. Martinis, A. Marusczyk, S. McArdle, J. R. McClean, T. McCourt, M. McEwen, A. Megrant, C. Mejuto-Zaera, X. Mi, M. Mohseni, W. Mruczkiewicz, J. Mutus, O. Naaman, M. Neeley, C. Neill,

H. Neven, M. Newman, M. Y. Niu, T. E. O'Brien, E. Ostby, B. Pató, A. Petukhov, H. Putterman, C. Quintana, J.-M. Reiner, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, D. Strain, K. J. Sung, P. Schmitteckert, M. Szalay, N. M. Tubman, A. Vainsencher, T. White, N. Vogt, Z. J. Yao, P. Yeh, A. Zalcman, and S. Zanker, Observation of separated dynamics of charge and spin in the fermi-hubbard model (2020), arXiv:2010.07965 [quant-ph] .

[18] B. Rost, L. D. Re, N. Earnest, A. F. Kemper, B. Jones, and J. K. Freericks, Demonstrating robust simulation of driven-dissipative problems on near-term quantum computers (2021), arXiv:2108.01183 [quant-ph] .

[19] R. N. Tazhigulov, S.-N. Sun, R. Haghshenas, H. Zhai, A. T. K. Tan, N. C. Rubin, R. Babbush, A. J. Minnich, and G. K.-L. Chan, Simulating challenging correlated molecules and materials on the sycamore quantum processor (2022), arXiv:2203.15291 [quant-ph] .

[20] S. Xu, Z.-Z. Sun, K. Wang, L. Xiang, Z. Bao, Z. Zhu, F. Shen, Z. Song, P. Zhang, W. Ren, X. Zhang, H. Dong, J. Deng, J. Chen, Y. Wu, Z. Tan, Y. Gao, F. Jin, X. Zhu, C. Zhang, N. Wang, Y. Zou, J. Zhong, A. Zhang, W. Li, W. Jiang, L.-W. Yu, Y. Yao, Z. Wang, H. Li, Q. Guo, C. Song, H. Wang, and D.-L. Deng, Digital simulation of non-abelian anyons with 68 programmable superconducting qubits (2022), arXiv:2211.09802 [quant-ph] .

[21] S. Guo, J. Sun, H. Qian, M. Gong, Y. Zhang, F. Chen, Y. Ye, Y. Wu, S. Cao, K. Liu, C. Zha, C. Ying, Q. Zhu, H.-L. Huang, Y. Zhao, S. Li, S. Wang, J. Yu, D. Fan, D. Wu, H. Su, H. Deng, H. Rong, Y. Li, K. Zhang, T.-H. Chung, F. Liang, J. Lin, Y. Xu, L. Sun, C. Guo, N. Li, Y.-H. Huo, C.-Z. Peng, C.-Y. Lu, X. Yuan, X. Zhu, and J.-W. Pan, Experimental quantum computational chemistry with optimised unitary coupled cluster ansatz (2023), arXiv:2212.08006 [quant-ph] .

[22] A. Erhard, H. Poulsen Nautrup, M. Meth, L. Postler, R. Stricker, M. Stadler, V. Negnevitsky, M. Ringbauer, P. Schindler, H. J. Briegel, R. Blatt, N. Friis, and T. Monz, Nature **589**, 220 (2021).

[23] R. Acharya, I. Aleiner, R. Allen, T. I. Andersen, M. Ansmann, F. Arute, K. Arya, A. Asfaw, J. Atalaya, R. Babbush, D. Bacon, J. C. Bardin, J. Basso, A. Bengtsson, S. Boixo, G. Bortoli, A. Bourassa, J. Bovaird, L. Brill, M. Broughton, B. B. Buckley, D. A. Buell, T. Burger, B. Burkett, N. Bushnell, Y. Chen, Z. Chen, B. Chiaro, J. Cogan, R. Collins, P. Conner, W. Courtney, A. L. Crook, B. Curtin, D. M. Debroy, A. Del Toro Barba, S. Demura, A. Dunsworth, D. Eppens, C. Erickson, L. Faoro, E. Farhi, R. Fatemi, L. Flores Burgos, E. Forati, A. G. Fowler, B. Foxen, W. Giang, C. Gidney, D. Gilboa, M. Giustina, A. Grajales Dau, J. A.

Gross, S. Habegger, M. C. Hamilton, M. P. Harrigan, S. D. Harrington, O. Higgott, J. Hilton, M. Hoffmann, S. Hong, T. Huang, A. Huff, W. J. Huggins, L. B. Ioffe, S. V. Isakov, J. Iveland, E. Jeffrey, Z. Jiang, C. Jones, P. Juhas, D. Kafri, K. Kechedzhi, J. Kelly, T. Khattar, M. Khezri, M. Kieferová, S. Kim, A. Kitaev, P. V. Klimov, A. R. Klots, A. N. Korotkov, F. Kostritsa, J. M. Kreikebaum, D. Landhuis, P. Laptev, K.-M. Lau, L. Laws, J. Lee, K. Lee, B. J. Lester, A. Lill, W. Liu, A. Locharla, E. Lucero, F. D. Malone, J. Marshall, O. Martin, J. R. McClean, T. McCourt, M. McEwen, A. Megrant, B. Meurer Costa, X. Mi, K. C. Miao, M. Mohseni, S. Montazeri, A. Morvan, E. Mount, W. Mruczkiewicz, O. Naaman, M. Neeley, C. Neill, A. Nersisyan, H. Neven, M. Newman, J. H. Ng, A. Nguyen, M. Nguyen, M. Y. Niu, T. E. O'Brien, A. Opremcak, J. Platt, A. Petukhov, R. Potter, L. P. Pryadko, C. Quintana, P. Roushan, N. C. Rubin, N. Saei, D. Sank, K. Sankaragomathi, K. J. Satzinger, H. F. Schurkus, C. Schuster, M. J. Shearn, A. Shorter, V. Shvarts, J. Skruzny, V. Smelyanskiy, W. C. Smith, G. Sterling, D. Strain, M. Szalay, A. Torres, G. Vidal, B. Villalonga, C. Vollgraff Heidweiller, T. White, C. Xing, Z. J. Yao, P. Yeh, J. Yoo, G. Young, A. Zalcman, Y. Zhang, N. Zhu, and Google Quantum AI, Nature **614**, 676 (2023).

[24] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, Phys. Rev. A **86**, 032324 (2012).

[25] B. Lekitsch, S. Weidt, A. G. Fowler, K. Mølmer, S. J. Devitt, C. Wunderlich, and W. K. Hensinger, Science Advances **3**, e1601540 (2017).

[26] A. Bermudez, X. Xu, R. Nigmatullin, J. O'Gorman, V. Negnevitsky, P. Schindler, T. Monz, U. G. Poschinger, C. Hempel, J. Home, F. Schmidt-Kaler, M. Biercuk, R. Blatt, S. Benjamin, and M. Müller, Phys. Rev. X **7**, 041061 (2017).

[27] E. Farhi, J. Goldstone, and S. Gutmann, A Quantum Approximate Optimization Algorithm (2014), arXiv:1411.4028 [quant-ph] .

[28] D. Wecker, M. B. Hastings, and M. Troyer, Phys. Rev. A **92**, 042303 (2015).

[29] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, New Journal of Physics **18**, 023023 (2016).

[30] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Nature Communications **9**, 4812 (2018).

[31] J. M. Kübler, A. Arrasmith, L. Cincio, and P. J. Coles, Quantum **4**, 263 (2020).

[32] R. Sweke, F. Wilde, J. Meyer, M. Schuld, P. K. Faehrmann, B. Meynard-Piganeau, and J. Eisert, Quantum **4**, 314 (2020).

[33] J. J. Meyer, J. Borregaard, and J. Eisert, npj Quantum Information **7**, 89 (2021).

[34] T. Piskor, J.-M. Reiner, S. Zanker, N. Vogt, M. Marthaler, F. K. Wilhelm, and F. G. Eich, Phys. Rev. A **105**, 062415 (2022).

[35] M. H. Devoret and R. J. Schoelkopf, Science **339**, 1169 (2013).

[36] C. Rigetti, J. M. Gambetta, S. Poletto, B. L. T. Plourde, J. M. Chow, A. D. Córcoles, J. A. Smolin, S. T. Merkel, J. R. Rozen, G. A. Keefe, M. B. Rothwell, M. B. Ketchen, and M. Steffen, Phys. Rev. B **86**, 100506 (2012).

[37] C. D. Bruzewicz, J. Chiaverini, R. McConnell, and J. M. Sage, Applied Physics Reviews **6**, 021314 (2019).

[38] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge, 2010).

[39] D. M. Greenberger, M. A. Horne, and A. Zeilinger, Going beyond bell's theorem (2007), arXiv:0712.0921 [quant-ph] .

[40] D. Gottesman, The heisenberg representation of quantum computers (1998), arXiv:quant-ph/9807006 [quant-ph] .

[41] J. J. S. Sakurai and J. J. Napolitano, *Modern Quantum Mechanics* (2014).

[42] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, Reviews of Modern Physics **92** (2020).

[43] R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, 1985).

[44] S. Machnes and M. B. Plenio, Surprising interactions of markovian noise and coherent driving (2014), arXiv:1408.3056 [quant-ph] .

[45] C. King, IEEE Transactions on Information Theory **49**, 221 (2003).

[46] R. Schutjens, F. A. Dagga, D. J. Egger, and F. K. Wilhelm, Phys. Rev. A **88**, 052330 (2013).

[47] N. M. Linke, D. Maslov, M. Roetteler, S. Debnath, C. Figgatt, K. A. Landsman, K. Wright, and C. Monroe, Proceedings of the National Academy of Sciences **114**, 3305 (2017).

[48] S. Novikov, J. E. Robinson, Z. K. Keane, B. Suri, F. C. Wellstood, and B. S. Palmer, Phys. Rev. B **88**, 060503 (2013).

[49] L. B. Nguyen, Y.-H. Lin, A. Somoroff, R. Mencia, N. Grabon, and V. E. Manucharyan, Phys. Rev. X **9**, 041041 (2019).

[50] R. Ditchfield, W. J. Hehre, and J. A. Pople, The Journal of Chemical Physics **54**, 724 (2003).

[51] J. Dunning, Thom H., The Journal of Chemical Physics **90**, 1007 (1989).

[52] D. A. Mazziotti, Phys. Rev. A **85**, 062507 (2012).

[53] D. A. Mazziotti, Phys. Rev. Lett. **108**, 263002 (2012).

[54] K. Burke and L. O. Wagner, International Journal of Quantum Chemistry **113**, 96 (2013).

[55] A. D. Becke, The Journal of Chemical Physics **98**, 1372 (1993).

[56] B. O. Roos, P. R. Taylor, and P. E. Sigbahn, Chemical Physics **48**, 157 (1980).

[57] S. Chmiela, H. E. Sauceda, I. Poltavsky, K.-R. Müller, and A. Tkatchenko, Computer Physics Communications **240**, 38 (2019).

[58] K. T. Schütt, P.-J. Kindermans, H. E. Sauceda, S. Chmiela, A. Tkatchenko, and K.-R. Müller, Schnet: A continuous-filter convolutional neural network for modeling quantum interactions (2017), arXiv:1706.08566 [stat.ML] .

[59] Training, validation, and test data sets, `https://en.wikipedia.org/wiki/Training,_validation,_and_test_data_sets`, accessed: 2023-09-16.

[60] J.-M. Reiner, F. Wilhelm-Mauch, G. Schön, and M. Marthaler, Quantum Science and Technology **4**, 035005 (2019).

[61] N. Vogt, S. Zanker, J.-M. Reiner, M. Marthaler, T. Eckl, and A. Marusczyk, Quantum Science and Technology **6**, 035003 (2021).

[62] J. Romero, R. Babbush, J. R. McClean, C. Hempel, P. J. Love, and A. Aspuru-Guzik, Quantum Science and Technology **4**, 014008 (2018).

[63] A. L. Fetter and J. D. Valecka, *Quantum Theory of Many-Particle Systems* (New York, 1971).

[64] G. F. Giuliani and G. Vignale, The perturbative calculation of linear response functions, in *Quantum Theory of the Electron Liquid* (Cambridge, 2005) Chap. 6, pp. 275–326.

[65] M. J. Powell, Mathematics Today-Bulletin of the Institute of Mathematics and its Applications **43**, 170 (2007).

[66] M. J. D. Powell, A direct search optimization method that models the objective and constraint functions by linear interpolation, in *Advances in Optimization and Numerical Analysis*, edited by S. Gomez and J.-P. Hennart (Dordrecht, 1994) pp. 51–67.

[67] J. Nocedal and S. Wright, *Numerical optimization* (2006).

[68] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, Phys. Rev. A **99**, 032331 (2019).

[69] A. Mari, T. R. Bromley, and N. Killoran, Phys. Rev. A **103**, 012405 (2021).

[70] G. E. Crooks, Gradients of parameterized quantum gates using the parameter-shift rule and gate decomposition (2019), arXiv:1905.13311 [quant-ph] .

[71] L. Banchi and G. E. Crooks, Quantum **5**, 386 (2021).

[72] C. G. Broyden, IMA Journal of Applied Mathematics **6**, 76 (1970).

[73] R. Fletcher, The Computer Journal **13**, 317 (1970).

[74] D. Goldfarb, Mathematics of Computation **24**, 23 (1970).

[75] D. F. Shanno, Mathematics of Computation **24**, 647 (1970).

[76] B. Bauer, D. Wecker, A. J. Millis, M. B. Hastings, and M. Troyer, Phys. Rev. X **6**, 031045 (2016).

[77] F. Vatan and C. Williams, Phys. Rev. A **69**, 032315 (2004).

[78] P. Jordan and E. Wigner, Zeitschrift für Physik **47**, 631 (1928).

[79] M. Matlak, J. Aksamit, B. Grabiec, and W. Nolting, Annalen der Physik **12**, 304–319 (2003).

[80] E. Grant, L. Wossnig, M. Ostaszewski, and M. Benedetti, Quantum **3**, 214 (2019).

[81] M. Ostaszewski, E. Grant, and M. Benedetti, Quantum **5**, 391 (2021).

[82] J. Romero, J. P. Olson, and A. Aspuru-Guzik, Quantum Science and Technology **2**, 045001 (2017).

[83] E. Farhi and H. Neven, arXiv:1802.06002 [quant-ph] (2018).

[84] V. Havlicek, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, Nature **567**, 209 (2019).

[85] A. Pesah, M. Cerezo, S. Wang, T. Volkoff, A. T. Sornborger, and P. J. Coles, Physical Review X **11**, 041011 (2021).

[86] J. E. Mayer, Phys. Rev. **100**, 1579 (1955).

[87] A. J. Coleman, Rev. Mod. Phys. **35**, 668 (1963).

[88] C. Lanssens, P. W. Ayers, D. Van Neck, S. De Baerdemacker, K. Gunst, and P. Bultinck, J. Chem. Phys. **148**, 084104 (2018).

[89] N. C. Rubin, R. Babbush, and J. McClean, New J. Phys. **20**, 053020 (2018).

[90] GitHub - HQSquantumsimulations/qoqo.

[91] S. B. Bravyi and A. Y. Kitaev, Annals of Physics **298**, 210 (2002).

[92] J. R. McClean, N. C. Rubin, K. J. Sung, I. D. Kivlichan, X. Bonet-Monroig, Y. Cao, C. Dai, E. S. Fried, C. Gidney, B. Gimby, P. Gokhale, T. Häner, T. Hardikar, V. Havlíček, O. Higgott, C. Huang, J. Izaac, Z. Jiang, X. Liu, S. McArdle, M. Neeley, T. O'Brien, B. O'Gorman, I. Ozfidan, M. D. Radin, J. Romero, N. P. D. Sawaya, B. Senjean, K. Setia, S. Sim, D. S. Steiger, M. Steudtner, Q. Sun, W. Sun, D. Wang, F. Zhang, and R. Babbush, Quantum Sci. Technol. **5**, 034014 (2020).

[93] H. P. Hratchian and H. B. Schlegel, in *Theory and Applications of Computational Chemistry*, edited by C. E. Dykstra, G. Frenking, K. S. Kim, and G. E. Scuseria (Amsterdam, 2005) pp. 195–249.

[94] D. G. Truhlar, B. C. Garrett, and S. J. Klippenstein, The Journal of Physical Chemistry **100**, 12771 (1996).

[95] K. Fukui, The Journal of Physical Chemistry **74**, 4161 (1970).

[96] A. Fernández-Ramos, J. A. Miller, S. J. Klippenstein, and D. G. Truhlar, Chemical Reviews **106**, 4518 (2006).

[97] T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic Structure Theory* (Chichester, 2000).

[98] M. Bursch, J.-M. Mewes, A. Hansen, and S. Grimme, Angewandte Chemie International Edition **61**, e202205735 (2022).

[99] A. A. Chamkin and E. S. Serkova, Journal of Computational Chemistry **41**, 2388 (2020).

[100] L. A. Curtiss, P. C. Redfern, and K. Raghavachari, The Journal of Chemical Physics **126**, 10.1063/1.2436888 (2007), 084108.

[101] G. da Silva, Chemical Physics Letters **558**, 109 (2013).

[102] B. Chan, J. Deng, and L. Radom, Journal of Chemical Theory and Computation **7**, 112 (2011).

[103] J. Montgomery, J. A., M. J. Frisch, J. W. Ochterski, and G. A. Petersson, The Journal of Chemical Physics **112**, 6532 (2000).

[104] A. Karton and L. Goerigk, Journal of Computational Chemistry **36**, 622 (2015).

[105] O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko, and K.-R. Müller, Chemical Reviews **121**, 10142 (2021).

[106] J. Behler, The Journal of Chemical Physics **145**, 170901 (2016).

[107] P. O. Dral, M. Barbatti, and W. Thiel, The Journal of Physical Chemistry Letters **9**, 5660 (2018).

[108] J. Westermayr and P. Marquetand, Chemical Reviews **121**, 9873 (2021).

[109] J. Lan, V. Kapil, P. Gasparotto, M. Ceriotti, M. Iannuzzi, and V. V. Rybkin, Nature Communications **12**, 766 (2021).

[110] V. Kapil and E. A. Engel, Proceedings of the National Academy of Sciences **119**, e2111769119 (2022).

[111] T. E. Gartner, P. M. Piaggi, R. Car, A. Z. Panagiotopoulos, and P. G. Debenedetti, Phys. Rev. Lett. **129**, 255702 (2022).

[112] S. Brickel, A. K. Das, O. T. Unke, H. T. Turan, and M. Meuwly, Electronic Structure **1**, 024002 (2019).

[113] S. J. Ang, W. Wang, D. Schwalbe-Koda, S. Axelrod, and R. Gómez-Bombarelli, Chem **7**, 738 (2021).

[114] M. Yang, L. Bonati, D. Polino, and M. Parrinello, Catalysis Today **387**, 143 (2022).

[115] M. de la Puente, R. David, A. Gomez, and D. Laage, Journal of the American Chemical Society **144**, 10524 (2022).

[116] T. A. Young, T. Johnston-Wood, H. Zhang, and F. Duarte, Phys. Chem. Chem. Phys. **24**, 20820 (2022).

[117] K. Töpfer, S. Käser, and M. Meuwly, Phys. Chem. Chem. Phys. **24**, 13869 (2022).

[118] J. Li, P. Reiser, B. R. Boswell, A. Eberhard, N. Z. Burns, P. Friederich, and S. A. Lopez, Chem. Sci. **12**, 5302 (2021).

[119] T. A. Young, T. Johnston-Wood, V. L. Deringer, and F. Duarte, Chem. Sci. **12**, 10944 (2021).

[120] Y. Cao, J. Romero, J. P. Olson, M. Degroote, P. D. Johnson, M. Kieferová, I. D. Kivlichan, T. Menke, B. Peropadre, N. P. D. Sawaya, S. Sim, L. Veis, and A. Aspuru-Guzik, Chemical Reviews **119**, 10856 (2019).

[121] B. Bauer, S. Bravyi, M. Motta, and G. K.-L. Chan, Chemical Reviews **120**, 12685 (2020).

[122] R. R. Jones and R. G. Bergman, Journal of the American Chemical Society **94**, 660 (1972).

[123] C. Ingold, *Structure and Mechanism in Organic Chemistry* (1969).

[124] H. Dong, B.-Z. Chen, M.-B. Huang, and R. Lindh, Journal of Computational Chemistry **33**, 537.

[125] Z. Kerekes, D. A. Tasi, and G. Czakó, The Journal of Physical Chemistry A **126**, 889 (2022).

[126] H. Jónsson, G. Mills, and K. W. Jacobsen, Nudged elastic band method for finding minimum energy paths of transitions, in *Classical and Quantum Dynamics in Condensed Phase Simulations*, pp. 385–404.

[127] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[128] F. Weigend and R. Ahlrichs, Phys. Chem. Chem. Phys. **7**, 3297 (2005).

[129] E. Aprà, E. J. Bylaska, W. A. de Jong, N. Govind, K. Kowalski, T. P. Straatsma, M. Valiev, H. J. J. van Dam, Y. Alexeev, J. Anchell, V. Anisimov, F. W. Aquino, R. Atta-Fynn, J. Autschbach, N. P. Bauman, J. C. Becca, D. E. Bernholdt, K. Bhaskaran-Nair, S. Bogatko, P. Borowski, J. Boschen, J. Brabec, A. Bruner, E. Cauët, Y. Chen, G. N. Chuev, C. J. Cramer, J. Daily, M. J. O. Deegan, T. H. Dunning, M. Dupuis, K. G. Dyall, G. I. Fann, S. A. Fischer, A. Fonari, H. Früchtl, L. Gagliardi, J. Garza, N. Gawande, S. Ghosh, K. Glaesemann, A. W. Götz, J. Hammond, V. Helms, E. D. Hermes, K. Hirao, S. Hirata, M. Jacquelin, L. Jensen, B. G. Johnson, H. Jónsson, R. A. Kendall, M. Klemm, R. Kobayashi, V. Konkov, S. Krishnamoorthy, M. Krishnan, Z. Lin, R. D. Lins, R. J. Littlefield, A. J. Logsdail, K. Lopata, W. Ma, A. V. Marenich, J. Martin del Campo, D. Mejia-Rodriguez, J. E. Moore, J. M. Mullin, T. Nakajima, D. R. Nascimento, J. A. Nichols, P. J. Nichols, J. Nieplocha, A. Otero-de-la Roza, B. Palmer, A. Panyala, T. Pirojsirikul, B. Peng, R. Peverati, J. Pittner, L. Pollack, R. M. Richard, P. Sadayappan, G. C. Schatz, W. A. Shelton, D. W. Silverstein, D. M. A. Smith, T. A. Soares, D. Song, M. Swart, H. L. Taylor, G. S. Thomas, V. Tipparaju, D. G. Truhlar, K. Tsemekhman, T. Van Voorhis, A. Vázquez-Mayagoitia, P. Verma,

O. Villa, A. Vishnu, K. D. Vogiatzis, D. Wang, J. H. Weare, M. J. Williamson, T. L. Windus, K. Woliński, A. T. Wong, Q. Wu, C. Yang, Q. Yu, M. Zacharias, Z. Zhang, Y. Zhao, and R. J. Harrison, The Journal of Chemical Physics **152**, 184102 (2020).

[130] B. O. Roos, P. R. Taylor, and P. E. Sigbahn, Chemical Physics **48**, 157 (1980).

[131] R. Lindh and B. J. Persson, Journal of the American Chemical Society **116**, 4963 (1994).

[132] Q. Sun, J. Yang, and G. K.-L. Chan, Chemical Physics Letters **683**, 291 (2017).

[133] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters, and G. K.-L. Chan, WIREs Computational Molecular Science **8**, e1340 (2018).

[134] G. D. Purvis and R. J. Bartlett, The Journal of Chemical Physics **76**, 1910 (1982).

[135] T. H. Dunning, The Journal of Chemical Physics **90**, 1007 (1989).

[136] D. E. Woon and T. H. Dunning, The Journal of Chemical Physics **98**, 1358 (1993).

[137] A. K. Wilson, D. E. Woon, K. A. Peterson, and T. H. Dunning, The Journal of Chemical Physics **110**, 7667 (1999).

[138] S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, and K.-R. Müller, Science Advances **3**, e1603015 (2017).

[139] S. Chmiela, H. E. Sauceda, K.-R. Müller, and A. Tkatchenko, Nature Communications **9**, 3887 (2018).

[140] H. E. Sauceda, S. Chmiela, I. Poltavsky, K.-R. Müller, and A. Tkatchenko, The Journal of Chemical Physics **150**, 114102 (2019).

[141] J. Steinmetzer, S. Kupfer, and S. Gräfe, International Journal of Quantum Chemistry **121**, e26390 (2021).

[142] M. Bursch, J.-M. Mewes, A. Hansen, and S. Grimme, Angewandte Chemie International Edition **61**, e202205735 (2022).

[143] G. Imbalzano, A. Anelli, D. Giofré, S. Klees, J. Behler, and M. Ceriotti, The Journal of Chemical Physics **148**, 241730 (2018).

[144] C. Bannwarth, S. Ehlert, and S. Grimme, Journal of Chemical Theory and Computation **15**, 1652 (2019).

[145] P. Atkins and J. Paula, *Atkins' physical chemistry* (2008).