IMPERIAL

Applications of Quantum Computing to Real Materials: Quantum Simulations of Conducting Polymers

Lorenzo Sticher

Supervisors:
Professor Peter Haynes
Professor Myungshik Kim
Professor Johannes Knolle

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

Imperial College London
Department of Materials – Faculty of Engineering

January 31st, 2026

Declaration of Originality

I, Lorenzo Sticher, declare that the work in this thesis is my own. The work of others has been appropriately referenced. A full list of references is given in the bibliography.

Copyright

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Abstract

Electronic structure theory is considered an extremely promising application of both near-term and fault-tolerant quantum computers. The variational quantum eigensolver (VQE) has been studied extensively as an algorithm for solving simple chemical systems on noisy quantum devices. However, given the current constraints of these devices, benchmarks have focused mostly on lattice and small molecular systems. The gap between realistic, useful electronic structure simulations and current idealised benchmark systems remains large. With this thesis we aim to bridge this gap, by considering simulations of *trans*-polyacetylene (TPA) as benchmark system. This system is industrially relevant and displays non-trivial physics, but is also importantly modelled by the relatively simple Su–Schrieffer–Heeger (SSH) model.

Using VQE simulations of the SSH model, we assess the feasibility of quantum computing from the viewpoint of an electronic structure practitioner. A particular emphasis rests on the calculation of molecular forces, an understudied application, but of great importance in realistic electronic structure calculations. Throughout our work, we find that within our implementation the shot- and circuit-noise present on realistic quantum devices impose limitations on VQE simulations of SSH that put them beyond the scope of current algorithmic and hardware capabilities. Analysing our calculations with a particular focus on the (in-)applicability of the variational principle in realistic calculations, we find that the consideration of molecular forces is crucial to appropriately evaluating the feasibility of VQE for realistic electronic structure calculations. Furthermore, we find that the SSH description of TPA provides a valuable benchmark for assessing the performance of and guiding future improvements in both algorithms and hardware.

Acknowledgements

I would like to extend my gratitude to everyone who supported me throughout my PhD. This thesis would certainly not have been possible without the help and encouragement of many people.

First and foremost, I would like to thank my supervisor, Prof. Peter Haynes, for his invaluable and continued guidance and support throughout the past four years, without which this PhD and thesis would not have been possible. I would also like to thank my co-supervisors, Prof. Myungshik Kim and Prof. Johannes Knolle, for their useful input in shaping this thesis from the perspectives of both quantum information theory and condensed matter physics.

I would also like to thank all my friends and colleagues on Level 12 of the EEE building, not only for many valuable academic discussions, but also for every lunchtime chat and coffee break, and for the welcoming, friendly, and supportive environment created by each and every one of them.

I am also deeply grateful to all my friends, both within and outside Imperial, as well as within and outside the UK, who have made these years such an enjoyable and memorable experience.

Over the past three years, I have also had the privilege of living with my amazing friends and flatmates, Alex and Amy. I honestly don't think I would have made it to this point in the PhD without them. Thank you for making our place feel like home, and for filling the past few years with so many cups of tea, biscuits, dinnertime chats, pub quiz "wins", and plenty of much-needed laughter. I'm incredibly grateful for the friendship, kindness, and everyday joy you brought to life outside the PhD.

Last but certainly not least, I have immense gratitude for my family's loving support and endless patience. Without their encouragement and belief in me throughout every part of this PhD, this thesis would not have been possible.

Contents

1 Introduction **8**

1.1 Thesis Outline 11

2 Background **12**

2.1 Electronic Structure Theory 12

2.1.1 The Electronic Structure Problem 12

2.1.2 Slater Determinants 14

2.1.3 Second Quantisation 14

2.2 The Variational Principle 15

2.2.1 Exact Energy 16

2.2.2 Exact Molecular Forces 17

2.2.3 Variational Principle in Practice 18

2.2.4 Generalised Functionals and Variance 19

2.3 Density Functional Theory 20

2.3.1 Hohenberg-Kohn DFT 20

2.3.2 Kohn-Sham DFT 21

2.3.3 DFT in Practice 21

2.4 Wave-function Methods 22

2.4.1 Hartree-Fock 22

2.4.2 Post-HF Methods 24

2.4.3 WF Methods in Practice 25

2.5 Molecular Forces 25

2.5.1 Hellmann–Feynman Theorem 26

2.5.2 Approximate Hellmann-Feynman Theorem 27

2.5.3 Pulay Forces 28

2.5.4 Non-variational/Response Forces 30

2.6 Quantum Computation 32

2.6.1 Qubits 33

2.6.2 Gates 34

2.6.3 Measurements . . . 34
2.6.4 Hardware Constraints . . . 35

3 Modelling *Trans*-polyacetylene 36
3.1 TPA and Conducting Polymers . . . 36
3.2 The Su–Schrieffer–Heeger Model . . . 37
3.2.1 The Tight-binding Model . . . 37
3.2.2 Bloch States . . . 38
3.2.3 Dimerisation and Peierls Distortion . . . 39
3.2.4 Topology and Solitons . . . 40
3.2.5 Electronic Interactions . . . 41
3.2.6 Model Parametrisation . . . 42
3.3 DFT Simulations of TPA . . . 42
3.3.1 DFT Implementation . . . 42
3.3.2 Numerical Results . . . 43

4 The Variational Quantum Eigensolver 49
4.1 The VQE Algorithm . . . 50
4.2 Fermion to Qubit Mapping . . . 50
4.2.1 Jordan–Wigner Encoding . . . 52
4.2.2 Alternative Encodings . . . 53
4.3 Ansatz Circuit . . . 54
4.3.1 Hamiltonian Variational Ansatz . . . 55
4.3.2 Adaptive Ansatz . . . 57
4.4 Energy Measurement . . . 58
4.4.1 Measurable Fragments . . . 59
4.4.2 Quantum Phase Estimation . . . 62
4.5 Classical Optimisation . . . 63
4.5.1 Circuit Gradients . . . 64
4.5.2 L-BFGS . . . 66
4.5.3 SPSA . . . 68

5 Ideal VQE Simulations of *Trans*-polyacetylene 70
5.1 The SSH Model . . . 71
5.2 VQE Implementation . . . 74
5.2.1 Fermion to Qubit Mapping . . . 74
5.2.2 Ansatz Circuit . . . 79
5.2.3 Energy Measurement . . . 89

5.2.4 Classical Optimisation . . . 90
5.3 Molecular Forces in VQE . . . 92
5.3.1 Implementation . . . 93
5.4 Numerical Results . . . 95
5.4.1 Simulation Methodology . . . 96
5.4.2 Energy Accuracy . . . 97
5.4.3 Molecular Force Accuracy . . . 101
5.5 Conclusion . . . 104

6 Realistic VQE Simulations of *Trans*-polyacetylene 105
6.1 VQE Implementation . . . 106
6.1.1 Shot Noise . . . 107
6.1.2 Circuit Noise . . . 109
6.2 The Variational Principle in Realistic Calculations . . . 111
6.2.1 Shot Noise . . . 113
6.2.2 Circuit Noise . . . 125
6.3 Numerical Results: Shot Noise . . . 127
6.3.1 Simulation Methodology . . . 127
6.3.2 SPSA . . . 128
6.3.3 L-BFGS . . . 131
6.3.4 Discussion . . . 133
6.4 Numerical Results: Circuit Noise . . . 140
6.4.1 Simulation Methodology . . . 140
6.4.2 Discussion . . . 141
6.5 Conclusion . . . 147

7 Conclusion 150

Bibliography 153

A Bond-centred TPA Model 172
A.1 Independent Electron Models . . . 173
A.2 Wannier Functions . . . 180
A.3 Interacting Electron Models . . . 181

B Shot Noise: Further Numerical Results 182

List of Figures

3.1 Skeletal structure of TPA . . . 36
3.2 Band-structure of TPA at fixed geometry using LDA functional . . . 44
3.3 Geometry optimised band-structure of TPA at fixed geometry using LDA and PBE functional . . . 45
3.4 Geometry optimised band-structure of TPA using PBE+U functional 47

5.1 Orbital ordering used for the Jordan-Wigner encoding of the SSH model 76
5.2 Hamiltonian fragments used for constructing HVA for SSH model . . 81
5.3 Convergence of energy and energy error with number of layers in ansatz circuit . . . 98
5.4 Convergence of probability of reaching chemical accuracy with number of layers in ansatz circuit . . . 99
5.5 Number of optimisation iterations versus number of layers in ansatz circuit . . . 100
5.6 Representative VQE optimisation traces at varying number of layers in ansatz circuit . . . 101
5.7 Convergence of molecular force error with number of layers in ansatz circuit . . . 102
5.8 Molecular force error against energy error for VQE calculations across a range of number of layers in ansatz circuit . . . 103

6.1 SPSA hyper-parameter grid search . . . 129
6.2 Energy error of SPSA results with varying number of shots . . . 130
6.3 Molecular force error of SPSA results with varying number of shots . 131
6.4 Energy error of L-BFGS results with varying number of shots . . . 132
6.5 Molecular force error of L-BFGS results with varying number of shots 132
6.6 Representative SPSA convergence traces at low and high shot noise . 133
6.7 Representative L-BFGS convergence traces at low and high shot noise 135
6.8 Energy error as function of the total cost of SPSA and L-BFGS calculations . . . 136

6.9 Molecular force error as function of the total cost of L-BFGS calculations 137
6.10 Stochastic versus variational energy error contributions at varying shot noise . . . 138
6.11 Stochastic versus variational molecular force error contributions at varying shot noise . . . 139
6.12 Energy error as a function of noise model components . . . 142
6.13 Energy error heat-map as function of 1- and 2-qubit gate errors . . . 143
6.14 Energy error as function of either only 1- or only 2-qubit gate errors . 144
6.15 Energy error as function of simultaneous 1- and 2-qubit gate relative error probabilities . . . 145
6.16 Force error as function of simultaneous 1- and 2-qubit gate relative error probabilities . . . 145
6.17 Energy and molecular force error as function of simultaneous 1- and 2-qubit gate relative error probabilities . . . 146

A.1 Wannier functions for SSH model . . . 178
A.2 Symmetrised Wannier functions for SSH model . . . 179

B.1 Representative SPSA convergence traces of energy error, ΔE, for a selected VQE calculation at different shot budgets, M . . . 183
B.1 Representative SPSA convergence traces of energy error, ΔE, for a selected VQE calculation at different shot budgets, M (continued). . 184
B.2 Representative L-BFGS convergence traces of energy error, ΔE, for a selected VQE calculation at different shot budgets, M . . . 185
B.2 Representative L-BFGS convergence traces of energy error, ΔE, for a selected VQE calculation at different shot budgets, M (continued). 186

List of Tables

6.1 Discrete hyper-parameter grid used for SPSA tuning 128
6.2 Optimal SPSA hyper-parameters found from grid-search 130

Chapter 1

Introduction

Electronic structure theory is the discipline of predicting properties of molecular and material systems of interest using numerical solutions or approximations to the quantum mechanical equations determining the motions of electrons. Throughout the last century the field has evolved significantly and a wide variety of methods and applications have been developed. They have been useful both in predicting properties, such as the calculation of molecular energy levels and structural properties of enzymes [1], and in understanding the underlying physics, such as the study of superconductors and two-dimensional layered materials [2, 3]. Over the past decades continuous increases in computational power and improvements in numerical methods have begun to make a "Materials by Design" approach possible in which computational simulations inform experimental implementations [4]. However, electronic structure calculations on classical computers are inherently limited by the exponential growth of the computational complexity with the size of the system, a limitation inherent to the quantum mechanical nature of the system. Quantum computing has long been proposed as a possible solution to this exponential scaling, since the quantum mechanical nature of such a device provides a natural setting to tackle this exponential complexity growth [5]. In the current era of "Noisy Intermediate Scale Quantum" (NISQ) [6], hybrid quantum-classical variational algorithms have established themselves as a promising avenue to make use of the currently available hardware [7]. The variational quantum eigensolver (VQE) is an example of such an algorithm aimed at solving the electronic structure problem [8].

VQE has been explored extensively for the calculation of ground-state energies of quantum mechanical systems [9]. The often cited long term goals are the simulations of large, complex molecules such as those encountered in the design of pharmaceutical drugs [10]. However, due to the current qubit count and noise related limitations of real quantum computers, and the high cost associated with emulating them on

classical hardware, VQE calculations have been restricted to the simplest possible systems. Frequently considered problems include spin and fermionic lattice models, such as Heisenberg or Hubbard models [11–14]. Compared to real materials and molecules these lattice models are much simpler in structure as each site is represented by a single orbital, Hamiltonians are highly local and symmetric, and dimensionality is frequently restricted to one or two dimensions. These simplifications allow for simulations of much larger systems given the current computational limitations [15]. At the same time solving these systems is still of scientific interest to the physics community, as many questions remain unanswered [16]. When it comes to materials and molecules, standard systems considered include H_2, LiH, other simple diatomic molecules, and small hydrogen chains [17–20]. Compared to lattice models, simulating these systems using quantum computers requires considerably more quantum resources as each atom usually requires multiple orbitals to describe and Hamiltonians are long-ranged, unstructured and three-dimensional. While compared to lattice models these investigations are a much better representation of the systems of interest in materials science and chemistry, they are ultimately restricted to the simplest possible molecular systems, for which properties can frequently be predicted correctly at a qualitative level using only mean-field calculations.

Furthermore, neither the lattice models nor the very small molecular systems are particularly representative of "realistic" electronic structure calculations of "useful" materials and properties. While predicting ground-state energies is the most fundamental step in any electronic structure calculation, electronic structure practitioners are usually interested in predicting physical properties and observables beyond the ground-state energy. More appropriate calculations would for example involve prediction of the ground-state molecular structure and its properties by using VQE as part of a geometry optimisation. The simple lattice models mentioned previously are inherently unable to capture this part of materials simulations. Similarly, while the small molecular systems often considered do technically allow for the inclusion of a geometry optimisation, the systems are too simplistic to warrant or allow such calculations; the qualitative nature of the physics in these systems does not depend on a correct description of the ground-state geometry, and the number of molecular coordinates in these systems is too small to require geometry optimised calculations. When the molecular geometry is in fact considered in these calculations, it is oftentimes done at a very rudimentary level. A potential energy surface (PES) is usually constructed explicitly, most of the time one-dimensional and with respect to a singular molecular coordinate, the bond length [17, 18, 20]. This is in stark contrast with more realistic calculations of larger molecules, in which molecular forces

have to be evaluated. This is a well studied discipline in electronic structure theory which has allowed the prediction of ground-state geometries, reaction pathways and vibrational frequencies of molecules [21].

With this thesis we aim to approach VQE calculations from a viewpoint of a practitioner of standard materials simulations, who is trying to perform "useful" and "realistic" simulations on quantum computers. However, as discussed, the most commonly considered lattice/Hubbard models and small molecular systems are not well suited for such an investigation, due to their limited chemical relevance and small system size respectively. Since we are inherently limited by the available quantum devices and classical simulability of quantum algorithms, in our work we instead focus on simulations of *trans*-polyacetylene (TPA). TPA as a molecule presents a great opportunity to explore some of these more realistic aspects of electronic structure calculations. It is an industrially relevant material that exhibits rich and interesting physics. However, its properties cannot be described using (semi-)local density functional theory (DFT) calculations, a strong motivation for performing more accurate and exact quantum mechanical calculations; DFT predicts not only quantitatively inaccurate but also qualitatively incorrect results such as the presence of a fictitious spin density wave (SDW) instead of the correct charge density wave (CDW) structure. Most importantly it can also be described using a very simple one-dimensional Hubbard-like model, the Su–Schrieffer–Heeger (SSH) model. The reduction of the complexity and dimensionality in this model system makes it highly amenable to simulations using VQE, especially given the current constraints imposed by NISQ devices [15]. Furthermore, unlike standard Hubbard models, the molecular-derived SSH model still makes it possible to explore the above described aspects of "realistic" quantum simulations of molecules and materials, since the correct prediction of the properties of TPA – not only quantitatively but also qualitatively – requires an accurate determination of the ground-state geometry using a molecular force based geometry optimisation.

We aim to exemplify the simulation of the SSH model using VQE, quantifying the quantum resources required to predict its properties to sufficient accuracy while simultaneously determining the limitations imposed by, and necessary improvements required to, the currently available NISQ devices to allow such calculations to be performed. Since a full geometry optimised calculation is beyond the scope of this thesis and the available classical and quantum resources, we instead focus on the calculation of the molecular forces in the SSH model, a crucial component of geometry optimisations and hence important for correctly predicting the properties of the system. We also aim to use our calculations and results to draw more general con-

clusions about the prospects of using quantum computers for realistic simulations of molecules and materials, in particular with a view on the calculation of molecular forces using VQE, since this application has received relatively little attention so far, and since it is well established in standard electronic structure theory that predicting sufficiently accurate molecular forces is in general considerably more difficult than predicting sufficiently accurate ground-state energies.

The Principle of Generality Since with our work we aim to approach the VQE calculation from a viewpoint of an electronic structure practitioner, we will in many instances consider "standard" VQE implementations which, while potentially suboptimal for our particular problem instance, more closely represent realistic electronic structure calculations.

1.1 Thesis Outline

In Chapter 2 we discuss the theoretical background necessary to understand the general electronic structure theory and quantum computing concepts used and discussed throughout this thesis. In Chapter 3 we introduce TPA, discuss the SSH model and present numerical results for DFT calculations of TPA. Chapter 4 presents the general concept of VQE and discusses the specific methods relevant for our work at an abstract level. In Chapter 5 we first discuss the VQE implementation for our calculations, specialising the previously discussed abstract concepts to the specific model considered. We then present numerical results for ideal VQE simulations of our model. In Chapter 6 we analyse the expected impact of shot- and circuit-noise on VQE calculations. We also present numerical results for our VQE calculations under the influence of shot- and circuit-noise.

Chapter 2

Background

2.1 Electronic Structure Theory

2.1.1 The Electronic Structure Problem

Electronic structure theory is concerned with determining the properties of molecular and condensed matter systems. As quantum mechanical system they are determined by the time-dependent Schrödinger equation (TDSE) [1] [22]:

$$\hat{H}|\psi\rangle = i\frac{\partial}{\partial t}|\psi\rangle \tag{2.1}$$

where the electronic structure Hamiltonian, $\hat{H}_{\mathrm{ES}}$, takes the form:

$$\begin{aligned}\hat{H}_{\mathrm{ES}} = &\overbrace{-\frac{1}{2}\sum_i \nabla_i^2}^{\hat{T}_e} \quad \overbrace{-\sum_{i,I}\frac{Z_I}{|\boldsymbol{r}_i - \boldsymbol{R}_I|}}^{\hat{V}_{ne}} \quad \overbrace{+\frac{1}{2}\sum_{i\neq j}\frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|}}^{\hat{V}_{ee}} \\ &\underbrace{-\sum_I \frac{1}{2M_I}\nabla_I^2}_{\hat{T}_n} \quad \underbrace{+\frac{1}{2}\sum_{I\neq J}\frac{Z_I Z_J}{|\boldsymbol{R}_I - \boldsymbol{R}_J|}}_{\hat{V}_{nn}} .\end{aligned} \tag{2.2}$$

It describes a general system of electrons and atomic nuclei with positions $\boldsymbol{r}_i$ and $\boldsymbol{R}_I$, charges e and $Z_I e$, and masses m_e and M_I respectively, all interacting through Coulomb repulsion. The individual contributions to T_e, V_{ne}, V_{ee}, T_n, and V_{nn} account for the electron kinetic energies, electron-nucleus interactions, electron-electron interactions, nucleus kinetic energies, and nucleus-nucleus interactions respectively. In the absence of any external interactions the physical system is further determined

[1] Unless otherwise specified any equations assume atomic units: $\hbar = m_e = e = 4\pi\varepsilon_0 = 1$

by the ground-state solution to the time-independent Schrödinger equation (TISE):

$$\hat{H}|\psi\rangle = E|\psi\rangle \tag{2.3}$$

The complexity of the equations grows exponentially with the system size N – since equ. (2.3) defines a differential equation of $3N$ variables, the coordinates of electrons and nuclei, a direct solution requires exponentially many resources in N. This makes analytic solutions intractable for all but the simplest systems. Approximate solutions to the electronic structure problem have therefore been sought out since the beginning of quantum mechanics. Of particular importance throughout electronic structure theory is the Born-Oppenheimer (BO) approximation. As part of this approximation and based on the difference in mass between the electrons and nuclei, phonon-electron interactions are ignored, making it possible to factorise the complete many-body wavefunction into electronic and nuclear parts:

$$\tilde{\psi}(\{\boldsymbol{r}_i\}, \{\boldsymbol{R}_I\}) = \psi(\{\boldsymbol{r}_i\}; \{\boldsymbol{R}_I\})\, \Phi(\{\boldsymbol{R}_I\}) \tag{2.4}$$

This factorisation allows the nuclear coordinates to be treated as classical parameters of the system that determine the electronic wavefunction and energy through their electro-static potential only. An effective time-independent electronic Hamiltonian under the BO approximation is given by [22]:

$$\hat{H}_{\mathrm{ES,e}}(\{\boldsymbol{R}_I\}) = \hat{T}_e + \hat{V}_{ne}(\{\boldsymbol{R}_I\}) + \hat{V}_{ee} + \left[\hat{V}_{nn}(\{\boldsymbol{R}_I\})\right] \tag{2.5}$$

with solutions determined by the corresponding TISE:

$$\hat{H}_{\mathrm{ES,e}}(\{\boldsymbol{R}_I\})|\psi_e(\{\boldsymbol{R}_I\})\rangle = E_e(\{\boldsymbol{R}_I\})|\psi_e(\{\boldsymbol{R}_I\})\rangle \tag{2.6}$$

and where the dependence on the nuclear positions as parameters has been made explicit. The electrostatic interaction of the nuclear configuration, $\hat{V}_{nn}(\{\boldsymbol{R}_I\})$, in equ. (2.5) does not need to be explicitly included when solving equ. (2.6), since in the BO approximation it is a purely classical contribution that acts as a constant energy shift to $E_e(\{\boldsymbol{R}_I\})$ and can be accounted for at the end of the computation. Solving the electronic TISE under the BO approximation is the main focus of electronic structure theory and of the work in our thesis and we will drop the explicit electronic subscripts e from here on.

2.1.2 Slater Determinants

Since electrons are indistinguishable fermions, physically relevant solutions to the electronic structure problem are required to be fully antisymmetric under the exchange of any two electrons, $\boldsymbol{r}_i \leftrightarrow \boldsymbol{r}_j$. This requirement can be naturally expressed and incorporated when many-body wavefunctions are constructed from a basis of orthonormal single-particle orbitals $\{\phi_p(\boldsymbol{r})\}$ as is frequently done in electronic structure theory. Given a set of N such spin orbitals an N-electron wave function can be directly formed as the product of each individual occupied orbital:

$$|\psi(\{\boldsymbol{r}_i\})\rangle = |\phi_{p_1}(\boldsymbol{r}_1)\phi_{p_2}(\boldsymbol{r}_2)\dots\phi_{p_N}(\boldsymbol{r}_N)\rangle \tag{2.7}$$

However, this many-body wavefunction is not antisymmetric under exchange of electrons. The antisymmetry of the wavefunction requires antisymmetry under exchange of electronic orbitals $\phi_p(\boldsymbol{r}) \leftrightarrow \phi_q(\boldsymbol{r})$ for any electrons i and j occupying any orbitals p and q respectively. An appropriately (anti-)symmetrised wavefunction can be constructed as the Slater determinant (SD) of the occupied orbitals [22]:

$$|\psi(\{\boldsymbol{r}_i\})\rangle = \frac{1}{\sqrt{N!}}\begin{vmatrix} \phi_{p_1}(\boldsymbol{r}_1) & \phi_{p_2}(\boldsymbol{r}_1) & \cdots & \phi_{p_N}(\boldsymbol{r}_1) \\ \phi_{p_1}(\boldsymbol{r}_2) & \phi_{p_2}(\boldsymbol{r}_2) & \cdots & \phi_{p_N}(\boldsymbol{r}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{p_1}(\boldsymbol{r}_N) & \phi_{p_2}(\boldsymbol{r}_N) & \cdots & \phi_{p_N}(\boldsymbol{r}_N) \end{vmatrix} \equiv |\det(\phi_{p_1}\phi_{p_2}\dots\phi_{p_N})(\{\boldsymbol{r}_i\})\rangle \tag{2.8}$$

If the single-particle basis is complete, the set of all unique SDs that can be constructed from this basis also forms a complete basis for the space of anti-symmetric many-body wavefunctions.

2.1.3 Second Quantisation

The fermionic antisymmetry requirement and Slater determinant construction can be naturally expressed and incorporated in the formalism of second quantisation. Rather than working with explicit, antisymmetric N-electron wavefunctions $\psi(\boldsymbol{r}_1, \dots, \boldsymbol{r}_N)$, the problem is reformulated in terms of occupations of a basis of orthonormal single-particle spin-orbitals $\{\phi_p(\boldsymbol{r})\}$. The relevant Hilbert space is the fermionic Fock space, spanned by occupation-number states $|n_0 n_1 \dots\rangle$, where on account of the Pauli exclusion principle $n_p \in \{0, 1\}$. The occupation-number states correspond to the SDs, equ. (2.8), formed by occupying the specified orbitals. The fundamental operators in this representation are the fermionic creation and annihi-

lation operators, $\hat{c}_p^\dagger$ and $\hat{c}_p$, which act on orbital occupations as:

$$\hat{c}_p^\dagger|\underbrace{00\dots0}_{p-1}0\dots\rangle = |00\dots01\dots\rangle \qquad \hat{c}_p^\dagger|00\dots10\dots\rangle = \mathbf{0} \tag{2.9}$$

$$\hat{c}_p|00\dots01\dots\rangle = |00\dots00\dots\rangle \qquad \hat{c}_p|00\dots00\dots\rangle = \mathbf{0} \tag{2.10}$$

The completely unoccupied state $|00\dots\rangle$, which is annihilated by all $\hat{c}_p$, is commonly referred to as the vacuum state, $|\Omega\rangle$. In this formalism the antisymmetry of the wavefunction is incorporated directly into the fermionic operators through their anti-commutation relations:

$$\{\hat{c}_p, \hat{c}_q\} = \hat{c}_p\hat{c}_q + \hat{c}_q\hat{c}_p = 0, \tag{2.11}$$

$$\{\hat{c}_p^\dagger, \hat{c}_q^\dagger\} = \hat{c}_p^\dagger\hat{c}_q^\dagger + \hat{c}_q^\dagger\hat{c}_p^\dagger = 0, \tag{2.12}$$

$$\{\hat{c}_p, \hat{c}_q^\dagger\} = \hat{c}_p\hat{c}_q^\dagger + \hat{c}_q^\dagger\hat{c}_p = \delta_{pq}. \tag{2.13}$$

In the second quantisation formalism the electronic structure Hamiltonian, $\hat{H}_{\mathrm{ES}}$, can be expressed as:

$$\hat{H}_{\mathrm{ES}} = \sum_{pq} h_{pq}\hat{c}_p^\dagger\hat{c}_q + \frac{1}{2}\sum_{pqrs} h_{pqrs}\hat{c}_p^\dagger\hat{c}_q^\dagger\hat{c}_r\hat{c}_s \tag{2.14}$$

with the electronic one- and two-body integrals defined as [2] [22]:

$$\begin{aligned} h_{pq} &= \langle\phi_p|\hat{T}_e + \hat{V}_{ne}|\phi_q\rangle = \int \mathrm{d}\boldsymbol{r}\, \phi_p^*(\boldsymbol{r})\left(-\frac{1}{2}\nabla_i^2 - \sum_I \frac{Z_I}{|\boldsymbol{r}-\boldsymbol{R}_I|}\right)\phi_q(\boldsymbol{r}) \\ h_{pqrs} &= \langle\phi_p\phi_q|\hat{V}_{ee}|\phi_r\phi_s\rangle = \iint \mathrm{d}\boldsymbol{r}_1\,\mathrm{d}\boldsymbol{r}_2\, \phi_p^*(\boldsymbol{r}_1)\phi_q^*(\boldsymbol{r}_2)\frac{1}{|\boldsymbol{r}_1-\boldsymbol{r}_2|}\phi_r(\boldsymbol{r}_1)\phi_s(\boldsymbol{r}_2) \end{aligned} \tag{2.15}$$

where from now on we drop the explicit dependence $\hat{H}_{\mathrm{ES}} = \hat{H}_{\mathrm{ES}}(\{\boldsymbol{R}_I\})$.

2.2 The Variational Principle

The variational principle lies at the heart of modern electronic structure theory. It establishes the conceptually simple numerical procedure of minimising the energy as a computational method used for finding the ground state of the electronic structure Hamiltonian, or any excited state if orthogonality relative to lower energy eigenstates is enforced. In addition to establishing this variational method, the variational principle has important consequences for the computational effort required to calculate accurate energies and molecular forces within this method.

[2] We adhere to the physicist's notation and convention for electronic repulsion integrals [22]

In this section we review the theory behind the variational principle, since most electronic structure methods, including Hartree-Fock (HF), density functional theory (DFT), configuration interaction (CI), and most importantly for our work, the variational quantum eigensolver (VQE), are based on the variational method. We further discuss the implications for the relative scaling of errors in energies and molecular forces with the accuracy of the computational method and how they broadly apply in standard electronic structure theory. This sets the context in which we will analyse the accuracy of energies and molecular forces in our VQE calculations in sec. 5 and sec. 6. Lastly, we also discuss an extension of the variational principle to the variance, which presents an alternative minimisation target for approximating ground and excited states. While not directly used as variational method in our work, we discuss the consequences for the stochastic noise in quantum Monte Carlo simulations, which will be of importance for the analysis of stochastic noise in realistic VQE calculations considered in sec. 6.

2.2.1 Exact Energy

The variational principle establishes a direct correspondence between eigenstates, ψ_i, of any observable $\hat{O}$, and stationary states of the Rayleigh-Ritz quotient, $O[\psi]$ [23]. $O[\psi]$ is defined as the expectation value $\langle\hat{O}\rangle$ considered as a functional of the wavefunction ψ:

$$O[\psi] \equiv \frac{\langle\psi|\hat{O}|\psi\rangle}{\langle\psi|\psi\rangle} \tag{2.16}$$

Specialising the discussion to the Hamiltonian $\hat{H}$ and the energy functional $E[\psi] = \frac{\langle\psi|\hat{H}|\psi\rangle}{\langle\psi|\psi\rangle}$, a direct expansion of the functional in terms of small deviations of the wavefunction, $\psi \to \psi + \delta\psi$, leads to:

$$\begin{aligned} E[\psi+\delta\psi] &= \frac{\langle\psi+\delta\psi|\hat{H}|\psi+\delta\psi\rangle}{\langle\psi+\delta\psi|\psi+\delta\psi\rangle} && (2.17)\\ &= E[\psi] + \frac{1}{\langle\psi|\psi\rangle}\Big[\langle\delta\psi|\hat{H}|\psi\rangle + \langle\psi|\hat{H}|\delta\psi\rangle\Big] && (2.18)\\ &\quad - \frac{E[\psi]}{\langle\psi|\psi\rangle}\left[\langle\delta\psi|\psi\rangle + \langle\psi|\delta\psi\rangle\right] + \mathcal{O}(\delta\psi^2) \\ &= E[\psi] + \frac{2}{\langle\psi|\psi\rangle}\,\mathrm{Re}\langle\delta\psi|\hat{H} - E[\psi]|\psi\rangle + \mathcal{O}(\delta\psi^2) && (2.19) \end{aligned}$$

The stationary condition $\frac{\delta E}{\delta\psi} = 0$ requires that the linear term vanishes for any deviation $\delta\psi$, and hence that ψ is an eigenstate of $\hat{H}$, $\hat{H}|\psi\rangle = E|\psi\rangle$, thereby establishing the sought after correspondence. An explicit parametrisation of the wavefunction

as a linear combination of the Hamiltonian eigenstates,

$$|\psi\rangle = \sum_{k=0}^{\infty} c_k |\psi_k\rangle \qquad \hat{H}|\psi_k\rangle = E_k|k\rangle \qquad \langle\psi_k|\psi_l\rangle = \delta_{kl} \tag{2.20}$$

allows us to explicitly show that the energy of any wavefunction is an upper bound to the exact ground-state energy, E_0, and that the error in the energy is quadratic in the error in the wavefunction with respect to the exact ground state, ψ_0

$$E[\psi] = \frac{\sum_{k=0}^{\infty} |c_k|^2 E_k}{\sum_{k=0}^{\infty} |c_k|^2} = \frac{E_0 + \sum_{k=1}^{\infty} \frac{|c_k|^2}{|c_0|^2} E_k}{1 + \sum_{k=1}^{\infty} \frac{|c_k|^2}{|c_0|^2}} \tag{2.21}$$

$$= E_0 + \sum_{k=1}^{\infty} \frac{|c_k|^2}{|c_0|^2}(E_k - E_0) + \mathcal{O}\left(\frac{|c_k|^4}{|c_0|^4}\right) \geq E_0 \tag{2.22}$$

with equality $E[\psi] = E_0$ only if $c_k = 0$ for $k \geq 1$ and hence $|\psi\rangle \propto |\psi_0\rangle$.

The variational principle has been attributed to the immense success that electronic structure methods have seen for finding approximate solutions to the full electronic structure problem [23]. Not only does it establish the variational method – minimise the energy across some considered wavefunction space – as a valid method for finding approximate wavefunctions, but it also shows that since the error in the energy is quadratic in the error in the wavefunction, even relatively inaccurate wavefunction approximations can provide sufficiently accurate energies.

2.2.2 Exact Molecular Forces

Frequently in electronic structure calculations, the goal is to predict properties of interest beyond the (exact) ground-state energy, such as polarisabilities, electronic densities, or, as the main focus of our work, nuclear gradients and molecular forces. These quantities will, in general, not be stationary points of their corresponding functionals for the exact ground-state wavefunction [24]. As an example, we consider the Hellmann-Feynman force discussed in sec. 2.5.1, for which the corresponding functional is given by:

$$\boldsymbol{F}[\psi] \equiv \frac{\langle\psi|\hat{\boldsymbol{F}}|\psi\rangle}{\langle\psi|\psi\rangle} \qquad \hat{\boldsymbol{F}} \equiv -\frac{\partial \hat{H}}{\partial \boldsymbol{R}} \tag{2.23}$$

By applying the variational principle, in particular equ. (2.19), to $\hat{\boldsymbol{F}}$ and $\boldsymbol{F}[\psi]$ we find that $\boldsymbol{F}[\psi]$ will only be stationary at the eigenstates of $\hat{\boldsymbol{F}}$. Unless the ground state of $\hat{H}$, ψ_0, for which we want to calculate $\hat{\boldsymbol{F}}[\psi_0]$, happens to coincide with an

eigenstate of $\frac{\partial \hat{H}}{\partial \boldsymbol{R}}$ – a highly unlikely scenario for a general $\hat{H}$ – the force at the ground state will not be variational, since to first order:

$$\boldsymbol{\delta F} = \boldsymbol{F}[\psi_0 + \delta\psi] - \boldsymbol{F}[\psi_0] = \frac{2}{\langle\psi|\psi\rangle} \mathrm{Re}\langle\delta\psi|\hat{\boldsymbol{F}} - \boldsymbol{F}[\psi]|\psi\rangle \neq 0 \tag{2.24}$$

Given an error in the wavefunction of $\delta\psi$, the corresponding errors in the energy, $\delta E = E[\psi_0] - E[\psi_0 + \delta\psi]$, and the molecular force, $\boldsymbol{\delta F} = \boldsymbol{F}[\psi_0] - \boldsymbol{F}[\psi_0 + \delta\psi]$, scale as:

$$\delta E = \mathcal{O}(\delta\psi^2) \qquad\qquad \boldsymbol{\delta F} = \mathcal{O}(\delta\psi) \tag{2.25}$$

with a relative scaling of:

$$\delta F \sim \sqrt{\delta E} \tag{2.26}$$

Consequently, it is expected that molecular forces will, in general, converge to the exact ground-state result more slowly than energies as the accuracy of the approximate solution is improved [25]. This is commonly observed in a wide variety of practical electronic structure calculations as further discussed below.

2.2.3 Variational Principle in Practice

While the above discussions consider the energy functional applied to the whole Hilbert space and with respect to the exact eigenstate, any computational method based on variationally optimising some (not necessarily exact or correct) energy functional will exhibit the same scaling in errors of energies and molecular forces, provided that the errors are calculated relative to the exact solution within the computational approach chosen. In DFT calculations for example, an approximate energy functional of the density such as LDA, $E_{\mathrm{LDA}}[\rho]$, is minimised (see sec. 2.3), and the minimum of this functional, $\rho_{0,\mathrm{LDA}}$, will generally not correspond to the exact ground state of the full electronic structure Hamiltonian, equ. (2.5). Nonetheless, the same variational arguments apply to the accuracy in the energy, $\delta E_{\mathrm{LDA}} = E_{\mathrm{LDA}}[\rho] - E_{\mathrm{LDA}}[\rho_{0,\mathrm{LDA}}]$ and the accuracy of the force, $\delta \boldsymbol{F}_{\mathrm{LDA}} = \boldsymbol{F}_{\mathrm{LDA}}[\rho] - \boldsymbol{F}_{\mathrm{LDA}}[\rho_{0,\mathrm{LDA}}]$, as we approach the minimum of the functional, $\rho_{0,\mathrm{LDA}}$ [26].

As a result, the variational scaling of energies has wide ranging consequences for electronic structure calculations, where it can be observed as calculations are being converged with respect to the computational parameters determining the accuracy of the solution such as the energy cut-off, E_{cut}, in plane-wave DFT (e.g. [27], or the basis set size, N_{b} in CI singles and doubles (e.g. [28]). It is well known amongst electronic structure practitioners that sufficiently converged molecular forces gener-

ally require more accurate and more expensive calculations, with the exact details set by the method, the system studied, and the accuracy targeted.

2.2.4 Generalised Functionals and Variance

The variational principle can be further generalised to any analytic function of the Hamiltonian, $f(\hat{H}) \equiv \sum_k \frac{f^{(k)}}{k!}\hat{H}^k$ [29], with corresponding functional:

$$f_E[\psi] = \frac{\langle\psi|f(\hat{H})|\psi\rangle}{\langle\psi|\psi\rangle} \tag{2.27}$$

where a correspondence can again be established between eigenstates ψ_i of $\hat{H}$ and stationary points of $f_E[\psi]$ by expanding to first order in $\delta\psi$,

$$f_E[\psi + \delta\psi] = f_E[\psi] + \frac{2}{\langle\psi|\psi\rangle}\,\mathrm{Re}\langle\delta\psi|f(\hat{H}) - f_E[\psi]|\psi\rangle + \mathcal{O}(\delta\psi^2) \tag{2.28}$$

and observing that eigenstates of $\hat{H}$ are also eigenstates of $\hat{H}^n$ and hence of $f(\hat{H}) = \sum_k \frac{f^{(k)}}{k!}\hat{H}^k$.

A functional that has seen interest in electronic structure theory as an alternative minimisation target to the energy functional is the energy variance of a state:

$$\mathrm{Var}_E[\psi] = \frac{\langle\psi|(\hat{H} - E[\psi])^2|\psi\rangle}{\langle\psi|\psi\rangle} = \frac{\langle\psi|\hat{H}^2|\psi\rangle}{\langle\psi|\psi\rangle} - \left(\frac{\langle\psi|\hat{H}|\psi\rangle}{\langle\psi|\psi\rangle}\right)^2 \tag{2.29}$$

referred to as such since it corresponds to the variance of the random variable associated with measurement outcomes in the basis of the energy eigenstates, as further discussed in sec. 4.4. By applying the generalised variational principle to each of the terms individually, the variance at any exact eigenstate, ψ_i can be shown to be stationary, $\mathrm{Var}_E[\psi_i + \delta\psi] = \mathrm{Var}_E[\psi_i] + \mathcal{O}(\delta\psi^2)$. However, unlike the energy functional, the variance also vanishes at each eigenstate $\mathrm{Var}_E[\psi] = 0$, which together with the positivity of the functional, $\mathrm{Var}_E[\psi] \geq 0$, establishes each eigenstate, ground or excited, to be a minimum of the functional. This result, commonly referred to as the zero-variance property, allows any eigenstate to be targeted with a direct minimisation of $\mathrm{Var}_E[\psi]$, in contrast to $E[\psi]$ where a direct minimisation only targets the ground state. Minimisation of the variance functional has found wide ranging applications, particularly in variational quantum Monte Carlo (QMC) where it can often lead to improved convergence of the minimisation [30], but has recently also been applied to excited state VQE calculations to circumvent the requirement of orthogonalisation relative to the ground state [31].

It should be noted that while $\mathrm{Var}_E[\psi]$ does correspond to the variance of a measurement of the energy (via projection onto the energy eigenstates), first and foremost it is simply another property of ψ, which can in theory be calculated exactly, provided we are working with a representation of ψ for which $(\hat{H} - E[\psi])|\psi\rangle$ can easily/efficiently be evaluated (e.g. in a CI method evaluation of $\mathrm{Var}_E[\psi]$ reduces to simple linear algebra). As such, even if the variance is being used as alternative minimisation objective, the energies and molecular forces at ψ are at this stage still given by $E[\psi]$ and $\boldsymbol{F}[\psi]$ and the errors δE and $\delta \boldsymbol{F}$ are therefore still expected to scale according to equ. (2.25) and (2.26).

In standard QMC simulations, however, $E[\psi]$ and $\boldsymbol{F}[\psi]$ are in fact estimated by stochastic methods where the estimators $\bar{E}[\psi]$ and $\bar{\boldsymbol{F}}[\psi]$ have variance $\mathrm{Var}_E[\psi]$ and $\mathrm{Var}_{\boldsymbol{F}}[\psi]$ respectively. As with the energy functional it can be shown that in general the molecular forces are not expected to obey a variational principle for the variance near the ground state ψ_0. The variances of the estimators of energies and molecular forces near the ground state ψ_0 of $\hat{H}$ are therefore expected to scale as:

$$\mathrm{Var}_E[\psi_0 + \delta\psi] = \mathcal{O}(\delta\psi^2) \tag{2.30}$$

$$\mathrm{Var}_{\boldsymbol{F}}[\psi_0 + \delta\psi] = \mathrm{Var}_{\boldsymbol{F}}[\psi_0] + \mathcal{O}(\delta\psi) = \mathcal{O}(1) \tag{2.31}$$

The zero-variance property of $E[\psi]$ at ψ_0 has been attributed to the success of QMC methods for estimating energies, as it leads to a reduced resource requirement when the exact solution is approached, while similarly, the absence of such a property for $\mathrm{Var}_{\boldsymbol{F}[\psi]}$ leads to difficulties in estimating molecular forces and other non-variational quantities in QMC [32].

2.3 Density Functional Theory

2.3.1 Hohenberg-Kohn DFT

Density functional theory (DFT) is one of the main computational approaches used to (approximately) solve the electronic structure problem. The Hohenberg-Kohn (HK) theorems demonstrate a unique correspondence between electronic densities and ground-state wavefunctions, $\rho(\boldsymbol{r}) \leftrightarrow |\psi_0(\{\boldsymbol{r}_i\})\rangle$, and based on this correspondence and the variational principle presented in sec. 2.2 establish the existence of a universal, i.e. system independent, energy functional of the density, $E[\rho]$, that is minimised by the exact ground-state density, ρ_0. This formulation allows the electronic structure problem to be expressed solely in terms of the electronic density, greatly reducing the dimensionality (and computational complexity) of the central

mathematical object from $3N$ for the full wavefunction to three for the electronic density. In the HK formalism the functional can be formally expressed as:

$$\begin{aligned} E_{\mathrm{HK}}[\rho] &= T[\rho] + E_{\mathrm{int}}[\rho] + \int \mathrm{d}\boldsymbol{r}\ V_{\mathrm{ext}}(\boldsymbol{r})\rho(\boldsymbol{r}) \\ &\equiv \mathcal{F}_{\mathrm{HK}}[\rho] + \int \mathrm{d}\boldsymbol{r}\ V_{\mathrm{ext}}(\boldsymbol{r})\rho(\boldsymbol{r}) \end{aligned} \tag{2.32}$$

where the functionals $T[\rho]$ and $E_{\mathrm{int}}[\rho]$ correspond to the kinetic and Coulomb interaction energies of a system with ground-state density ρ, and V_{ext} is most commonly identified with V_{nn}. However, a mathematically tractable form of the functional $\mathcal{F}_{\mathrm{HK}}[\rho]$ is not known and implementations of the DFT formalism require the development and use of approximate energy functionals.

2.3.2 Kohn-Sham DFT

Kohn-Sham (KS) DFT makes accurate approximations and efficient implementations tractable by introducing a fictitious system of non-interacting electrons with the same ground-state density:

$$E_{\mathrm{KS}}[\rho] = T_s[\rho] + \int \mathrm{d}\boldsymbol{r}\ V_{\mathrm{ext}}(\boldsymbol{r})\rho(\boldsymbol{r}) + E_{\mathrm{H}}[\rho] + E_{\mathrm{XC}}[\rho] \tag{2.33}$$

where $T_s[\rho]$ and $E_{\mathrm{H}}[\rho]$ represent the kinetic energy and classical electrostatic (Hartree) energy of the non-interacting electron system and $E_{\mathrm{XC}}[\rho]$ is referred to as the Exchange-Correlation (XC) functional. Solving for the ground-state density and energy can be shown to be equivalent to solving the KS system with a Hamiltonian given by:

$$\hat{H}_{\mathrm{KS}} = -\frac{1}{2}\sum_i \nabla_i^2 + \sum_i V_{\mathrm{ext}}(\boldsymbol{r}_i) + \sum_i V_{\mathrm{H}}[\rho](\boldsymbol{r}_i) + \sum_i \frac{\delta E_{\mathrm{XC}}[\rho]}{\delta\rho(\boldsymbol{r}_i)} \tag{2.34}$$

This system needs to be solved self-consistently; since V_{H} and $\delta E_{\mathrm{XC}}/\delta\rho$ are functionals of ρ the solution requires an output density, ρ_{out}, that is consistent with the input density of the calculation ρ_{in}, i.e. $\rho_{\mathrm{out}} = \rho_{\mathrm{in}}$.

2.3.3 DFT in Practice

While KS DFT is in theory only a formally exact restatement of HK DFT, the form of the KS equations has made accurate implementations of DFT possible, since it was found that $E_{\mathrm{XC}}[\rho]$ is more amenable to accurate approximations. Over the last decades a hierarchy of approximate XC functionals has been developed, termed "Ja-

cob's ladder", increasing in both complexity and accuracy [33]. The simplest of these are the local and semi-local functionals termed local-density approximation (LDA) and generalised gradient approximation (GGA). LDA considers the functional to be purely local, depending only on the charge density at a specific point. GGA generalises this notion to a semi-local functional, allowing it to also depend on the gradient of the charge density at each point:

$$E_{\mathrm{LDA}}[n(\boldsymbol{r})] = \int d\boldsymbol{r}\ n(\boldsymbol{r})\ \epsilon_{\mathrm{LDA}}(n(\boldsymbol{r})) \tag{2.35}$$

$$E_{\mathrm{GGA}}[n(\boldsymbol{r})] = \int d\boldsymbol{r}\ n(\boldsymbol{r})\ \epsilon_{\mathrm{GGA}}(n(\boldsymbol{r}), |\nabla n(\boldsymbol{r})|) \tag{2.36}$$

While increasingly accurate exchange-correlation functional approximations continue to be developed, DFT suffers from the absence of a systematically improvable approximate functional [33]. In practice, DFT-based methods trade accuracy for computational efficiency and are generally able to produce qualitatively correct results for relatively large system sizes. As a result, they find extensive use in materials science and condensed-matter physics, where qualitative reliability and favourable scaling are the primary requirements, for simulations of comparatively large systems, involving structures consisting of hundreds of atoms (e.g. [34]). Nevertheless, the accuracy achieved is strongly system- and functional-dependent, particularly as the degree of electronic correlation increases [35]. Common failures include the systematic underestimation of band gaps in crystalline materials [36] and the self-interaction error caused by approximate functionals allowing spurious interactions of electrons with themselves [37].

2.4 Wave-function Methods

2.4.1 Hartree-Fock

Wave-function based methods still consider quantum mechanics at the level of the exponentially complex many-body wave-function of the system. Tractability of the problem is achieved by restricting the wave-function to a small, specific subspace of the full Hilbert space over which an approximate solution is found. The Hartree-Fock (HF) method is one of the simplest wave-function based methods for approximately solving the electronic structure problem. In HF the many body wavefunction is restricted to a single Slater determinant built from a set of orthonormal single-

particle orbitals, $\{\phi_p(\boldsymbol{r})\}$ [23]:

$$|\psi_{\mathrm{HF}}\rangle = |\det(\phi_1\phi_2\dots)\rangle \tag{2.37}$$

The orbitals used to construct the SD are commonly known as molecular orbitals (MO) and chosen in accordance with the variational principle by minimising the energy of the resulting wavefunction:

$$E_{\mathrm{HF}} = \min_{\{\phi_p\}} \langle\psi_{\mathrm{HF}}|\hat{H}|\psi_{\mathrm{HF}}\rangle \tag{2.38}$$

While solving HF directly in the position basis is possible by setting up a set of integro-differential equations, the most common framework for HF instead considers an expansion of the MOs in terms of a set of (generally non-orthogonal) atomic orbitals (AO), $\{\chi_\mu\}$:

$$\phi_p(\boldsymbol{r}) = \sum_\mu C_{\mu p}\chi_\mu(\boldsymbol{r}) \tag{2.39}$$

The task of HF is to then determine the expansion coefficients $C_{\mu p}$ to minimise the SD energy and best approximate the exact many-body ground state.

Restricted Hartree-Fock (RHF), the simplest variant of HF, further limits the space of SDs to be searched by requiring that the two spin species for each orbital are treated the same, enforcing that both spin species have identical spatial orbitals and are also always simultaneously occupied [23]:

$$|\phi_{p\alpha}(\boldsymbol{r})\rangle = |\phi_p(\boldsymbol{r})\rangle|\uparrow\rangle \qquad |\phi_{p\beta}(\boldsymbol{r})\rangle = |\phi_p(\boldsymbol{r})\rangle|\downarrow\rangle \tag{2.40}$$

$$|\psi_{\mathrm{RHF}}\rangle = |\det(\phi_{1\alpha}\phi_{1\beta}\phi_{2\alpha}\phi_{2\beta}\dots)\rangle \tag{2.41}$$

The minimisation task, equ. (2.38), for RHF can be expressed as the solution to the Roothaan-Hall equations, a generalised, i.e. $\mathbf{S} \neq \mathbf{I}$, non-linear, i.e. $\mathbf{f} = \mathbf{f}(\mathbf{C})$, eigenvalue equation for the Fock operator $\mathbf{f}$ [23]:

$$\mathbf{fC} = \mathbf{SC}\epsilon \tag{2.42}$$

where $\mathbf{C}$ is the matrix of MO expansion coefficients, $\mathbf{S}$ is the matrix of AO overlap

integrals $\langle\chi_\mu|\chi_\nu\rangle$, ϵ is a diagonal matrix of the orbital energies, and $\mathbf{f}$ is defined as:

$$f_{\mu\nu} = h_{\mu\nu} + \sum_{ij} P_{ij}\left(h_{\mu i\nu j} - \frac{h_{\mu ij\nu}}{2}\right) \tag{2.43}$$

$$\text{with} \quad P_{\mu\nu} = 2\sum_{a}^{N/2}\langle\chi_\mu|\phi_a\rangle\langle\phi_a|\chi_\nu\rangle = 2\sum_{a}^{N/2} C_{\mu a}C_{\nu a} \tag{2.44}$$

where the orbitals are assumed to be real. The non-linearity of these equations requires them to be solved in a self-consistent manner, most frequently via repeated diagonalisation until convergence is reached. In unrestricted HF (UHF) the spatial orbitals belonging to the two spin species are allowed to differ, and no double occupancy is enforced. Other common HF variants include restricted open-shell HF (ROHF), which enforces a shared spatial-orbital set while permitting open-shell occupations, and generalised HF (GHF), which allows spin mixing. However, while they differ in complexity, all HF variants aim to produce a set of single particle MOs that are deemed best within the specific restrictions imposed.

2.4.2 Post-HF Methods

While HF treats exchange exactly, its single-determinant form neglects most electron correlation beyond mean-field, motivating post-HF methods that systematically enlarge the variational space or introduce correlated parametrisations of the wavefunction [23]. In this sense, HF is not only a stand-alone mean-field approximation, but also a standard starting point that supplies both a reference determinant and an MO basis that enables systematic, hierarchical improvements towards the exact solution (within the given basis set).

In configuration interaction (CI) the wavefunction is expanded as a linear combination of excited determinants, e.g. in CI singles and doubles (CISD) [23]:

$$|\psi_{\text{CISD}}\rangle = c_0|\Phi_0\rangle + \sum_{ia} c_i^a|\Phi_i^a\rangle + \sum_{\substack{i<j\\a<b}} c_{ij}^{ab}|\Phi_{ij}^{ab}\rangle \tag{2.45}$$

where the reference determinant $|\Phi_0\rangle$ is typically taken as the HF solution, and the excited determinants $|\Phi_i^a\rangle$, $|\Phi_{ij}^{ab}\rangle$ are formed by promoting electrons from occupied orbitals $i, j, \ldots$ into virtual orbitals $a, b, \ldots$ of the HF MO basis. This highlights a central practical role of HF: once a set of orthonormal MOs has been obtained, it provides a convenient one-particle basis in which correlated methods can build controlled, particle–hole type expansions of the exact wavefunction. In the limit that all excitation ranks are included, CI becomes full CI (FCI), which is exact within

the chosen finite orbital basis. Closely related constructions underlie many other post-HF methods. In coupled cluster (CC) theory the correlated state is written as an exponential of excitation operators acting on the HF reference, e.g. in CCSD [23]:

$$|\psi_{\mathrm{CC}}\rangle = e^{\hat{T}}|\psi_{\mathrm{HF}}\rangle \quad \text{with} \quad \hat{T} = \sum_{ia} t_i^a \hat{c}_a^\dagger \hat{c}_i + \sum_{\substack{i<j \\ a<b}} t_{ij}^{ab} c_a^\dagger c_b^\dagger c_i c_j \tag{2.46}$$

with $\hat{T} = \hat{T}_1 + \hat{T}_2$ defined with respect to the same occupied/virtual partition.

2.4.3 WF Methods in Practice

Ultimately, all above mentioned methods suffer from the trade-off between accuracy and computational complexity. Wave-function based methods can be very accurate, but are generally also very computationally expensive. As such, they mainly find applications in the field of chemistry, in which relatively small molecules are considered and highly accurate energies are required. Despite continuous effort in the development of wavefunction based algorithms and the increase in computational power, they are still limited to small to moderately sized systems. Exact FCI calculations have been pushed to include up to 24 orbitals, expanding the wavefunction in more than one trillion SDs [38]. More sophisticated algorithms have been able to deal with systems including up to 77 orbitals [39]. Many large molecules, especially strongly correlated ones, however have remained intractable for high-accuracy calculations using wavefunction based methods.

2.5 Molecular Forces

Applications of electronic structure theory frequently require the evaluation of derivatives of the energy with respect to parameters of the Hamiltonian and/or system, since many interesting and important properties are defined in terms of the system's response to an applied perturbation. The practice of calculating analytic derivatives and response quantities in electronic structure theory has therefore become a commonplace and well studied discipline [21, 40, 41]. Electric dipole moments, for example, require evaluating derivatives with respect to an applied external electric field $\boldsymbol{\mu} = -\frac{dE(\boldsymbol{\mathcal{E}})}{d\boldsymbol{\mathcal{E}}}\Big|_{\boldsymbol{\mathcal{E}}=\mathbf{0}}$ since they quantify the system response to a small electric field perturbation $E(\boldsymbol{\mathcal{E}}) \approx E(\boldsymbol{\mathcal{E}} = 0) - \boldsymbol{\mu} \cdot \boldsymbol{\mathcal{E}}$. However, of particular importance in electronic structure theory, and the focus of our work, are the molecular forces, the derivatives of the energy with respect to the nuclear coordinates

$\boldsymbol{R}$:

$$\boldsymbol{F} \equiv -\frac{dE(\boldsymbol{R})}{d\boldsymbol{R}} \tag{2.47}$$

Accurate molecular forces are crucial to enable the calculation of equilibrium geometries, reaction pathways and vibrational frequencies of molecules [21]. While a direct evaluation of gradients using numerical differentiation/finite-difference approximations is possible, it requires $\mathcal{O}(N_p)$ independent perturbations and hence independent electronic structure calculations to estimate a response to N_p possible perturbations. Applied to molecular forces this incurs a cost $\mathcal{O}(3N_A)$ for a system of N_A atoms. The theory of analytic derivatives makes efficient evaluation of molecular forces possible at a generally negligible overhead compared to a single evaluation of $E(\boldsymbol{R})$. As such, the calculation of analytic molecular forces has become a well-studied area of research [21] and a commonplace method in numerical calculations [42].

In this section we review the theory behind the Hellmann-Feynman theorem, a well known and frequently applied method for calculating molecular forces. We also discuss the frequently encountered correction terms of Pulay and non-variational / response forces, which arise when the conditions required for applications of the HF theorem break, focusing in particular on expressions encountered when a Hartree-Fock molecular orbital basis is used. This section sets the context in which we will discuss the calculation of molecular forces in VQE calculations in sec. 5.

2.5.1 Hellmann–Feynman Theorem

The Hellmann-Feynman (HF) theorem is a well known result that allows one to greatly simplify the calculation of molecular forces and derivatives with respect to nuclear coordinates (as well as any other response quantities and energy derivatives) [23]. Given a Hamiltonian $\hat{H}(\boldsymbol{R})$ that depends on the nuclear positions $\boldsymbol{R}$, we can calculate the ground state $|\psi(\boldsymbol{R})\rangle$ and ground-state energy $E(\boldsymbol{R})$, where we stress that both the state and the energy depend on $\boldsymbol{R}$ as a result of the dependence $\hat{H}(\boldsymbol{R})$. A direct differentiation of the ground-state energy $E(\boldsymbol{R}) = \langle\psi(\boldsymbol{R})|\hat{H}(\boldsymbol{R})|\psi(\boldsymbol{R})\rangle$ leads to:

$$\frac{dE}{d\boldsymbol{R}} = \langle\frac{d\psi}{d\boldsymbol{R}}|\hat{H}|\psi\rangle + \langle\psi|\frac{d\hat{H}}{d\boldsymbol{R}}|\psi\rangle + \langle\psi|\hat{H}|\frac{d\psi}{d\boldsymbol{R}}\rangle \tag{2.48}$$

Since $|\psi(\boldsymbol{R})\rangle$ is the ground state, we know that $\hat{H}(\boldsymbol{R})|\psi(\boldsymbol{R})\rangle = E(\boldsymbol{R})|\psi(\boldsymbol{R})\rangle$ (as well as the hermitian conjugate of this equation), and since we further assume it to

be normalised, $\langle\psi(\boldsymbol{R})|\psi(\boldsymbol{R})\rangle = 1$, the gradient simplifies:

$$\frac{dE}{d\boldsymbol{R}} = \langle\psi|\frac{d\hat{H}}{d\boldsymbol{R}}|\psi\rangle + E(\boldsymbol{R})\frac{d}{d\boldsymbol{R}}\langle\psi|\psi\rangle = \langle\psi|\frac{d\hat{H}}{d\boldsymbol{R}}|\psi\rangle \tag{2.49}$$

The above result shows that the evaluation of first order gradients does not explicitly require the calculation and inclusion of the wavefunction response $|\frac{d\psi}{d\boldsymbol{R}}\rangle$. Applied to the nuclear gradient, we find that the molecular forces can be calculated by evaluating the expectation of a suitably defined HF force observable on the ground-state wavefunction:

$$\boldsymbol{F} = \langle\psi|\hat{\boldsymbol{F}}|\psi\rangle \qquad \text{with} \qquad \hat{\boldsymbol{F}} \equiv -\frac{d\hat{H}}{d\boldsymbol{R}} \tag{2.50}$$

The HF theorem is a direct result of the variational principle for exact ground states discussed in sec. 2.2, since to first order a perturbation $|\psi\rangle \to |\psi\rangle + |\delta\psi\rangle$ does not affect the energy.

2.5.2 Approximate Hellmann-Feynman Theorem

Direct application of the HF theorem is limited due to its assumption that the energy is evaluated at the exact ground state of the system. In realistic calculations this is practically never the case, since the wavefunction accuracy is limited by the computational cost through e.g. the energy cut-off E_{cut} or the basis set size N_b. However, similarly to the variational principle, the HF theorem can be shown to also apply to inexact wavefunctions provided they are variationally optimised [23]. Such a wavefunction solution fulfils the stationary condition:

$$\langle\delta\psi|\hat{H}|\psi\rangle + \langle\psi|\hat{H}|\delta\psi\rangle = 0 \tag{2.51}$$

which, when applied to the perturbation induced by a variation $\boldsymbol{R} \to \boldsymbol{R} + \delta\boldsymbol{R}$, $\delta\psi = \frac{d\psi}{d\boldsymbol{R}} \cdot \delta\boldsymbol{R}$, allows us to conclude that the HF theorem still holds, since for any $\delta\boldsymbol{R}$ we find:

$$\left(\langle\frac{d\psi}{d\boldsymbol{R}}|\hat{H}|\psi\rangle + \langle\psi|\hat{H}|\frac{d\psi}{d\boldsymbol{R}}\rangle\right) \cdot \delta\boldsymbol{R} = 0 \tag{2.52}$$

and hence that:

$$\frac{dE}{d\boldsymbol{R}} = \langle\psi|\frac{d\hat{H}}{d\boldsymbol{R}}|\psi\rangle + \left[\langle\frac{d\psi}{d\boldsymbol{R}}|\hat{H}|\psi\rangle + \langle\psi|\hat{H}|\frac{d\psi}{d\boldsymbol{R}}\rangle\right] = \langle\psi|\frac{d\hat{H}}{d\boldsymbol{R}}|\psi\rangle \tag{2.53}$$

The (approximate) HF theorem is not limited to energy functionals of the form $\langle\psi|\hat{H}|\psi\rangle$. A common framework to discuss energy derivatives and the HF theorem

which puts exact and approximate wavefunctions on more equal footing considers the general energy functional $E(\boldsymbol{\theta};\boldsymbol{R})$ that depends on the nuclear positions $\boldsymbol{R}$ (e.g. through the Hamiltonian dependence $\hat{H}(\boldsymbol{R})$) and a set of parameters $\boldsymbol{\theta}$ (e.g. parametrising the wavefunction $|\psi\rangle = |\psi(\boldsymbol{\theta})\rangle$) [43]. The ground state is specified by a set of solution parameters θ_* that will in general depend on the specified molecular geometry, $\boldsymbol{\theta}_* = \boldsymbol{\theta}_*(\boldsymbol{R})$, and are determined by optimising the energy with respect to $\boldsymbol{\theta}$ such that the stationary condition holds:

$$\left.\frac{\partial E(\boldsymbol{\theta};\boldsymbol{R})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_*} \equiv \frac{\partial E(\boldsymbol{\theta}_*;\boldsymbol{R})}{\partial \boldsymbol{\theta}} = \mathbf{0} \tag{2.54}$$

A direct application of the chain rule and equ. (2.54) allows us to re-derive the HF theorem once we identify $E(\boldsymbol{\theta}_*,\boldsymbol{R}) \to \langle\psi(\boldsymbol{\theta}_*)|\hat{H}(\boldsymbol{R})|\psi(\boldsymbol{\theta}_*)\rangle$:

$$\frac{dE(\boldsymbol{\theta}_*;\boldsymbol{R})}{d\boldsymbol{R}} = \frac{\partial E(\boldsymbol{\theta}_*;\boldsymbol{R})}{\partial \theta_i}\frac{\partial \theta_{*,i}}{\partial \boldsymbol{R}} + \frac{\partial E(\boldsymbol{\theta}_*;\boldsymbol{R})}{\partial \boldsymbol{R}} = \frac{\partial E(\boldsymbol{\theta}_*;\boldsymbol{R})}{\partial \boldsymbol{R}} = \langle\psi|\frac{d\hat{H}}{d\boldsymbol{R}}|\psi\rangle \tag{2.55}$$

2.5.3 Pulay Forces

A common way in which the assumptions underlying the HF theorem can fail in subtle ways is when atom-centred basis sets are being used. Perturbing the nuclear positions perturbs not only the nuclei and the electrostatic potential they create, but also the location of the basis functions used to construct the wavefunction. Even if at a fixed geometry $\boldsymbol{R}$ the wavefunction is fully variationally optimised, varying $\boldsymbol{R} \to \boldsymbol{R} + \delta\boldsymbol{R}$ causes the variational space itself to vary. A wavefunction expressed in terms of the basis at $\boldsymbol{R}$ can therefore in general not be expressed exactly using the basis at $\boldsymbol{R} + \delta\boldsymbol{R}$. As a result the ground-state wavefunction perturbation induced by $\boldsymbol{R} \to \boldsymbol{R} + \delta\boldsymbol{R}$ cannot be described within the variational space at $\boldsymbol{R}$ only so that in general:

$$\left(\langle\frac{d\psi}{d\boldsymbol{R}}|\hat{H}|\psi\rangle + \langle\psi|\hat{H}|\frac{d\psi}{d\boldsymbol{R}}\rangle\right)\cdot\delta\boldsymbol{R} \neq 0 \tag{2.56}$$

and the HF theorem does in general not hold [23]. In the more general framework of the energy functional $E(\boldsymbol{\theta},\boldsymbol{R})$ this breaking of the HF theorem is captured by recognising that a moving/variable basis set will similarly prevent us from expressing the wavefunction perturbation $|\psi\rangle + |\delta\psi\rangle$ due to $\boldsymbol{R} \to \boldsymbol{R} + \delta\boldsymbol{R}$ in terms of a perturbation of the variational parameters $\boldsymbol{\theta} \to \boldsymbol{\theta} + \delta\boldsymbol{\theta}$. An explicit dependence on $\boldsymbol{R}$ is therefore necessary, $|\psi\rangle = |\psi(\theta_*(\boldsymbol{R}),\boldsymbol{R})\rangle$, which prevents the identification $\frac{\partial E(\boldsymbol{\theta}_*;\boldsymbol{R})}{\partial \boldsymbol{R}} \to \langle\psi|\frac{d\hat{H}}{d\boldsymbol{R}}|\psi\rangle$ and hence the application of the HF theorem. Any additional terms that arise due to a variable basis set (and hence

vanish when the basis is stationary) are commonly referred to as Pulay forces $\boldsymbol{F} = -\langle\psi|\frac{d\hat{H}}{d\boldsymbol{R}}|\psi\rangle + \boldsymbol{F}_{\text{Pulay}} = \boldsymbol{F}_{\text{HF}} + \boldsymbol{F}_{\text{Pulay}}$, the importance of which was first noted in Pulay [44]. Evaluation of Pulay forces has become a standard task and is widely available in electronic structure software packages [42]. However, the exact form of and method for evaluating $\boldsymbol{F}/\boldsymbol{F}_{\text{Pulay}}$ varies depending on the electronic structure method and wavefunction representation used. In DFT, in which a real space representation is natural, explicit Pulay density response terms need to be included [45]:

$$\boldsymbol{F}_{\text{Pulay}} = -\int \mathrm{d}\boldsymbol{r} \frac{\delta F[n]}{\delta n(\boldsymbol{r})} \frac{\partial n(\boldsymbol{r})}{\partial \boldsymbol{R}}, \tag{2.57}$$

However, since most quantum chemistry methods are naturally presented in second quantisation, in this case the wavefunction response can be formally accounted for within the Hamiltonian matrix element derivatives, since the variational space is usually defined as (a fixed subset of) the m-orbital Fock space and moving basis functions simply lead to a variation of the matrix elements, but not the variational space itself [23]. The Hellmann-Feynman theorem therefore still holds, provided the Hamiltonian derivative is identified as:

$$\frac{d\hat{H}}{d\boldsymbol{R}} = \sum_{pq} \frac{dh_{pq}}{d\boldsymbol{R}} \hat{c}_p^\dagger \hat{c}_q + \frac{1}{2} \sum_{pqrs} \frac{dh_{pqrs}}{d\boldsymbol{R}} \hat{c}_p^\dagger \hat{c}_q^\dagger \hat{c}_r \hat{c}_s \tag{2.58}$$

where $\frac{dh_{pq}}{d\boldsymbol{R}}$ is calculated from equ. (2.15) with the inclusion of "orbital-following" as:

$$\frac{dh_{pq}}{d\boldsymbol{R}} = \langle\phi_p|\frac{d\hat{h}}{d\boldsymbol{R}}|\phi_q\rangle + \langle\frac{d\phi_p}{d\boldsymbol{R}}|\hat{h}|\phi_q\rangle + \langle\phi_p|\hat{h}|\frac{d\phi_q}{d\boldsymbol{R}}\rangle \tag{2.59}$$

(and similarly for $\frac{dh_{pqrs}}{d\boldsymbol{R}}$) and can hence be seen to implicitly account for the wavefunction response explicitly required in a real space representation. The above expression holds provided that the perturbation-dependent orbitals are chosen to be orthonormal $\langle\phi_p(\boldsymbol{R})|\phi_q(\boldsymbol{R})\rangle = I_{pq}$ at any $\boldsymbol{R}$. In most practical quantum chemistry calculations orthonormal molecular orbitals (MO) $|\chi_p\rangle$ are constructed from non-orthonormal atomic orbitals (AO) $|\phi_\mu\rangle$ using e.g. Hartree-Fock or Kohn-Sham calculations (see sec. 2.3 and 2.4). Such a construction introduces additional terms resulting from the orthonormalisation requirement for the molecular orbitals, which can be expressed in terms of the familiar correction terms involving the atomic

orbital overlap $S_{\mu\nu} \equiv \langle \phi_\mu | \phi_\nu \rangle$ as [43]:

$$\begin{aligned} \frac{dh_{pq}}{d\boldsymbol{R}} &= h^{(\boldsymbol{R})}_{pq} - \frac{1}{2}\sum_r \left(h_{rq}\, S^{(\boldsymbol{R})}_{rp} + h_{pr}\, S^{(\boldsymbol{R})}_{rq} \right) \\ \frac{dg_{pqrs}}{d\boldsymbol{R}} &= g^{(\boldsymbol{R})}_{pqrs} - \frac{1}{2}\sum_t \left(g_{tqrs}\, S^{(\boldsymbol{R})}_{tp} + g_{ptrs}\, S^{(\boldsymbol{R})}_{tq} + g_{pqts}\, S^{(\boldsymbol{R})}_{tr} + g_{pqrt}\, S^{(\boldsymbol{R})}_{ts} \right) \end{aligned} \tag{2.60}$$

where a superscript $(\boldsymbol{R})$ denotes differentiation at fixed molecular orbital coefficients $C_{\mu p}$, i.e.:

$$\begin{aligned} h^{(\boldsymbol{R})}_{pq} &= \sum_{\mu v} C_{\mu p} C_{vq} \frac{dh_{\mu v}}{d\boldsymbol{R}} \\ g^{(\boldsymbol{R})}_{pqrs} &= \sum_{\mu v \sigma \rho} C_{\mu p} C_{vq} C_{\sigma r} C_{\rho s} \frac{dg_{\mu v \sigma \rho}}{d\boldsymbol{R}} \\ S^{(\boldsymbol{R})}_{pq} &= \sum_{\mu v} C_{\mu p} C_{vq} \frac{dS_{\mu v}}{d\boldsymbol{R}} \end{aligned} \tag{2.61}$$

and we have assumed real atomic orbitals (and hence overlaps) and real molecular orbital coefficients. These correction terms by themselves are frequently referred to as the Pulay force terms in the quantum chemistry literature. Equ. (2.60) also allows for a direct evaluation using well established AO electronic and overlap integral derivatives without incurring any $\mathcal{O}(N_A)$ overhead compared to an evaluation of the electronic integrals/energy functional [43, 46]. In Hartree-Fock calculations the overlap integral correction assumes the well known form [22]:

$$\boldsymbol{F}_{\text{Pulay}} = \sum_{\mu\nu} Q_{\nu\mu} \frac{\partial S_{\mu v}}{\partial \boldsymbol{R}} \tag{2.62}$$

expressed in terms of the energy-weighted density matrix $Q_{\nu\mu} = \sum_a^N \varepsilon_a C_{\mu a} C_{\nu a}$. It is well established that Pulay-forces vanish in the limit of a complete basis set, since all variational spaces at any configuration $\boldsymbol{R}$ become equivalent, as a result of the fact that an equivalent stationary atomic orbital basis can be defined consistently for all $\boldsymbol{R}$ [43]. However, Pulay forces will in general not vanish in the FCI limit as long as the basis set is finite/incomplete and perturbation dependent.

2.5.4 Non-variational/Response Forces

Apart from Pulay forces, additional (non Hellmann-Feynman) forces can arise as a result of a wavefunctions that is explicitly non-variational with respect to the perturbation, requiring an explicit inclusion of wavefunction response terms. A general electronic structure method is specified by an energy functional $E(\boldsymbol{\theta}, \boldsymbol{\Theta}; \boldsymbol{R})$, where

the ground state is determined by variationally optimising only some of the parameters, $\frac{\partial E(\boldsymbol{\theta}_*)}{\partial \boldsymbol{\theta}} = 0$, while other parameters are determined non-variationally based on other more general conditions, $\boldsymbol{f}(\boldsymbol{\Theta}_*; \boldsymbol{R}) = 0$. The non-variational parameters $\boldsymbol{\Theta}_* = \boldsymbol{\Theta}_*(\boldsymbol{R})$ prevent a direct application of the (approximate) HF theorem and require inclusion and evaluation of the response term $\frac{\partial \Theta_{*,i}}{\partial \boldsymbol{R}}$:

$$\frac{dE(\boldsymbol{\theta}_*, \boldsymbol{\Theta}_*; \boldsymbol{R})}{d\boldsymbol{R}} = \frac{\partial E(\boldsymbol{\theta}_*, \boldsymbol{\Theta}_*; \boldsymbol{R})}{\partial \Theta_i} \frac{\partial \Theta_{*,i}}{\partial \boldsymbol{R}} + \frac{\partial E(\boldsymbol{\theta}_*, \boldsymbol{\Theta}_*; \boldsymbol{R})}{\partial \boldsymbol{R}} \tag{2.63}$$

Similar to Pulay forces, evaluation of the non-variational/response forces is a well-established and common task in electronic structure software packages [42]. A particularly common example are the Hartree-Fock response forces, which are encountered in post-Hartree-Fock methods which do not re-optimise the MOs. In standard CI, for example, the wavefunction/energy functional is parametrised by the MO coefficients $\boldsymbol{C}$ and the CI amplitudes $\boldsymbol{c}$. The MO coefficients $\boldsymbol{C}_*$ are determined by performing a Hartree-Fock calculation, minimising the Hartree-Fock energy $E_{\mathrm{HF}}(\boldsymbol{C}; \boldsymbol{R})$ by solving $\frac{\partial E_{\mathrm{HF}}}{\partial \boldsymbol{C}} = 0$. The optimal CI amplitudes $\boldsymbol{c}_*$ are then solved for by variationally optimising the CI energy $E_{\mathrm{CI}}(\boldsymbol{c}, \boldsymbol{C}; \boldsymbol{R})$ such that $\frac{\partial E_{\mathrm{CI}}}{\partial \boldsymbol{c}} = 0$. While the MOs do optimise E_{HF}, they will in general not correspond to a variational minimum of the "full" energy E_{CI} since E_{CI} is optimised with fixed MOs, $\boldsymbol{C}_*$. Molecular forces in CI therefore require evaluation of the MO coefficient response induced by the perturbation, $\frac{\partial C_{*,i}}{\partial \boldsymbol{R}}$, identifying $\boldsymbol{\Theta} \to \boldsymbol{C}$. The MO response can be obtained as a solution to the well-known coupled-perturbed Hartree-Fock equations given by [43]:

$$\frac{\partial^2 E_{\mathrm{HF}}(\boldsymbol{C}_*; \boldsymbol{R})}{\partial C_i \partial C_j} \frac{\partial C_{*,j}}{\partial \boldsymbol{R}} = -\frac{\partial^2 E_{\mathrm{HF}}(\boldsymbol{C}_*; \boldsymbol{R})}{\partial C_i \, \partial \boldsymbol{R}} \tag{2.64}$$

which result from differentiating the Hartree-Fock condition $\frac{\partial E_{\mathrm{HF}}}{\partial \boldsymbol{C}} = 0$ used to solve for $\boldsymbol{C}_*$, with respect to the perturbation, $\frac{d}{d\boldsymbol{R}}(\frac{\partial E_{\mathrm{HF}}}{\partial \boldsymbol{C}}) = 0$. Non-variational response terms are generally costly to evaluate; while a single response equation can usually be solved at a cost equivalent to a single evaluation of the energy E_{HF} (or equivalent energy functionals for other methods) by employing iterative Krylov subspace methods, the full response requires the solution of an independent equation for each perturbation, and hence scales linearly with the number of perturbations – $\mathcal{O}(3N_A)$ for molecular forces. A common approach for deriving relatively simple and computationally more efficient expressions for analytic gradients in non-variational methods is the Lagrangian/Lagrange multiplier approach. A fully variational Lagrangian

is constructed by introducing Lagrange multipliers $\tilde{\boldsymbol{\Theta}}$:

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\Theta}, \tilde{\boldsymbol{\Theta}}; \boldsymbol{R}) = E(\boldsymbol{\theta}, \boldsymbol{\Theta}; \boldsymbol{R}) + \tilde{\Theta}_i f_i(\boldsymbol{\Theta}; \boldsymbol{R}) \tag{2.65}$$

The ground state, $(\boldsymbol{\theta}_*, \boldsymbol{\Theta}_*)$, corresponds to the (fully variational) minimum $\mathcal{L}_* \equiv \mathcal{L}(\boldsymbol{\theta}_*, \boldsymbol{\Theta}_*, \tilde{\boldsymbol{\Theta}}_*; \boldsymbol{R})$ of $\mathcal{L}$, defined by $\frac{\partial \mathcal{L}_*}{\partial \boldsymbol{\theta}} = \frac{\partial \mathcal{L}_*}{\partial \boldsymbol{\Theta}} = \frac{\partial \mathcal{L}_*}{\partial \tilde{\boldsymbol{\Theta}}} = 0$. Since at the ground state $\mathcal{L}_* = E_*$, and since the full stationarity of $\mathcal{L}$ implies $\frac{d\mathcal{L}_*}{d\boldsymbol{R}} = \frac{\partial \mathcal{L}_*}{\partial \boldsymbol{R}}$, the analytic gradient can be identified as:

$$\frac{dE_*}{d\boldsymbol{R}} = \frac{\partial \mathcal{L}_*}{\partial \boldsymbol{R}} = \frac{\partial E_*}{\partial \boldsymbol{R}} + \tilde{\Theta}_{*,i} \frac{\partial f_{*,i}}{\partial \boldsymbol{R}} \tag{2.66}$$

The Lagrange multipliers $\tilde{\boldsymbol{\Theta}}_*$ are determined from the stationary condition $\frac{\partial \mathcal{L}}{\partial \tilde{\boldsymbol{\Theta}}} = 0$, which requires solving only a single response equation:

$$\frac{\partial f_{*,j}}{\partial \Theta_i} \tilde{\Theta}_{*,j} = -\frac{\partial E_*}{\partial \Theta_i} \tag{2.67}$$

The computational cost of evaluating $\tilde{\boldsymbol{\Theta}}_*$ therefore scales independently of the number of perturbations. While $\mathcal{O}(3N_A)$ evaluations of $\frac{\partial f_{*,i}}{\partial \boldsymbol{R}}$ are still required, they appear in equ. (2.66) as a vector product with $\tilde{\Theta}_{*,j}$, and can usually be evaluated at a negligible cost and independent of the number of perturbations (e.g. [47]). As such, the Lagrangian approach generally allows for evaluation of molecular forces at a cost approximately equivalent to the evaluation of the energy itself. As a specific example, applied to the case of CI calculations the analytic gradient is given by [43]:

$$\frac{dE_{\mathrm{CI}}}{d\boldsymbol{R}} = \frac{\partial E_{\mathrm{CI}}}{\partial \boldsymbol{R}} + \tilde{C}_i \frac{\partial^2 E_{\mathrm{HF}}}{\partial C_i \, \partial \boldsymbol{R}} \quad \text{with} \quad \frac{\partial^2 E_{\mathrm{HF}}}{\partial C_i \partial C_j} \tilde{C}_j = -\frac{\partial E_{\mathrm{CI}}}{\partial C_i} \tag{2.68}$$

In the same way that Pulay forces vanish in the complete basis set limit, the non-variational/response forces will vanish in the FCI limit, as any perturbation of the non-variational parameters, $\boldsymbol{\Theta} \to \boldsymbol{\Theta} + \delta\boldsymbol{\Theta}$ can be described by/absorbed into a perturbation of the variational parameters, $\boldsymbol{\theta} \to \boldsymbol{\theta} + \delta\boldsymbol{\theta}$ (e.g. [47]).

2.6 Quantum Computation

A quantum computer is a device that manipulates a quantum mechanical system to store and process information. Over the past decades many proposals have been made for realisations of quantum computers based on varying underlying principles. The most promising and most frequently considered type of device in the current NISQ era of quantum computing is the digital/gate-based quantum computer based on qubit systems [6]. It constructs a quantum system out of a register of individual

qubits and allows the quantum system to be manipulated through implementations of gates between subsets of qubits to implement arbitrary unitary evolutions of the system. Information is extracted from the quantum computer by performing measurements on individual qubits. Such a device is frequently referred to as a quantum processing unit (QPU). Contemporary QPUs implement qubits, quantum gates and measurements using a variety of physical hardware platforms, most prominently superconducting circuits and trapped ions, with other approaches such as neutral atoms, photonics, and semiconductor spin qubits also being actively developed. In this section we briefly review the theory underlying such a qubit-based digital quantum computer.

2.6.1 Qubits

A qubit is a two-level quantum system with a basis of states frequently denoted by $\{|0\rangle, |1\rangle\}$, which are referred to as computational basis states [48]. Its Hilbert space is given by $\mathcal{H} = \mathbb{C}^2$ and pure states of the system, $|\psi\rangle \in \mathcal{H}$, can be written as

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \qquad \text{with} \qquad \alpha, \beta \in \mathbb{C} \qquad \text{and} \qquad |\alpha|^2 + |\beta|^2 = 1. \tag{2.69}$$

Unlike a classical bit, which is either 0 or 1, quantum mechanics allows the qubit to occupy a coherent superposition of basis states. A quantum computer containing N such qubits can be described by states of Hilbert space $\mathcal{H} = (\mathbb{C}^2)^{\otimes N} \simeq \mathbb{C}^{2N}$, where in accordance with the standard theory of quantum mechanics Hilbert spaces are combined into composite systems using the tensor product, $\mathcal{H}_1 \otimes \mathcal{H}_2$. The computational basis states similarly consist of tensor products of individual computational basis states, $|n\rangle \equiv |n_0\rangle \otimes |n_1\rangle \cdots \otimes |n_{N-1}\rangle$ with $n_i \in \{0, 1\}$ and hence $n \in \{0, 1\}^N$. A general state is given by [3]:

$$|\psi\rangle = \sum_{n \in \{0,1\}^N} c_n |n\rangle \qquad \text{with} \qquad \sum_n |c_n|^2 = 1. \tag{2.70}$$

The tensor-product structure enables entanglement, i.e. correlations that cannot be represented as products of single-qubit states, and which play a central role in quantum algorithms.

[3] We note the use of zero indexing, a big endian convention in which the first qubit corresponds to the left-most index, and a standard enumeration scheme of basis states according to their binary value, i.e. $|00\rangle$, $|01\rangle$, $|10\rangle$, $|11\rangle$

2.6.2 Gates

Since the quantum computer is a quantum mechanical system, the many-body state of the quantum computer, $|\psi\rangle$, can be manipulated to perform computations by applying unitary evolutions, $\hat{U}$. In the gate model, $\hat{U}$ is implemented by applying a sequence of quantum gates to a small subset of qubits, most frequently one and two qubit gates [48]. If the register is in state $|\psi\rangle$, the application of a gate $\hat{U}$ yields:

$$|\psi\rangle \;\mapsto\; \hat{U}|\psi\rangle \tag{2.71}$$

A particularly important set of single-qubit operators are the Pauli matrices, X, Y and Z:

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{2.72}$$

as well as the single-qubit rotations generated by them, e.g. $R_Z(\theta) = \exp(-i\theta\hat{Z}/2)$. Another frequently encountered single-qubit gate is the Hadamard gate, H:

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \tag{2.73}$$

on account of its basis changing action $HZH = X$. The prototypical two-qubit entangling gate is the controlled-NOT (CNOT / CX), flipping qubit 1 controlled on the state of qubit 0:

$$\text{CNOT / CX} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{2.74}$$

Arbitrary unitaries can be implemented if the quantum device supports a universal gate set, allowing any U to be approximated to any desired accuracy by an appropriate digital quantum circuit [48].

2.6.3 Measurements

The many-body wavefunction and any information "stored" within it can be accessed by measurements on the state in accordance with the postulates of quantum mechanics. Digital quantum computers can measure individual qubits to produce classical bit strings, usually at the end of a circuit but potentially also during execution, by performing projective measurements on the computational basis states

[48]:

$$\hat{\Pi}_0 = |0\rangle\langle 0| \qquad \hat{\Pi}_1 = |1\rangle\langle 1| \tag{2.75}$$

If the measured qubit is in state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, the Born rule gives

$$p(0) = \langle\psi|\hat{\Pi}_0|\psi\rangle = |\alpha|^2 \qquad p(1) = \langle\psi|\hat{\Pi}_1|\psi\rangle = |\beta|^2 \tag{2.76}$$

and the post-measurement state is projected onto the corresponding eigenstate. For multi-qubit registers, measuring all qubits in the computational basis yields a bit string $x \in \{0,1\}^n$ with probability $p(x) = |c_x|^2$ for $|\psi\rangle = \sum_n c_n|n\rangle$. Such projective measurement on the computational basis correspond to measurements of Pauli-Z observables and can be directly used to estimate their expectation value.

2.6.4 Hardware Constraints

Ideal circuit descriptions frequently assume a specific set of implementable gates and that any gate may be applied between any pair of qubits. Realistic devices are frequently restricted to a specific set of native gates, which are based on natural and easy to implement gates in the specific device architecture. Similarly, specific device architectures may only allow a direct implementation of gates between subsets of qubits based on the physical arrangement of the qubits on the device. Quantum circuits must therefore be compiled into the available operations and routed in accordance with the device connectivity graph by inserting additional gates (most commonly SWAP gates).

Chapter 3

Modelling *Trans*-polyacetylene

3.1 TPA and Conducting Polymers

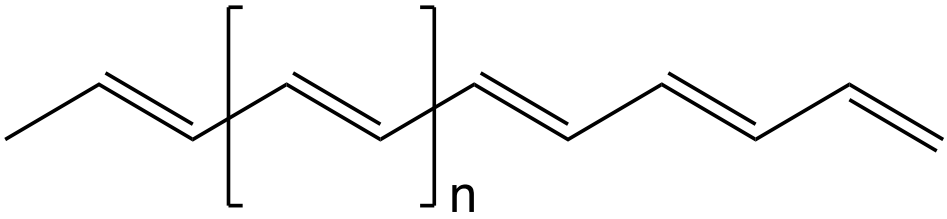


Figure 3.1: Skeletal structure of *trans*-polyacetylene: A periodic, quasi-linear chain of carbon atoms with alternating single and double bonds

The discovery of *trans*-polyacetylene (TPA) and the investigation of its high conductivity when doped created the field of conducting polymers and was awarded the Nobel Prize in 2000 [49–51]. In one dimension any metallic system is unstable against a distortion, a result known as Peierls instability [52]. As a result these polymers – quasi-one-dimensional systems – are generally semi-conducting in the ground state. Similar to crystalline semiconductors, their conductivity is highly tunable through doping with either reducing or oxidising agents. They combine this desirable electronic property of semiconductors with the high processability and desirable mechanical properties of polymers. Conducting polymers have found a variety of applications including flexible electronics, solar cells and organic light emitting diodes [53]. TPA in particular has only found limited use in industry since its insolubility in most solvents and instability when exposed to air make it unsuitable for many industrial applications [54]. In spite of this it has been the focus of extensive research since, as the simplest but simultaneously prototypical example, it provides a great opportunity for studying and better understanding the physics underlying the behaviour and properties of conducting polymers [55].

Like many other conducting polymers, TPA is a conjugated (i.e. dimerised) system, and consists of a backbone of carbon atoms with alternating short (double) and long (single) bonds (see fig. 3.1). It's a semi-conducting material in which delocalisation of the π-bonds allows conduction of electricity upon doping. The symmetric structure allows TPA to form two configurations degenerate in energy. Boundaries between the two phases are associated with defect states known as "solitons", quasiparticles with a wide variety of interesting physical properties. They display reversed spin-charge relations, are highly mobile but can also be pinned by impurities, and exhibit pair confinement in the presence of Coulomb interactions. As they are the lowest lying excited states they have been attributed to directly contribute to a lot of the macroscopic properties observed in TPA [55].

3.2 The Su–Schrieffer–Heeger Model

Many relatively simple systems have been modelled successfully using tight-binding Hamiltonians. In the case of *trans*-polyacetylene (TPA) the Su-Schrieffer-Heeger (SSH) model has been shown to appropriately model and explain many key electronic properties of the system [56, 57], the most important of which for our work we will reproduce here. However, the model has recently also found applications in the study of symmetry protected topological phases, as it is the simplest example of a topological insulator [58].

3.2.1 The Tight-binding Model

The central approximation in this model is to treat the π electrons explicitly in a nearest-neighbour tight-binding picture (one p_z orbital per carbon), while the σ electrons are incorporated implicitly via an elastic energy associated with bond distortions. The p orbitals are centred on each carbon atom and are assumed to be bound tightly enough that only nearest neighbour hoppings $t_{n,n+1}$ have to be considered. The σ electrons are not treated explicitly; instead each σ bond is assumed to contribute an energy E_σ to the system. Both $t_{n,n+1}$ and E_σ will depend on the amount of dimerisation present as both depend on the separation between neighbouring carbon atoms. The main coordinate driving the dimerisation is u_n, the displacement of the n-th carbon atom (and corresponding orbital) along the chain axis (relative to an undistorted geometry). Assuming small displacements around the equilibrium geometry determined by the σ-bond structure, $t_{n,n+1}$ and E_σ can be

expanded to the lowest order in the bond-length change:

$$t_{n,n+1} \approx t + \alpha\left(u_{n+1} - u_n\right) \qquad E_\sigma \approx \frac{K}{2}\left(u_{n+1} - u_n\right)^2 \tag{3.1}$$

Accounting for the (classical) kinetic energy associated with the motion of the nuclei in each carbon-hydrogen site in the system, a necessary inclusion in case of a time-dependent distortion, yields the full SSH Hamiltonian:

$$\begin{aligned}\hat{H}_{\text{SSH}} = &-\sum_{i=0}^{N}\left(t + \alpha\left(u_{i+1} - u_i\right)\right)\left(\hat{c}_i^\dagger \hat{c}_{i+1} + \hat{c}_{i+1}^\dagger \hat{c}_i\right) \\ &\left[+\sum_{i=0}^{N}\frac{K}{2}\left(u_{i+1} - u_i\right)^2 + \sum_{i=0}^{N}\frac{M_{\text{CH}}}{2}\dot{u}_i^2\right]\end{aligned} \tag{3.2}$$

where $c_i^\dagger$ and c_i correspond to creation and annihilation operators for the p-orbital on the i-th carbon atom. In our work we will be concerned with steady-state solutions, $\dot{u} = 0$ and note that (similar to V_{nn} in the electronic structure Hamiltonian, equ. (2.5)) the elastic energy $\frac{K}{2}\left(u_{i+1} - u_i\right)^2$ corresponds to a constant energy shift of all eigenstates and can simply be accounted for at the end of any quantum mechanical calculation.

3.2.2 Bloch States

Two limiting cases are of particular interest: the undistorted chain $u_n = 0$ and a perfectly dimerised chain $u_n = (-1)^n u$. In both cases the electronic Hamiltonian is a periodic, independent particle tight-binding model of the form:

$$\hat{H} = \sum_{\boldsymbol{R},\boldsymbol{R}',i,j} H_{ij}(\boldsymbol{R} - \boldsymbol{R}')c_{i,\boldsymbol{R}}^\dagger c_{j,\boldsymbol{R}'} \tag{3.3}$$

where i and j label orbitals within the same unit cell (UC) and $\boldsymbol{R}$ and $\boldsymbol{R}'$ label the position of each unit cell. As a result of Bloch's theorem they can be block diagonalised using a basis of Bloch states, $|\phi_{\boldsymbol{k}}\rangle$, eigenstates with well defined crystal momentum $\boldsymbol{k}$ under translation by a lattice site $\boldsymbol{R}$, $\hat{T}_{\boldsymbol{R}}|\phi_{\boldsymbol{k}}\rangle = e^{i\boldsymbol{k}\cdot\boldsymbol{R}}|\phi_{\boldsymbol{k}}\rangle$:

$$\hat{c}_{i,\boldsymbol{k}}^\dagger = \sum_{\boldsymbol{R}} e^{i\boldsymbol{k}\cdot\boldsymbol{R}}\hat{c}_{i,\boldsymbol{R}}^\dagger \qquad\qquad \hat{c}_{i,\boldsymbol{R}}^\dagger = \frac{V_{\text{UC}}}{(2\pi)^3}\int_{BZ} e^{-i\boldsymbol{k}\cdot\boldsymbol{R}}\hat{c}_{i,\boldsymbol{k}}^\dagger \, d\boldsymbol{k} \tag{3.4}$$

where $\boldsymbol{k} \in \mathrm{BZ}$ is restricted to lie in the first Brillouin Zone (BZ). In the Bloch basis the Hamiltonian takes the form:

$$\hat{H} = \sum_{\boldsymbol{k},i,j} H_{ij}(\boldsymbol{k}) c^{\dagger}_{i,\boldsymbol{k}} c_{j,\boldsymbol{k}} \qquad \text{with} \qquad H_{ij}(\boldsymbol{k}) = \sum_{\boldsymbol{R}} e^{i\boldsymbol{k}\cdot\boldsymbol{R}} H_{ij}(\boldsymbol{R}) \tag{3.5}$$

The states at different wavevectors can be explicitly seen to be completely decoupled and the system can be solved by diagonalising the momentum-space single-particle matrix at fixed $\mathbf{H}(\boldsymbol{k})$ for fixed $\boldsymbol{k}$, an $n \times n$ matrix for a system containing n orbitals per unit cell.

3.2.3 Dimerisation and Peierls Distortion

In the completely undimerised case the system is periodic about translation by a single lattice site and the formation of Bloch states, as required by translational symmetry, completely determines the eigenstates of the system. Its bandstructure consists of a single half-filled band with dispersion:

$$E(k) = -2t\cos(kd) \qquad -\frac{\pi}{d} < k \leq \frac{\pi}{d} \tag{3.6}$$

In the perfectly dimerised state, $u_n = (-1)^n u$ the periodicity of the underlying lattice is halved compared to the undimerised state, doubling the unit cell to contain two sites with a lattice constant $a = 2d$. The real-space Hamiltonian for this case takes the form:

$$\hat{H} = -\sum_{n} \left(v \cdot c^{\dagger}_{n,A} c_{n,B} + w \cdot c^{\dagger}_{n,B} c_{n+1,A} \right) + \text{h.c.} \tag{3.7}$$

where the two sublattices of the system have been made explicit in the subscripts A and B and where we have defined $v = t + \alpha u$ and $w = t - \alpha u$. We note that this specialised, perfectly dimerised Hamiltonian is sometimes also referred to as the SSH model. As a result the Brillouin zone is reduced to the range $-\frac{\pi}{2d} < k \leq \frac{\pi}{2d}$ and instead contains two bands: a fully filled valence band and an empty conduction band. In terms of a perturbation picture, the undimerised sinusoidal bandstructure folds onto itself and the dimerisation of the lattice presents a perturbation of wavevector $\frac{\pi}{2d}$ that opens up a gap at the zone-boundaries. This can be seen to reduce the total energy of the valence band, and as a result the total electronic energy. This reduction in energy can also be shown explicitly, again by forming Bloch states and solving the Hamiltonian, resulting in valence and conduction band dispersions and

a momentum space Hamiltonian given by:

$$\mathbf{H}(k) = \begin{bmatrix} 0 & v + we^{-ik\cdot 2d} \\ v + we^{+ik\cdot 2d} & 0 \end{bmatrix} \tag{3.8}$$

$$E(k) = \pm\sqrt{(2t\cos(kd))^2 + (2\alpha u \sin(kd))^2} \qquad -\frac{\pi}{2d} < k \leq \frac{\pi}{2d} \tag{3.9}$$

The dimerisation not only lowers the electronic valence-band energy, but remains energetically favourable once the elastic penalty associated with deforming the σ-bond backbone is included [57]. For small distortions one finds that the electronic energy gain scales non-analytically as $E_v \sim u^2 \ln(1/u)$, whereas the elastic cost scales as $E_\sigma \sim u^2$. Consequently, $u = 0$ is unstable and the total energy develops minima at a finite distortion magnitude, implying that a dimerised lattice is energetically preferred over the undimerised chain. This is the Peierls distortion underlying the bond alternation in TPA.

3.2.4 Topology and Solitons

In TPA there are two symmetry-related bond-alternation patterns, $u_n = \pm(-1)^n u$, corresponding to an interchange of the strong and weak bonds (equivalently $v \leftrightarrow w$). At half filling both patterns are insulating and separated from the metallic state by a gap, but they are not equivalent once one accounts for the chiral (sublattice) symmetry of the SSH Hamiltonian. The two dimerisation patterns cannot be smoothly and continuously transformed into one-another without closing the gap in the bandstructure along the path. This is the definition of a topological insulator and the SSH model provides the simplest example of such a system in one dimension [58]. In such systems the bulk spectrum is gapped, yet the bulk bandstructure carries a topological invariant which cannot be changed by smooth deformations that preserve the protecting symmetry without closing the gap. A direct consequence is the bulk-boundary correspondence: when a topologically non-trivial phase terminates (or interfaces with a trivial phase), in-gap boundary states must appear.

The topology of the SSH model is most transparently exposed by writing the Bloch Hamiltonian and dispersion in the form

$$\mathbf{H}(k) = \boldsymbol{d}(k)\cdot\boldsymbol{\sigma} \qquad E(k) = \pm|\boldsymbol{d}(k)| \tag{3.10}$$

$$\boldsymbol{d}(k) = \big(v + w\cos(ka),\, w\sin(ka),\, 0\big)^T \tag{3.11}$$

with $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ the vector of Pauli matrices. The band gap closes only if $|\boldsymbol{d}(k)| = 0$ for some k, i.e. when the trajectory of $\boldsymbol{d}(k)$ in the d_x–d_y plane passes

through the origin. Since $\boldsymbol{d}(k)$ traces a circle of radius $|w|$ centred at $(v, 0)$, the origin is enclosed if and only if $|v| < |w|$ (for the present choice of unit cell and with $v, w \in \mathbb{R}$). The two insulating phases are therefore distinguished by whether this circle winds around the origin: the trivial phase corresponds to $|v| > |w|$ (no winding), while the topological phase corresponds to $|v| < |w|$ (non-zero winding).

The topological non-triviality manifests in the existence of edge states in finite systems, which occur when a double bond is split in half due to the termination to the trivial vacuum. The same mechanism yields domain-wall states: if the bond pattern changes between the two phases somewhere along the chain a topological defect known as a "soliton" is formed. Associated with and localised on this topological soliton is an electronic mid-gap state. These electronic mid-gap states are particularly interesting as they present a reversed spin-charge relation: Charged states, $q = \pm e$, carry no spin, $s = 0$, while neutral states $q = 0$, carry spin half, $s = 1/2$. These excitations provide a natural explanation for several anomalous electronic properties of doped TPA, including its high conductivity at substantial doping levels [55].

3.2.5 Electronic Interactions

The SSH model as introduced above is an independent-particle model: electrons couple to the lattice distortion through the bond-dependent hoppings, but explicit electron-electron repulsion is neglected. However, a both qualitatively and quantitatively correct description of TPA requires some consideration of Coulomb interactions between electrons in the p_z orbitals. A minimal and widely used extension is obtained by augmenting $\hat{H}_{\mathrm{SSH}}$ with Hubbard-like density–density interactions, retaining only the dominant local and short-range contributions: the standard on-site Hubbard repulsion U, penalising double occupancy of an orbital, and a nearest-neighbour repulsion V capturing the leading order inter-site Coulomb term along the chain. This yields the interacting (extended) SSH Hamiltonian

$$\hat{H} = \hat{H}_{\mathrm{SSH}} + U \sum_i \hat{n}_{i\uparrow}\, \hat{n}_{i\downarrow} + V \sum_{i,\sigma,\sigma'} \hat{n}_{i,\sigma}\, \hat{n}_{i+1,\sigma'} \tag{3.12}$$

with $\hat{n}_{i,\sigma} = \hat{c}^{\dagger}_{i,\sigma}\hat{c}_{i,\sigma}$. This model has been shown to capture the physics of TPA both qualitatively and quantitatively and has therefore been used extensively as model Hamiltonian for the system [55, 59, 60]. We note that within this interacting model V is assumed to be constant and geometry independent, which neglects a first-principles dependence of interaction strengths on the bond length.

3.2.6 Model Parametrisation

In practice the SSH (and extended SSH) Hamiltonian is not an ab initio model and its couplings should be interpreted as effective, screened parameters. In the literature these are typically obtained by fitting to a small set of experimental observables (e.g. bandwidth, bond alternation, and a gap scale). In this work we adopt a standard, commonly used parameter set for TPA that is representative of these fits and widely used in prior studies of the SSH and extended SSH model [55, 59, 60]:

$$\begin{aligned} t &= 2.5\,\mathrm{eV} & U &= 2.5\,t_0 = 6.25\,\mathrm{eV} \\ \alpha &= 4.1\,\mathrm{eV/\AA} & V &= \frac{U}{4} = 1.5625\,\mathrm{eV} \\ u &= 0.04\,\mathrm{\AA} & K &= 40\,\mathrm{eV/\AA}^2 \end{aligned} \tag{3.13}$$

3.3 DFT Simulations of TPA

DFT is oftentimes the first approach taken to simulate a system of interest due to its relatively low computational cost and its ability to generally produce results in qualitative agreement with experiments. As such, it should be the first step in trying to understand and model the structure and properties of TPA. However, as we will show, (semi-)local DFT is unable to capture the physics of TPA, in particular its bond-alternation structure.

3.3.1 DFT Implementation

A variety of well established DFT packages exist implementing many different flavours of DFT. A natural choice for the simulation of extended, periodic systems is plane-wave DFT, which solves the electronic-structure problem under the assumption of periodic boundary conditions [26]. As such, it is also well suited to the simulation of infinite chain polymers, quasi-one-dimensional periodic systems. For our calculations we use the CASTEP plane-wave DFT code [61].

An important step in DFT calculations is the convergence of numerical parameters to ensure that a sufficiently accurate solution (within the mathematical approximations of DFT itself) is attained. The main parameters in plane-wave DFT include the cutoff-energy, E_{cut}, determining the size of the plane-wave basis set or equivalently the maximum frequency at which Fourier transforms are sampled, and the density of k-point grid, which determines the accuracy with which integrals over the Brillouin zone (BZ) are approximated [26]. If the system being simulated is not truly periodic in all three dimensions, such as a defect within a crystal, a planar ma-

terial, or a quasi-one-dimensional polymer chain, additional numerical parameters are introduced. The system has to be embedded within a larger super-cell, creating an artificial three-dimensional periodic system. The super-cell size has to be large enough for the results to converge to the isolated system, where the periodic copies no longer interact [26]. In the following calculations, infinitely long TPA chains, periodic along the x-axis and oriented with their zig-zag plane along the y-axis, are arranged in a rectangular grid along the y- and z-directions.

3.3.2 Numerical Results

Numerical parameters are considered converged when a property of interest of the system is within a specified threshold of the "true" result. The specific property of interest can greatly affect the convergence of the numerical parameters. A common property used is the total energy of the system. The defining feature of TPA, however, is its dimerisation. As discussed in sec. 3.2, this dimerisation is driven by the opening of the band-gap and a resulting reduction in the energy of the valence band in the vicinity of the BZ edge. The decrease in the total energy of the system due to this effect will be relatively small. As such, a convergence using the total energy might underestimate the numerical parameters required to properly converge the dimerisation of TPA. A property which more accurately reflects the dimerisation is the difference between the forces acting on neighbouring carbon atoms, the effective force driving the dimerisation. Explicit convergence tests confirm this. Converging the total energy to within a fractional threshold of 10^{-4} results in $E_{\text{cut}} = 750\,\text{eV}$, a k-point grid of dimensions $12 \times 2 \times 2$ (for k_x, k_y and k_z respectively), and super-cell lattice constants of $b = 16\,\text{Å}$, $c = 9\,\text{Å}$. Instead converging the dimerisation force to within a fractional threshold of 10^{-3} yields $E_{\text{cut}} = 560\,\text{eV}$, $80 \times 2 \times 2$, and $b = 30\,\text{Å}$, $c = 15\,\text{Å}$. The total energy convergence greatly underestimates the required BZ sampling grid density, while overestimating E_{cut}. Dimerisation force convergence leads to a much denser sampling, which is required to properly capture the band-gap opening in the small region next to the BZ edge and as a result the dimerisation of TPA. We used the numerical parameters found by converging the dimerisation force for all of the following calculations, additionally checking that convergence was maintained when other parameters, such as the XC-functional, were changed. All calculations were performed using the ultrasoft pseudopotentials available within CASTEP.

We performed initial bandstructure calculations using the LDA functional at a fixed geometry roughly equivalent to the experimental geometry of TPA. The results, shown in fig. 3.2, clearly exhibit the expected band-gap at the Brillouin zone edge

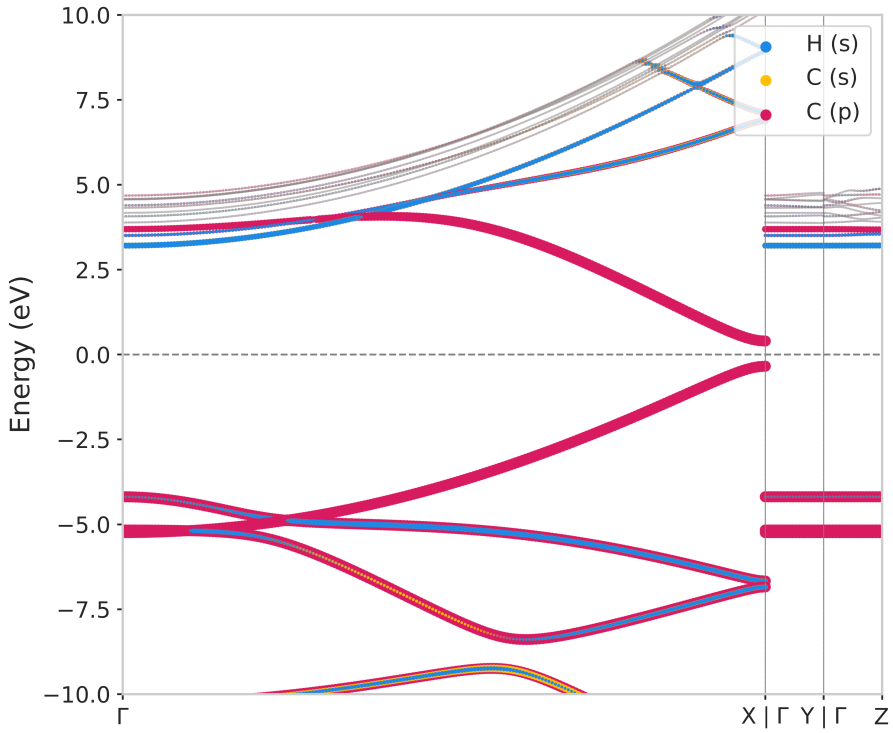
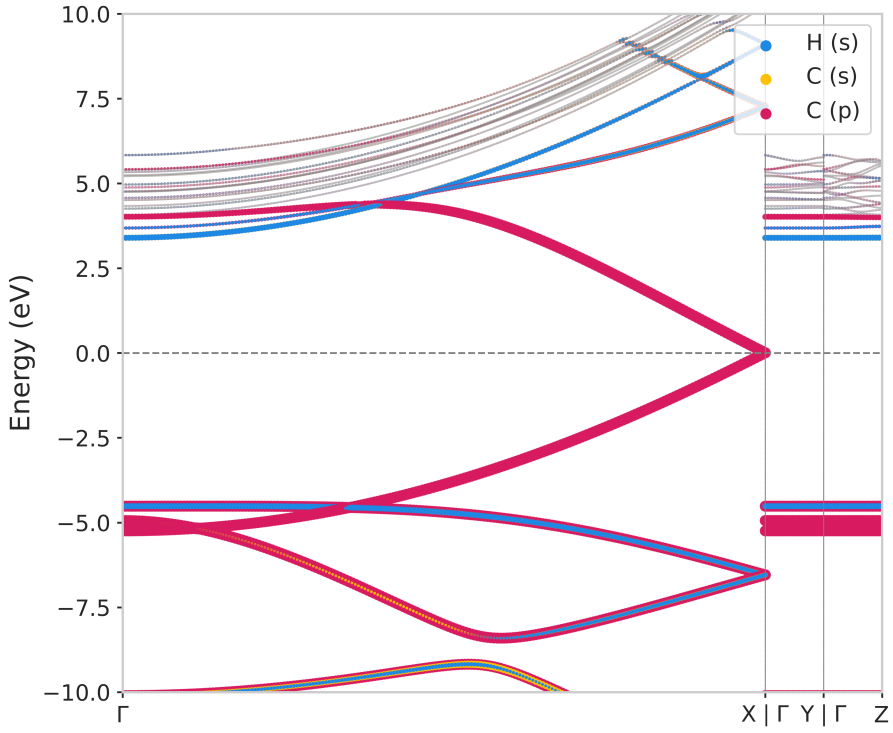


Figure 3.2: Orbital projected band-structure of TPA calculated using LDA functional at fixed geometry roughly equivalent to experimental measurements, with a bond-length alternation of $d_{\mathrm{BLA}} = d_{\mathrm{C\text{–}C}} - d_{\mathrm{C=C}} = 0.084\,\text{Å}$. The band gap was found to be $E_{\mathrm{gap}} = 0.74\,\mathrm{eV}$. The projection of the band-structure onto relevant atomic orbitals, indicated by colour, clearly shows the π character of valence and conduction band. The flat band-structure along the super-cell directions, $\Gamma - Y$ and $\Gamma - Z$, indicates good convergence.

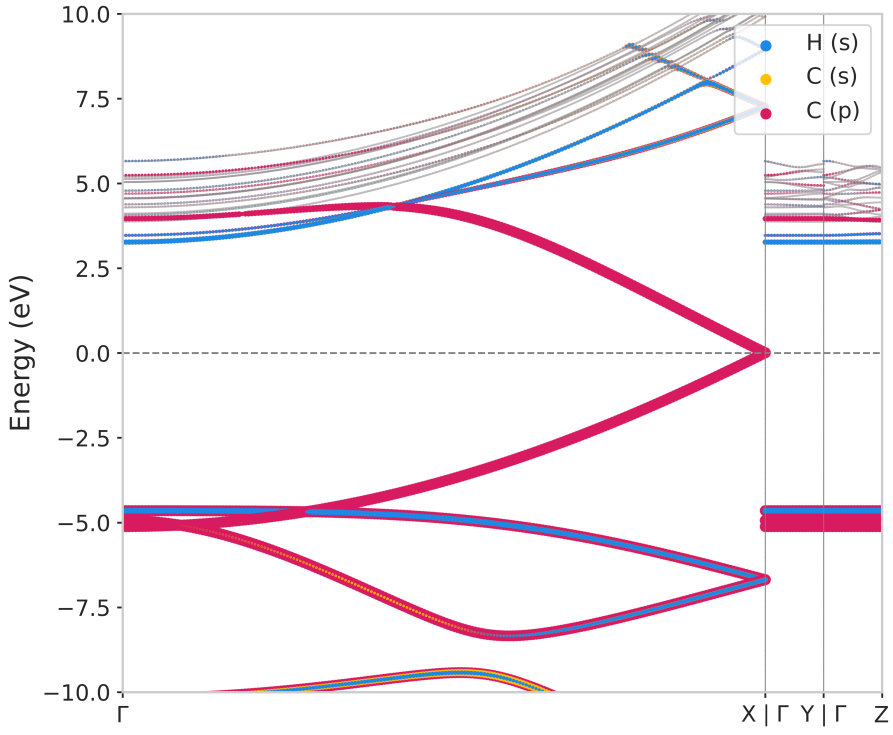
separating the π-electron bands and leading to a semi-conducting ground state.

However, within DFT it is important to perform full geometry optimisations for the ground-state structure. Unlike in high-accuracy quantum chemistry calculations, the qualitative nature of DFT results and the non-systematic approximate nature of the XC-functionals present no guarantee that qualitatively correct fixed geometry calculations translate into qualitatively correct potential energy surfaces (PES). We therefore also performed geometry optimisations for the TPA chain starting from the experimental geometry. To improve convergence the super-cell lattice parameters were constrained to their initial values to prevent physically irrelevant deformations along the supercell dimensions. The bandstructure at the optimised geometry is shown in fig. 3.3a for calculations using the LDA functional. LDA results in a geometry with negligibly small bond-alternation and a metallic ground state with no band-gap. Even using the Perdew-Burke-Ernzerhof (PBE) functional, a GGA type-functional frequently used to improve the accuracy of results compared to LDA, no bond-alternation or band-gap was observed (see fig. 3.3b).

While it is well known that (semi-)local DFT functionals such as LDA and PBE



(a)



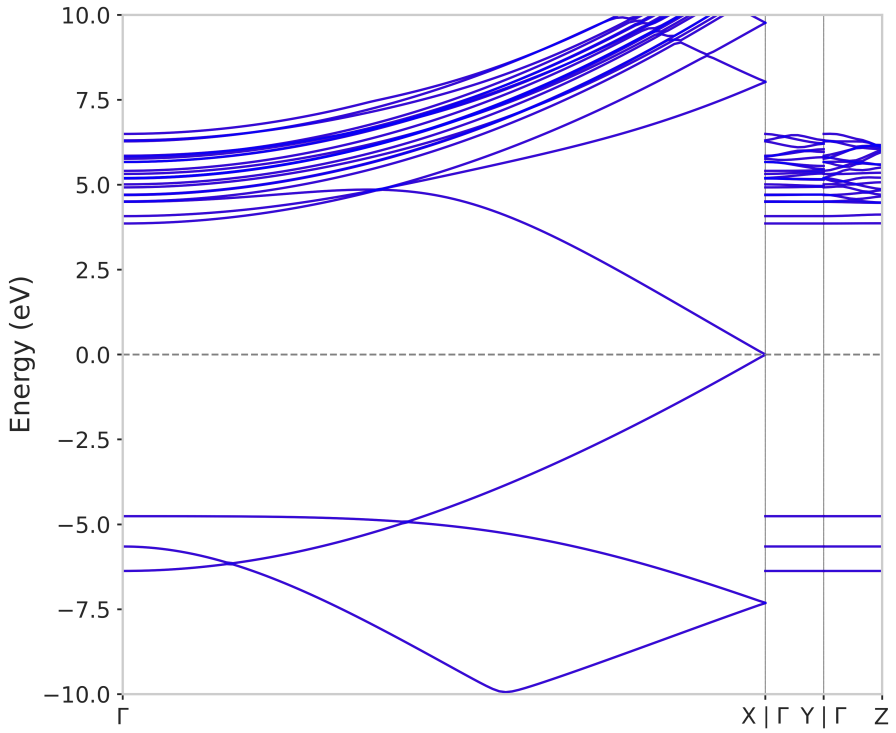
(b)

Figure 3.3: Orbital projected band-structure of TPA after geometry optimisation calculated using (a) LDA and (b) PBE functional. A negligible bond length alternation and band gap were found in both cases: (a) $d_{\mathrm{BLA}} = 6.7 \times 10^{-5}$ Å, $E_{\mathrm{gap}} = 0.59\,\mathrm{meV}$ and (b) $d_{\mathrm{BLA}} = 2.1 \times 10^{-4}$ Å, $E_{\mathrm{gap}} = 1.8\,\mathrm{meV}$

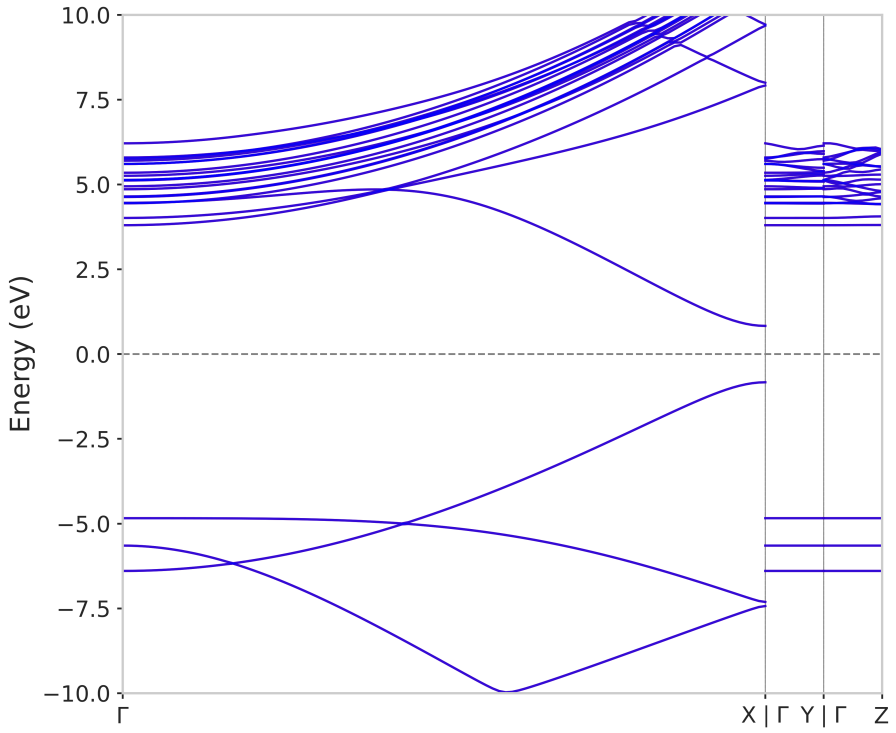
tend to underestimate the band-gap of insulators and semi-conductors [36], our results are not within the scope of this systematic underestimation. The absence of a band-gap might be attributable to the self-interaction error of DFT which tends to disproportionately delocalise electrons by allowing them to interact with themselves. The dimerisation structure of TPA localises electrons more strongly along the double bonds in bonding and anti-bonding orbitals. As such, the self-interaction error and the dimerisation structure are directly competing forces.

The DFT+U method is oftentimes used to correct for part of the self-interaction error and artificially widen band-gaps [62]. In this scheme a Hubbard-like term is added to the XC-functional, inspired by the form of the Hubbard Hamiltonian $\sum_i U n_{i,\uparrow} n_{i,\downarrow}$, which aims to replace the erroneous self-interaction error of the XC-functional with a more accurate treatment. The parameter U can in theory be calculated from first principles by considering variations of the functional with fractional occupation numbers and correcting the behaviour of the functional to reproduce discontinuities of the exact XC-functional [63]. However, oftentimes it is simply chosen in a heuristic fashion to produce results in agreement with experiments. We choose the latter approach and performed calculations using a PBE+U functional with a heuristically chosen $U = 4\,\mathrm{eV}$ on each C-atom. These calculations were found to be highly sensitive to the initial conditions chosen, in particular the initial spin on each C-atom. If the calculations were initialised with the C-atoms in a spin-unpolarised state, results similar to those of simple LDA or PBE calculations were attained: a negligible bond-alternation and band-gap as can be seen in fig. 3.4a. If instead they were initialised with C-atoms of alternating spin polarisation, a band gap was observed as shown in fig. 3.4b. However, bond-alternation was nonetheless not observed. Instead a spin density wave was found to be the ground state, a state in which neighbouring C-atoms are spin polarised in opposite directions.

It is well established from prior DFT studies of TPA and related conducting polymers that local and semi-local DFT functionals fail to appropriately describe the electronic and structural properties of TPA [64–71]. We find similar results in our calculations, where, as shown above, LDA and PBE completely suppress the dimerisation, while PBE+U does not restore the Peierls distortion and instead converges to an SDW state. However, this is a stronger failure than is typically observed in many previous DFT studies, where local and semi-local functionals generally retain a small but finite dimerisation, underestimating experimental values by $60-80\%$ [64, 66–70]. For some of these studies this difference may partly reflect the way in which the infinite-chain limit is approached: finite-chain oligomers are considered using open boundary conditions, with the infinite-chain limit obtained by extrapolation



(a)



(b)

Figure 3.4: Band-structure of TPA after geometry optimisation calculated using PBE+U with $U_{\mathrm{C}} = 4\,\mathrm{eV}$. Initial conditions were found to greatly affect the calculation. (a) A spin unpolarised initial condition leads to negligible bond-length alternation and band gap $d_{\mathrm{BLA}} = 8.7 \times 10^{-4}\,\text{Å}$, $E_{\mathrm{gap}} = 9.2\,\mathrm{meV}$. (b) An alternating spin polarised initial condition still leads to a negligible bond-length alternation $d_{\mathrm{BLA}} = 8.3 \times 10^{-4}\,\text{Å}$, but finds a band-gap $E_{\mathrm{gap}} = 1.7\,\mathrm{eV}$ opening due to the presence of a spin-density wave of amplitude $s_{x,\mathrm{C}_1} = -s_{x,\mathrm{C}_2} = 0.36$

[67, 68, 70]. In contrast, the calculations presented here directly consider the infinite chain with periodic boundary conditions, with numerical parameters converged with respect to the force driving the dimerisation. More direct comparisons are provided by studies using explicitly periodic calculations of the infinite-chain system, where a small but finite dimerisation has again been observed [64, 66]. Insufficient k-point convergence has been suggested as one possible source of this finite dimerisation, reflecting the sensitivity of the results to Brillouin-zone sampling [65, 71, 72]. The complete suppression of the dimerisation observed in the present calculations is, however, consistent with other earlier LDA calculations, which found no appreciable minimum at non-zero dimerisation amplitude [65], as well as with more recent plane-wave DFT calculations, which find the same undimerised structure using PBE and a similar SDW state using PBE+U [71]. The failure observed here and in other studies can be understood as arising from the limitations of local and semi-local XC-functionals in describing electronic interactions and correlations in this system. Approaches that reduce the delocalisation error of these functionals have been found to describe the Peierls distortion in TPA more accurately. The inclusion of exact exchange has been shown to increase both the dimerisation and the band-gap, with hybrid and range-separated hybrid functionals giving geometries closer to experiment [67, 68, 73, 74]. Self-interaction-corrected functionals similarly lead to an increase in the dimerisation relative to local and semi-local DFT calculations [69, 70].

While, as discussed above, a direct comparison with previous studies is complicated by differences in numerical details, such as basis set, boundary conditions, finite-size extrapolation, and convergence parameters, the results presented here agree with previous findings that local and semi-local functionals fail to appropriately describe the Peierls distortion in TPA, while also providing evidence for a stronger failure mode than is observed in most other studies. In particular, these functionals do not merely underestimate the dimerisation in the present calculations, but fail to stabilise a dimerised TPA structure at all. Combined with the established interpretation of this failure as arising from self-interaction error, this indicates that an accurate description of electronic interactions and correlations is essential for recovering the correct structural physics of TPA, providing an important motivation for the VQE simulations presented in later chapters.

Chapter 4

The Variational Quantum Eigensolver

Quantum computing has long been considered a natural setting for electronic structure theory since the underlying task requires the description and manipulation of quantum many-body states whose classical representations generically scale exponentially with system size [5]. While classical electronic-structure methods have been able to achieve remarkable accuracy with controlled approximations such as those described in sec. 2.4 – as a result of both the continuous advancements in the electronic structure algorithms themselves, as well as the ever increasing classical computational resources available – a direct solution to the electronic structure problem on classical computers is fundamentally limited by the exponential scaling of the Hilbert space.

Early proposals for quantum computational chemistry focused on fault-tolerant algorithms, most prominently quantum phase estimation (QPE) and Hamiltonian simulation [75]. However, the deep, fully coherent circuits and long logical gate sequences required in these algorithms presupposes the availability of fault-tolerant quantum computers. Current devices instead operate in the noisy intermediate-scale quantum (NISQ) regime, where limited qubit counts, finite coherence times, and imperfect gate operations constrain circuit depth and measurement fidelity [6]. This has motivated a focus on hybrid quantum-classical strategies that trade asymptotic guarantees for implementations compatible with near-term hardware.

The variational quantum eigensolver (VQE) [8] has emerged as a promising hybrid alternative for quantum chemistry calculations in the NISQ regime. Compared to QPE algorithms, VQE is able to produce results with relatively low qubit counts and shallow circuits, and correspondingly less stringent requirements on coherence times and gate fidelities. Since its introduction VQE has been studied extensively;

experimental implementations of VQE have been successfully performed for a variety of very small problem instances, enabled by continuous algorithmic developments reducing the circuit-depth of VQE and improving error mitigation techniques [8, 76–80]. However, practical quantum advantage using VQE for quantum chemistry has not yet been demonstrated. Even given the shallow circuits required by VQE, calculations beyond classical simulability have so far remained intractable for the currently available NISQ devices.

With this context in place, the remainder of this chapter develops the components required to implement VQE for electronic structure problems. We do not present a complete overview of all available implementations of VQE here; instead we focus here on presenting the general theory behind the methods used directly as part of our work, comparing and contrasting them only at a relatively high level with alternative approaches. While the most appropriate choice of methods depends strongly on the system studied, we delay a more detailed motivation for our choice of methods – as well as a specialised discussion of methods relative to our problem instance – to sec. 5.

4.1 The VQE Algorithm

The variational quantum eigensolver (VQE) [8, 9], like many other electronic structure methods, solves for the ground state and its energy by invoking the variational principle of sec. 2.2, and hence minimising the energy over a suitably defined search space to approximate the true ground state, while using a quantum computer to construct the states and evaluate their energies. It achieves this by:

1. Preparing a trial state $|\psi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})|0\rangle$ on the quantum computer using an ansatz circuit $U(\boldsymbol{\theta})$ parametrised by a parameter vector $\boldsymbol{\theta}$ applied to an initial state $|0\rangle$

2. Evaluating the expectation value of the energy $\langle\psi(\boldsymbol{\theta})|\hat{H}|\psi(\boldsymbol{\theta})\rangle$ using an appropriate measurement scheme on the quantum computer

3. Minimising the energy $E_* = \min_{\boldsymbol{\theta}}\langle\psi(\boldsymbol{\theta})|\hat{H}|\psi(\boldsymbol{\theta})\rangle$ using a classical optimisation over the parameters of the trial state

4.2 Fermion to Qubit Mapping

While VQE can in theory be applied to any Hamiltonian, $\hat{H}$, implementation on quantum computers requires $\hat{H}$ to be an observable on the Hilbert space of an N-

qubit quantum computer, $\mathcal{H}_q = \mathbb{C}^{2N}$, in order for $\langle\psi(\boldsymbol{\theta})|\hat{H}|\psi(\boldsymbol{\theta})\rangle$ to be measurable. A VQE simulation of a general Hamiltonian therefore requires first identifying an encoded qubit Hamiltonian $\hat{H}_q$ that shares the same spectrum of eigenvalues, and hence allows simulation of the ground state of $\hat{H}$ by using the quantum computer to perform VQE calculations of $\hat{H}_q$. Furthermore, while the ansatz circuit $\hat{U}(\boldsymbol{\theta})$ can in theory be any unitary on $\mathcal{H}_q$, many ansatz circuits are constructed not on $\mathcal{H}_q$, but on the natural Hilbert space of the problem considered. A general encoding is specified by an isometry $\hat{U} : \mathcal{H} \mapsto \mathcal{H}_q$ between states of the natural problem Hilbert space and the qubit Hilbert space, which allows encoding the Hamiltonian as $\hat{H}_q = \hat{U}\hat{H}\hat{U}^\dagger$ and similarly any other operators as $\hat{A}_q = \hat{U}\hat{A}\hat{U}^\dagger$.

On a quantum computer observables/Hamiltonians are most naturally expressed in terms of tensor products of single-qubit Pauli matrices $\{I, X, Y, Z\}$ acting on the basis of computational basis states $|n_0 n_1 \cdots n_{N-1}\rangle \in \mathcal{H}_q = \mathbb{C}^{2N}$, since they form a basis for the Hermitian operators on this space:

$$\hat{H}_{\text{qubit}} = \sum_\alpha w_\alpha \hat{P}_\alpha, \qquad \hat{P}_\alpha \in \{I, X, Y, Z\}^{\otimes n}, \tag{4.1}$$

However, as discussed in sec. 2.1.3, the electronic structure Hamiltonian, $\hat{H}_{\text{ES}}$ (equ. (2.5)), on m orbitals is naturally expressed in second quantisation in terms of fermionic creation and annihilation operators $\{\hat{c}_p^\dagger, \hat{c}_q\}$ acting on the Fock space of occupation number states $|n_0 n_1 \cdots n_{m-1}\rangle \in \mathcal{H}_f = \mathbb{C}^{2m}$, equ. (2.14). Similarly, parametrised unitary circuits on a quantum computer are naturally expressed in terms of exponentiated Pauli gates (e.g. $\hat{U}(\theta) = e^{i\theta\hat{P}}$), while ansatz circuits for electronic structure are oftentimes constructed from creation and annihilation operators (e.g. $\hat{U}(\theta) = e^{i\theta(\hat{c}_p^\dagger \hat{c}_q + \text{h.c.})}$). A fermion to qubit encoding specifies a mapping from fermion to qubit states as well as from fermion to qubit operators, allowing us to identify corresponding encoded states of the system, to express $\hat{H}_{\text{ES}}$ as an equivalent encoded operator while ensuring the energy spectrum is preserved, and to implement the fermionic ansatz circuit $\hat{U}(\boldsymbol{\theta})$ on the quantum computer.

In the following section, we describe the Jordan-Wigner encoding, the most common choice of encoding for quantum chemistry calculations and the main encoding chosen for our work due to both its wide ranging use in practical calculations and its relative simplicity for our model as further discussed in sec. 5. However, we will also briefly summarise the main motivations behind the development of alternative existing encodings to allow for better discussion of potential improvements and shortcomings of our work in sec. 5.

4.2.1 Jordan–Wigner Encoding

The Jordan–Wigner (JW) encoding [81] is the oldest and conceptually simplest mapping. It constructs an encoding by realising that the m-orbital Fock space $\mathcal{H}_f = \mathbb{C}^{2m}$ and the m-qubit Hilbert space $\mathcal{H}_q = \mathbb{C}^{2m}$ are isomorphic, and directly identifying the occupation number of spin-orbital p with the $|0\rangle, |1\rangle$ states of qubit p,

$$|n_0 n_1 \cdots n_{m-1}\rangle_f \longleftrightarrow |n_0 n_1 \cdots n_{m-1}\rangle_q \tag{4.2}$$

Encoding the fermionic creation and annihilation operators therefore simply amounts to expanding the operators in an alternative operator basis. In this encoding the spin ladder operators $\{\hat{\sigma}_p^-, \hat{\sigma}_p^+\}$ can be recognised to directly act like the fermionic creation and annihilation operators $\{\hat{c}_p^\dagger, \hat{c}_p\}$ on the orbital occupation, where we note that $\sigma_p^- \leftrightarrow c_p^\dagger$ due to the standard convention for computational basis states $|0\rangle \leftrightarrow |\uparrow\rangle$:

$$\sigma_p^-|\ldots 0_p \ldots\rangle = |\ldots 1_p \ldots\rangle \qquad \sigma_p^-|\ldots 1_p \ldots\rangle = \mathbf{0} \tag{4.3}$$

$$\sigma_p^+|\ldots 1_p \ldots\rangle = |\ldots 0_p \ldots\rangle \qquad \sigma_p^-|\ldots 0_p \ldots\rangle = \mathbf{0} \tag{4.4}$$

However, unlike the fermionic operators they do not preserve the required anti-symmetry properties of the fermionic many body wavefunction, equ. (2.8). The JW-encoding remedies this by including a JW-string, which explicitly enforces the fermionic sign structure. Writing $Z^{<p} \equiv Z_0 Z_1 \ldots Z_{p-1}$, the JW encoded fermionic operators $\{c_p^\dagger, c_p\}$ can be shown to be equivalent to the spin ladder operators :

$$\hat{c}_p^\dagger \;\mapsto\; \mathcal{E}_{\mathrm{JW}}(\hat{c}_p^\dagger) \;=\; Z^{<p}\,\hat{\sigma}_p^- \;=\; Z^{<p}\,\frac{X_p - iY_p}{2}, \tag{4.5}$$

$$\hat{c}_p \;\mapsto\; \mathcal{E}_{\mathrm{JW}}(\hat{c}_p) \;=\; Z^{<p}\,\hat{\sigma}_p^+ \;=\; Z^{<p}\,\frac{X_p + iY_p}{2}, \tag{4.6}$$

From eqs. (4.5)–(4.6) one directly obtains useful identities, e.g. for the number operator

$$\hat{n}_p \equiv \hat{c}_p^\dagger \hat{c}_p \;\mapsto\; \frac{I - Z_p}{2}, \tag{4.7}$$

and for the – ubiquitous in electronic structure – single and double excitations ($p \geq q$ and $r \geq s$):

$$\hat{c}_p^\dagger \hat{c}_q \;\mapsto\; \frac{1}{4} Z^{q<p} (X_p - iY_p)(X_q + iY_q) \tag{4.8}$$

$$\hat{c}_p^\dagger \hat{c}_q^\dagger \hat{c}_r \hat{c}_s \;\mapsto\; \frac{1}{16} Z^{q<p} Z^{s<r} (X_p - iY_p)(X_q - iY_p)(X_r + iY_r)(X_s + iY_s) \tag{4.9}$$

where $Z^{q<p} \equiv Z_{q+1}Z_{q+2}\dots Z_{p-1}$. In general, under the JW-encoding a k-body fermionic term maps to a sum of at most 2^k Pauli strings, with support (i.e. the set of qubit indices on which a Pauli acts non-trivially in the Pauli string) that can in general span the interval from the smallest to the largest orbital index appearing in the term.

The JW-encoding (as well as most other commonly used fermionic encodings) has the important property that gates involving individual fermionic Hermitian and anti-Hermitian excitations, $\hat{h}^{pq}_{rs} \equiv \hat{c}^\dagger_p\hat{c}^\dagger_q\hat{c}_r\hat{c}_s + \text{h.c.}$ and $\hat{\tau}^{pq}_{rs} \equiv \hat{c}^\dagger_p\hat{c}^\dagger_q\hat{c}_r\hat{c}_s - \text{h.c.}$ respectively, can be implemented exactly using a simple Pauli product decomposition, e.g.:

$$\hat{h}^{pq}_{rs} = \hat{c}^\dagger_p\hat{c}^\dagger_q\hat{c}_r\hat{c}_s + \text{h.c.} \rightarrow \sum_k \hat{P}_k \tag{4.10}$$

$$\exp(\hat{h}^{pq}_{rs}) = \exp(i(\hat{c}^\dagger_p\hat{c}^\dagger_q\hat{c}_r\hat{c}_s + \text{h.c.})) \rightarrow \exp(i\sum_k \hat{P}_k) = \prod_k \exp(i\hat{P}_k) \tag{4.11}$$

and similarly for $\hat{h}^p_q/\hat{\tau}^{pq}_{rs}\hat{\tau}^{pq}_{rs}$. These gates are frequently used to construct ansatz circuits for fermionic systems. This decomposition is exact since the constituent Paulis, $\hat{P}_k$, in the decomposition of any such term are all mutually commuting, $[\hat{P}_k, \hat{P}_l] = 0$, e.g. $\hat{c}^\dagger_p c_{p+1} + \text{h.c.} \rightarrow \frac{1}{2}(X_pX_{p+1} + Y_pY_{p+1})$ and $\hat{c}^\dagger_p c_{p+1} - \text{h.c.} \rightarrow \frac{i}{2}(X_pY_{p+1} - Y_pX_{p+1})$, and hence $[XX, YY] = 0$ and $[XY, YX] = 0$ [82].

4.2.2 Alternative Encodings

The non-locality of the Pauli operators encountered in the JW-encoding as a result of the JW-strings can present a problem, particularly for NISQ applications. Implementing gates involving fermionic excitations between two orbitals p, q – gates commonly encountered in fermionic ansatz circuits such as UCCSD – will generally require two qubit gates between any intermediate qubits on account of the presence of $Z^{q<p}$ [83–85]. Such a construction leads to high depth circuits, both as a result of the depth of each excitation operator, as well as due to the additional depth resulting from a reduced ability to parallelise the application of multiple fermionic excitation gates acting on disjoint subsets of orbitals [84, 86]. In NISQ applications this can present problems due to the relatively high gate errors and low decoherence times. In addition, it can also lead to worsening of effective readout errors since more qubits need to be measured to construct the relevant fermionic expectation values [87]. Smart ansatz design and circuit implementations can lessen the cost of implementing the JW-strings by e.g. introducing ancilla qubits, choosing specific operator orderings or introducing more complex compilation techniques, or by completely removing the need to account for them in the ansatz circuit in the first place using a

variety of different ansatz design techniques [11, 77, 84, 88–95]. Alternative encodings have also been developed that ameliorate the problematic JW-strings, such as the parity [96], Bravyi-Kitaev (BK) [97] or ternary tree (TT) mappings [98]. Some encodings are also able to fully remove the problematic JW-strings using additional ancilla qubits [97, 99, 100]. These are generally tailored to encode a specific problem instance and its geometry, for example mapping local fermionic interactions to local qubit interactions specifically for fermions interacting with nearest neighbours on a square lattice.

In addition to the problematic JW-strings, the JW-encoding is agnostic to any symmetries that are frequently present in fermionic systems and does not exploit them to reduce the effective size of the Hilbert space and hence the number of qubits required for the encoding. To this end, other encodings have been developed which are able to reduce the qubit requirement by explicitly incorporating symmetries and conserved quantities into the mapping. If the Hamiltonian possesses symmetries, it is sufficient to only encode a specific symmetry sector, leading to a reduced qubit requirement. This allows taking advantage of parity and point symmetries [101, 102] or the conservation of particle number [103] to reduce the size of the encoded space.

4.3 Ansatz Circuit

The choice of parametrised circuit $U(\boldsymbol{\theta})$ is a central design decision in VQE. It not only determines the reachable family of states $|\psi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})|\psi_i\rangle$ (known as the expressivity of the ansatz circuit) and hence the best attainable energy, but also affects the trainability, the implementation cost, and the robustness to noise. It mirrors the choice of wavefunction parametrisation in standard electronic structure theory where states can be approximated in terms of e.g. either a configuration interaction or a coupled cluster expansion. While in theory any parametrised ansatz circuit can be used to perform a VQE calculation, useful and successful applications of VQE to electronic structure theory generally use ansatz circuits that satisfy some or all of the below requirements. The ansatz circuit should be:

chemically motivated Since $\hat{H} = \hat{H}_{\mathrm{ES}}$ chemically motivated ansatz circuits can greatly increase the accuracy/performance of the calculation

symmetry compatible If $\hat{H}$ is known to conserve symmetries, ansatz circuits can be designed to preserve the corresponding quantum numbers

expressive The ansatz should be expressive enough to approximate the target ground state to sufficient accuracy

easily trainable The parameter landscape of the ansatz circuit should allow for easy and fast optimisation

hardware efficient The ansatz circuit should be implementable on the quantum computer with a low circuit depth and gate count

In this section, we outline the theory behind the Hamiltonian variational ansatz, an ansatz commonly chosen for applications to spin and fermion lattice models and the main ansatz chosen as part of our work, due to its relatively simple implementation for the Hubbard-like model under consideration as further discussed in sec. 5.2.2. We also briefly discuss the family of adaptively constructed ansatz circuits, which, while not used in our work, is a common choice for more general quantum chemistry calculations to avoid some of the road-blocks encountered in VQE.

4.3.1 Hamiltonian Variational Ansatz

The Hamiltonian variational ansatz (HVA) [104] has emerged as a both popular and successful ansatz choice for the simulation of spin-lattice and Hubbard models [11, 13, 14, 90, 105–116]. It is a problem specific ansatz circuit that uses the form of the Hamiltonian to construct the ansatz circuit by implementing an approximate adiabatic evolution towards the true ground state. The L-layered HVA is given by:

$$|\psi(\boldsymbol{\theta})\rangle = \hat{U}(\boldsymbol{\theta})|\psi_i\rangle = \prod_{k=1}^{L} \prod_{\alpha} e^{-i\,\theta_{k,\alpha}\,\hat{H}_\alpha}|\psi_i\rangle \tag{4.12}$$

where $\hat{H}_\alpha$ are Hermitian fragments of the Hamiltonian satisfying $\sum_\alpha \hat{H}_\alpha = \hat{H}$ for which $e^{i\hat{H}_\alpha}$ can easily be implemented on a quantum computer. Usually this is achieved by choosing each $\hat{H}_\alpha$ to be constructed from mutually commuting terms of the Hamiltonian, $\hat{H}_\alpha = \sum_i \hat{h}_{\alpha,i}$ with $[\hat{h}_{\alpha,i}, \hat{h}_{\alpha,j}] = 0$ for all i, j, which allows an exact decomposition of the evolution gate as:

$$e^{-i\theta_{k,\alpha}\hat{H}_\alpha} = \prod_j e^{-i\theta_{k,\alpha}\hat{h}_{\alpha,j}} \tag{4.13}$$

where each $\exp(-i\theta_{k,\alpha}\hat{h}_{\alpha,i})$ can be implemented directly since $\hat{h}_{\alpha,i}$ will usually correspond to an individual Hermitian fermionic excitation, $\hat{h}^p_q \equiv \hat{c}^\dagger_p\hat{c}_q + \text{h.c.}$ and $\hat{h}^{pq}_{rs} \equiv \hat{c}^\dagger_p\hat{c}^\dagger_q\hat{c}_r\hat{c}_s + \text{h.c.}$, or a Pauli string directly, $\hat{P}$. Fermionic operators are first mapped to the corresponding Pauli operators $\hat{h} = \sum_l \hat{P}_l$ and then implemented

using a product decomposition as:

$$e^{-i\theta\hat{h}} = \prod_l e^{-i\theta\hat{P}_l} \tag{4.14}$$

which is usually exact, since the Pauli operators corresponding to $\hat{h}_q^p/\hat{h}_{rs}^{pq}$ under most fermionic encodings (and in particular the JW-encoding) are mutually commuting $[\hat{P}_l, \hat{P}_m] = 0$ (see sec. 4.2.1). However, as discussed further in sec. 5.2.2, this decomposition is not guaranteed to be exact if $\hat{h}$ is identified as a more complex fermionic "building block", such as spin-adapted fermionic excitations. The initial state, $|\psi_i\rangle$ is chosen to be an efficiently constructable ground state of an easily diagonalisable part $\hat{H}_0$ of the Hamiltonian $\hat{H} = \hat{H}_0 + \hat{V}$. The equivalence to an approximate adiabatic evolution can be observed by considering an interpolation between $\hat{H}_i$ and $\hat{H}_f$ with ground states $|\psi_i\rangle$ and $|\psi_f\rangle$ respectively:

$$\hat{H}[s] = (1-s)\,\hat{H}_i + s\,\hat{H}_f, \qquad s(0) = 0,\ s(T) = 1 \tag{4.15}$$

We can prepare $|\psi_f\rangle$ from $|\psi_i\rangle$, since the adiabatic theorem guarantees that, if the evolution is sufficiently slow, i.e. as $T \to \infty$, the state remains in the instantaneous ground state throughout and hence that

$$|\psi_f\rangle \simeq \mathcal{T}\exp\left(-i\int_0^T \hat{H}[s(t)]\,dt\right)|\psi_i\rangle \tag{4.16}$$

where $\mathcal{T}$ represents the time-ordering operator. For small enough time-steps Δt, $(\Delta t \to 0)$, and hence also a large enough number of steps, $(T = L\Delta t \to \infty)$, this evolution can be approximated as:

$$\mathcal{T}\exp\left(-i\int_0^T \hat{H}[s(t)]\,dt\right) \simeq \prod_{k=1}^{L}\prod_{\alpha} e^{-i\,\hat{H}_\alpha[s(k\Delta t)]\,\Delta t} \tag{4.17}$$

where the $\hat{H}[s(t)]$ has been split into sub-observables $\sum_\alpha \hat{H}_\alpha[s(t)] = \hat{H}[s(t)]$. By considering $\hat{H}_i \to \hat{H}_0$ and $\hat{H}_f \to \hat{H}_0 + \hat{V}$ and hence $\hat{H}[s(t)] = \hat{H}_0 + s\hat{V}$, the HVA, equ. (4.12), can be seen to approximate this time evolution. By incorporating variational parameters $\hat{H}_\alpha[s(k\Delta t)]\Delta t \to \theta_{k,\alpha}\hat{H}_\alpha$, the ansatz circuit can better approximate the true adiabatic evolution reducing the errors introduced by the digitisation and trotterisation, even if the number of repetitions is not sufficiently large to approximate it directly using equ. (4.17). While the HVA is in theory guaranteed to be able to reproduce the exact adiabatic evolution in the limit of $L \to \infty$ within its parameter space, it is ultimately still a heuristic ansatz when it comes to the choice of $\hat{H}_0$ and

$|\psi_i\rangle$, the splitting $\hat{H}_\alpha$ and the initialisations of $\theta_{k,\alpha}$, in particular at limited L, and should be considered distinct from the digitised adiabatic evolution circuit. The "derivation" from a digitised and trotterised adiabatic evolution presented above should be considered simply a motivation and a successful ansatz circuit is by no means required to be able to implement the adiabatic evolution exactly (in the limit $L \to \infty$), in particular since an adiabatic evolution is not even guaranteed to be able to prepare the correct ground state if the energy gap closes along the path [117].

4.3.2 Adaptive Ansatz

Adaptive ansatz construction has emerged as a promising strategy for general quantum-chemistry simulations. The prototypical approach is the Adaptive Derivative-Assembled Pseudo-Trotter (ADAPT) ansatz [20]; numerous refinements and variants [93, 94, 118–120] have been shown to deliver accurate energies for small molecules with reduced resource requirements. Adaptive algorithms grow the ansatz circuit iteratively: at each step a VQE optimisation is performed on the currently considered ansatz $|\psi^{(n-1)}(\boldsymbol{\theta})\rangle$ and the result of the optimisation is then used to propose an adapted/extended ansatz circuit $|\psi^{(n)}(\boldsymbol{\theta})\rangle$ to improve the accuracy of the calculation.

In the ADAPT-VQE algorithm, each update selects one anti-Hermitian operator $\hat{A}_{\alpha_n}$ from a predefined pool $\mathcal{P} = \{\hat{A}_\alpha\}$ and appends the corresponding parametrised unitary evolution gate:

$$|\psi^{(n)}(\boldsymbol{\theta})\rangle = e^{\theta_n \hat{A}_{\alpha_n}} |\psi^{(n-1)}(\boldsymbol{\theta})\rangle = \prod_n e^{\theta_n \hat{A}_{\alpha_n}} |\psi_i\rangle \tag{4.18}$$

The $\hat{A}_\alpha$ are usually chosen as, or constructed from either the anti-Hermitian unitary coupled cluster (UCC) operators, $\hat{\tau}^p_q = \hat{c}^\dagger_p \hat{c}_q - \text{h.c.}$ and $\hat{\tau}^{pq}_{rs} = \hat{c}^\dagger_p \hat{c}^\dagger_q \hat{c}_r \hat{c}_s - \text{h.c.}$, or Pauli strings directly, $\hat{P}$. Again, in the fermionic case operators are usually implemented using a product decomposition over the encoded Pauli strings, $\hat{A} = \sum_l \hat{P}_l$:

$$e^{\theta \hat{A}} = \prod_l e^{\theta \hat{P}_l} \tag{4.19}$$

which, as in the HVA, is an exact decomposition under most fermionic encodings (in particular the JW-encoding) if $\hat{A}$ are identified as individual $\hat{\tau}^p_q / \hat{\tau}^{pq}_{rs}$ (see sec. 4.2.1), where we again stress that for more complicated generators this decomposition is not necessarily exact (see sec. 5.2.2). The original ADAPT-VQE pool consists of the generalised (i.e. p,q,r,s can correspond to both occupied or virtual orbitals) and spin complemented (i.e. spin flip $\alpha \leftrightarrow \beta$ symmetrised linear combination)

singles and doubles UCC operators $\mathcal{P} = \left\{\hat{\tau}_{q_\alpha}^{p_\alpha} + \hat{\tau}_{q_\beta}^{p_\beta}, \hat{\tau}_{r_\alpha s_\alpha}^{p_\alpha q_\alpha} + \hat{\tau}_{r_\beta s_\beta}^{p_\beta q_\beta}, \hat{\tau}_{r_\alpha s_\beta}^{p_\alpha q_\beta} + \hat{\tau}_{r_\beta s_\alpha}^{p_\beta q_\alpha}\right\}$. The appended operator is chosen based on the derivative of the energy with respect to the newly introduced operator/parameter at $\theta_k = 0$:

$$g_k = \frac{\partial}{\partial \theta_k} \langle \psi^{(n-1)} | e^{-\theta_k \hat{A}_k} \hat{H} e^{\theta_k \hat{A}_k} | \psi^{(n-1)} \rangle \Big|_{\theta_k = 0} = \langle \psi^{(n-1)} | [\hat{H}, \hat{A}_k] | \psi^{(n-1)} \rangle \quad (4.20)$$

The operator which maximises $|g_k|$ is appended to the ansatz since to first order it is expected to lead to the largest improvement in the energy. At each step ADAPT-VQE (i) re-optimises the current parameters, (ii) evaluates energy gradients for all candidates in an operator pool, and (iii) appends the operator with the largest gradient magnitude. This is repeated until sufficient convergence is established. The original ADAPT-VQE has since been modified and improved upon by considering more hardware efficient implementations, such as alternative operator pools consisting of Pauli strings that reduce circuit depths required to implement the ADAPT layers [93, 94], or by simultaneously adding multiple operators in a circuit depth efficient way [119, 121]. Variants have also refined the selection rule [94, 118] to improve the choice of operators at each step.

4.4 Energy Measurement

Implementing a VQE calculation requires access to the expectation value of the energy, $\langle \hat{H} \rangle = \langle \psi | \hat{H} | \psi \rangle$, for the state, $|\psi\rangle$, prepared on the quantum computer. However, unlike in electronic structure theory on classical computers, expectation values of observables cannot be directly evaluated, since the wavefunction is not directly accessible, hence preventing us from evaluating $\langle \hat{O} \rangle = \langle \psi | \hat{O} | \psi \rangle$ using straightforward linear algebra as might be done classically. Instead, information about $|\psi\rangle$ and $\langle \hat{O} \rangle$ can only be accessed via projective measurements on the computational basis states. Since these measurements are inherently stochastic the expectation value is calculated by constructing a stochastic estimator $\bar{H}$ for $\langle \hat{H} \rangle$ based on the measurement results.

Expectation values $\langle \hat{O} \rangle$ in quantum mechanics correspond by definition to the expectation value of the random variable associated with the result of a projective measurement on the eigenstate $|o_i\rangle$ with eigenvalue o_i of $\hat{O}$. Expanding $|\psi\rangle = \sum_i c_i |o_i\rangle$ this correspondence becomes immediately clear as $\langle \hat{O} \rangle = \sum_i |c_i|^2 o_i$ where $|c_i|^2$ is the probability of measuring o_i for a projective measurement on the basis $\{|o_i\rangle\}$ by the Born rule. If $\mathbf{O}$ is the random variable associated with the measurement outcome, we therefore have $\mathbb{E}[\mathbf{O}] = \langle \hat{O} \rangle$. Similarly, the quantum operator variance

$\mathrm{Var}(\hat{O}) = \langle(\hat{O} - \langle\hat{O}\rangle)^2\rangle$, is by definition also the variance of the projective measurement $\mathrm{Var}[\mathbf{O}] = \mathrm{Var}(\hat{O})$, since $\mathrm{Var}(\hat{O}) = \sum_i |c_i|^2 o_i^2 - (\sum_i |c_i|^2 o_i)^2$. We use $\mathrm{Var}(\cdot)$ to distinguish the quantum operator variance from the variances of random variables $\mathrm{Var}[\cdot]$ and stress that $\mathrm{Var}(\hat{O})$ is not directly a variance of a random variable but only via its association with $\mathrm{Var}[\mathbf{O}]$.

Expectation values for a physically realised wavefunction $|\psi\rangle$ (such as on a quantum computer) can in theory be calculated by repeatedly preparing $|\psi\rangle$ and measuring on the basis $\{|o_i\rangle\}$, recording the measurement results $o_{i(k)}$, $k \in [1, 2, \cdots, M]$ and forming the sample mean $\bar{O} = \sum_k^M \frac{o_{i(k)}}{M}$ as an unbiased estimator of the population mean and hence expectation value $\mathbb{E}[\bar{O}] = \langle\hat{O}\rangle$. The variance of this estimator will be given by $\mathrm{Var}[\bar{O}] = \frac{\mathrm{Var}(\hat{O})}{M}$. Such repeated executions of the same quantum circuit required by the inherent stochasticity of quantum measurements are frequently referred to as "shots".

Quantum computers, however, are not able to directly perform measurements in any arbitrary basis $\{|o_i\rangle\}$, but can only implement projections onto the computational basis states $\{|000\ldots\rangle, |100\ldots\rangle, \ldots\}$ and hence can only directly estimate expectations of diagonal Paulis, i.e. Pauli-Z products $\langle Z_i Z_j Z_k \ldots\rangle$. Calculating arbitrary expectation values therefore requires more complicated measurement schemes, involving rotations of the state using unitaries $\hat{U}|\psi\rangle$ in such a way that computational basis state measurements can be used to construct appropriate estimators $\bar{O}$ of $\langle\hat{O}\rangle$.

In this section, we discuss the theory behind two measurement strategies. First we will describe measurements performed by splitting observables into easily measurable fragments. This measurement strategy is the standard choice for NISQ devices and VQE applications and the main measurement scheme considered for our work. We discuss extensions to this scheme both in terms of generalised fragments and optimised measurement allocation for later discussion. However, we will also briefly discuss the theory behind quantum phase estimation, an algorithm for fault-tolerant quantum computing, since it in theory allows implementing a direct projective measurement on the operator eigenbasis as discussed above and as considered in sec. 6.2.1.

4.4.1 Measurable Fragments

Since the Pauli products $\{\hat{P}_i\}$ form a basis for Hermitian operators, any observable (and in particular the Hamiltonian) can be decomposed as a sum over N_p Pauli

products:

$$\hat{H} \;=\; \sum_{\alpha}^{N_P} w_\alpha \, \hat{P}_\alpha, \qquad \hat{P}_\alpha \in \{I, X, Y, Z\}^{\otimes N} \tag{4.21}$$

The simplest route to make expectation value estimation compatible with computational basis measurements is to take advantage of the linearity of quantum expectation values $\langle \hat{H} \rangle \;=\; \sum_\alpha w_\alpha \langle \hat{P}_\alpha \rangle$ and estimate each $\langle \hat{P}_\alpha \rangle$ individually [8]. Since an individual Pauli product $\hat{P}_\alpha$ can be efficiently diagonalised, $\hat{U}_\alpha \hat{P}_\alpha \hat{U}_\alpha^\dagger = Z_i Z_j Z_k \ldots$, and the diagonalising unitary circuit $\hat{U}_\alpha$ efficiently constructed ($Z = HXH$ and $Z = HS^\dagger YSH$), we can calculate $\langle \hat{P}_\alpha \rangle$ by applying $\hat{U}_\alpha$ to the prepared state $|\psi\rangle$ and then performing computational basis measurements to estimate $\langle Z_i Z_j Z_k \ldots \rangle$ on the transformed state $\hat{U}_\alpha |\psi\rangle$. This constructs the sought after expectation value since $\langle \psi | \hat{U}_\alpha^\dagger Z_i Z_j Z_k \ldots \hat{U}_\alpha | \psi \rangle = \langle \psi | \hat{P}_\alpha | \psi \rangle$. The estimator $\bar{H}$ can be explicitly constructed using a total shot budget M by using $m \equiv \frac{M}{N_p}$ repeated measurements for each term $\hat{P}_\alpha$, recording measurement results $P_{\alpha,i}$:

$$\bar{H} = \sum_{\alpha}^{N_P} w_\alpha \bar{P}_\alpha = \sum_{\alpha}^{N_P} w_\alpha \sum_{i}^{m} \frac{P_{\alpha,i}}{m} \tag{4.22}$$

While this measurement procedure is conceptually simple, it requires individually estimating the N_p operators present in the decomposed Hamiltonian. Since a general electronic structure Hamiltonian on N_f orbitals in the JW-encoding scales as $N_p = \mathcal{O}(N_f^4)$ and since each $\hat{P}_\alpha$ requires a different measurement unitary $\hat{U}_\alpha$ to be appended to $|\psi\rangle$, such a measurement scheme can be costly to implement in practice. More sophisticated measurement schemes consider splitting $\hat{H}$ into a set of N_A more general fragments $\{\hat{A}_\alpha\}$:

$$\hat{H} \;=\; \sum_{\alpha}^{N_A} \hat{A}_\alpha, \tag{4.23}$$

However, we still require that each $\hat{A}_\alpha$ can be efficiently diagonalised

$$\hat{A}_\alpha = \hat{U}_\alpha^\dagger \left(\sum_i a_{i,\alpha} \, Z_i \;+\; \sum_{ij} b_{ij,\alpha} \, Z_i Z_j \;+\; \cdots \right) \hat{U}_\alpha \tag{4.24}$$

and that a circuit implementing the diagonalising unitary $\hat{U}_\alpha$ can be efficiently constructed. Similar to above, $\langle \hat{A}_\alpha \rangle$ can be calculated by simultaneously estimating the diagonal Paulis $\{Z_i,\, Z_i Z_j,\, \ldots\}$ using computational basis measurements on $\hat{U}_\alpha |\psi\rangle$

and constructing $\bar{H}$ using the recorded measurement results $A_{\alpha,i}$:

$$\bar{H} = \sum_{\alpha}^{N_A} \bar{A}_\alpha = \sum_{\alpha}^{N_A} \sum_{i}^{m} \frac{A_{\alpha,i}}{m} \tag{4.25}$$

Since $\hat{A}_\alpha$ is no longer restricted to be a single Pauli term, the number of terms required to be measured can be greatly reduced. The most commonly considered approaches build easily diagonalisable fragments out of mutually commuting Pauli operators, since mutually commuting operators can be simultaneously diagonalised. The simplest approach considers Pauli operators that are mutually qubit-wise commuting, i.e. two Paulis, $\hat{P}_1$ and $\hat{P}_2$, that not only commute in general, $[\hat{P}_1, \hat{P}_2] = 0$, but also commute on each individual qubit operator in the Pauli string, $\hat{P}_{1/2} = \prod_i \hat{P}_{1/2,i}$ with $\hat{P}_{1/2,i} \in \{I_i, X_i, Y_i, Z_i\}$ and $[\hat{P}_{1,i}, \hat{P}_{2,i}] = 0$ [122, 123]. Such a set of Pauli operators allows for a simple and efficient simultaneous diagonalisation using only single-qubit gates and $\mathcal{O}(1)$ circuit depth. Generally commuting Pauli operators, $[\hat{P}_1, \hat{P}_2] = 0$, can also be measured simultaneously, however at an increased cost resulting from a more complex simultaneous diagonalisation circuit [123, 124]. Alternative schemes also consider sums of anti-commuting Pauli terms [125] or work directly in the fermionic space by factorising the fermionic interaction term [126].

An important consideration for any measurement scheme is the variance of the constructed estimator $\bar{H}$ of $\langle \hat{H} \rangle$, as it will determine the expected stochastic error and the number of measurements, M, required to estimate $\langle \hat{H} \rangle$ to within an accuracy ϵ. The root mean square error (RMSE) is a frequently used metric to quantify the error of an estimator $\bar{\phi}$ used to determine an unknown parameter ϕ

$$\mathrm{RMSE}(\bar{\phi}) \equiv \sqrt{\mathbb{E}\left[(\bar{\phi} - \phi)^2\right]} = \sqrt{\mathrm{Bias}[\bar{\phi}]^2 + \mathrm{Var}[\bar{\phi}]} \tag{4.26}$$

as a function of the variance $\mathrm{Var}[\bar{\phi}]$ and bias $\mathrm{Bias}[\bar{\phi}] \equiv \mathbb{E}[\bar{\phi}] - \phi$ of the estimator. For the unbiased estimators $\bar{H}$ considered here, the variance therefore directly determines the RMSE:

$$\mathrm{RMSE}(\bar{H}) = \sqrt{\mathrm{Var}[\bar{H}]} \tag{4.27}$$

and minimising the variance of the estimator at a fixed shot budget M minimises the expected error of our estimate of $\langle \hat{H} \rangle$. For the above considered measurement schemes, simple variance propagation can be used to evaluate $\mathrm{Var}[\bar{\mathrm{H}}]$, by recognising that repeated measurements are independent, $\mathrm{Cov}[A_{\alpha,i}, A_{\beta,j}] = 0$ for any α, β and

i, j:

$$\mathrm{Var}[\bar{H}] = \sum_{\alpha} \mathrm{Var}[\bar{A}_{\alpha}] = \sum_{\alpha} \frac{\mathrm{Var}(\hat{A}_{\alpha})}{m} \tag{4.28}$$

We stress that it is now important to distinguish between the quantum variance of an operator $\mathrm{Var}(\hat{A}_{\alpha})$ and the variance of the measurement of an operator $\mathrm{Var}[A_{\alpha,i}]$ (and similarly for covariances). Whereas $\mathrm{Cov}[A_{\alpha,i}, A_{\beta,j}] = 0$ since individual measurements are independent, we will in general have $\mathrm{Cov}(\hat{A}_{\alpha}, \hat{A}_{\beta}) \neq 0$. As we are usually interested in estimating $\langle \hat{H} \rangle$ to within some fixed accuracy ε (in RMSE), the true sampling cost M of a measurement scheme is determined by the variance of $\bar{H}$ as:

$$M = N_A\, m = N_A\, \frac{\sqrt{\sum_{\alpha} \mathrm{Var}\, \hat{A}_{\alpha}}}{\varepsilon^2} = \mathcal{O}\left(\frac{1}{\varepsilon^2}\right) \tag{4.29}$$

The measurement scheme with lowest sampling cost is not necessarily given by the scheme with the smallest number of observable fragments N_A, and methods have been developed that try to minimise the cost directly by minimising the variance of the constructed estimator [126–128]. However, these methods are generally limited by their ignorance of $\mathrm{Var}(\hat{A}_{\alpha})$. The scaling of $M \sim \varepsilon^{-2}$ is an inherent statistical limitation of estimators constructed from repeated sampling, and is considered to be a considerable roadblock for applications of VQE to larger, chemically and industrially relevant molecules [129].

4.4.2 Quantum Phase Estimation

Quantum Phase Estimation (QPE) [48] provides an alternative route to energy measurement that is conceptually different from the fragment-based approach. Since it uses very deep circuits, it requires low error rates and long coherence time, making it unfeasible to implement on NISQ hardware. In its ideal form, QPE realises a projective measurement in the eigenbasis of the Hamiltonian and is therefore naturally a fault-tolerant (rather than NISQ) measurement scheme. It performs this measurement by encoding the eigenphase $e^{-iE_n t}$ of an eigenstate $\hat{H}|\psi_n\rangle = E_n|\psi_n\rangle$ in a k-bit ancilla register using repeated applications of controlled powers of the time evolution operator

$$U(t) \;=\; e^{-i\hat{H}t}, \tag{4.30}$$

Applying an inverse quantum Fourier transform on the ancilla register and subsequently performing a computational basis measurement returns a k-bit estimate of the phase $\phi_n = E_n t \bmod 2\pi$ and hence the energy E_n. If instead of an exact eigenstate QPE is performed on a general state $|\psi\rangle = \sum_n c_n|\psi_n\rangle$ ancilla register

measurement will return E_n with probability $|c_n|^2$ and collapse the state to $|\psi_n\rangle$.

In contrast to fragment-based measurement schemes, QPE can be seen to effectively implement a projective measurement directly on the energy eigenbasis (in the limit of $k \to \infty$), which can in turn be used to estimate expectation values $\langle \hat{H} \rangle$ for general states $|\psi\rangle$. The estimator $\bar{H}$ for $\langle \hat{H} \rangle$ formed from M repeated QPE measurements therefore has a variance given by:

$$\mathrm{Var}\big[\bar{H}\big] \;=\; \frac{\mathrm{Var}(\hat{H})}{M}. \tag{4.31}$$

4.5 Classical Optimisation

The choice of classical optimisation routine is pivotal to VQE performance. As is well known from standard optimisation tasks, the optimisation algorithm can greatly affect how efficiently the minimisation landscape is navigated and how quickly the candidate solution converges to the true minimum. An incredibly wide range of optimisation algorithms have been developed that are tailored to the specific problem instance and accuracy requirements [130]. For example, limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) [131] is often considered the state of the art for optimising deterministic black box functions to high accuracy if the gradient $\boldsymbol{g} = \frac{\partial E(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ can be analytically and efficiently calculated. In electronic structure theory conjugate gradients (CG) minimisation [132] is frequently employed to optimise the (energy of the) wavefunction in the SCF part of the calculation [26]. In machine learning, however, the adaptive moment estimation (ADAM) optimiser [133] with stochastic gradient evaluations has become a popular choice, since it not only avoids calculating expensive and exact gradients, but also prevents the optimiser from easily over-fitting the model on the training data.

The optimisation in VQE inherently encounters two main issues. As discussed in sec. 4.4, the expectation values, and hence the objective function of VQE the energy expectation $E(\boldsymbol{\theta}) = \langle \psi(\boldsymbol{\theta}) | \hat{H} | \psi(\boldsymbol{\theta}) \rangle$, cannot be evaluated exactly. Instead, we only have access to an unbiased statistical estimator $\mathbb{E}[\bar{H}] = \langle \hat{H} \rangle$ and as a result only an unbiased stochastic approximation of the true objective function $\mathbb{E}[\bar{E}(\boldsymbol{\theta})] = E(\boldsymbol{\theta})$. While in theory we recover the exact objective function $\bar{E}(\boldsymbol{\theta}) \to E(\boldsymbol{\theta})$ in the limit $M \to \infty$, in practice this limit is unreachable [129] and stochastic optimisation algorithms have been specifically developed to work directly with the stochastic approximation $\bar{E}(\boldsymbol{\theta})$. Furthermore, on classical computers derivatives $\frac{\partial E(\boldsymbol{\theta})}{\partial \theta_k}$ are often easily and efficiently computable – in electronic structure theory the electronic gradient can usually be calculated directly from $|\psi\rangle$ at negligible overhead compared

to an evaluation of $\langle\psi(\boldsymbol{\theta})|\hat{H}|\psi(\boldsymbol{\theta})\rangle$ – allowing the use of gradient-based algorithms with improved convergence rates such as CG in SCF calculations [26]. In VQE, however, the parametrisation of the ansatz circuit as a product of parametrised unitary operators, combined with our ignorance of $|\psi(\boldsymbol{\theta})\rangle$ makes calculating analytic gradients non-trivial and introduces additional overheads. Both of these complications make the choice of optimisation algorithm in VQE non-trivial, and many classical, as well as bespoke quantum minimisation algorithms have been developed and benchmarked on a variety of problem instances [134–136].

In this section we discuss finite difference gradients, as well as the parameter shift method for evaluating analytic gradients. We summarise the L-BFGS algorithm, a non-stochastic quasi-Newton optimisation algorithm, which we use in conjunction with parameter shift gradients for VQE calculations in our work. We also discuss the theory behind the SPSA optimisation algorithm, a stochastic gradient-free algorithm and hence a particularly common choice for optimising VQE circuits in realistic applications where shot budgets are limited, which we also consider in our calculations.

4.5.1 Circuit Gradients

Finite Differences

Finite-difference (FD) gradients are the simplest method of calculating derivatives of a black box objective function, since they do not require any additional knowledge or ability to perform calculations beyond evaluation of the objective function. Applied to the VQE objective function the FD gradient vector can be approximated using a symmetric finite difference formula as:

$$\frac{\partial}{\partial\theta_k}E(\boldsymbol{\theta}) \approx \frac{E\left(\boldsymbol{\theta}+h\cdot\boldsymbol{e}_k\right)-E\left(\boldsymbol{\theta}-h\cdot\boldsymbol{e}_k\right)}{2h}+\mathcal{O}(h^2) \tag{4.32}$$

where $\boldsymbol{e}_k$ are unit vectors corresponding to the direction θ_k. FD gradients become exact in the limit of $h\to 0$. Evaluation of symmetric FD gradients incurs a relatively large cost of $2P$ objective function evaluations for the full gradient vector for P parameters. This cost generally hinders application of FD gradients in optimisation tasks with high dimensional objective functions. Evaluation of FD gradients is further complicated by the presence of stochastic noise in the evaluation of the objective function $\bar{E}(\boldsymbol{\theta})$. While the estimator of the objective function is unbiased, $\mathbb{E}[\bar{E}(\boldsymbol{\theta})]=E(\boldsymbol{\theta})$, the finite step size leads to the FD gradient being a biased estimator

of the true gradient with RMSE given by:

$$\mathrm{RMSE}^2[\bar{g}_k] \approx \left(\frac{\partial_k^3 E(\boldsymbol{\theta}) \cdot h^2}{3!}\right)^2 + \frac{\sigma^2}{2h^2} \tag{4.33}$$

where $\sigma^2(\boldsymbol{\theta}) = \mathrm{Var}[\bar{E}(\boldsymbol{\theta})]$ is the variance of the objective function at $\boldsymbol{\theta}$ and we have assumed that $\sigma^2(\boldsymbol{\theta} + h \cdot \boldsymbol{e}_k) \approx \sigma^2(\boldsymbol{\theta} - h \cdot \boldsymbol{e}_k) \approx \sigma^2(\boldsymbol{\theta}) \equiv \sigma^2$. A smaller step size can be seen to trade-off a decrease in the systematic finite-difference approximation for an increase in the stochastic error. The requirement of tuning the step size h to minimise errors further hinders the application of FD gradients in optimisation tasks.

Parameter Shift

Evaluating exact/analytic derivatives of the energy $E(\boldsymbol{\theta}) = \langle\psi(\boldsymbol{\theta})|\hat{H}|\psi(\boldsymbol{\theta})\rangle$ – the objective function in the VQE optimisation – requires casting the derivative in terms of expressions that can be easily evaluated on the quantum computer. The parameter-shift (PS) rule [137] is a method for evaluating this gradient analytically. Most commonly considered ansatz circuits, and the ones considered in our work, are of the form $\hat{U}(\boldsymbol{\theta}) = \prod_k \hat{U}_k(\theta_k)$ where each individual unitary is a parametrised evolution of a Hermitian generator $\hat{G}_k$, $\hat{U}_k(\theta_k) = \exp(i\theta_k \hat{G}_k)$. For an ansatz circuit consisting of a single parameter, a direct differentiation of $\langle\psi(\theta_k)|\hat{H}|\psi(\theta_k)\rangle$ leads to:

$$\frac{\partial}{\partial\theta_k} E(\theta_k) = \langle\psi(\theta_k)|[-i\,\hat{G}_k\,\hat{H} + i\,\hat{H}\,\hat{G}_k]|\psi(\theta_k)\rangle = i\langle[\hat{H}, \hat{G}_k]\rangle \tag{4.34}$$

The PS rule assumes that the generator $\hat{G}_k$ has only two equally spaced eigenvalues $\pm r$. Under this assumption, the gradient can be expressed using expectation values of $\hat{H}$ of the original ansatz circuit with a shifted set of parameters:

$$\frac{\partial}{\partial\theta_k} E(\theta_k) = r\,\left(E\left(\theta_k + \frac{\pi}{4r}\right) - E\left(\theta_k - \frac{\pi}{4r}\right)\right) \tag{4.35}$$

The same expression can be shown to hold true for a multi-parameter ansatz circuit:

$$\frac{\partial}{\partial\theta_k} E(\boldsymbol{\theta}) = r\,\left(E\left(\boldsymbol{\theta} + \frac{\pi}{4r}\boldsymbol{e}_k\right) - E\left(\boldsymbol{\theta} - \frac{\pi}{4r}\boldsymbol{e}_k\right)\right) \tag{4.36}$$

The PS rule applies in particular to Pauli rotation gates $\exp(i\theta\hat{P})$ since Pauli products $\hat{P}$ are Hermitian involutory operators $\hat{P}^2 = \mathbb{I}$ and hence have eigenvalues ± 1. Since Pauli rotation gates are the most frequently encountered parametrised gates this allows us to differentiate the VQE objective function $E(\boldsymbol{\theta})$. Even if the

ansatz circuit is defined in terms of more general parametrised gates $\exp(i\theta\hat{H})$, implementing this gate generally requires transpilation to Pauli rotation gates first. The PS rules are therefore generally applicable to (transpiled) ansatz circuits. Repeated/correlated parameters in the ansatz circuit can be trivially accounted for by treating each parameter as unique and applying the chain rule to combine gradients appropriately.

Since PS gradients are conceptually very similar to FD gradients, both calculated from discrete evaluations of the objective functions, they also have a similar computational cost. Calculating the full PS gradient vector for the P parameters in the circuit requires 2 evaluations for each of the P_G Pauli rotation gates in the (transpiled) ansatz circuit. Since generally $P_G \geq P$ (e.g. $\exp(i\theta(c_i^\dagger c_j + \text{h.c.}))$ transpiles to 2 Pauli rotation gates in the JW-encoding) the PS gradient can be seen to be at least as expensive as the FD gradient, requiring $2P_G$ calls to the objective function. In general the transpilation overhead is constant and $P_G = \mathcal{O}(P)$. However, unlike FD gradients, PS gradients are analytic/exact since the PS rule, equ. (4.36), does not involve any approximation. They also do not require small step sizes h and hence do not suffer from the issue of amplifying the stochastic noise in the objective function $\bar{E}(\theta)$ if h is not tuned appropriately.

Extensions of the parameter-shift rules to more general gates/generators have been developed that are oftentimes able to reduce the implementation cost by a constant factor over a naive application of the standard parameter-shift rules to the Pauli gates in the transpiled circuit [138, 139]. Higher-order derivatives can also be calculated by nested application of parameter-shift rules [140]. While it is technically possible to implement arbitrary parametrised unitaries to which parameter shift rules in general do not apply, these are not frequently encountered in NISQ applications and are generally not considered as potential ansatz circuits.

4.5.2 L-BFGS

The limited-memory Broyden–Fletcher–Goldfarb–Shanno(L-BFGS) algorithm [130, 131] is a quasi-Newton gradient-based optimiser. Newton's method forms a local quadratic model $\tilde{E}(\boldsymbol{\theta})$ of the true objective $E(\boldsymbol{\theta})$ using the gradient $\boldsymbol{g}^{(t)} \equiv \boldsymbol{g}(\boldsymbol{\theta}^{(t)})$ and Hessian $\boldsymbol{H}^{(t)} \equiv \boldsymbol{H}(\boldsymbol{\theta}^{(t)})$ at the current candidate iterate $\boldsymbol{\theta}^{(t)}$:

$$\tilde{E}(\boldsymbol{\theta}) = E(\boldsymbol{\theta}^{(t)}) + \boldsymbol{g}^{(t)}\delta\boldsymbol{\theta} + \frac{1}{2}\delta\boldsymbol{\theta}^T\boldsymbol{H}^{(t)}\delta\boldsymbol{\theta} \tag{4.37}$$

A new candidate iterate $\boldsymbol{\theta}^{(t+1)}$ is formed by minimising the quadratic model under the assumption of a positive definite $\boldsymbol{H}^{(t)}$:

$$\boldsymbol{\theta}^{(t+1)} = \text{argmin}(\tilde{E}(\boldsymbol{\theta})) = \boldsymbol{\theta}^{(t)} - \left(\boldsymbol{H}^{(t)}\right)^{-1} \boldsymbol{g}^{(t)} \tag{4.38}$$

L-BFGS (and quasi-Newton methods in general) reduce the computational cost of this method by iteratively approximating $\boldsymbol{H}^{(t)}$ and hence the exact Newton step, equ. (4.38), using gradient information at the previous iterates $\{\boldsymbol{g}^{(j)}; j = 0, \ldots, t-1\}$ instead of calculating it directly. The update rule for the Hessian approximation $\tilde{H}^{(t)}$ in L-BFGS is based on the standard BFGS update rule:

$$\tilde{\boldsymbol{H}}^{(t+1)} = (\boldsymbol{V}^{(t)})^T \, \tilde{\boldsymbol{H}}^{(t)} \, \boldsymbol{V}^{(t)} + \boldsymbol{\rho}^{(t)} \, \boldsymbol{s}^{(t)} \, (\boldsymbol{s}^{(t)})^T \tag{4.39}$$

$$\boldsymbol{\rho}^{(t)} = \frac{1}{(\boldsymbol{y}^{(t)})^T \boldsymbol{s}^{(t)}} \qquad \boldsymbol{V}^{(t)} = I - \boldsymbol{\rho}^{(t)} \boldsymbol{y}^{(t)} (s^{(t)})^T \tag{4.40}$$

In L-BFGS only a limited number of steps are kept in the working memory and the update $\tilde{\boldsymbol{H}}^{(k+1)}$ is reconstructed at each step based only on the last m iteration steps. This greatly decreases the memory requirements of the optimisation algorithm without generally affecting its performance [130, 131]. L-BFGS (and quasi-Newton algorithms in general) also do not perform the exact Newton step of equ. (4.38) within the approximation $\boldsymbol{H}^{(t)} \rightarrow \tilde{\boldsymbol{H}}^{(t)}$, but instead perform a line search over possible step lengths $a^{(t)}$ to find a step with sufficient objective function and gradient reduction to improve the stability and convergence of the iterates:

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - a^{(t)} \left(\tilde{\boldsymbol{H}}^{(t)}\right)^{-1} \boldsymbol{g}^{(t)} \tag{4.41}$$

where $a^{(t)}$ is chosen such that the Wolfe conditions are satisfied for some $0 < c_1 < c_2 < 1$:

$$E\left(\boldsymbol{\theta}^{(t+1)}\right) \leq E(\boldsymbol{\theta}^{(t)}) + c_1 \, (\boldsymbol{\theta}^{(t+1)} - \boldsymbol{\theta}^{(t)})^T \, \boldsymbol{g}^{(t)} \tag{4.42}$$

$$-(\boldsymbol{\theta}^{(t+1)} - \boldsymbol{\theta}^{(t)})^T \, \boldsymbol{g}^{(t+1)} \leq -c_2 \, (\boldsymbol{\theta}^{(t+1)} - \boldsymbol{\theta}^{(t)})^T \, \boldsymbol{g}^{(t)} \tag{4.43}$$

As a quasi-Newton algorithm, L-BFGS generally displays super-linear convergence speeds and is able to efficiently minimise a wide variety of objective functions. It achieves this with negligible overhead compared to other gradient-based algorithms, since it avoids direct evaluation of the high-dimensional Hessian required for Newton's method. It becomes especially efficient when the required gradients can be calculated using reverse mode automatic differentiation techniques at a constant overhead over a single call to the objective function. It is therefore generally con-

sidered the work-horse of numerical optimisation when high-accuracy solutions are sought after and analytic gradients can be efficiently evaluated [130, 131].

L-BFGS is frequently applied in ideal simulations of VQE calculations, where $E(\boldsymbol{\theta})$ can be evaluated analytically, i.e. without stochastic noise, and where the cost of calculating $\boldsymbol{g}$ from PS gradients is not of concern [82]. In realistic VQE applications, however, the use of L-BFGS can encounter issues. As discussed in sec. 4.5.1, on quantum hardware gradients can only be calculated using the PS rules with a cost scaling as $2P_G$, a similar scaling to an application of FD gradients. This makes an application of L-BFGS considerably more expensive than might be expected from standard classical applications. Furthermore, the stochastic nature of the cost function is not particularly well suited for the deterministic L-BFGS algorithm [136].

4.5.3 SPSA

The simultaneous perturbation stochastic approximation (SPSA) algorithm [141] is a stochastic optimiser that forms a stochastic approximation of the gradient using only two function evaluations per iteration, and uses this gradient approximation to progress towards the minimum, in a similar way as the well known stochastic gradient descent (SGD) algorithm. At each iteration SPSA forms the gradient estimate by using a central FD approximation in each component using only a single step along a randomly chosen direction $\boldsymbol{\Delta}^{(k)}$, perturbing all parameters simultaneously

$$\bar{\boldsymbol{g}}^{(t)} = \frac{E(\boldsymbol{\theta}^{(t)} + c^{(t)}\boldsymbol{\Delta}^{(k)}) - E(\boldsymbol{\theta}^{(t)} - c^{(t)}\boldsymbol{\Delta}^{(t)})}{2\,c^{(t)}}(\boldsymbol{\Delta}^{(t)})^{-1} \tag{4.44}$$

where $(\boldsymbol{\Delta}^{(t)})^{-1}$ is the element-wise inverse of $\boldsymbol{\Delta}^{(t)}$. In contrast, a standard FD approximation of the gradient applies a central FD approximation to each element individually, perturbing each parameter independently. The perturbation direction $\boldsymbol{\Delta}^{(k)}$ is generally chosen by sampling each entry from ± 1 with equal probability. The candidate iterate is then updated by taking a standard gradient descent step along the gradient estimate $\bar{\boldsymbol{g}}^{(t)}$:

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - a^{(t)}\hat{\boldsymbol{g}}(\boldsymbol{\theta}_k), \tag{4.45}$$

The standard SPSA algorithm introduces parametrised forms for the perturbation and descent step size, $c^{(t)}$ and $a^{(t)}$ respectively, as sequences decaying as a power law

in the iteration step t, in terms of a set of tunable hyper-parameters a, A, α, c, γ::

$$a^{(t)} = \frac{a}{(t+1+A)^{\alpha}}, \quad c_k = \frac{c}{(t+1)^{\gamma}}, \tag{4.46}$$

SPSA has in particular been developed and found applications as an algorithm for minimising stochastic problems for which only the stochastic objective function can be evaluated and for which a direct method for evaluating a (stochastic) gradient does not exist. Similarly to SGD, it generally only achieves sub-linear rates of convergence since it not only does not use curvature information, but is also limited by the stochastic noise in the objective function. However, in comparison to an FD approximation of the gradient, the direct competitor to SPSA for the considered setting of only being able to evaluate stochastic objective functions, it becomes asymptotically more efficient [141]. SPSA only requires 2 function evaluations per iteration step, in stark contrast to the $2P$ evaluations required for a more standard FD approximation for the gradient, in which each element of $(\boldsymbol{g})_k$ is approximated using an independent perturbation of θ_k. Furthermore, the use of an explicitly stochastic algorithm to minimise a stochastic objective function generally guarantees convergence to a local minimum in expectation in spite of the stochastic objective function, since stochastic errors average out over multiple iterations.

SPSA is frequently applied in realistic VQE calculations, where only the stochastic energy estimate $\bar{E}(\boldsymbol{\theta})$ is available and where the cost of calculating $\boldsymbol{g}$ from PS gradients is not negligible [135, 136]. In this setting SPSA is attractive because it is able to avoid an explicit, expensive PS gradient evaluation, as well as being inherently well suited to the minimisation of a stochastic objective function.

Chapter 5

Ideal VQE Simulations of *Trans*-polyacetylene

In this chapter we study idealised, noiseless variational quantum eigensolver (VQE) simulations of *trans*-polyacetylene (TPA), i.e. simulations in which circuits are executed with perfect fidelity and hence no circuit noise is present, and where expectation values can be measured exactly and hence no shot noise is present.

As previously discussed, we approach these VQE calculations from the viewpoint of an electronic structure practitioner trying to perform useful simulations of realistic systems. However, near-term noisy intermediate-scale quantum (NISQ) devices remain limited by qubit number, coherence, and noise, which constrains ab initio VQE to very small molecules and minimal basis sets. In contrast to the standard small and simple chemical benchmark systems frequently considered in VQE, TPA can be captured by the Su–Schrieffer–Heeger (SSH) model, a compact, chemically motivated, one-dimensional Hubbard-like lattice model, as discussed in sec. 3.2. This greatly reduces the complexity of the VQE calculation and makes it possible to consider comparatively large systems within a NISQ-relevant qubit and coherence budget. At the same time it is not a trivial benchmark system: its key physical properties depend on correlation and on the interplay between electronic structure and geometry, and, as has been shown in sec. 3.3, cannot be reliably captured, neither quantitatively nor qualitatively, by mean-field treatments such as density functional theory (DFT). Finally, TPA is a system where the determination of the correct ground-state geometry is not optional. As discussed in sec. 3, even an only qualitatively correct prediction of its properties requires an accurate prediction of the ground-state geometry. As such, simulations of TPA and the SSH model not only allow but necessitate the evaluation of molecular forces in the VQE framework, an application of VQE that has received relatively little attention so far, but, as

discussed in sec. 2.5, is of great interest to electronic structure practitioners. A full geometry optimised calculation is beyond the scope of this thesis and the available computational classical and quantum resources. We will therefore instead focus on the calculation of the molecular force for a fixed, perfectly dimerised instance of the SSH model using VQE.

The goal of the work in this chapter is twofold. Firstly, we quantify the quantum resources required for accurate simulations of TPA using the SSH model in the noiseless VQE calculations. We consider not only the prediction of accurate ground-state energies, but in particular also accurate molecular forces, with a viewpoint from standard electronic structure theory, where, as discussed in sec. 2.2, it is well established that, based on the variational principle, predicting sufficiently accurate molecular forces is in general considerably more difficult than predicting sufficiently accurate ground-state energies. Secondly, the same results establish a noiseless baseline that isolates algorithmic effects from circuit and shot noise induced errors. This baseline is the reference point for the noisy calculations conducted in sec. 6.

We begin by clearly restating the problem that we are trying to solve, the SSH model previously discussed in sec. 3.2. We then continue by presenting the specific VQE implementation chosen: we survey and discuss commonly considered approaches for simulations of Hubbard-like models, motivate our choice of methodology, and discuss the implementation specialised to our model. Following this, we discuss the evaluation of molecular forces in VQE, surveying relevant publications, before discussing the specific approach chosen for our calculations, as well as detailing the implementation within VQE. Lastly we present the numerical results of our VQE calculations for energies and molecular forces for our model using the described implementation.

5.1 The SSH Model

As discussed in sec. 3.2, the physics of TPA is appropriately described by the SSH model under the inclusion of Hubbard U and V terms, and it is therefore this SSH model that we are trying to simulate. A complete electronic structure calculation of the SSH model necessitates the use of an analytic molecular force based geometry optimisation for all but the smallest model instances, since the computational complexity scaling of a direct scan of the potential energy surface (PES) would scale as $\mathcal{O}(2^N)$ or similarly finite-difference (FD) energy derivatives scale as $\mathcal{O}(N)$ for an N-site SSH model. While such approaches could be made feasible for predicting the ground-state geometry of a symmetrically dimerised SSH model by restricting the

parameter space to a single bond-length coordinate, such approaches would fail to capture any symmetry broken states such as the solitons described in sec. 3.2. It would also defeat part of the motivation for the choice of TPA for the work in this thesis, since it is a model system that specifically allows more realistic molecular force calculations to be explored. In realistic simulations of materials, systems will typically have much lower symmetry and necessitate molecular force calculations. However, a full geometry optimised calculation is beyond the scope of this thesis and the available computational classical and quantum resources. We will therefore instead be focusing solely on the calculation of the molecular force for a fixed, perfectly dimerised instance of the SSH model using VQE.

The model instance that we are considering is given by the even site number ($N = 2N_{\text{uc}}$, with N_{uc} the number of unit cells in the system), static ($\dot{u} = 0$), perfectly dimerised ($u_n = (-1)^n u$), periodic ($c_N^\dagger \equiv c_0^\dagger$ and $u_N \equiv u_0$), interacting ($U \neq 0, V \neq 0$), atom-centred SSH model presented in sec. 3. Defining the staggered hoppings as $t_{e/o} \equiv t \pm 2\alpha u$ this model can be expressed in terms of a more standard Hubbard-like notation as:

$$\begin{aligned}\hat{H} = &- \sum_{\substack{i=0,\sigma \\ i \text{ even}}}^{N-2} t_{\text{e}} \left(\hat{c}_{i,\sigma}^\dagger \hat{c}_{i+1,\sigma} + \hat{c}_{i+1,\sigma}^\dagger \hat{c}_{i,\sigma} \right) - \sum_{\substack{i=1,\sigma \\ i \text{ odd}}}^{N-1} t_{\text{o}} \left(\hat{c}_{i,\sigma}^\dagger \hat{c}_{i+1,\sigma} + \hat{c}_{i+1,\sigma}^\dagger \hat{c}_{i,\sigma} \right) \\ &+ \sum_{i=0}^{N-1} U \hat{n}_{i,\uparrow} \hat{n}_{i,\downarrow} + \sum_{i=0,\sigma,\sigma'}^{N-1} V \hat{n}_{i,\sigma} \hat{n}_{i+1,\sigma'}\end{aligned} \tag{5.1}$$

Although, as discussed, we consider calculations at a fixed molecular geometry, the atoms will still experience a molecular force (and adjust according to this molecular force if the geometry was allowed to vary). While the above considered form of $\hat{H}$, equ. (5.1), hides the dependence on the atomic positions u_i, it is implicit in the hopping terms $t_{\text{e/o}} = t_{\text{e/o}}(\{u_i\})$. As discussed in sec. 2.5, if the exact ground state, $|\psi_0\rangle$ of $\hat{H}$ is known, the molecular force can be evaluated according to the HF theorem, equ. (2.50), as:

$$F_i = \langle \psi_0 | \hat{F}_i | \psi_0 \rangle \tag{5.2}$$

using the HF force operator, $\hat{\boldsymbol{F}}$, which for the SSH model is given by:

$$\hat{F}_i \equiv -\frac{d\hat{H}}{du_i} = \sum_\sigma \alpha \left(\hat{c}_{i-1,\sigma}^\dagger \hat{c}_{i,\sigma} + \hat{c}_{i-1,\sigma}^\dagger \hat{c}_{i,\sigma} \right) - \sum_\sigma \alpha \left(\hat{c}_{i,\sigma}^\dagger \hat{c}_{i+1,\sigma} + \hat{c}_{i+1,\sigma}^\dagger \hat{c}_{i,\sigma} \right) \tag{5.3}$$

We note that as discussed in sec. 2.5.3, the Pulay forces do not enter our expression for the molecular force. While our basis could be considered moving – the p orbitals

that are used to derive the SSH model are centred on each C-atom and are therefore displaced together with the atomic positions – the SSH model is an orthogonal tight-binding model in which the orbitals are by assumption orthogonal at any geometry, hence causing the Pulay force contribution to vanish. We also note that $\hat{F}_i$ is independent of the molecular geometry as a result of the linear parametrisation of the model with respect to u_i.

The model is parametrised using the constants:

$$\begin{aligned} t &= 2.5\,\text{eV} & U &= 2.5\,t_0 = 6.25\,\text{eV} \\ \alpha &= 4.1\,\text{eV/Å} & V &= \frac{U}{4} = 1.5625\,\text{eV} \\ u &= 0.04\,\text{Å} & K &= 40\,\text{eV/Å}^2 \end{aligned} \tag{5.4}$$

which as discussed in sec. 3.2 are a set of well known and frequently used, previously parametrised values that approximately represent the experimentally observed dimerisation and molecular properties of TPA.

Furthermore, as discussed in sec. 3.2 the SSH model is symmetric under a number of operations which we will summarise here again, as they will be relevant in following discussion:

Particle number symmetry As a physical electronic-structure Hamiltonian constructed from number conserving fermionic one- and two-body interaction terms it conserves the total particle number $\hat{N} = \sum_{i,\sigma} \hat{n}_{i,\sigma}$, i.e. $[\hat{H}, \hat{N}] = 0$.

Spin symmetry Since all hoppings are spin-independent and the interaction terms depend only on number operators, the Hamiltonian is invariant under spin rotations and $\hat{S}^z = \frac{1}{2}\sum_i (\hat{n}_{i,\uparrow} - \hat{n}_{i,\downarrow})$ is conserved, as well as more generally $[\hat{H}, \hat{S}^2] = 0$.

Translation symmetry Owing to the perfect dimerisation, uniform interaction terms, even number of sites and periodic boundary conditions, the model is periodic under translation by two lattice sites $\hat{c}_{i,\sigma} \mapsto \hat{c}_{i+2,\sigma}$, and hence commutes with the translation operator $[\hat{H}, \hat{T}_2] = 0$.

Inversion symmetry Similarly, as a result of perfect dimerisation, uniform interaction terms and even number of sites, the system is symmetric under inversion about any bond, e.g. for the central bond $\hat{c}_{i,\sigma} \mapsto \hat{c}_{N-1-i,\sigma}$ and hence $[\hat{H}, \hat{\mathcal{P}}_I] = 0$ for the inversion operator $\hat{\mathcal{P}}_I$.

5.2 VQE Implementation

5.2.1 Fermion to Qubit Mapping

As discussed in sec. 4.2, the Jordan-Wigner (JW) encoding is one of the most commonly used fermion to qubit encodings in quantum chemistry applications. However, it suffers from the presence of JW-strings, highly non-local Pauli operators required to enforce the anti-symmetry of the fermionic wavefunction. While a judicious choice of JW-encoding (in conjunction with specialised circuit implementations and ansatz designs), or even the use of more complex encodings can help overcome these limitations for specific models, problem instances, ansatz circuits, and/or device connectivities (see sec. 4.2), these tend to be highly specific and specialised. In the spirit of The Principle of Generality we use the JW-encoding with what can be considered the default implementation details (both in terms of approaches found in the literature as well as in quantum simulation software packages), which we describe in detail for our specific model instance below. We stress that this choice is by no means optimal, and that encodings enabling better VQE results are expected to be possible – particularly when considering circuit and shot noise – by specialising the encoding to the specific model or problem instance, ansatz circuit, circuit decomposition, device topology, and noise characteristics, and vice versa.

As we will discuss, the JW-encoding turns out to be considerably more efficient for our particular model and ansatz circuit considered compared to a more general electronic structure system and ansatz, avoiding most of the JW-string overhead that is problematic in NISQ applications due to the presence of gate noise. However, we also discuss potential issues that are encountered with this approach as well as potential mitigations using more carefully chosen alternative encodings. In this context we will also briefly discuss other approaches mentioned in sec. 4.2 and how they are expected to lose their comparative advantage for our problem instance. While we do discuss and contrast these encodings in terms of the approximate CNOT count complexity of standard circuit implementations, we delay a more explicit discussion of the effects of noise to sec. 6, noting that as far as the noiseless calculations in this chapter are concerned, all encodings produce equivalent results due to the absence of gate noise.

Jordan-Wigner Strings

The structure of our TPA model – one-dimensional with highly local interactions – allows for an efficient application of the JW-encoding to the Hamiltonian in which JW-strings are unproblematic, which combined with our choice of ansatz circuit

(see sec. 5.2.2) also allows for an efficient mapping of the VQE circuit using the JW-encoding. The fundamental issue of the JW-strings arises from the requirement of specifying an ordering of orbitals and an ordering of qubits. Under the mapping of equ. (4.5) only fermionic operators acting on orbitals local within this ordering will get mapped to qubit operators acting on qubits local within their respective ordering, since (assuming an even product of fermionic operators) the JW-strings will cancel (see equ. (4.8) noting that $Z^{n<n+1} = I$):

$$\hat{c}_n^\dagger \hat{c}_{n+1} \;\mapsto\; \frac{1}{4}(X_n - iY_n)(X_{n+1} + iY_{n+1}) \tag{5.5}$$

Any other fermionic operators get mapped to qubit operators acting on non-local qubits resulting in highly non-local JW-strings (see. equ. (4.8)):

$$\hat{c}_n^\dagger \hat{c}_{n+r} \;\mapsto\; \frac{1}{4}(Z_{n+1}Z_{n+2}\dots Z_{n+r-1})(X_n - iY_n)(X_{n+r} + iY_{n+r}) \tag{5.6}$$

The degree of non-locality is quantified by the Pauli weight, w – the number of qubits on which the Pauli string acts non-trivially. The cost to implement a gate involving a general Pauli string scales directly with w, requiring $\mathcal{O}(w)$ CNOT gates and a resulting circuit of depth $\mathcal{O}(w)$ using the standard "CNOT ladder" circuit implementation of exponentiated Pauli strings $\exp(i\theta\hat{P})$ [83]. The JW-encoding on a general electronic structure Hamiltonian will map some fermionic local operators to Pauli operators with high Pauli weight; for a general electronic structure ansatz circuit such as UCCSD in which excitations involving all possible $\mathcal{O}(M^4)$ orbital combinations of fermionic operators occur and need to be implemented, the average Pauli weight of the UCCSD excitation operators scales as $\mathcal{O}(M)$, hence introducing an average $\mathcal{O}(M)$ cost to implement each fermionic operator and leading to a standard, unoptimised scaling of gate count and depth for the ansatz circuit of $\mathcal{O}(M^5)$ [82]. Even in systems in which only spatially local fermionic orbitals interact and need to be implemented as part of the circuit construction, costly JW-strings will appear, since in two (or higher) dimensional systems the two (or higher) dimensional locality cannot be represented locally in the one dimensional orbital ordering. For a 2D Hubbard square of side length L, for example, some local fermionic operators will inevitably map to non-local Pauli operators resulting in an average Pauli weight of $\mathcal{O}(L)$, again leading to an unoptimised scaling of gate count and depth of $\mathcal{O}(L^3)$ [86]. It should be stressed that this is an issue independent of the device connectivity and will arise even on an all-to-all connected device, since JW-strings are not an issue of limited connectivity, but rather of enforcing the wavefunction anti-symmetry and the resulting requirement for implementing/measuring highly non-local parity

operators. It is worth noting that qubit operators that are local in the JW-encoding are not inherently required to be physically local on the quantum computer; limited connectivity on the device can introduce additional complications. However, since any physical device is required to have at least one-dimensional connectivity, it is always possible to choose a JW-encoding that also respects the locality of the device connectivity. As such, the locality constraints introduced by the JW-encoding can be considered to be the limiting constraints for a worst-case/naive implementation cost. However, since our SSH model, equ. (5.1), is a one-dimensional locally interacting model, and since only these local interactions need to be encoded as part of the ansatz circuit (see sec. 5.2.2) these issues of preserving locality are in principle not present since the dimensionality of the fermionic system and the orbital ordering match and the locality can hence be preserved through the mapping.

Encoding the TPA Model

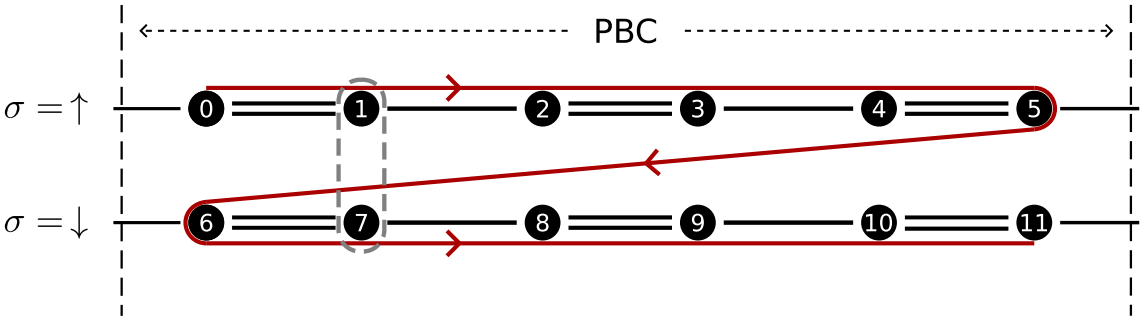


Figure 5.1: Jordan-Wigner (JW) orbital ordering used to encode the $N = 6$ SSH Hamiltonian on qubits. Upper and lower orbitals correspond to spin-up and spin-down orbitals for each lattice site. The red "string" and the numbering indicate the ordering chosen.

The specific JW-ordering of fermionic orbitals chosen for our model is depicted in fig. 5.1. First all spin up ($\sigma = \uparrow$) fermionic orbitals are enumerated according to their natural ordering, followed by all spin down ($\sigma = \downarrow$) again according to their natural ordering, so that we identify fermionic and qubit indices as $(i, \uparrow) \to i$ and $(i, \downarrow) \to N + i$. This is the default implementation provided in QISKIT Nature [142] and a commonly chosen ordering in the literature (if it is explicitly specified) when applying the JW-encoding [86, 90, 95, 106, 112, 143]. Under this mapping the Hamiltonian can be transformed using the JW-transformations specific to our model:

$$\hat{c}_i^\dagger \hat{c}_{i+1} + \hat{c}_{i+1}^\dagger \hat{c}_i \;\mapsto\; \frac{1}{2}(X_i X_{i+1} + Y_i Y_{i+1}) \tag{5.7}$$

$$\hat{n}_i \;\mapsto\; \frac{1}{2}(I - Z_i) \tag{5.8}$$

As can be seen in fig. 5.1, this encoding does not technically preserve the locality of all interaction terms, since the Hubbard U and V interaction terms are local in the fermionic picture but get mapped to non-local qubit operators, e.g. for $U\hat{n}_{i,\uparrow}\hat{n}_{i,\downarrow}$ we have to map the indices as $(i,\uparrow) \to i$ and $(i,\downarrow) \to N+i$. However, since the interaction terms in our model take the form of products of occupation number operators $\hat{n}$, the JW-strings for these terms cancel. As alluded to previously, since these terms are nonetheless not local in the JW-ordering, it is possible for the implementation of these to incur an $\mathcal{O}(N)$ SWAP gate overhead if the device connectivity is limited and further assumed to be congruent with the JW-ordering. Similarly, since we are working with periodic boundary conditions, we are required to identify $\hat{c}_N^\dagger \equiv \hat{c}_0^\dagger$. As a result, the Hamiltonian contains a PBC hopping term $\hat{c}_{N-1}^\dagger \hat{c}_0 + \text{h.c.}$ (see fig. 5.1), whose locality is not preserved and which therefore gets mapped to non-local qubits, incurring a high Pauli weight JW-string:

$$\hat{c}_{N-1}^\dagger \hat{c}_0 + \hat{c}_0^\dagger \hat{c}_{N-1} \;\mapsto\; \frac{1}{2}(Z_1 Z_2 \dots Z_{N-2})(X_0 X_{N-1} + Y_0 Y_{N-1}) \tag{5.9}$$

While, as discussed, one-dimensional model locality can be represented appropriately by locality in the one-dimensional orbital ordering, this only applies to open boundary conditions since the JW-encoding is not defined periodically. PBCs fundamentally change the topology of the system and prevent us from mapping every term to a local qubit operator, even in a 1D system. Unlike the cross-spin interactions, this term incurs a $\mathcal{O}(N)$ gate and depth overhead independent of device connectivity as a result of the non-local JW-string. The encoding of the PBC introduces an apparent asymmetry in the treatment of the hopping terms. As previously mentioned, in the noiseless regime all encodings are equivalent and exact, and in this regime no symmetries of the model get broken as a result of this asymmetric encoding. The Hamiltonian (or any other symmetric operators) will remain symmetric with respect to correctly encoded symmetry operators. We note that since the (fermionic) translation operator is built from a series of fermionic swap gates [144]:

$$\hat{F}_{pq} = 1 + \hat{c}_p^\dagger \hat{c}_q + \hat{c}_q^\dagger \hat{c}_p - \hat{n}_p - \hat{n}_q, \tag{5.10}$$

the encoded translation operator does not encode directly to a cyclic qubit permutation, since a simple qubit translation can be seen to not commute with the JW-encoded Hamiltonian, equ. (5.11). However, in a noisy simulation setting the asymmetry in the encoding is likely to introduce an asymmetry in the system, since the highly non-local PBC hopping will have a considerably higher error rate than all other local interaction terms. In a simple model in which the error rate when im-

plementing a term is directly proportional to its Pauli-weight, $p \sim w$, the PBC term will carry an error rate $p_{\text{PBC}} = \mathcal{O}(N)$, while all other terms will carry a constant error $p = \mathcal{O}(1)$. Lastly, we note that while it might be expected that the impact of the PBC term, both in terms of the cost to implement and its noise-induced asymmetry, becomes negligible in the thermodynamic limit of $N \to \infty$ – in the same way that the effects of the choice of BC on the energy of a model become negligible [52] – this is not guaranteed, since both the cost and noise asymmetry induced by the PBC hopping increase with N .

Under this encoding the fully encoded qubit Hamiltonian for our model can be expressed as:

$$\begin{aligned}
\hat{H} = &- \sum_{\substack{i=0 \\ i \text{ even}}}^{N-2} \frac{t_e}{2}(X_iX_{i+1} + Y_iY_{i+1}) - \sum_{\substack{i=1 \\ i \text{ odd}}}^{N-3} \frac{t_o}{2}(X_iX_{i+1} + Y_iY_{i+1}) \\
&- (Z_1Z_2 \dots Z_{N-2})(X_{N-1}X_0 + Y_{N-1}Y_0) \\
&- \sum_{\substack{i=N \\ i \text{ even}}}^{2N-2} \frac{t_e}{2}(X_iX_{i+1} + Y_iY_{i+1}) - \sum_{\substack{i=N+1 \\ i \text{ odd}}}^{2N-3} \frac{t_o}{2}(X_iX_{i+1} + Y_iY_{i+1}) \\
&- (Z_{N+1}Z_{N+2} \dots Z_{2N-2})(X_{2N-1}X_N + Y_{2N-1}Y_N) \\
&+ \sum_{i=0}^{N-1} \frac{U}{2}(I - Z_i)(I - Z_{i+N}) \\
&+ \sum_{i=0}^{N-1} \frac{V}{2}(I - Z_i)(I - Z_{i+1}) + \sum_{i=0}^{N-1} \frac{V}{2}(I - Z_{i+N})(I - Z_{i+N+1}) \\
&+ \sum_{i=0}^{N-1} \frac{V}{2}(I - Z_i)(I - Z_{i+N+1}) + \sum_{i=0}^{N-1} \frac{V}{2}(I - Z_{i+N})(I - Z_{i+1})
\end{aligned} \tag{5.11}$$

Alternative Encodings

As discussed in sec. 4.2.2, more complex encodings to reduce circuit depths and/or qubit count have been developed. However, in addition to our choice of the JW-encoding following The Principle of Generality, the alternative encodings are also expected to lose their comparative advantage due to the absence of non-local JW-terms in our model, as well as the relative simplicity of our model Hamiltonian. On the one hand, the ansatz/circuit designs and fermionic encodings specifically developed to ameliorate the JW-strings [11, 88–90, 96–100] are, in addition to being highly specialised, also unlikely to provide any significant circuit depth reduction due to the absence of the relevant, costly terms, potentially even increasing the Pauli weight of operators encountered in the model. On the other hand, while encodings

that inherently account for and preserve the symmetries of the system, e.g. particle number conservation or spatial symmetries, are generally applicable and not specific to certain problem instances, they will break the advantages that come from the locality of the fermionic Hamiltonian under consideration. Particle number conserving encodings [103] generally come at the expense of encoding each fermionic operator in a much larger number of qubit operators, removing any advantage resulting from the favourable $N_P = \mathcal{O}(M)$ sparsity/locality induced scaling of the Hamiltonian for implementing the ansatz circuit and measuring the energy (see sec. 5.2.2 and 5.2.3). However, for our model in particular, it would in addition greatly increase the Pauli weight of each qubit operator since JW-strings are originally mostly absent as a result of the locality of the JW-encoding, $w = \mathcal{O}(1)$. Symmetry conserving encodings [101, 102] first require the formation of symmetrised orbitals, which in our case would break the locality of the fermionic Hamiltonian, since although the Hamiltonian is symmetric, the tight-binding orbitals themselves are not (see sec. 5.1), and formation of symmetrised orbitals would lead to delocalisation of the orbitals and hence the interactions. Again, this removes any advantage from the original sparsity/locality of the Hamiltonian for circuit implementation and energy measurement, as well as the original resulting simplifications/cancellations of most JW-strings. It should be noted that these issues would however be less severe/not present in simulations of a general, i.e. non-local/dense, electronic structure Hamiltonians, where the Hamiltonian is considerably more dense to begin with, $N_P = \mathcal{O}(M^4)$, and where high Pauli weight strings are already present in the JW-encoding, $w = \mathcal{O}(M)$. The cost incurred by a particle conserving encoding is hence greatly reduced. Similarly, in this case symmetric orbitals can be formed without affecting the complexity of the Hamiltonian terms present (and usually are already formed as part of the molecular orbital construction from e.g. Hartree-Fock calculations).

5.2.2 Ansatz Circuit

The ansatz circuit which we use to simulate our model is the Hamiltonian variational ansatz (HVA). As discussed in sec. 4.3, the HVA has been successfully applied to a variety of spin-lattice and Hubbard-like models, not only because it provides a model specific and physically motivated ansatz that can be designed to conserve many of the symmetries of the system, but also because the sparse nature of these models allows for a relatively cheap implementation of each layer of the ansatz compared to a full electronic structure Hamiltonian – all while the repeated, layered structure enables a systematic convergence to the ground state. As discussed in sec. 4.3.1 the

general HVA framework for a Hamiltonian $\hat{H}$ implements a circuit of the form:

$$|\psi(\boldsymbol{\theta})\rangle = \hat{U}(\boldsymbol{\theta})|\psi_i\rangle = \prod_{k=1}^{L} \prod_{\alpha} e^{-i\,\theta_{k,\alpha}\,\hat{H}_\alpha}|\psi_i\rangle \tag{5.12}$$

where the parametrised unitary evolution $U(\boldsymbol{\theta})$ is constructed from Hamiltonian fragments chosen to satisfy $\sum_\alpha \hat{H}_\alpha = \hat{H}$ and the initial state $|\psi_i\rangle$ chosen to be the ground state of an easily diagonalisable part $\hat{H}_0$ of $\hat{H}$. As previously mentioned, since the HVA is a heuristic ansatz circuit there is no "correct" way to design and implement the circuit, and the form of equ. (5.12) contains a lot of inherent flexibility in the choice of $U(\boldsymbol{\theta})/\hat{H}_\alpha$ and $|\psi_i\rangle/\hat{H}_0$.

Parametrised Unitary Evolution

As discussed in sec. 4.3.1, Hamiltonian fragments $\hat{H}_\alpha$ are typically constructed to ensure that ansatz layers permit an efficient and exact circuit implementation by grouping contributions into sets of mutually commuting terms. For fermionic systems these fragments often correspond to specific classes or spatial partitions of fermionic excitation operators. [11, 90, 106, 115, 116, 145, 146]. Similarly, in spin-lattice problems fragmentation is frequently guided by spatial partitions of the lattice structure [13, 14]. Appropriate choices for the Hamiltonian fragments are highly dependent on the problem structure and desired properties of the ansatz circuit such as efficiency of implementation, a minimised parameter count, and conservation of symmetries. In the spirit of The Principle of Generality, we chose the most standard and most natural Hamiltonian splitting for our model, into disjoint subsets of even- and odd-bond hopping terms, followed by electronic interaction terms, using what can be considered the standard implementation of the resulting ansatz circuit in terms of quantum gates. This splitting is a commonly chosen approach for implementing an HVA circuit for Hubbard models, since it is not only a particularly natural choice for splitting the Hamiltonian, but has also been found to allow for and relatively efficient and parallelised implementation of the ansatz circuit in the JW-encoding [11, 90, 105–107, 109, 112]. However, in contrast to higher-dimensional systems, for our one-dimensional model the resulting ansatz circuit is simple and efficient to implement without the need for any complex, potentially problem tailored circuit compilation techniques. Furthermore, for our particular model this HVA construction will also naturally encode all symmetries of our model Hamiltonian, ensuring that the parametrised state is of a well defined symmetry (provided an appropriately symmetric initial state is prepared as further discussed later), an analytic requirement fulfilled by any exact ground state (or more generally, any eigenstate) of a

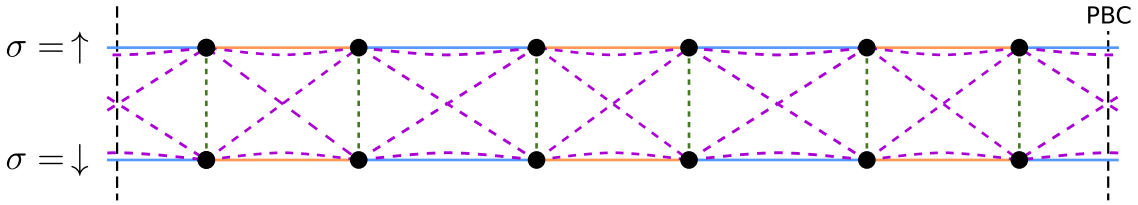


Figure 5.2: Hamiltonian fragments used to construct the HVA for the $N = 6$ SSH Hamiltonian. Upper and lower orbitals correspond to spin-up and spin-down species for each lattice site. $\hat{H}_\mathrm{e}$, $\hat{H}_\mathrm{o}$, $\hat{H}_U$ and $\hat{H}_V$, correspond to the blue, orange, green and purple coloured interactions, respectively.

Hamiltonian. Such symmetry-adapted ansatz circuits are not only useful because they prevent unphysical symmetry breaking such as spin contamination, and enable calculations of excited states of a specific symmetry without having to perform a general excited state calculation [95, 143, 147, 148], but have also been shown to directly improve the accuracy of the VQE calculation [147, 149–151]. However, we once again stress that the choice of splitting is likely not optimal both in terms of the gate count itself and the accuracy achieved by the VQE calculation, when it is considered in conjunction with alternative encodings, ansatz circuits and device topologies, and in particular when circuit and shot noise are also accounted for. It should be noted that more general splittings of the Hamiltonian are sometimes considered in order to reduce the overall depth of the ansatz circuit [13, 107].

The particular splitting of $\hat{H}$ chosen is depicted in fig. 5.2. The hopping terms are naturally split into two sets of all even-bond and all odd-bond hopping terms, respectively:

$$\hat{H}_\mathrm{e} = -\sum_{\substack{i \text{ even} \\ \sigma}}^{N} t_\mathrm{e} \left(\hat{c}^\dagger_{i,\sigma} \hat{c}_{i+1,\sigma} + \hat{c}^\dagger_{i+1,\sigma} \hat{c}_{i,\sigma} \right) \tag{5.13}$$

$$\hat{H}_\mathrm{o} = -\sum_{\substack{i \text{ odd} \\ \sigma}}^{N} t_\mathrm{o} \left(\hat{c}^\dagger_{i,\sigma} \hat{c}_{i+1,\sigma} + \hat{c}^\dagger_{i+1,\sigma} \hat{c}_{i,\sigma} \right) \tag{5.14}$$

In each group interactions commute with each other since they always act on disjoint subsets of fermionic orbitals. The two spin species can be and are considered simultaneously since $\hat{H}$ contains no cross-spin hopping terms. The Hubbard(-like) interaction terms $\hat{n}_{i,\sigma}\hat{n}_{j,\rho}$ are considered separately since they all trivially commute with each other:

$$\hat{H}_\mathrm{ee} = \hat{H}_U + \hat{H}_V = \sum_{i=0}^{N} U \hat{n}_{i,\uparrow}\hat{n}_{i,\downarrow} + \sum_{i=0,\sigma,\sigma'}^{N} V \hat{n}_{i,\sigma}\hat{n}_{i+1,\sigma'} \tag{5.15}$$

All groups are built from mutually commuting fermionic excitations, $\hat{c}^\dagger_p \hat{c}^\dagger_q \hat{c}_r \hat{c}_s + \mathrm{h.c.}$,

which, as discussed in sec. 4.3.1, allows for an exact implementation of the evolution gates $\exp(i\theta\hat{H}_{\mathrm{e}})$, $\exp(i\theta\hat{H}_{\mathrm{o}})$ and $\exp(i\theta\hat{H}_{\mathrm{ee}})$ under the JW-encoding. We found a slight generalisation of the ansatz circuit that considers $\hat{H}_U$ and $\hat{H}_V$ separately to perform better. An individual HVA layer of our circuit is therefore given by:

$$U(\boldsymbol{\theta}) = U(\theta_{\mathrm{e}}, \theta_{\mathrm{o}}, \theta_{\mathrm{U}}, \theta_{\mathrm{V}}) = e^{-i\theta_{\mathrm{e}}\hat{H}_{\mathrm{e}}} e^{-i\theta_{\mathrm{o}}\hat{H}_{\mathrm{o}}} e^{-i\theta_U\hat{H}_U} e^{-i\theta_V\hat{H}_V} \tag{5.16}$$

or at the level of Pauli gates using the JW-encoding applied to the fragments $\hat{H}_{\mathrm{e}} \to \sum_i \alpha_{e,i}\hat{P}_{\mathrm{e},i}$ (and similarly for other terms):

$$\begin{aligned} U(\boldsymbol{\theta}) = &\prod_k \exp\left(-i\theta_{\mathrm{e}}\alpha_{e,k}\hat{P}_{\mathrm{e},k}\right) \prod_l \exp\left(-i\theta_{\mathrm{o}}\alpha_{o,l}\hat{P}_{\mathrm{o},l}\right) \\ &\prod_m \exp\left(-i\theta_U\alpha_{U,m}\hat{P}_{U,m}\right) \prod_n \exp\left(-i\theta_V\alpha_{V,n}\hat{P}_{V,n}\right) \end{aligned} \tag{5.17}$$

The complete ansatz circuit will consist of L repetitions of this circuit layer, each with its own set of parameters $\boldsymbol{\theta}_k = (\theta_{e,k}, \theta_{o,k}, \theta_{U,k}, \theta_{V,k})$.

This ansatz circuit is efficient to implement under the JW-encoding detailed in sec. 5.2.1. With the exception of the PBC term in the odd-bond group all terms in both the even- and odd-bond hopping groups, equ. (5.13) and (5.14) respectively, can be implemented simultaneously at a constant depth, $\mathcal{O}(1)$, due to their locality in the JW-encoding. However, as a result of its non-locality, the PBC term, equ. (5.9), requires an implementation involving $\mathcal{O}(N)$ CNOT gates arranged in a $\mathcal{O}(N)$ depth circuit. As previously mentioned, the cost associated with the implementation of the PBC term does not become negligible in the thermodynamic limit, $N \to \infty$, when compared to the circuit required to implement all other hopping terms. While the Hubbard-like interactions in the group equ. (5.15) cannot all be implemented fully simultaneously, they can be implemented at a constant depth, $\mathcal{O}(1)$, since they allow for an efficient parallelisation under the assumption of all-to-all connectivity. However, as discussed in sec. 5.2.1 circuit depths can potentially increase as a result of the limited connectivity of the device. In the case of linear connectivity congruent with the JW-ordering, a naive implementation of each of the $\mathcal{O}(N)$ Hubbard-like interaction terms requires $\mathcal{O}(N)$ local SWAP gates and hence depth $\mathcal{O}(N)$ leading to a potential $\mathcal{O}(N^2)$ gate count and circuit depth.

Furthermore, the HVA construction allows symmetries to be built into the ansatz circuit, since the operators $\hat{H}_\alpha$ can often be chosen to respect the symmetries of the Hamiltonian. In particular, if each $\hat{H}_\alpha$ commutes with a symmetry operator $\hat{G}$ (or, more generally, with all symmetry operations of a group G), i.e. $[\hat{H}_\alpha, \hat{G}] = 0$, then the corresponding parametrised ansatz unitary, $U(\boldsymbol{\theta})$, equ. (5.12), also com-

mutes with $\hat{G}$. Consequently, the parametrised ansatz state, $|\psi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})|\psi_i\rangle$, transforms according to the same representation of $\hat{G}$ as the initial state, $|\psi_i\rangle$, for all parameters $\boldsymbol{\theta}$. In practice, this means that by choosing $|\psi_i\rangle$ to have the desired symmetry quantum numbers (or to belong to a desired irreducible representation when applicable), the ansatz cannot mix in components from other symmetry sectors during the optimisation. For our model in particular, our choice of ansatz circuit can be shown to conserve all symmetries of the model (see sec. 5.1). Firstly, any electronic structure Hamiltonian, $\hat{H}_{\text{ES}}$, equ. (2.5), conserves the particle number $\sum_i \hat{n}_i = \hat{N}$, since it contains only particle number conserving terms, i.e. $[\hat{N}, \hat{c}_p^\dagger \hat{c}_q] = [\hat{N}, \hat{c}_p^\dagger \hat{c}_q^\dagger \hat{c}_r \hat{c}_s] = 0$. An HVA circuit of the form equ. (5.12) that is built directly from individual Hamiltonian terms and hence number conserving operators, i.e. $\hat{h}_{\alpha,i} \in \{\hat{c}_p^\dagger \hat{c}_q + \text{h.c.}, \hat{c}_p^\dagger \hat{c}_q^\dagger \hat{c}_r \hat{c}_s + \text{h.c.}\}$, will also conserve $\hat{N}$, since the implementation of these Hermitian fermionic excitations using equ. (4.14) is exact under most fermionic encodings, in particular the JW-encoding chosen in our implementation. Any HVA implementation will therefore by construction always conserve the particle number symmetry of $\hat{H}_{\text{ES}}$. Furthermore, any $\hat{H}_{\text{ES}}$ will simultaneously also conserve spin symmetry, i.e. $[\hat{S}^2, \hat{H}_{\text{ES}}] = 0$ and $[\hat{S}_z, \hat{H}_{\text{ES}}] = 0$ as well as $[\hat{S}^2, \hat{N}] = 0$ and $[\hat{S}_z, \hat{N}] = 0$, since $\hat{H}_{\text{ES}}$ can be expressed solely in terms of the spin-summed one- and two-body fermionic operators $\hat{E}_{pq} = \sum_\sigma \hat{c}_{p_\sigma}^\dagger \hat{c}_{q_\sigma}$ and $\hat{e}_{pqrs} = \hat{E}_{pr}\hat{E}_{qs}$, which all individually conserve spin $[\hat{S}^2, \hat{E}_{pq}] = [\hat{S}_z, \hat{E}_{pq}] = [\hat{S}^2, \hat{e}_{pqrs}] = [\hat{S}_z, \hat{e}_{pqrs}] = 0$ (in addition to commuting with $\hat{N}$) [23]. Similarly to particle number conservation, the HVA circuit, equ. (4.14), will then conserve spin symmetry provided that the $\hat{h}_{\alpha,i}$ fragments are now identified at the level of spin-adapted one- and two-body Hermitian fermionic excitation operators, i.e. $\hat{h}_{\alpha,i} \in \{\hat{E}_{pq} + \text{h.c.}, \hat{e}_{pqrs} + \text{h.c.}\}$ (or linear combinations thereof [93, 152]). While in principle this allows for a simple conservation of spin symmetries in ansatz circuits, in practice it is limited by the ability to implement gates involving spin adapted excitation operators exactly, a problem previously encountered and identified in UCC and ADAPT-VQE calculations [147, 150]. In these calculations evolution by the spin-adapted UCC operator, $\hat{e}_{pqrs} - \text{h.c.}$, was shown to introduce/require a Trotterisation approximation using equ. (4.19), since these terms break down into a sum of non-commuting UCC operators:

$$\hat{e}_{pqrs} - \text{h.c.} = \hat{E}_{pr}\hat{E}_{qs} - \text{h.c.} = \hat{\tau}_{rs}^{pq} + \hat{\tau}_{r\bar{s}}^{p\bar{q}} + \hat{\tau}_{\bar{r}s}^{\bar{p}q} + \hat{\tau}_{\bar{r}\bar{s}}^{\bar{p}\bar{q}} \tag{5.18}$$

This approximation will break the spin symmetry of the spin-adapted excitations, and depending on the choice of ordering of the individual Pauli terms after a JW-encoding, $\hat{e}_{pqrs} - \text{h.c.} = \sum_i \hat{P}_i$, could potentially break further symmetries such as particle number conservation. Implementing these spin adapted operators exactly

is possible but incurs a considerable gate overhead [152]. Similarly, in the HVA implementing evolutions by spin adapted Hermitian terms $\hat{e}_{pqrs}+\text{h.c.}$ would conserve the spin symmetry of the system, but is limited by the approximation introduced in the product formula, equ. (4.14). However, our particular model Hamiltonian only includes a limited set of Hubbard-like interactions, $\hat{n}_p\hat{n}_q$, for which the spin adapted versions, $\hat{e}_{pqpq} = \sum_{\sigma\rho}\hat{n}_{p_\sigma}\hat{n}_{q_\rho}$ can be implemented exactly, and for which spin symmetry can hence be conserved exactly. Since both spin species versions of each hopping term are always implemented simultaneously and hence with the same parameters, and similarly since all spin species versions of each U and V interaction term can and are implemented simultaneously, we are effectively working with $\hat{E}_{pq}$ and $\hat{e}_{pqrs}$ excitations and hence conserving $\hat{S}^2$ and $\hat{S}_z$. For our model in particular the ansatz circuit also conserves all spatial symmetries of the model. The Hamiltonian splitting can be seen to respect the translation symmetry of the system, since all $\hat{H}_\alpha$ for our case are symmetric about translation by a unit cell/two lattice sites, $[\hat{T}_2, \hat{H}_{\text{e/o}/U/V}] = 0$, and will therefore conserve the crystal momentum $\hbar k$ of a translation symmetric state $\hat{T}_2|\psi\rangle = e^{ika}|\psi\rangle$. Furthermore, all $\hat{H}_\alpha$ commute with the inversion symmetry of the Hubbard model $[\hat{\mathcal{P}}_I, \hat{H}_{\text{e/o}/U/V}]$ and will therefore conserve the parity of the initial state, $P_I = \pm 1$ with $\hat{\mathcal{P}}_I|\psi\rangle = P_I|\psi\rangle$.

Initial State

Similarly to the parametrised unitary evolution, the initial state $|\psi_i\rangle$ can be chosen from a variety of different options, with the only real requirement being that it can be efficiently prepared on the quantum computer. Commonly considered initial states are computational basis states [90, 108, 111], simple, physically inspired tensor product states [13, 14, 105, 112, 113, 115, 146, 153], ground states of non-interacting spin or free fermion Hamiltonians [11, 90, 104, 106, 107, 109, 112, 114], as well as Hartree-Fock Slater determinants (SD) [106, 115, 116, 146, 154]. However, appropriate choices of initial states are highly dependent on the problem instance considered. Following The Principle of Generality, we chose the simplest, but still natural initial state for our calculations, the many-body ground state of the non-interacting/single-particle part of our Hamiltonian, i.e. equ. (5.1) with $U = V = 0$:

$$\hat{H}_0 = \sum_{pq} h_{pq} c_p^\dagger \hat{c}_q = -\sum_{\substack{i=0,\sigma \\ i \text{ even}}}^{N-2} t_\text{e}\left(\hat{c}_{i,\sigma}^\dagger \hat{c}_{i+1,\sigma} + \hat{c}_{i+1,\sigma}^\dagger \hat{c}_{i,\sigma}\right) - \sum_{\substack{i=1,\sigma \\ i \text{ odd}}}^{N-1} t_\text{o}\left(\hat{c}_{i,\sigma}^\dagger \hat{c}_{i+1,\sigma} + \hat{c}_{i+1,\sigma}^\dagger \hat{c}_{i,\sigma}\right) \tag{5.19}$$

implemented using the standard quantum circuit for preparing arbitrary SDs on a quantum computer. This initial state has been successfully used in HVA calcula-

tions of the Hubbard model to provide a high-fidelity starting point that accurately captures electron localisation and correlation effects [11, 90, 105–107, 109]. This many-body ground state, $|\psi_0\rangle$, consists of a single SD, $|\psi_0\rangle = |\mathrm{SD}_0\rangle$, formed by occupying the N lowest energy single-particle eigenstates, $|\tilde{\phi}_p\rangle$, of h_{pq}:

$$|\mathrm{SD}_0\rangle = \prod_{p\in\mathrm{occ}} \hat{\tilde{c}}_p^\dagger |\Omega\rangle \tag{5.20}$$

Not only can $|\tilde{\phi}_p\rangle$ be efficiently computed classically by diagonalising h_{pq}, but $|\mathrm{SD}_0\rangle$ itself can also be efficiently prepared directly on the quantum device. It is one of the simplest initial states that improves upon a simple computational basis state by approximately treating some of the model physics. Additionally, since $\hat{H}_0$ conserves the same symmetries as the full model Hamiltonian (for spatial symmetries this is satisfied by any electronic structure Hamiltonians, $\hat{H}_{\mathrm{ES}}$, equ. (2.5), since the Coulomb kernel, $1/|\boldsymbol{r}_i - \boldsymbol{r}_j|$, is totally symmetric under any spatial symmetry), the single-particle eigenstates – and as a consequence the many-body SD – can be constructed with well defined symmetry, which, in conjunction with the symmetry conserving structure of the parametrised circuit discussed above, implements an ansatz circuit that targets/parametrises only a specific symmetry sector of the model. For our particular model we also find that a HF solution does not necessarily provide a better initial state for the ansatz circuit.

For our model $|\mathrm{SD}_0\rangle$ can be formed explicitly from the single-particle eigenstates presented in sec. 3.2 by fully occupying the valence band:

$$|\mathrm{SD}_0\rangle = \prod_{k,\sigma} c_{k,v,\sigma}^\dagger |\Omega\rangle \tag{5.21}$$

This SD can be shown to be totally symmetric under all symmetries of the Hamiltonian. Since the non-interacting orbitals have a well defined spin component with equal spatial components for both spin species, and since the non-interacting ground state SD consists of only doubly occupied spin orbitals, $|\mathrm{SD}_0\rangle$ is a closed-shell state satisfying $S = S_z = 0$. In our particular case this closed-shell SD can be shown to also be totally symmetric under the spatial symmetries, $\hbar k = 0$ and $P_I = +1$, although we note that for general systems such an SD is not guaranteed to be totally symmetric or might even require forming a symmetry adapted linear combination of degenerate SDs [23, 155]. Since we are working with a half-filled system with an isolated and hence fully occupied valence band, the inversion symmetric pairs $\hbar k$ and $-\hbar k$, which transform as partners of an irreducible representation of the symmetry group, are always simultaneously occupied. An anti-symmetrised product of

the $\hbar k$ and $-\hbar k$ single-particle states,

$$\frac{1}{\sqrt{2}}\left(|\phi_k\rangle|\phi_{-k}\rangle - |\phi_{-k}\rangle|\phi_k\rangle\right) \tag{5.22}$$

is totally symmetric under translation symmetry, $k = 0$, and antisymmetric under inversion symmetry, $P_I = -1$, which, when combined with its spin pair, forms a totally symmetric contribution, $k = 0$ and $P_I = (-1)^2 = +1$. The momentum of the state at the Brillouin zone (BZ) boundary, $k = \pi$, cancels with its spin pair, $\pi+\pi = 2\pi \equiv 0 \mod 2\pi$, and since the valence band has a bonding orbital character, i.e. it is symmetric under inversion $\hat{\mathcal{P}}_I|\phi_{k=0/\pm\pi,v}\rangle = +1 \cdot |\phi_{k=0/\pm\pi,v}\rangle$, both states at the BZ centre and boundary, $k = 0$ or π, contribute totally symmetric phases, $P_I = +1$. Since all molecular orbital contributions are fully symmetric, $|\mathrm{SD}_0\rangle$ itself will also be fully symmetric.

Targeting other symmetry sectors would require initial states of appropriate symmetry which can also be easily created using appropriately constructed SDs, or symmetry adapted linear combinations thereof. The momentum of any SD is the total momentum of all occupied states, $\hbar k_{\mathrm{SD}} = \sum_i \hbar k_i$, which is trivially seen by considering $|\mathrm{SD}\rangle = \prod_i \hat{c}^\dagger_{k_i}|\Omega\rangle$, which under translations transforms as:

$$\hat{T}_2|\mathrm{SD}\rangle = \exp(\sum_i ik_i a)|\mathrm{SD}\rangle \tag{5.23}$$

While for generic $\hbar k_{\mathrm{SD}}$ this state already transforms as a partner of an irreducible representation of the symmetry group (in conjunction with its inversion symmetric pair at $-\hbar k_{\mathrm{SD}} = \sum_i -\hbar k_i$), when the total momentum lies at the BZ centre or boundary, $\hbar k_{\mathrm{SD}} = 0$ or $\pm\pi$, an inversion (anti-)symmetric sum which satisfies $\hat{\mathcal{P}}_I|\mathrm{SD}_\pm\rangle = \pm 1 \cdot |\mathrm{SD}_\pm\rangle$ needs to be formed, given by:

$$|\mathrm{SD}_\pm\rangle = \frac{1}{\sqrt{2}}(|\mathrm{SD}\rangle \pm \hat{\mathcal{P}}_I|\mathrm{SD}\rangle) \tag{5.24}$$

While initial states with general well defined spin quantum numbers, S and S_z, will consist of more complicated linear combinations of SDs, a particularly simple state of non-zero spin is the high-spin open-shell state in which all singly occupied orbitals in the SD belong to the same (e.g. α) spin species. This SD can be shown to satisfy $S = S_z = n_\alpha/2$ for n_α unpaired spin-α electrons [23].

Under the JW-encoding $|\mathrm{SD}_0\rangle$ can be efficiently prepared directly on the quantum computer, requiring only a $\mathcal{O}(N^2)$ gate count and a $\mathcal{O}(N)$ circuit depth on an architecture with local connectivity [88, 156]. The state preparation circuit imple-

ments the unitary

$$\hat{U} = \exp(-\sum_{pq} \kappa_{pq} \hat{c}_p^\dagger \hat{c}_q) \tag{5.25}$$

which applies a single-particle basis change between a reference and new basis

$$|\tilde{\phi}_p\rangle = \sum_{pq} (\exp(-\kappa))_{pq} |\phi_q\rangle \tag{5.26}$$

acting in second quantisation as

$$\hat{U}\hat{c}_p^\dagger\hat{U}^\dagger = \hat{\tilde{c}}_p^\dagger \qquad \text{and} \qquad \hat{U}\hat{c}_p^\dagger\hat{c}_q^\dagger \ldots |\Omega\rangle = \hat{\tilde{c}}_p^\dagger\hat{\tilde{c}}_q^\dagger \ldots |\Omega\rangle \tag{5.27}$$

on operators and states respectively. It allows us to prepare $|\mathrm{SD}_0\rangle$ starting from a SD in the orbital basis, which in the JW-encoding corresponds to a simple computational basis state. If the single-particle eigenstates are enumerated ascending in energy such that the valence band states correspond to orbitals $p = 0, 1, \ldots, N-1$, we implement $|\mathrm{SD}_0\rangle = \hat{U}|11\ldots100\ldots0\rangle$. The symmetry adapted linear combinations of SDs required for targeting specific symmetry sectors can also be easily implemented, since linearity allows us to prepare an appropriate linear combination of orbital basis SDs/computational basis states, a simple task for the small number of constituent SDs encountered in symmetry adapted excitations in our model.

While a Hartree-Fock (HF) solution, $|\psi_{\mathrm{HF}}\rangle$, could likely provide an improved initial starting state without requiring any additional quantum resources to prepare – it also takes the form of a single SD, $|\psi_{\mathrm{HF}}\rangle = |\mathrm{SD}_{\mathrm{HF}}\rangle$ – other investigations have found the impact of the initial state to be negligible [115]. Furthermore, as we discuss below, for our model the symmetries of the Hamiltonian would constrain a symmetry-adapted restricted closed-shell HF (RHF) solution to be qualitatively very similar to the non-interacting SD. The eigenstates of the non-interacting SSH model presented in sec. 3.2 are determined by the symmetry of the Hamiltonian. The formation of Bloch waves, as required by translational symmetry, restricts the eigenstates to have well defined crystal momentum $\hbar k$, while inversion symmetry maps the degenerate Bloch states at $\hbar k$ and $-\hbar k$ into each other. Since we are working with a half-filled system in which the valence band is fully occupied, and hence the inversion symmetric orbital pairs at $\hbar k$ and $-\hbar k$ are always simultaneously occupied, the Fock operator built from these orbitals can be shown to be totally symmetric under the same symmetry operations of the original Hamiltonian [157]. As a result, the eigenstates of this Fock operator are determined by the same symmetries of the Hamiltonian, and symmetry adapted orbitals are therefore an internally consistent solution to the RHF equations, equ. (2.42). This can be con-

firmed from an explicit evaluation of the Fock operator, $f_{\mu\nu}$ equ. (2.43), using the single particle eigenstates, $|\phi_{k,v}\rangle$, of sec. 3.2. The relevant components of the HF density matrix are explicitly shown to be totally symmetric under all symmetries of the original Hamiltonian:

$$P_{ii} = 2\sum_{k}^{N/2} \frac{1}{N} = 1 \qquad P_{2i,2i+1} \equiv B_{\mathrm{e}} \qquad P_{2i+1,2i+2} \equiv B_{\mathrm{o}} \tag{5.28}$$

As a result, the only effects of RHF are to introduce a constant energy shift, $U/2 + 2V$, and to renormalise the hopping terms, $t_{\mathrm{e}} \to t_{\mathrm{e}} - VB_{\mathrm{e}}$ and $t_{\mathrm{o}} \to t_{\mathrm{o}} - VB_{\mathrm{o}}$ so that it is symmetric under the same transformations as the original non-interacting SSH model:

$$F_{ij} = -\sum_{i \text{ even}} (t_{\mathrm{e}} - VB_{\mathrm{e}})|i\rangle\langle i+1| - \sum_{i \text{ odd}} (t_{\mathrm{o}} - VB_{\mathrm{o}})|i\rangle\langle i+1| + \left(\frac{U}{2} + 2V\right)\frac{\delta_{ij}}{2} + \text{h.c.} \tag{5.29}$$

The solutions of a symmetry adapted RHF calculation can therefore be explicitly seen to be qualitatively similar to the non-interacting SD as a result of the high symmetry of the model.

While more general, potentially symmetry broken, HF solutions – such as those likely to arise in RHF without symmetry adaptation or unrestricted HF (UHF) – can provide qualitatively different initial states with a higher ground-state overlap, this comes at the cost of breaking the symmetry of the parametrised ansatz state, leading to potential symmetry contamination and preventing the parametrisation of a specific symmetry sector. Furthermore, performing a VQE calculation with a symmetry broken initial state but a symmetry preserving unitary, $U(\boldsymbol{\theta})$, is likely to be detrimental for the accuracy of the VQE calculation. If an initial state with partial overlap on, e.g., both even, $|\psi_i^+\rangle$, and odd, $|\psi_i^-\rangle$, symmetry sectors is prepared:

$$|\psi_i\rangle = \alpha|\psi_i^+\rangle + \beta|\psi_i^-\rangle \tag{5.30}$$

the relative weights on the two symmetry sectors, α and β, will remain constant, since such a circuit is unable to couple different symmetry sectors. The lowest energy state that can be reached is therefore limited to $E = |\alpha|^2 E_0^+ + |\beta|^2 E_0^-$, with $E_0^\pm$ the lowest energy even/odd states. Recovering the true ground state would require either also breaking the symmetry of the circuit resulting in a general non-symmetry adapted calculation, or implementing an expensive symmetry projection procedure to restore the symmetry of the state [143, 147, 150, 151].

5.2.3 Energy Measurement

As discussed in sec. 4.4, estimation of the energy expectation value, $\langle \hat{H} \rangle$, requires a measurement protocol compatible with the projective computational basis state measurements available on a real quantum device. The procedure that we use to estimate $\langle \hat{H} \rangle$ is based on grouping generally commuting Pauli terms, as described in sec. 4.4.1. We choose this method in the spirit of The Principle of Generality since it is a standard measurement procedure that is applicable to any electronic structure system and widely supported and used in applications. While for general systems this measurement procedure requires classical heuristic algorithms to find an appropriate partitioning of the Hamiltonian into such groups, as well as requiring potentially deep circuits to implement the diagonalising basis rotations, for Hubbard models, and our model in particular, both of these issues are greatly simplified [90]. However, we note that, as far as the calculations in the current chapter are concerned, all measurement schemes are equivalent due to the absence of shot noise in the infinite shot number limit, $M \to \infty$.

A natural choice for the commuting groups used in the measurement procedure is given by the same splitting considered in the ansatz circuit construction (see sec. 5.2.2), in which the Hamiltonian is grouped (as depicted in fig. 5.2) according to:

$$\hat{H} = \hat{H}_{\mathrm{e}} + \hat{H}_{\mathrm{o}} + \hat{H}_{\mathrm{ee}} \tag{5.31}$$

with $\hat{H}_{\mathrm{e}}$, $\hat{H}_{\mathrm{o}}$ and $\hat{H}_{\mathrm{ee}}$ as defined previously, equ. (5.13), (5.14) and (5.15) respectively. As discussed earlier, each group consists only of mutually commuting fermionic operators. Under the JW-encoding all mutually commuting fermionic operators in each of these groups also map to a set of mutually commuting Pauli operators. As discussed in sec. 4.4.1 this allows for a simultaneous measurement of all Paulis in each group, and hence a direct measurement of any of the three Hamiltonian terms using a single measurement. An estimator, $\bar{H}$, can therefore be constructed using only three individual measurements, one for each term.

For $\hat{H}_{\mathrm{e/o}}$ the mutual commutativity arises since all hopping terms in each group act on disjoint fermionic orbitals. With the exception of the PBC term, all hoppings in each group therefore also act on a disjoint set of qubits under the JW-encoding (see equ. (5.7)) and hence trivially commute. The sole exception is the PBC term, equ. (5.9), which, due to its non-local nature, does not trivially commute with other terms. However, it nonetheless commutes with all terms in its respective group, $\hat{H}_{\mathrm{e}}$, since it anti-commutes with other even-bond hopping terms on an even number of qubit sites. For $\hat{H}_{\mathrm{ee}}$ the commutativity arises since the encoded Pauli strings for

all terms are fully diagonal (see equ. (5.8)) and hence also trivially commute. The unitary transformation allowing simultaneous measurements of $\hat{H}_{\mathrm{e/o}}$ individually diagonalises each hopping term, equ. (5.7), and can be implemented with a simple $\mathcal{O}(1)$ depth circuit construction [90].

5.2.4 Classical Optimisation

As previously mentioned in sec. 4.5, the choice of optimisation algorithm plays a crucial part in the performance of the VQE algorithm. Given that our objective function in the current chapter is evaluated deterministically, i.e. noise-free and without stochastic sampling error, we adopt the limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) algorithm described in sec. 4.5.2 for our minimisation, following The Principle of Generality. Furthermore, as described in sec. 4.5.2, it is a standard choice for, and widely successful at optimising a large range of objective functions, both in general and in VQE calculations in particular.

Circuit Gradients

The evaluation of the objective function gradients, as required by L-BFGS, is performed using the parameter-shift (PS) rules described in sec. 4.5.1. Despite the constant number of parameters present in each layer of the ansatz circuit, estimating the full circuit gradient $\boldsymbol{g}_k = \frac{\partial E(\boldsymbol{\theta})}{\partial \theta_k}$ will require $\mathcal{O}(N)$ circuit evaluations under the PS scheme. This results from the requirement to apply a product rule to evaluate the repeated parameters that occur as a result of the Pauli rotation decomposition of the HVA circuit, equ. (4.14). While finite-difference gradients are possible in principle and come at substantially reduced cost since no application of the chain rule for repeated parameters is required, they introduce an arbitrary step-size hyperparameter and can be numerically brittle when used with the L-BFGS algorithm, especially when the circuit contains many parameters or when the landscape has mixed curvature scales; they are therefore not considered in these calculations.

Parameter Initialisation

The initialisation of the optimisation parameters is similarly a highly heuristic choice that can greatly affect the performance of the VQE algorithm. In practice, many implementations adopt simple, relatively problem-agnostic prescriptions such as sampling each independent parameter uniformly from a narrow interval around the origin, e.g. $\theta \in [-\pi/5, +\pi/5]$ [106, 115], sometimes also accounting for the number of circuit layers, L, e.g. $\theta \in [0, 1/L]$ [14, 90]. Such small angle parameter initiali-

sations have also been shown to help avoid issues related to barren plateaus [153]. Parameters transferred from solutions to smaller model instances have also been considered [158]. While such approaches could improve optimisation stability, they are typically more model dependent: their efficacy depends on physically meaningful notions of locality, an appropriate continuation path in system size, or the availability of reliable classical solutions to seed from. Furthermore, while the avoidance of barren plateaus through restricted parameter initialisation is beneficial, this method has been linked to classical simulability of the VQE algorithm [159]. Following The Principle of Generality, we instead decide to initialise parameters uniformly over the valid parameter range, prioritising a generic initialisation scheme.

The periodicity of variational parameters is frequently stated to be $\theta \in [-\pi, +\pi]$ [9]. For simple Pauli rotation gates, $\exp(i\theta\hat{P})$, this is true since the gate action can be written as $\cos(\theta)\hat{I} + i\sin(\theta)\hat{P}$. However, the more complex Hamiltonian evolution gates, $\exp(i\theta\hat{H}_\alpha)$, considered in the HVA can have varying or completely absent periodicity, depending on the choice of $\hat{H}_\alpha$. For a general $\hat{H}_\alpha$ the periodicity is determined by the spectrum, $\sigma(\hat{H}_\alpha) = \{\lambda_1, \lambda_2, \dots\}$, since, when diagonalised by a suitable basis transformation Λ, the action of the gate is given by:

$$\Lambda^\dagger \exp(i\theta\hat{H}_\alpha)\Lambda = \mathrm{diag}(e^{i\theta\lambda_1}, e^{i\theta\lambda_2}, \dots) \tag{5.32}$$

If any two λ_i and λ_j are incommensurate real numbers, i.e. if the ratio $\lambda_i/\lambda_j \notin \mathbb{Q}$, then no finite value of θ exists for which all phase factors $e^{i\theta\lambda_k}$ simultaneously recur. In this case $\exp(i\theta\hat{H}_\alpha)$ is aperiodic and the parameter θ does not possess a well-defined periodicity. More generally, a finite period, T, exists if and only if there is a non-zero T such that:

$$T\lambda_k = 2\pi n_k \qquad \forall k. \tag{5.33}$$

or equivalently all λ_i are commensurate for any pair (i, j). This condition is trivially satisfied for any $\hat{P}$, whose spectrum is $\sigma(\hat{P}) = \pm 1$, but is violated for general $\hat{H}_\alpha$. As a consequence, the common practice of initialising variational parameters uniformly over $\theta \in [-\pi, +\pi]$ is not formally justified for general Hamiltonian evolution gates used in the HVA.

However, for our choice of ansatz circuit described in sec. 5.2.2 the HVA circuit gates can be seen to be periodic with an easily identifiable period. The Hamiltonian is split into $\hat{H}_\mathrm{e}$ (equ. (5.13)), $\hat{H}_\mathrm{o}$ (equ. (5.14)), $\hat{H}_U$ and $\hat{H}_V$ (equ. (5.15)), with the corresponding evolution gates $\exp(i\theta\hat{H})$ each having their own independent parameter (see equ. (5.16)). Under this splitting all fermionic operators in each of the four terms share the same interaction strength, t_e, t_o, U and V respectively.

Under the JW-encoding, equ. (5.7) and (5.8), all Pauli terms will also share the same coefficients, i.e. $\hat{H}_e \rightarrow \sum_i \alpha_{e,i}\hat{P}_i = \alpha_e \sum_i \hat{P}_i$. The resulting trotterised circuit implementation, equ. (5.17), can therefore be seen to be clearly periodic with a period of $T_e = \frac{2\pi}{\alpha_e}$, or equivalently $T_e = 2\pi$ under an appropriate redefinition of parameters, $\theta_e\alpha_e \rightarrow \theta_e$. This allows a uniform initialisation of the ansatz circuit over the whole parameter space to be easily implemented by choosing $\theta \in [-\pi, +\pi]$.

5.3 Molecular Forces in VQE

As discussed in sec. 2.5, the calculation of analytic gradients, and in particular molecular forces, is a well studied discipline in electronic structure theory. However, within VQE and its applications to electronic structure theory it has received relatively little attention from the quantum computing community. Given the current limitations of NISQ devices the molecular systems usually studied within VQE are simply too small and too simplistic to warrant or require investigations into the evaluation of molecular forces. Furthermore, unlike TPA, the qualitatively correct description of the systems does not depend on an accurate description of the ground-state geometry. Frequently considered systems for VQE simulations include H_2, LiH, NH_3, H-chains and similar small standard benchmark systems [17–20, 93, 119]. While the PES of these molecules is sometimes considered as part of the VQE calculations, this generally occurs at a level of a simple one-dimensional molecular coordinate, such as a canonical bond-length, which is scanned across a reasonable set of parameter values [17, 20, 93, 94, 160]. The aim of these calculations is usually not a high-accuracy prediction of molecular properties but a demonstration of the performance of the VQE algorithm as the system becomes more or less correlated. While a prediction of the correct molecular geometry is still an important part of a full electronic structure calculation for these molecules, the simplicity of the systems and the small dimensionality of the molecular coordinates and PES obviates the need for a molecular force based geometry optimisation. Furthermore, the current state of NISQ devices already prevents accurate evaluation of molecular energies for larger molecules on real quantum computers. As such, investigations focusing on VQE performance on real machines rarely consider any observables or predictions beyond the ground-state energies.

In spite of the limitations presented by the standard benchmark system, a growing subset of publications has considered the calculation of molecular forces in VQE more explicitly. However, this literature spans a range of ansatz circuits and assumptions and, consequently, a range of methods for calculating molecular forces.

Most works treat forces primarily at the approximate Hellmann-Feynman (HF) level, using VQE to prepare an approximate ground state, $|\psi(\boldsymbol{\theta}_*)\rangle$, and then estimating the molecular forces as:

$$\boldsymbol{F} = \langle\psi(\boldsymbol{\theta}_*)|\hat{\boldsymbol{F}}|\psi(\boldsymbol{\theta}_*)\rangle \tag{5.34}$$

by invoking the approximate HF theorem of sec. 2.5.2. This is valid since VQE prepares an approximate wavefunction by minimising the energy $E(\boldsymbol{\theta})$, and is therefore variationally optimised with respect to the parameters $\boldsymbol{\theta}$, i.e. $\frac{\partial E}{\partial \boldsymbol{\theta}} = 0$. The expectation value of $\hat{\boldsymbol{F}}$ is estimated, just like $\hat{H}$ itself and any other observable, using the techniques presented in sec. 4.4. As discussed in sec. 2.5.3 the HF theorem in second quantisation inherently accounts for any moving basis set and resulting Pulay forces. This method has been considered for the calculation of molecular geometries and ab initio molecular dynamics of very simple molecules [161–167]. Some works have also considered the inclusion of non-variational response molecular forces in VQE, an important consideration when the ansatz circuit contains non-variational parameters [168–172]. As discussed in sec. 2.5.4, when non-variational parameters are present response terms need to be included for a consistent calculation of the molecular force. However, the evaluation of the response molecular force was in general found to be relatively expensive, especially compared to standard electronic structure theory, where, as previously discussed, the Lagrangian formalism allows a, both conceptually and computationally, simple framework for the evaluation of the non-variational response.

For our work we decide to focus on the calculation of the HF force as estimator of the true molecular force experienced by the system. As we will discuss, this does neglect non-variational contributions of the molecular force in our specific ansatz circuit implementation. However, since both the aim and the strength of quantum computing and VQE calculations is the preparation of the exact ground state to high precision, the exact HF theorem, as discussed in sec. 2.5, is expected to be relevant and applicable. Molecular forces therefore should not require the invocation of the approximate HF theorem resulting from the stationarity of the ansatz circuit parameters, $\frac{\partial E}{\partial \boldsymbol{\theta}} = 0$, nor should they require an inclusion of response terms based on non-variational ansatz circuit parameters.

5.3.1 Implementation

Non-variational Response

The ansatz circuit used in our work is not fully variational. Non-variational response terms in the HVA can occur in two ways. Firstly, as discussed in sec. 4.3.1, the

HVA constructs the parametrised unitary, $\hat{U}(\boldsymbol{\theta})$, from Hamiltonian evolution gates, $\exp(i\theta\hat{H}_\alpha)$, where $\hat{H}_\alpha$ are chosen as fragments of the full Hamiltonian, $\hat{H}$. The ansatz circuit will therefore in general incur an explicit dependence on the molecular geometry through the dependence of the Hamiltonian, $\hat{H} = \hat{H}(\boldsymbol{R})$, and as a result its fragments, $\hat{H}_\alpha = \hat{H}_\alpha(\boldsymbol{R})$:

$$\hat{U}(\boldsymbol{\theta}) = \hat{U}(\boldsymbol{\theta}; \boldsymbol{R}) = \prod_k e^{i\theta_{k,\alpha}\hat{H}_\alpha(\boldsymbol{R})} \tag{5.35}$$

This is an explicit dependence on $\boldsymbol{R}$ beyond the implicit dependence through the optimised variational parameters $\theta_* = \theta_*(\boldsymbol{R})$, which will in general not vanish since the stationary conditions for the VQE solution only hold for $\boldsymbol{\theta}$. As explained in sec. 2.5.4, the molecular force will therefore incur a response term. Secondly, the preparation of the initial state, $|\psi_i\rangle$, in the HVA occurs directly on the quantum computer using an appropriate unitary circuit, as discussed in sec. 4.3.1. While in general this unitary circuit does not need to be geometry dependent – a fixed initial state such as a computational basis state is a common choice – our particular choice of initial state as the ground state of the non-interacting Hamiltonian, $\hat{H}_0$, as discussed in sec. 5.2.2, introduces an explicit dependence of $|\psi_i\rangle$ on the molecular geometry through the dependence of $\hat{H}_0$ on $\boldsymbol{R}$, $\hat{H}_0 = \hat{H}_0(\boldsymbol{R})$. As a result, the unitary circuit preparing $|\psi_i\rangle$ will depend explicitly on $\boldsymbol{R}$ through the molecular orbital coefficients determining $|\psi_i\rangle$:

$$\hat{U} = \hat{U}(\boldsymbol{R}) = \exp(-\sum_{pq} \kappa_{pq}(\boldsymbol{R})\hat{c}_p^\dagger \hat{c}_q) \tag{5.36}$$

However, as discussed, both of these contributions to the molecular force are expected to vanish in the limit where an exact ground-state wavefunction, $|\psi_0\rangle$ of $\hat{H}$ is prepared on the quantum computer.

Hellmann-Feynman Force

The HF force in VQE simulations of the SSH model can be evaluated by measuring the expectation value of the force operator, $\hat{F}_i$, equ. (5.3), on the ground state $|\psi(\boldsymbol{\theta}_*)\rangle$ after optimising the ansatz circuit:

$$F_i = \langle\psi(\boldsymbol{\theta}_*)|\hat{F}_i|\psi(\boldsymbol{\theta}_*)\rangle \tag{5.37}$$

This measurement is implemented using the same approach as the evaluation of the energy, $\langle\psi(\boldsymbol{\theta})|\hat{H}|\psi(\boldsymbol{\theta})\rangle$, presented in sec. 5.2.3. The same measurement scheme used

to measure $\hat{H}_e$ and $\hat{H}_o$ can also be used to construct all components F_i of $\boldsymbol{F}$, since they contain the same set of fermionic operators. An estimate $\bar{\boldsymbol{F}}$ can therefore be constructed using only two individual measurements.

5.4 Numerical Results

At this point we turn to the numerical results of the idealised, noiseless VQE simulations of the SSH model for TPA. As discussed, the aim of this chapter is to quantify the quantum resources required for chemically meaningful predictions of both energies and molecular forces, while also focusing on analysing the standard electronic structure theory viewpoint that molecular forces are considerably more expensive to calculate accurately than energies, as discussed in sec. 2.2.

In our calculations we specifically consider an instance of the SSH model with $N = 6$ sites. For the finite-chain geometry studied here this is the smallest size that admits a non-trivial bulk unit cell, in the sense that not all unit cells are simultaneously adjacent to the boundaries. While ideal simulations of VQE calculations with larger systems are possible, the noisy simulations presented in sec. 6 are naturally limited to $N = 6$. In practice classical computational resources impose a limit since noisy simulations are substantially more expensive than their noiseless counterparts, and already the $N = 6$ instance is expected to be beyond the capabilities of currently available NISQ hardware once realistic noise and measurement overheads are accounted for.

The construction of the HVA circuit allows its expressivity to be systematically improved by increasing the number of layers, L. Determining the accuracy with which energies and molecular forces can be predicted using the HVA requires identifying an appropriate value of L, and in these results we assess this convergence with respect to L, in direct analogy to standard convergence studies in electronic structure theory (e.g., with respect to basis-set size or plane-wave cut-off), in order to determine the quantum resources required for chemically relevant calculations using VQE. We are specifically interested in the value of L necessary for chemically relevant predictions of energies and molecular forces in our model, as well as the extent to which the variational principle affects the relative quantum resources required for each observable.

The utility of electronic-structure calculations requires a target (or benchmark) accuracy relative to a set of reference results that defines when a computed observable is "chemically meaningful". For energies, a widely adopted reference point is chemical accuracy, ΔE_{chem}, conventionally taken as 1 kcal/mol, or approximately

0.04 eV. For our calculation we naturally compare our results to the exact ground-state energy, E_0, calculated from a (classically computed) exact diagonalisation (ED) result of $\hat{H}$. While we stress that in realistic scenarios E_0 would not be classically computable, our model instance is small enough to allow ED to be performed. Sufficient accuracy is achieved for our VQE calculations if they are within an error ΔE_{chem} of E_0.

However, for molecular forces no widely adopted chemical accuracy equivalent exists. The accuracy required to predict meaningful geometries generally depends on the molecular system considered; if the bonds are particularly "stiff" even relatively low molecular force accuracies can be used to predict reasonably accurate geometries. Based on ED calculations we found an accuracy of $\Delta F_{\text{chem}} = 0.01\,\text{eV/Å}$ relative to the exact result, $\boldsymbol{F}_0$, to be sufficiently accurate for determining ground-state geometries of the SSH model. This is also supported by a simple estimate based on the spring constant, K, of our model, for which we find an expected accuracy in an estimate of the bond length based on $\Delta u \sim \frac{\Delta F_{\text{chem}}}{K}$ of $\Delta u \approx 3 \times 10^{-4}\,\text{Å}$.

5.4.1 Simulation Methodology

The numerical calculations presented in this section were implemented using the QISKIT software stack [173]. The relevant package versions used were `qiskit~=0.46.0`, `qiskit-nature~=0.5.0`, and `qiskit-aer~=0.14.0`. All calculations discussed in this section were performed on the Imperial HPC cluster using the available AMD EPYC 7742 CPUs.

The second-quantised operators defined in sec. 5.1 were implemented explicitly as `FermionicOp` objects in QISKIT Nature and mapped to qubit operators using the built-in `JordanWignerMapper`. The default convention used by QISKIT Nature results in a mapping using the spin-orbital ordering described in sec. 5.2.1.

The ideal and noiseless VQE calculations were performed using the exact state vector estimator provided in QISKIT, `qiskit.primitives.Estimator`. The estimator evolves the state vector of the system according to the provided quantum circuit in the absence of any circuit noise and evaluates expectation values directly on the final state vector. We used this estimator to evaluate the observables for our ansatz circuit, thereby obtaining the relevant expectation values without finite-shot sampling error or circuit noise. The measurement procedure described in sec. 4.4 and sec. 5.2.3 does not need to be explicitly implemented in these simulations, since we directly evaluate the exact expectation values that the finite-shot measurement procedure estimates in the $M \to \infty$ limit.

For the simulations in this chapter, circuit gradients were evaluated us-

ing QISKIT's reverse-mode gradient implementation, `ReverseEstimatorGradient`. This is a differentiation method applicable to noiseless state vector simulations, which is able to evaluate exact gradients at a significantly reduced cost compared to a direct implementation of PS gradient techniques [174]. Since, as discussed in sec. 4.5.1, PS gradients also evaluate exact gradients, reverse-mode gradients are equivalent to hardware-compatible PS gradients for the present calculations. This enables a substantial reduction in the computational cost of evaluating the circuit gradients in quantum simulations using state vector simulators. Similar exact differentiation methods are available in several other quantum software frameworks [175]. However, since these methods rely on access to the simulated state vector, they are not a hardware-compatible gradient evaluation protocol. For this reason, the parameter-shift rule as described in sec. 4.5.1 and sec. 5.2.4 remains the reference procedure whenever hardware-relevant computational complexity, circuit-evaluation counts, or gate counts are discussed.

Although alternative and potentially more performant simulation methods such as optimised tensor-network simulators are available, we nonetheless opted for an exact state vector simulator. In the high-accuracy regime considered here an exact simulation is of great importance, since it avoids introducing any additional simulator-dependent approximation and inaccuracies into the results.

For the $N = 6$ system considered here, each individual VQE calculation is sufficiently small that parallelisation within a single optimisation gives little practical benefit. The calculations were therefore parallelised over independent VQE instances, corresponding to different random initialisations and different HVA layer numbers.

5.4.2 Energy Accuracy

For each value of L we perform $N_{\text{seed}} = 50$ independent VQE calculations according to the implementation detailed in sec. 5.2, and in particular with random parameter initialisations as described in sec. 5.2.4. The accuracy of the ground-state energy predicted by these VQE calculations is shown in fig. 5.3, where the energy error is defined relative to the ED result as:

$$\Delta E \equiv |E_{\text{VQE}} - E_0| \tag{5.38}$$

for the ED reference, E_0, and the energy predicted by the VQE calculation E_{VQE}. We also plot the probability of success in reaching ΔE_{chem} across the VQE seeds in fig. 5.4.

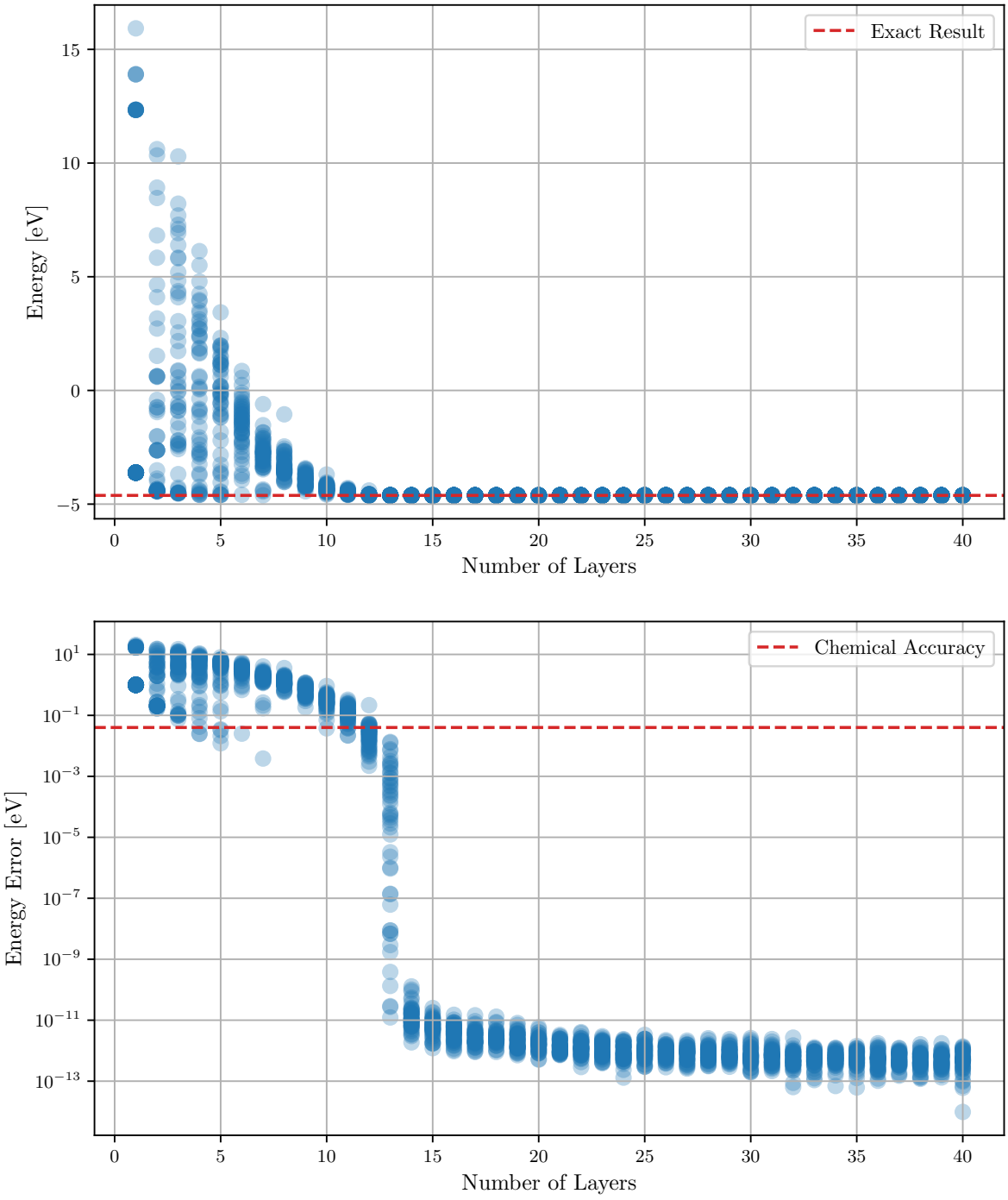


Figure 5.3: (a) Energy, E_{VQE}, and (b) Energy error, ΔE, as a function of the number of layers, L, in the ansatz circuit. Each marker corresponds to an individual VQE calculation. Markers are plotted with partial transparency; darker regions indicate higher point density due to overlap.

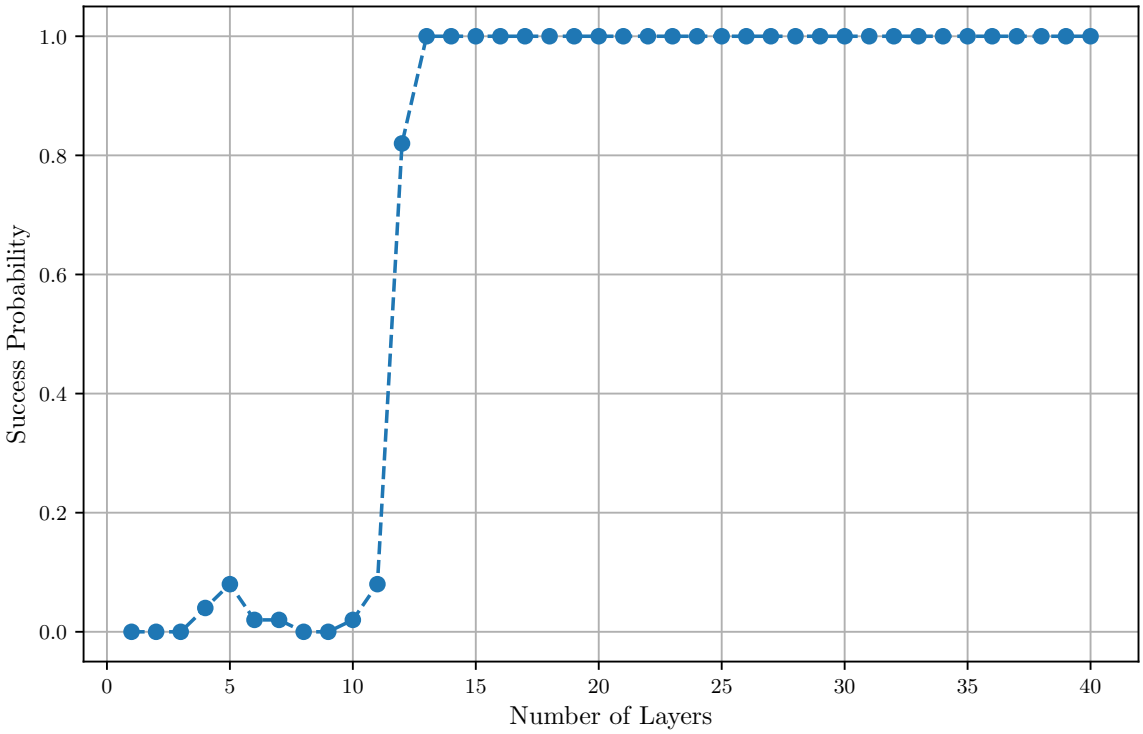


Figure 5.4: Probability of success, $\Delta E < \Delta E_{\text{chem}}$, as a function of the number of layers, L, evaluated over all VQE calculations performed at each L.

The results in fig. 5.3 and fig. 5.4 show that increasing L yields a pronounced improvement in the attainable accuracy for the ground-state energy. In particular, around a characteristic depth – in our case approximately $L \approx 13$ – the VQE accuracy rises sharply, reaching the chemically relevant regime, $\Delta E < \Delta E_{\text{chem}}$, after which it quickly saturates to an accuracy limited by the accumulation of machine precision errors. Chemical accuracy can be achieved reliably with an ansatz circuit starting at depth $L = 13$. The sharp onset of high accuracy with increasing L suggests two qualitatively distinct regimes: an under-parametrised regime in which the ansatz cannot reliably represent (or the optimiser cannot reliably reach) the ground state to within chemical accuracy, and an over-parametrised regime in which the exact ground state is readily accessible and the optimisation typically succeeds. Most importantly, there is a negligibly small transition region; the VQE calculations produce either chemically meaningless results for $L < 13$ or reproduce the ground state exactly for $L > 13$. This is unusual behaviour from a viewpoint of standard electronic structure theory where convergence with respect to computational parameters tends to be slow and systematic, and chemical accuracy does not necessarily require the preparation of an exact ground state.

This behaviour is indicative of a feature known as a computational phase transition [110]. As shown in fig. 5.5, a key practical feature of the transition is that opti-

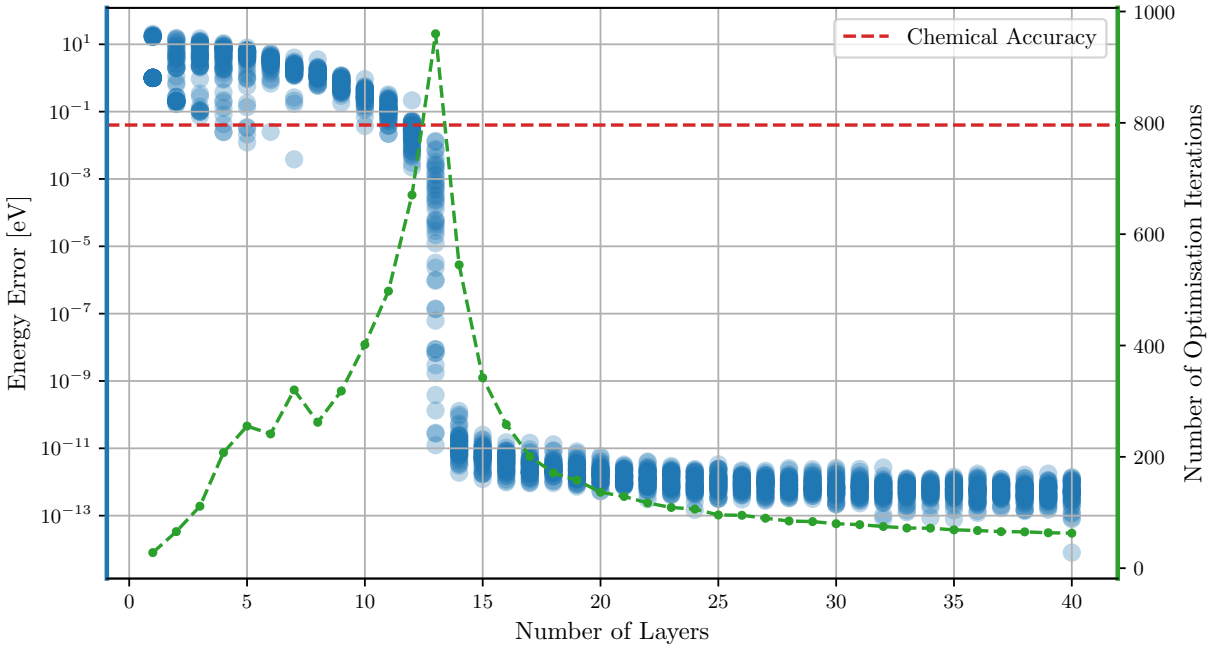


Figure 5.5: Same data as in fig. 5.3; now also showing the number of optimisation iterations required for the VQE calculation to converge as a function of the number of layers, L, averaged over all VQE calculations performed at each L.

misation is most challenging in its vicinity, where the number of iterations required for the optimisation algorithm to converge increases sharply, with the regimes either side converging considerably faster. This is further illustrated in fig. 5.6, where it can be seen that the convergence of the optimisation algorithm to the (local) minimum slows considerably at the transition. The optimisation algorithm used, L-BFGS, is a quasi-Newton algorithm that displays super-linear convergence for well-behaved, approximately quadratic parameter surfaces [130]. The considerably slowed convergence rate at $L = 13$ is therefore indicative of a highly non-quadratic objective function.

The behaviour for our problem instance is consistent with a growing body of literature analysing the parameter surface and optimisability of VQE ansatz circuits [19, 110, 176–178], where it has been found in general that VQE algorithms have highly non-quadratic parameter surfaces with many local minima of bad accuracy. Furthermore, it has been suggested that the onset of this over-parametrisation can in certain problem instances scale exponentially with the size of the system [177] and therefore present a substantial roadblock to VQE simulations of larger systems. ADAPT-VQE, and similar approaches, have been suggested as a possible solution for this phenomenon, since it has been found to be able to prepare an accurate ground state without the need for over-parametrising the ansatz circuit [19].

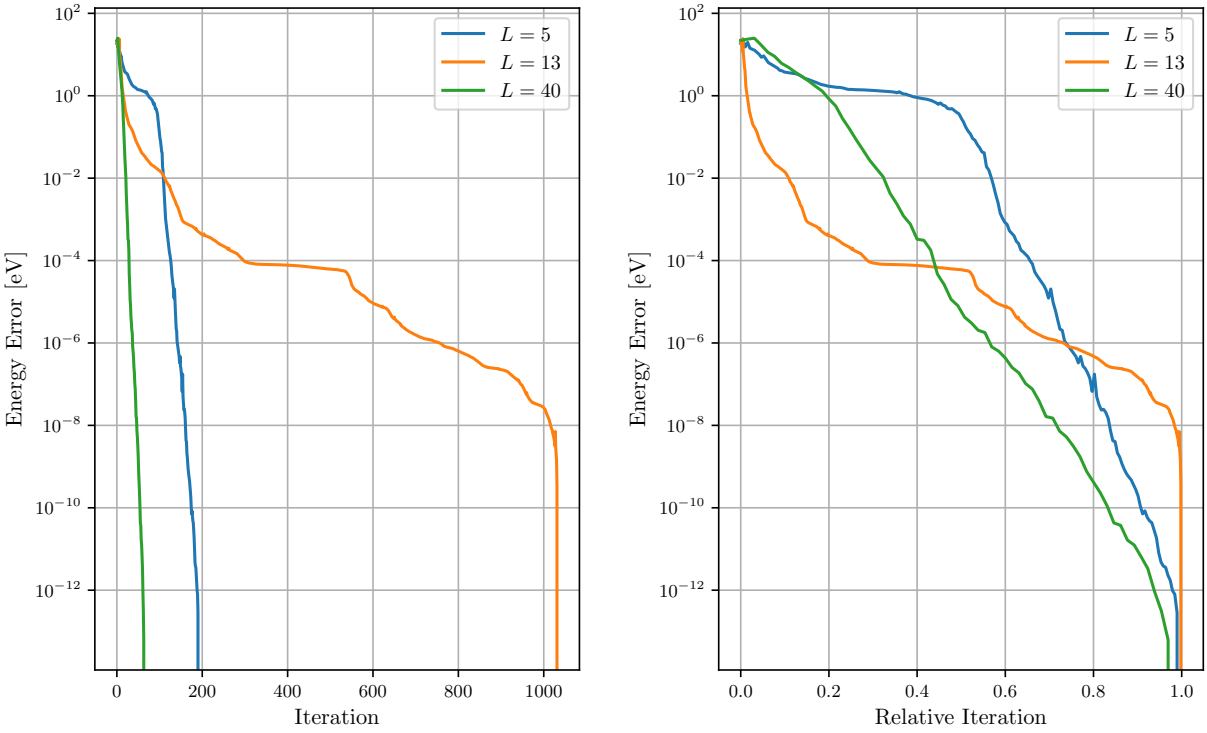


Figure 5.6: Representative convergence traces of the VQE optimisation displaying the energy error relative to the optimal VQE result for $L = 5, 13$ and 40 in terms of (a) absolute iteration count and (b) iteration count relative to final iteration count of each convergence trace.

5.4.3 Molecular Force Accuracy

We now assess convergence of molecular forces with respect to ansatz depth. Fig. 5.7 shows the accuracy of the molecular forces predicted by each VQE calculation, where we quantify the molecular force accuracy according to the root mean square (RMS) error relative to the ED result:

$$\Delta F \equiv \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(F_{\mathrm{VQE},i} - F_{0,i}\right)^2} \tag{5.39}$$

Similarly, $\boldsymbol{F}_0$ is the ED reference result, while $\boldsymbol{F}_{\mathrm{VQE}}$ is the molecular force predicted by the VQE calculation.

The data depicted in fig. 5.7 shows that force convergence tracks energy convergence closely: once the ansatz depth exceeds the computational transition at $L \approx 13$, the force error typically collapses to below the target threshold for all calculations. Again, after this transition molecular forces are converged to within machine precision. As a result of the computational phase transition exhibited by our VQE calculation, we therefore find that practically useful forces can be calculated

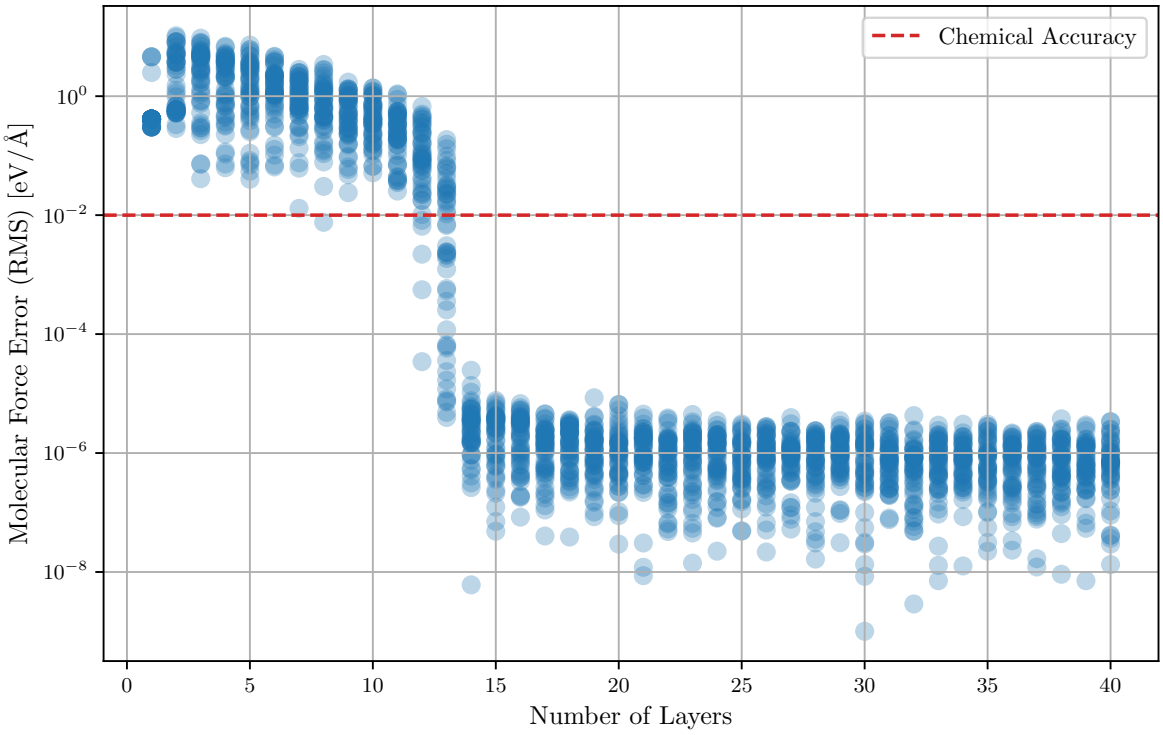


Figure 5.7: Molecular force error, ΔF, as a function of the number of layers, L, in the ansatz circuit. Each marker corresponds to an individual VQE calculations. Markers are plotted with partial transparency; darker regions indicate higher point density due to overlap.

in our problem instance without any increase in circuit depth and hence quantum resources compared to chemically accurate energies. Just as for the ground-state energy, molecular forces reach the relevant chemical accuracy targets for an ansatz starting at depth $L = 13$. This stands in stark contrast to standard electronic structure theory, where, as discussed in sec. 2.2, the variational principle ensures an in general much faster and easier convergence of energies compared to molecular forces.

Furthermore, the above results indicate that for our problem instance the non-variational response forces, which, as discussed in sec. 5.3.1, we have neglected in our calculations, do not need to be accounted for. Again, as a result of the computational phase transition exhibited by our VQE calculations, the exact HF force, which by virtue of the HF theorem includes any potential non-variational response, can be calculated without any additional quantum resources. In contrast, in standard electronic structure theory, the wave-function accuracy is generally too low to neglect these contributions.

However, the variational principle is still expected to apply to VQE calculations. To analyse the validity of the variational principle we additionally compare the

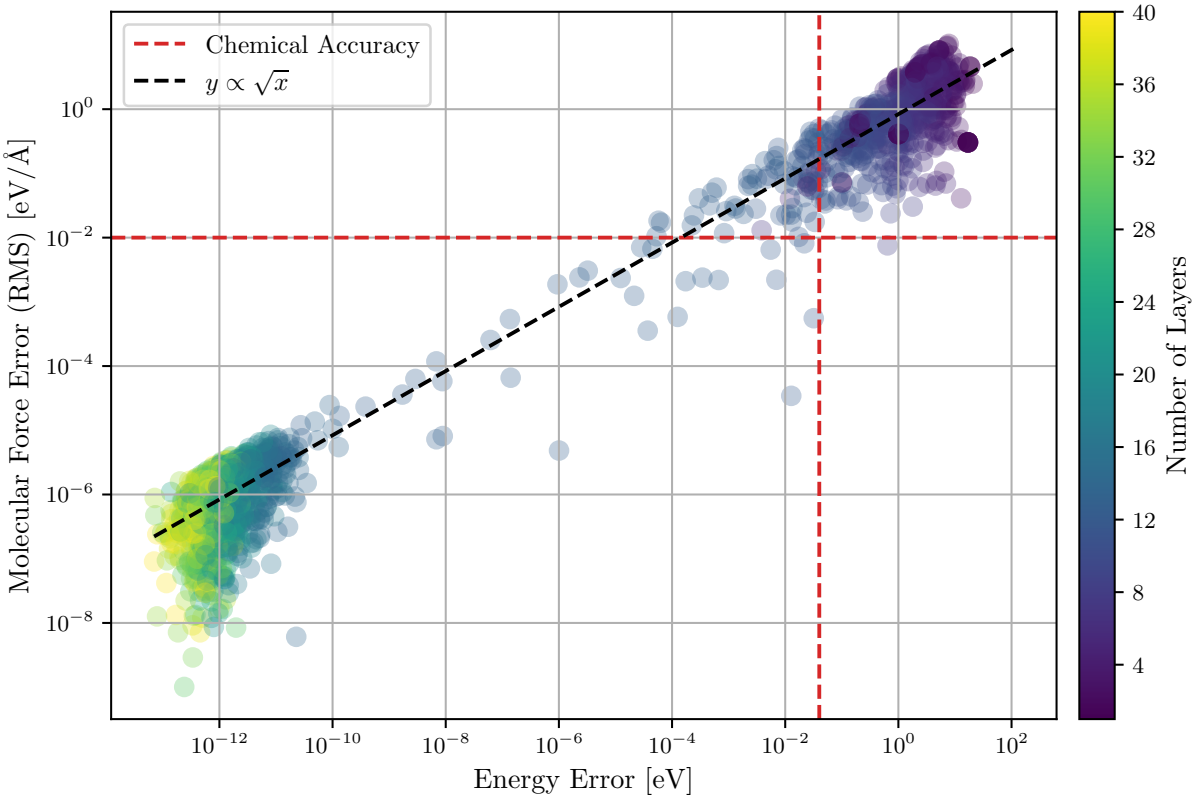


Figure 5.8: Molecular force error, ΔF, against energy error, ΔE, for all VQE calculations at any L. Each marker corresponds to an individual VQE calculations. Markers are plotted with partial transparency; darker regions indicate higher point density due to overlap.

relative accuracies of energies and molecular forces as L is varied in fig. 5.8.

As can be seen from this data, the variational principle holds for our VQE calculations, as expected. The accuracies of energies and molecular forces clearly follow the relationship:

$$\Delta F \propto \sqrt{\Delta E} \tag{5.40}$$

as predicted by the variational principle, equ. (2.26). However, as discussed, the convergence to chemical accuracy is dominated by the sharp computational phase transition; the negligible transition region – as evidenced by the clustering of points around complete (non-)convergence and the sparsity of points in-between – causes both ΔE and ΔF to converge at approximately the same L, and hence the effect of the variational principle on relative computational cost of ΔE and ΔF to be minimal.

5.5 Conclusion

In this chapter we presented idealised (noiseless, infinite-shot) VQE simulations of TPA within the interacting SSH model, with the goal of quantifying the algorithmic resources required for chemically meaningful predictions of both ground-state energies and molecular forces, and of establishing a clean baseline that isolates ansatz and optimiser behaviour from the sampling- and hardware-induced errors considered later. By working with a compact lattice Hamiltonian that nonetheless retains the essential interplay between correlation and geometry, we were able to study a non-trivial electronic-structure target while directly benchmarking molecular forces, an observable that is central to realistic electronic structure calculations but is still comparatively under-explored in VQE studies.

For the $N = 6$ instance studied, the HVA shows a clear depth-controlled transition: in the under-parametrised regime the energy improves only slowly, but around $L \approx 13$ the error drops rapidly, after which chemically accurate energies are obtained reliably and the exact-diagonalisation reference is quickly reached to machine precision. Molecular forces follow the same trend: once the circuit enters the accurate-energy regime, the Hellmann–Feynman force error simultaneously falls below chemical accuracy and rapidly approaches the exact result. Consistent with the variational scaling $\Delta F \propto \sqrt{\Delta E}$, this implies that in the idealised setting forces do not require appreciably greater depth than energies, and that any neglected non-variational response contributions become irrelevant once near-exact state preparation is achieved.

Taken together, these ideal simulations suggest that, for the present model instance, the dominant barrier to chemically meaningful predictions is not a slow, incremental variational convergence but the onset of a regime in which ansatz depth and optimisation jointly enable near-exact state preparation. This provides a clear baseline for the remainder of the thesis, where we examine how finite sampling and imperfect circuits modify this behaviour and reintroduce practical constraints on the feasibility of energy- and force-driven calculations.

Chapter 6

Realistic VQE Simulations of *Trans*-polyacetylene

In this chapter we study noisy variational quantum eigensolver (VQE) simulations of *trans*-polyacetylene (TPA). We consider two distinct error sources that limit the accuracy of VQE predictions: shot noise, arising from the statistical fluctuations inherent in estimating expectation values from a finite number of measurements, and circuit noise, arising from device imperfections such as finite gate fidelities, decoherence, and state-preparation and measurement errors.

This chapter builds directly on the idealised, noiseless VQE calculations presented in sec. 5. As discussed, TPA is an informative NISQ-era benchmark because, unlike the small molecular test sets often used in VQE studies, it admits a description in terms of a relatively simple one-dimensional Hubbard-like lattice model that remains chemically grounded — the Su–Schrieffer–Heeger (SSH) model. At the same time it constitutes a non-trivial benchmark: the ground state is highly correlated, and a qualitatively correct description of the model physics also requires an accurate prediction of the equilibrium geometry. Approaching from the viewpoint of an electronic structure practitioner interested in performing useful calculations of realistic materials, in sec. 5 we therefore studied VQE predictions of both ground-state energies and molecular forces within the SSH model, targeting chemical accuracy and quantifying the quantum resources required to achieve it.

The aim of the work in this chapter is to quantify the limitations that shot and circuit noise impose on chemically accurate predictions of ground-state energies and molecular forces for the SSH model – from the same electronic structure practitioner viewpoint. Circuit executions were previously assumed perfect and expectation values were evaluated exactly. Here we retain the same physical problem instance and VQE workflow so that any degradation in accuracy can be attributed

directly to shot noise or circuit noise. We treat the number of shots and the effective strength of circuit noise as computational parameters that control the accuracy of the simulation, and we consider what measurement budget and what noise levels are required before chemically accurate results become attainable. As in sec. 5 we are particularly interested in molecular forces, and in the extent to which they are more difficult to converge to chemical accuracy than ground-state energies. While the most realistic simulations would consider the combined effect of shot and circuit noise simultaneously, such calculations are beyond the scope of this thesis and the available classical and quantum resources. Instead, we study shot noise and circuit noise in turn, neglecting the other source in each set of simulations.

We begin by discussing the implementation of our VQE calculations in the noisy setting, for each shot and circuit noise. We then apply the variational principle with respect to both noise sources, and discuss how energies and molecular forces are expected to be affected and to what extent the variational principle is violated. Lastly, we present numerical results for the accuracy of energies and molecular forces in the presence of either shot noise or circuit noise.

6.1 VQE Implementation

In this section we describe the practical implementation details of the VQE calculations used throughout this chapter. The overall methodology follows the one introduced in sec. 5. The main difference is that, rather than performing a systematic convergence study with respect to the number of ansatz layers, L, we specialise to a fixed circuit depth of $L = 15$. This choice is motivated by the noiseless benchmark results presented previously, which indicate that $L = 15$ achieves chemical accuracy for the observables of interest with high consistency across random initialisations. Fixing the circuit depth in this way isolates the impact of stochastic shot noise and realistic circuit noise on an ansatz that is sufficiently expressive in the ideal limit, and hence allows us to determine the limits each noise source imposes on the ability to reach chemical accuracy in the VQE calculations.

At the same time, this specialisation introduces an important caveat: there is no guarantee that the depth $L = 15$ remains optimal once either of the two noise sources considered are present. In particular, the best-performing ansatz depth at a specific noise-level can depend non-trivially on the level of noise. Deeper circuits can offer increased expressivity and hence accuracy in the ideal setting, but also accumulate greater errors due to circuit noise [179]. Similarly, as the circuit depth increases, the optimisation landscape could become considerably more prone to the effects of

shot noise, due to a requirement to resolve gradients that may be exponentially small – a phenomenon known as barren plateaus [180]. However, as a result of the abrupt computational phase transition described in sec. 5.4, we expect these effects to be relatively minor. The depth of the ansatz circuit cannot be reduced much further in order to reduce the impact of noise or the effect of barren plateaus without completely preventing VQE from reaching close to chemical accuracy.

In the following, we describe the VQE implementation details that are specific to the calculations, considering shot noise and circuit noise separately.

6.1.1 Shot Noise

Modelling Shot Noise

As discussed in sec. 4.4, on quantum hardware expectation values, $\langle \hat{H} \rangle$ and similar, are estimated using a number of shots, M – repeated projective measurements probabilistically collapsing the state according to the measurement outcome. In principle, a faithful simulation of measurements and the associated stochasticity requires implementing such a collapse of the wavefunction for each measurement, and the circuit therefore needs to be simulated separately for each individual shot. This presents a severe limitation to classical simulations of quantum circuits, especially in the high-shot regime that is relevant when targeting chemically meaningful accuracies in VQE.

However, since VQE only requires the measurements to estimate expectation values of the final state, all measurements are performed at the end of the circuit and the collapsed state is immediately discarded. A separate circuit simulation for each shot is therefore not required. Instead, the state can be prepared only once, and all measurement outcomes are sampled directly according to Born's rule on this state for all required shots. However, even under this simplification the cost to sample the shots can be substantial [145].

Furthermore, in VQE we are not directly interested in the full distribution of bitstring outcomes, but only in an unbiased statistical estimator $\bar{H}$ of $\langle \hat{H} \rangle$ constructed from the measurements. As discussed in sec. 4.4, $\bar{H}$ has some associated variance $\mathrm{Var}[\bar{H}]$ depending on the measurement procedure and prepared state. In the regime of large shot counts in which chemically accurate VQE simulations operate, the central limit theorem (CLT) implies that the resulting energy estimator is well-approximated by a normal random variable. Rather than explicitly sampling bitstrings for each shot, it suffices to draw a single sample from a Gaussian distribution of mean $\mu = \langle \hat{H} \rangle$ and variance $\sigma^2 = \mathrm{Var}[\bar{H}]$. Crucially, this allows shot noise to be

simulated at a negligible overhead independent of the number of shots, and makes the high shot noise simulations of VQE presented in this work possible.

Concretely, the measurement procedure introduced in sec. 5.2.3 decomposes the Hamiltonian as:

$$\hat{H} = \hat{H}_{\mathrm{e}} + \hat{H}_{\mathrm{o}} + \hat{H}_{\mathrm{ee}} \tag{6.1}$$

The estimator variance can be evaluated as:

$$\mathrm{Var}[\bar{H}] = \frac{1}{M}\left(\mathrm{Var}(\hat{H}_{\mathrm{e}}) + \mathrm{Var}(\hat{H}_{\mathrm{o}}) + \mathrm{Var}(\hat{H}_{\mathrm{ee}})\right) \tag{6.2}$$

where the quantum variances $\mathrm{Var}(\cdot)$ can be calculated from the wavefunction $|\psi\rangle$ using $\mathrm{Var}(\hat{O}) = \langle \hat{O}^2 \rangle_\psi - \langle \hat{O} \rangle_\psi^2$, as discussed in sec. 4.4.

Classical Optimisation

As discussed in sec. 4.5, the classical optimisation and choice of algorithm plays a crucial role in determining the accuracy of the VQE calculation. In the presence of shot noise we are faced with a stochastic optimisation problem in which the objective function, $E(\boldsymbol{\theta})$, is approximated by a stochastic estimator, $\bar{E}(\boldsymbol{\theta})$. We note that, similarly, any circuit gradients evaluated from either the finite-difference (FD) approximations or the parameter shift (PS) rules described in sec. 4.5.1 will suffer from the same stochastic noise, since both methods are directly based on evaluations of $E(\boldsymbol{\theta})$.

As previously shown in sec. 5.4 for our problem instance limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) is well suited to the minimisation of the VQE energy in the noiseless regime, converging quickly and without a need for tuning the algorithm to the objective function. However, as discussed in sec. 4.5, it is less well suited to VQE in realistic calculations as a result of its susceptibility to noise and requirement for (within the VQE framework) expensive gradient evaluations. Nonetheless, following The Principle of Generality we consider it as a candidate for realistic calculations with shot noise. Its problem-agnostic, fast convergence is highly desirable, especially from a viewpoint of realistic calculations performed by an electronic structure practitioner. Following the same methodology of sec. 5.2.4, gradients are evaluated using the parameter shift (PS) rules. For the ansatz circuit considered we find that one evaluation of the full PS gradient requires $C_{\mathrm{step}} = 2400$ evaluations of $E(\boldsymbol{\theta})$.

Furthermore, we also consider the simultaneous perturbation stochastic approximation (SPSA) minimisation algorithm for our calculations. As discussed in sec. 4.5, the SPSA algorithm has found applications in VQE particularly for the minimi-

sation of the energy in realistic calculations, due to its inherent ability to optimise in the presence of stochasticity, and relative shot frugality per iteration compared to a direct evaluation of FD or PS circuit gradients. As such and following The Principle of Generality we also consider it as a candidate optimisation algorithm for our calculations.

All optimisations for both L-BFGS and SPSA are initialised using the same generic uniform initialisation scheme described in sec. 5.2.4.

Within our calculations we make the simplifying assumption that all measurements of expectation values within a single VQE calculation are performed with the same fixed number of shots, M – both the evaluations used during the optimisation as well as the evaluation of $E_{\mathrm{VQE}}/\boldsymbol{F}_{\mathrm{VQE}}$. We note that this neglects the potential for shot-frugal optimisation; algorithms have been developed which adjust the number of shots taken as the optimisation progresses, using fewer measurements at the start when the accuracy requirements are low, with the aim of minimising quantum resources used [145, 181]. Furthermore, SPSA and other stochastic algorithms have been shown to sometimes be able to beat the noise floor by implicitly averaging over previous iterates [135]; this could allow a lower shot count to be used during the optimisation than for the final measurement of $E_{\mathrm{VQE}}/\boldsymbol{F}_{\mathrm{VQE}}$, or vice versa an increased accuracy by allotting more shots to the final measurement of $E_{\mathrm{VQE}}/\boldsymbol{F}_{\mathrm{VQE}}$. However, in either case the accuracy of the final result is still expected to be limited by the number of shots used to estimate the final measurements of E_{VQE} and $\boldsymbol{F}_{\mathrm{VQE}}$ and the ability to resolve the minimum within this error.

6.1.2 Circuit Noise

Modelling Circuit Noise

Hardware imperfections of real quantum devices are usually considered as part of quantum simulations through the specification of a noise model for the gates comprising the ansatz circuit, frequently depolarising or amplitude damping noise-channels [179, 182]. However, calculations are impacted by realistic hardware and its limitations in many other ways. Not only are additional noise sources present, for example idling noise or state preparation and measurement (SPAM) noise, but restricted topologies and native gate sets, and the resulting requirement for transpilation of the circuit onto the hardware, can also affect the extent to which circuits are impacted by noise. In this work, we model circuit noise using a device-calibrated noise model based on the IBM Sherbrooke QPU [173]. The noise model is used here as a representative noise model for currently available superconducting devices,

rather than as an attempt to reproduce the behaviour of a particular experimental run on the device.

Since we are interested in simulating the circuit at varying levels of circuit noise reduction we have to consider the form of the noise model more closely. Realistic device noise models comprise several distinct contributions:

1. **Gate error**, capturing the imperfect implementation of quantum gates (single- and two-qubit) via effective noise channels applied to each instruction

2. **Idling error**, describing decoherence accumulated while qubits are idling during scheduling gaps or routing overhead

3. **State preparation error**, representing inaccuracies in initialisation and reset procedures

4. **Measurement error**, accounting for imperfections in the measurement procedure leading to an incorrect readout of bit-strings

For our calculations we decide to focus on the one- and two-qubit gate noise. These are in general – as well as for our VQE instance as discussed further in sec. 6.4 – the dominant noise sources. Furthermore, we also decide to ignore the noise introduced by the gate error due to the basis transformation required in the measurement protocol described in sec. 5.2.3. This allows for a simplified and cheaper simulation, since it obviates the need to simulate a separate circuit for each measurement group.

A practical challenge is that these error sources are not controlled by a single independent parameter, and scaling them "up" or "down" consistently by adjusting the gate properties is not straightforward. In particular, the gate errors in the QISKIT noise models account not only for thermal relaxation of qubits through an amplitude damping channel:

$$\mathcal{E}_{\mathrm{AD}}(\rho) = \begin{pmatrix} \rho_{00} + (1 - e^{-T_{\mathrm{gate}}/T_1})\rho_{01} & e^{-T_{\mathrm{gate}}/T_2}\rho_{01} \\ e^{-T_{\mathrm{gate}}/T_2}\rho_{10} & e^{-T_{\mathrm{gate}}/T_1}\rho_{11} \end{pmatrix} \tag{6.3}$$

but also for general infidelity through a depolarising noise channel:

$$\mathcal{E}_{\mathrm{DP}}(\rho) = (1 - \lambda)\,\rho \;+\; \frac{\lambda}{2}\,I \tag{6.4}$$

which are combined into the noise channel $\mathcal{E}(\rho) = \mathcal{E}_{\mathrm{DP}}\left(\mathcal{E}_{\mathrm{AD}}(\rho)\right)$. This noise channel is applied to the density matrix, ρ, of the quantum state after the quantum gate. T_{gate} is the gate duration, T_1 and T_2 are benchmarked properties of the qubit, and λ is

chosen heuristically to match the benchmarked (in-)fidelity of the gate. A particular natural and simple way to rescale errors is to form the convex combination:

$$\mathcal{E}_f := \mathcal{I} + f\,(\mathcal{E} - \mathcal{I}) \tag{6.5}$$

where f can be interpreted as the fractional reduction in error probabilities, or equivalently the error probability of the gates relative to the QPU baseline. Applied to $\mathcal{E}_{\text{AD}}$ and $\mathcal{E}_{\text{DP}}$ the new channel can be seen to rescale the error probability of the gates as $p \to f \cdot p$, provided one makes the natural identification of $p \to (1 - e^{-T_{\text{gate}}/T_1})$, $(1 - e^{-T_{\text{gate}}/T_2})$ and λ. Physically, a reduction of noise $\mathcal{E} \to \mathcal{E}_f$ can therefore be seen to be achieved through an improvement in $T_1/T_2/\lambda$. This is the method that we choose for our calculations.

Transferring Ideal Solutions

Benchmarking the accuracy of a VQE calculation under a realistic noise model requires performing the full optimisation under the considered noise model, as is done for example for the simulations involving shot noise. However, the inclusion of circuit noise makes simulations considerably more computationally expensive, and hence full VQE calculations for our problem intractable. We instead focus on analysing the degradation of the accuracy of the optimal solution found under the ideal VQE simulations of sec. 5.4. While this does not represent a fully realistic VQE framework, it is nonetheless expected to provide a reasonably cheap and controlled baseline for the accuracy achievable by the VQE calculations under the inclusion of circuit noise. Specifically, we consider:

$$E_{\text{noisy}}(\boldsymbol{\theta}_*) \qquad \text{with} \qquad \boldsymbol{\theta}_* = \arg\min_{\theta} E_{\text{ideal}}(\boldsymbol{\theta}) \tag{6.6}$$

where E_{noisy} and E_{ideal} refer to an evaluation of the objective function in the noisy and ideal setting respectively.

6.2 The Variational Principle in Realistic Calculations

As discussed in sec. 2.2, the variational principle of electronic structure theory leads to a favourable scaling of energy errors, δE, compared to molecular force errors, $\boldsymbol{\delta F}$, with respect to the accuracy with which the true ground state ψ_0 is approximated, $\delta\psi = \psi - \psi_0$. This principle applies quite broadly throughout (variational) electronic

structure theory since the relevant computational parameters affect the accuracy of the calculation implicitly through the accuracy of the wavefunction approximation. Specifically, in DFT the energy cut-off, E_{cut}, directly determines the variational space of electronic densities, ρ, over which we are searching for the minimum of the density functional, $E[\rho]$. As a result, the dominant effect of E_{cut} is through an implicit dependence $\delta\rho = \delta\rho(E_{\text{cut}})$ such that the variational principle directly applies to energy and molecular force errors induced by a finite E_{cut}:

$$\delta E(E_{\text{cut}}) = E[\rho(E_{\text{cut}})] - E[\rho_0] = \mathcal{O}(\delta\rho(E_{\text{cut}})^2) \tag{6.7}$$

$$\boldsymbol{\delta F}(E_{\text{cut}}) = \boldsymbol{F}[\rho(E_{\text{cut}})] - \boldsymbol{F}[\rho_0] = \mathcal{O}(\delta\rho(E_{\text{cut}})) \tag{6.8}$$

Similarly, in a CISD calculation the basis set size, N_b, determines the variational space of wavefunctions, ψ, over which the energy functional $E[\psi]$ is minimised, and affects the accuracy only through an implicit dependence $\delta\psi = \delta\psi(N_b)$ so that, again, the variational principle applies directly:

$$\delta E(N_b) = E[\psi(N_b)] - E[\psi_0] = \mathcal{O}(\delta\psi(N_b)^2) \tag{6.9}$$

$$\boldsymbol{\delta F}(N_b) = \boldsymbol{F}[\psi(N_b)] - \boldsymbol{F}[\psi_0] = \mathcal{O}(\delta\psi(N_b)) \tag{6.10}$$

The two main computational parameters limiting the accuracies of our realistic VQE calculation, shot and circuit noise, behave considerably differently, since they are in general not directly variational, as might be expected from standard electronic structure theory. As we will discuss in the following sections, both the shot and the circuit noise, each characterised by a noise strength λ, can be considered to affect the accuracy of the calculation through both an implicit dependence $\delta\psi = \delta\psi(\lambda)$, where the noise limits our ability to resolve the exact ground state, as well as an explicit error on the functionals, $\Delta E(\lambda)$ and $\Delta\boldsymbol{F}(\lambda)$:

$$\delta E(\lambda) = E[\psi(\lambda)] - E[\psi_0] + \Delta E(\lambda) \tag{6.11}$$

$$\boldsymbol{\delta F}(\lambda) = \boldsymbol{F}[\psi(\lambda)] - \boldsymbol{F}[\psi_0] + \Delta\boldsymbol{F}(\lambda) \tag{6.12}$$

A direct application of the variational principle is therefore not possible and a more careful analysis of the dependence of each quantity on the strength of the noise is required. In the following sections we will therefore discuss the expected scaling of $\delta\psi = \delta\psi(\lambda)$ with respect to shot and circuit noise, as well as the implications for the relative scaling of energies and molecular forces.

6.2.1 Shot Noise

As previously discussed in sec. 4.4, on a quantum computer expectation values, $\langle\hat{O}\rangle_\psi$, are evaluated by performing repeated measurements and forming an unbiased estimator $\bar{O}$, $\mathbb{E}_\psi[\bar{O}] = \langle\hat{O}\rangle_\psi$, where we have made explicit the dependence on the state being measured. For any finite number of shots, M, this estimator will have a non-zero variance, $\mathrm{Var}_\psi[\bar{O}]$, and introduce statistical noise to the expectation value. As such at any finite M we have a statistical error associated with the estimation of observable expectation values, which can be quantified by the root mean square error (RMSE)

$$\mathrm{RMSE}(\bar{O}) \equiv \sqrt{\mathbb{E}_\psi\left[(\bar{O} - \langle\hat{O}\rangle_\psi)^2\right]} = \sqrt{\left(\mathbb{E}_\psi[\bar{O}] - \langle\hat{O}\rangle_\psi\right)^2 + \mathrm{Var}_\psi[\bar{O}]} \tag{6.13}$$

with its well-known bias variance decomposition.

Combined Systematic and Stochastic Errors Since given a wavefunction, ψ, the accuracy of our calculation is determined by the error relative to the exact ground-state energy $\langle\psi_0|\hat{H}|\psi_0\rangle = \langle\hat{H}\rangle_0$ (evaluated exactly, without any statistical noise), we quantify the accuracy of our energy estimate as the RMSE relative to $\langle\hat{H}\rangle_0$:

$$\delta E_{\mathrm{RMSE},0}(\psi) = \sqrt{\mathbb{E}_\psi\left[(\bar{H} - \langle\hat{H}\rangle_0])^2\right]} = \sqrt{\left(\langle\hat{H}\rangle_\psi - \langle\hat{H}\rangle_0\right)^2 + \mathrm{Var}_\psi[\bar{H}]} \tag{6.14}$$

$$= \sqrt{\delta E_v(\psi)^2 + \delta E_s(\psi)^2} \tag{6.15}$$

where we have defined:

$$\delta E_v(\psi) = \langle\hat{H}\rangle_\psi - \langle\hat{H}\rangle_0 \qquad \delta E_s(\psi) = \sqrt{\mathrm{Var}_\psi[\bar{H}]} \equiv \sigma_\psi(\bar{H}) \tag{6.16}$$

and where we stress that, while we will often suppress explicit dependences on ψ for legibility, $\delta E_v(\psi)$ and $\delta E_s(\psi)$ depend on ψ via $\langle\hat{H}\rangle_\psi$ and $\mathrm{Var}_\psi[\bar{H}]$. Similar expressions hold for the molecular force:

$$\delta\boldsymbol{F}_{\mathrm{RMSE},0}(\psi) = \sqrt{\delta\boldsymbol{F}_v(\psi)^2 + \delta\boldsymbol{F}_s(\psi)^2} \tag{6.17}$$

$$\delta\boldsymbol{F}_v(\psi) = \langle\hat{\boldsymbol{F}}\rangle_\psi - \langle\hat{\boldsymbol{F}}\rangle_0 \qquad \delta\boldsymbol{F}_s(\psi) = \sqrt{\mathrm{Var}_\psi[\bar{\boldsymbol{F}}]} \equiv \sigma_\psi(\bar{\boldsymbol{F}}) \tag{6.18}$$

The error consists of two contributions. The variational contribution, δE_v ($\delta\boldsymbol{F}_v$), quantifies the systematic error in the energy (molecular forces) evaluated exactly, i.e. non-stochastically, arising as a result of an approximate ground state $\psi = \psi_0 + \delta\psi$.

This is the standard source of error in electronic structure theory and is the error analysed as part of the variational principle in sec. 2.2 for which we found:

$$\delta E_v = \delta E[\psi_0 + \delta\psi] = \mathcal{O}(\delta\psi^2) \qquad \delta \boldsymbol{F}_v = \delta \boldsymbol{F}[\psi_0 + \delta\psi] = \mathcal{O}(\delta\psi) \tag{6.19}$$

On the other hand, the statistical contribution δE_v ($\delta \boldsymbol{F}_v$) quantifies the stochastic error that arises from a finite shot budget and hence stochastic estimation of $\langle \hat{H} \rangle_\psi$ ($\langle \hat{\boldsymbol{F}} \rangle_\psi$), and limits our ability to resolve the systematic variational error.

The Variational Principle As argued in equ. (6.11) the contribution from the statistical part δE_s ($\delta \boldsymbol{F}_s$) prevents a direct application of the variational principle to the combined error $\delta E_{\mathrm{RMSE},0}$ ($\delta \boldsymbol{F}_{\mathrm{RMSE},0}$). In particular, as $\mathrm{Var}[\bar{H}]$ depends on M (usually $\propto \frac{1}{\sqrt{M}}$ as a result of the standard error on the mean and variance propagation), an explicit dependence of the energy accuracy on the shot noise is introduced, $\delta E_s = \delta E_s(M)$, beyond the implicit and directly variational dependence $\delta E_v = \delta E_v(\delta\psi)$. We expect the statistical noise to also directly limit the best achievable accuracy in the wavefunction $\delta\psi_* = \delta\psi_*(M)$. Therefore, whether the consequences of the variational principle for the relative scaling of error in energies and molecular forces (equ. (2.26)) can be observed as M is varied will depend on the relative scaling of $\delta\psi_*(M)$ and the resulting variational errors $\delta E_v(\delta\psi_*(M))$, $\boldsymbol{\delta F}_v(\delta\psi_*(M))$, and statistical errors $\delta E_s(M), \boldsymbol{\delta F}_s(M)$. The variational principle is expected to apply if the variational part of the error is sufficiently resolved, that is, if in the limit of large M we have $\delta E_s(M) \leq \delta E_v(\delta\psi_*(M))$ ($\boldsymbol{\delta F}_s(M) \leq \boldsymbol{\delta F}_v(\delta\psi_*(M))$). A slightly different viewpoint shifts the focus on to the number of shots that would be required to fully resolve the variational error. At M and $\delta\psi_*(M)$ the number of shots M_E required to resolve δE_v is determined by $\delta E_v(\delta\psi_*(M_E)) \sim \delta E_s(M_E)$ (similarly for the molecular force and $M_{\boldsymbol{F}}$). If $M \leq M_{E/\boldsymbol{F}}$, we will be limited by the statistical error of our estimators, since the number of shots is not sufficient to fully resolve the variational error.

The Noise Floor While the above discussion keeps the limiting wavefunction accuracy $\delta\psi_*(M)$ as general as possible, we expect it itself to be linked to and determined by δE_v and δE_s. In order to analyse its behaviour we introduce the concept of the noise floor. We define the noise floor as the hyper-surface in the Hilbert space satisfying

$$\delta E_v(\delta\psi_*(M)) = \sqrt{\mathrm{Var}_{\delta\psi_*(M)}[\bar{H}]} = \delta E_s(\delta\psi_*(M), M) \tag{6.20}$$

i.e. where the variational and statistical contributions to the energy error are equal. This noise floor has previously been used in a simplified form to compare the performance of different optimisation algorithms at a fixed number of shots with respect to δE [135]. In our work, we will use it to analyse the expected scaling with M of the performance of optimisation algorithms, and, as discussed above, the resulting implications for the variational principle and relative scaling of energy and molecular force accuracies. The motivation for this definition comes from considering at which point ψ the statistical noise in the estimator $\bar{H}$ prevents us from resolving any further differences/improvements in the energy with high probability. At a point ψ with energy $\langle \hat{H} \rangle_\psi$ the largest improvement possible is given by $\delta E_v = \langle \hat{H} \rangle_\psi - \langle \hat{H} \rangle_0$. Hence, given an observation $\bar{H}$ at ψ, if $\sqrt{\mathrm{Var}_\psi[\bar{H}]} \sim \delta E_v(\psi)$ we will not be able to resolve any improvement in the energy, since with high probability we are likely to misidentify ψ as the lowest energy state, even if ψ_0 was known/sampled. While the choice of optimisation algorithm, associated hyper-parameters, and inherent stochasticity is going to affect the exact achievable/resolvable accuracy, we nonetheless expect the noise floor as defined by equ. (6.20) to set an approximate limit to our ability to resolve $\delta\psi$ at any finite M and hence to determine the scaling of $\delta\psi_*(M)$. Since equ. (6.20) defines the noise floor at a given M by the requirement that the variational error in the energy is just resolved, $\delta E_v = \delta E_s$, the focus of our analysis shifts to the variational and statistical errors in the molecular force. The total error in the energy (equ. (6.14)) simplifies to:

$$\delta E_{\mathrm{RMSE},0}(\delta\psi_*(M)) = \sqrt{\delta E_v(\delta\psi_*(M))^2 + \delta E_s(\delta\psi_*(M), M)^2} \tag{6.21}$$

$$= \sqrt{2}\ \delta E_v(\delta\psi_*(M)) = \sqrt{2}\ \delta E_s(\delta\psi_*(M), M) \tag{6.22}$$

and the main consideration is whether the variational error in the molecular forces at $\delta\psi_*(M)$ is sufficiently resolved given M, or whether the molecular force is limited by its statistical error.

Outline In the following sections, we will discuss and derive expressions for $\delta\psi_*(M)$ and resulting expressions for $\delta E_v(\delta\psi_*(M))$ $(\boldsymbol{\delta F}_v(\delta\psi_*(M)))$ and $\delta E_s(M)(\boldsymbol{\delta F}_s(M))$ in two different measurement schemes, considering both a general expansion as well as the specific example of a two level system. We first look at the ideal, but unrealistic scenario of a direct measurement of the energy, where we prove that it is the optimal measurement scheme. We apply the variational principle for the variance and show how it leads to favourable scaling of $\delta\psi_*(M)$. As a result, we find that the variational principle for the relative scaling of energy and molecular force errors applies, since the variational errors of the molecular forces

are always sufficiently resolved. We then consider a realistic measurement scheme and argue that it breaks the variational principle for the variance, which leads to a worse scaling of $\delta\psi_*(M)$ and hence an increased sampling cost. We also find that, as a result, the variational error in the molecular forces is not necessarily always resolved and instead might be limited by the statistical error, thereby breaking the expected relative variational scaling of energies and forces.

Direct Energy Measurement

We will first discuss the case of an estimation of the energy by a direct measurement in the basis of energy eigenstates (and similarly a direct measurement of the molecular force in the basis of molecular force eigenstates). As discussed in sec. 4.4 such a measurement would require implementation of either a unitary diagonalising the Hamiltonian, or the QPE protocol. While neither are feasible on NISQ devices, we will nonetheless discuss this as an ideal scenario, to elucidate the ways in which a realistic measurement scenario breaks the ideal results, and how a QPE enabled measurement procedure might be able to benefit from the variational principle. The variance of a direct measurement of the energy, i.e. projection onto the Hamiltonian eigenstates, is as discussed in sec. 4.4, given by:

$$\mathrm{Var}[\bar{H}] = \frac{\mathrm{Var}(\hat{H})}{M} \tag{6.23}$$

A direct measurement of the energy is not only the conceptually simplest measurement scheme, but can also be shown to be the ideal measurement procedure, in the sense that it minimises the variance of the estimator $\mathrm{Var}[\bar{H}]$. Such a measurement scheme is therefore not only highly beneficial in the limit of small wavefunction deviations $\delta\psi$ as a result of the variational principle, as we will show later, but is also the best measurement scheme for any wavefunction ψ in the Hilbert space.

Optimality We first prove the optimality of the direct energy measurement. Within the framework of measurement schemes considered in sec. 4.4 the variance of an estimator for $\langle\hat{H}\rangle$ is given by:

$$\mathrm{Var}[\bar{H}] = \sum_\alpha \mathrm{Var}[\bar{A}_\alpha] = \sum_\alpha \frac{\mathrm{Var}(\hat{A}_\alpha)}{m_\alpha} \tag{6.24}$$

where we recall that the estimator $\bar{H}$ is constructed by measuring individual observables $\{\hat{A}_\alpha\}$ that satisfy $\hat{H} = \sum_\alpha \hat{A}_\alpha$. By generalising an argument presented in [126] it can be shown that $\{\hat{A}_\alpha\} = \{\hat{H}\}$, i.e. a direct measurement of the energy,

minimises $\mathrm{Var}[\bar{H}]$. We first minimise $\mathrm{Var}[\bar{H}]$ for a fixed decomposition $\{\hat{A}_\alpha\}$ with respect to the allocation of shots m_α under the constraint of a fixed total shot budget $\sum_\alpha m_\alpha = M$. The optimal choice can be shown to satisfy $m_\alpha \propto \sqrt{\mathrm{Var}(\hat{A}_\alpha)}$ [183], with a resulting optimal variance given by:

$$\mathrm{Var}[\bar{H}] = \frac{1}{M}\left(\sum_\alpha \sqrt{\mathrm{Var}(\hat{A}_\alpha)}\right)^2 \tag{6.25}$$

We now further minimise $\mathrm{Var}[\bar{H}]$ over the possible decompositions $\{\hat{A}_\alpha\}$. Recalling that quantum variance of an observable is given by

$$\mathrm{Var}(\hat{O}) = \langle(\hat{O} - \langle\hat{O}\rangle)^2\rangle = \langle(\Delta\hat{O})^2\rangle \tag{6.26}$$

we can apply the triangle inequality $\|u+v\| \leq \|u\| + \|v\|$ by identifying $u = \Delta\hat{A}\,|\psi\rangle$ and $v = \Delta\hat{B}\,|\psi\rangle$. Observing that $\Delta(\hat{A}+\hat{B}) = \Delta\hat{A} + \Delta\hat{B}$ we find

$$\sqrt{\mathrm{Var}(\hat{A}+\hat{B})} \leq \sqrt{\mathrm{Var}(\hat{A})} + \sqrt{\mathrm{Var}(\hat{B})} \tag{6.27}$$

where equality is reached when $\hat{A} \propto \hat{B}$. Applying this inequality to equ. (6.25) we can show that consolidating two individually measured observables $\{\hat{A}_i, \hat{A}_j\}$ into a single measurement of $\hat{A}_i + \hat{A}_j$ always leads to lowering of the variance provided the measurement shots are allocated optimally:

$$\frac{1}{M}\left(\sum_{\alpha\neq i,j} \sqrt{\mathrm{Var}(\hat{A}_\alpha)} + \sqrt{\mathrm{Var}(\hat{A}_i + \hat{A}_j)}\right)^2 \leq \frac{1}{M}\left(\sum_\alpha \sqrt{\mathrm{Var}(\hat{A}_\alpha)}\right)^2 \tag{6.28}$$

We therefore conclude that the optimal measurement scheme is given by direct measurement of the Hamiltonian, i.e. $\{\hat{A}_\alpha\} = \{\hat{H}\}$ with variance of equ. (6.23), since it is the most consolidated measurement scheme $\{\hat{A}_\alpha\} \to \{\hat{H} = \sum_\alpha \hat{A}_\alpha\}$.

Zero-variance Property We now turn back to the analysis of the noise floor and the resulting variational and statistical errors for a direct measurement of the energy. As discussed in sec. 2.2, a variational principle can be established not only for the energy but also for the variance of the energy for which it was found (equ. (2.30)):

$$\mathrm{Var}_\psi(\hat{H}) = \mathrm{Var}_E[\psi_0 + \delta\psi] = \mathcal{O}(\delta\psi^2) \tag{6.29}$$

where, as mentioned previously, $\mathrm{Var}_E[\psi]$ corresponds up to notation to the quantum variance as discussed in the current chapter $\mathrm{Var}_\psi(\hat{H}) = \mathrm{Var}_E[\psi]$ and we have made

explicit the dependence of the variance on ψ. Applying this result to the variance of the estimator of a direct energy measurement (equ. (6.23)) at a wavefunction near the ground state, $\psi_0 + \delta\psi$, we find

$$\mathrm{Var}[\bar{H}] = \frac{\mathrm{Var}(\hat{H})}{M} = \frac{\mathrm{Var}_E[\psi_0 + \delta\psi]}{M} = \mathcal{O}\left(\frac{\delta\psi^2}{M}\right) \tag{6.30}$$

We see that the variance of the estimator decreases quadratically as we approach the exact ground state, vanishing exactly when $\psi = \psi_0$. This is highly beneficial since a reduced variance of the estimator corresponds to a reduced stochastic error at a fixed shot budget, with the limiting case at the exact ground state resulting in an exact, non-stochastic measurement ($\mathrm{Var}(\bar{H}) = 0$). At first this might seem to indicate that given any finite M we are always able to reach the exact ground-state energy, since $\delta E_{\mathrm{RMSE},0}(\psi_0) = 0$ independent of M. However, this would require knowledge of ψ_0 to begin with.

Noise Floor As discussed previously, we expect the noise floor defined by equ. (6.20) to limit our ability to resolve ψ_0 to within an error $\delta\psi_*(M)$ at a given M. Introducing the expansion coefficients α_E and β_E:

$$\delta E_v = \alpha_E \left\|\delta\psi\right\|^2 + \mathcal{O}(\delta\psi^3) \qquad \mathrm{Var}[\hat{H}] = \beta_E \left\|\delta\psi\right\|^2 + \mathcal{O}(\delta\psi^3) \tag{6.31}$$

which in full generality should be understood to correspond to directional functional derivatives but are simplified for legibility, we find that to leading orders in $\delta\psi$ the noise floor condition is:

$$\alpha_E \left\|\delta\psi_*\right\|^2 = \sqrt{\frac{\beta_E \left\|\delta\psi_*\right\|^2}{M}} \tag{6.32}$$

resulting in a noise floor that scales with M as:

$$\left\|\delta\psi_*\right\| = \frac{\sqrt{\beta_E}}{\alpha_E}\frac{1}{\sqrt{M}} = \mathcal{O}\left(\frac{1}{\sqrt{M}}\right) \tag{6.33}$$

Having established an expression for the noise floor we can now analyse the full dependence of the energy error $\delta E_{\mathrm{RMSE},0}$ (equ. (6.14)) on M. At $\delta\psi_*$ we find that the variational and statistical contributions to the energy error are expected to scale to leading orders in $\delta\psi$ as:

$$\delta E_v(\delta\psi_*(M)) = \delta E_s(\delta\psi_*(M), M) = \frac{\beta_E}{\alpha_E}\frac{1}{M} = \mathcal{O}\left(\frac{1}{M}\right) \tag{6.34}$$

and hence the total error is given by:

$$\delta E_{\mathrm{RMSE},0}(\delta\psi_*(M), M) = \sqrt{2}\frac{\beta_E}{\alpha_E}\frac{1}{M} = \mathcal{O}\left(\frac{1}{M}\right) \tag{6.35}$$

This establishes the expected scaling of the energy accuracy as M is varied, $\delta E \sim \frac{1}{M}$. This scaling exhibits a quadratic speedup compared to what might be expected when considering just the statistical error at a fixed point ψ, $\sqrt{\mathrm{Var}[\bar{H}]} = \mathcal{O}\left(\frac{1}{\sqrt{M}}\right)$.

Molecular Force Error Similarly, we now consider the accuracy of the molecular force $\delta\boldsymbol{F}_{\mathrm{RMSE},0}$ (equ. (6.17)) given a finite M at the noise floor limiting the accuracy of the VQE calculations. While the same argument as before can be made to show that a direct measurement of the force is the optimal estimator for the molecular forces, no variational principle applies to the variance of the molecular force near the ground state ψ_0 as discussed in sec. 2.2, where it was argued that in general the variance approaches a constant non-vanishing value (equ. (2.31)):

$$\mathrm{Var}_\psi(\hat{\boldsymbol{F}}) = \mathrm{Var}_{\boldsymbol{F}}[\psi_0 + \delta\psi] = \mathcal{O}(1) \tag{6.36}$$

where, again, up to notation, $\mathrm{Var}_\psi(\hat{\boldsymbol{F}}) = \mathrm{Var}_{\boldsymbol{F}}[\psi]$. Introducing similar expansion coefficients as for the energy (equ. (6.31)):

$$\delta\boldsymbol{F}_v = \boldsymbol{\alpha}_{\boldsymbol{F}}\,\|\delta\psi\| + \mathcal{O}(\delta\psi^2) \qquad \mathrm{Var}[\hat{\boldsymbol{F}}] = (\sigma_0(\hat{\boldsymbol{F}}))^2 + \mathcal{O}(\delta\psi) \tag{6.37}$$

we find at $\delta\psi_*$ (equ. (6.33)) to first order in $\delta\psi$:

$$\delta\boldsymbol{F}_v(\delta\psi_*(M)) = \boldsymbol{\alpha}_{\boldsymbol{F}}\frac{\sqrt{\beta_E}}{\alpha_E}\frac{1}{\sqrt{M}} = \mathcal{O}\left(\frac{1}{\sqrt{M}}\right) \tag{6.38}$$

$$\delta\boldsymbol{F}_s(\delta\psi_*(M), M) = \sigma_0(\hat{\boldsymbol{F}})\frac{1}{\sqrt{M}} = \mathcal{O}\left(\frac{1}{\sqrt{M}}\right) \tag{6.39}$$

with a total error:

$$\delta\boldsymbol{F}_{\mathrm{RMSE},0}(\delta\psi_*(M), M) = \frac{1}{\sqrt{M}}\sqrt{\left(\boldsymbol{\alpha}_{\boldsymbol{F}}\frac{\sqrt{\beta_E}}{\alpha_E}\right)^2 + \sigma_0(\hat{\boldsymbol{F}})^2} = \mathcal{O}\left(\frac{1}{\sqrt{M}}\right) \tag{6.40}$$

We therefore find that the achievable accuracy in the molecular forces scales as $\delta\boldsymbol{F} \sim \frac{1}{\sqrt{M}}$. Comparing equ. (6.35) and (6.40) we see that despite the introduction of shot noise, the accuracy achievable in energies and molecular forces at a given M obeys a variational-like scaling $\delta E \sim \sqrt{\delta\boldsymbol{F}}$ (equ. (2.25)). This is because the variational principle for the energy variance ensures that whenever the energy

is just resolved, $\delta E_v \sim \delta E_s$, the corresponding molecular force is also always just resolved, $\delta \boldsymbol{F}_v \sim \delta \boldsymbol{F}_s$ (up to constant prefactors), and neither quantity is asymptotically limited by the statistical error. From a different viewpoint, the number of shots M_E required to resolve a systematic energy error $\delta E_v(\delta\psi)$ at $\delta\psi$ is also sufficient to just resolve the corresponding systematic molecular force error $\delta \boldsymbol{F}_v(\delta\psi)$ at $\delta\psi$. It is important to stress that it should not be concluded from this viewpoint that the variational principle does not hold – e.g. both energies and molecular forces require the same number of shots and hence computational effort to resolve – since such a conclusion only considers the statistical errors at a fixed $\delta\psi$ and neglects the variational contribution to the total error through $\delta\psi_*(M)$.

Two Level System To make the arguments presented above more explicit, we consider the case of a simple two–level system, for which analytic expressions of the quantities considered in our argument can easily be derived. Given ground and excited eigenstates $|g\rangle, |e\rangle$ and an energy gap $\Delta = E_e - E_g > 0$, the Hamiltonian is given by:

$$\hat{H} = E_g|g\rangle\langle g| + E_e|e\rangle\langle e| \tag{6.41}$$

we parametrise the state vector as:

$$|\psi\rangle = \sin(\theta)|g\rangle + \cos(\theta)\,|e\rangle, \tag{6.42}$$

where for simplicity we have assumed $\phi = 0$ and redefined $\frac{\theta}{2} \to \theta$ in the standard parametrisation of the Bloch sphere. The energy and variance of ψ can easily be derived:

$$\langle \hat{H} \rangle_\psi = \cos^2(\theta) E_g + \sin^2(\theta) E_e = E_g + \sin^2(\theta)\ \cdot \Delta \tag{6.43}$$

$$\mathrm{Var}_\psi(\hat{H}) = \cos^2(\theta)\sin^2(\theta)\,(E_e - E_g)^2 = \cos^2(\theta)\sin^2(\theta)\ \cdot \Delta^2 \tag{6.44}$$

Expanding in θ about the ground state at $\theta = 0$ we can exemplify the variational principle for the energy and energy variance:

$$\delta E_v = \langle \hat{H} \rangle_\psi - E_g \sim \Delta \cdot \theta^2 = \mathcal{O}(\theta^2) \tag{6.45}$$

$$\mathrm{Var}_\psi(\hat{H}) \sim \Delta^2 \cdot \theta^2 = \mathcal{O}(\theta^2) \tag{6.46}$$

where we identify the wavefunction accuracy $\|\delta\psi\| \to \theta$, and the constants $\alpha_E \to \Delta$ and $\beta_E \to \Delta^2$. The noise floor (equ. (6.33)) in this two level system is therefore

given by:

$$\theta_* = \frac{\sqrt{\Delta^2}}{\Delta}\frac{1}{\sqrt{M}} = \frac{1}{\sqrt{M}} \tag{6.47}$$

and we find that the combined variational and statistical error is given by:

$$\delta E_{\mathrm{RMSE},0}(\theta_*, M) = \sqrt{2}\frac{\Delta^2}{\Delta}\frac{1}{M} = \sqrt{2}\Delta\frac{1}{M} \tag{6.48}$$

The molecular force in the two level system is given by a non-diagonal observable $\hat{F}$ with matrix elements $F_{ij} = \langle i|\hat{F}|j\rangle$. For simplicity we consider $F_{00} = F_{11} \equiv F_d$ and $F_{01} = F_{10} \equiv F_c$. The energy and variance of this non-diagonal force operator are given by:

$$\langle\hat{F}\rangle_\psi = F_d + 2\sin(\theta)\cos(\theta)F_c \tag{6.49}$$

$$\mathrm{Var}_\psi(\hat{F}) = F_c^2\cos^2(2\theta) \tag{6.50}$$

where again, we can exemplify the absence of a variational principle for the molecular force:

$$\delta F_v = \langle\hat{F}\rangle_\psi - F_d \sim 2F_c\cdot\theta = \mathcal{O}(\theta) \tag{6.51}$$

$$\mathrm{Var}_\psi(\hat{F}) \sim F_c^2 - 4F_c^2\cdot\theta^2 = \mathcal{O}(1) \tag{6.52}$$

and identify $\alpha_F \to 2F_c$ and $\sigma_0(\hat{F}) \to F_c$. We therefore find that the errors on the molecular force are given by:

$$\delta F_v(\theta_*) = \frac{2F_c}{\sqrt{M}} \qquad \delta F_s(\theta_*, M) = \frac{F_c}{\sqrt{M}} \tag{6.53}$$

$$\Rightarrow \quad \delta F_{\mathrm{RMSE},0}(\theta_*, M) = \frac{\sqrt{5}F_c}{\sqrt{M}} \tag{6.54}$$

Realistic Energy Measurement

As discussed in sec. 4.4 a realistic measurement scheme for estimating expectation values involves splitting $\hat{H}$ into a set of easily diagonalisable and hence efficiently measurable observables $\{\hat{A}_\alpha\}$ satisfying $\hat{H} = \sum_\alpha \hat{A}_\alpha$, where we recall that the estimator $\bar{H}$ is constructed from m_α individual measurement results $A_{\alpha(k)}$ of the observable $\hat{A}_\alpha$ as:

$$\bar{H} = \sum_\alpha \bar{A}_\alpha = \sum_\alpha\sum_k^{m_\alpha}\frac{A_{\alpha(k)}}{m_\alpha} \tag{6.55}$$

with a variance given by:

$$\mathrm{Var}[\bar{H}] = \sum_\alpha \mathrm{Var}[\bar{A}_\alpha] = \sum_\alpha \frac{\mathrm{Var}(\hat{A}_\alpha)}{m_\alpha} \tag{6.56}$$

Zero-variance Property Breaking The linearity of quantum mechanical expectation values $\langle\hat{A}\rangle + \langle\hat{B}\rangle = \langle\hat{A}+\hat{B}\rangle$ ensures that this estimator is unbiased with expectation value $\mathbb{E}[\bar{H}] = \langle\hat{H}\rangle$. Even though we do not measure $\hat{H}$ directly, but instead by splitting it into sub-observables, in expectation we are nonetheless measuring the exact energy and the variational principle (equ. (2.25)) applies trivially, so that near ψ_0 we have:

$$\mathbb{E}_\psi[\bar{H}] = E[\psi_0 + \delta\psi] = E[\psi_0] + \mathcal{O}(\delta\psi^2) \tag{6.57}$$

However, unlike the expectation, the quantum mechanical variance is in general not linear $\mathrm{Var}(\hat{A}) + \mathrm{Var}(\hat{B}) \neq \mathrm{Var}(\hat{A}+\hat{B})$. Hence, as a result of the measurements of individual observables being uncorrelated, the estimator variance, equ. (6.56), is not proportional to the quantum mechanical variance of $\hat{H}$, $\mathrm{Var}[\bar{H}] \not\propto \mathrm{Var}[\hat{H}]$ (even if $m_\alpha = m$). Similar to the variance of the molecular force discussed in sec. 2.2, $\mathrm{Var}[\hat{A}_\alpha]$ will in general not obey a variational principle at/near ψ_0. We see that measuring $\hat{H}$ by splitting it into sub-observables breaks the variational principle for $\mathrm{Var}[\bar{H}]$ applied in the previous section (equ. (6.30)), since measurements of the individual observables are uncorrelated. Near ψ_0 the variance will therefore approach a constant, non-vanishing value:

$$\mathrm{Var}_\psi[\bar{H}] = \mathrm{Var}_0[\bar{H}] + \mathcal{O}(\delta\psi) = \mathcal{O}(1) \tag{6.58}$$

Noise Floor As in the previous section, we can use these results to derive an expression for the achievable wavefunction accuracy, $\delta\psi_*(M)$, using the noise floor condition, equ. (6.20), to leading order in $\delta\psi$:

$$\alpha_E \left\|\delta\psi_*\right\|^2 = \frac{\tilde{\sigma}_0(\bar{H})}{\sqrt{M}} \tag{6.59}$$

$$\Rightarrow \quad \left\|\delta\psi_*\right\| = \sqrt{\frac{\tilde{\sigma}_0(\bar{H})}{\alpha_E}} \frac{1}{\sqrt[4]{M}} = \mathcal{O}\left(M^{-1/4}\right) \tag{6.60}$$

where we have defined the single shot standard deviation $\frac{\tilde{\sigma}_0(\bar{H})}{\sqrt{M}}$ to make the M dependence of $\sigma_0(\hat{H})$ explicit. We see that as a result of breaking the variational principle for the variance within this measurement scheme, we expect worse scaling of

the achievable wavefunction accuracy with M compared to the direct measurement of the energy considered in the previous section (equ. (6.33)). Similarly, we find that the total error in the energy scales as:

$$\delta E_v(\delta\psi_*(M)) = \delta E_s(\delta\psi_*(M), M) = \frac{\tilde{\sigma}_0(\bar{H})}{\sqrt{M}} = \mathcal{O}\left(\frac{1}{\sqrt{M}}\right) \tag{6.61}$$

$$\Rightarrow \quad \delta E_{\mathrm{RMSE},0}(\delta\psi_*(M), M) = \sqrt{2}\frac{\tilde{\sigma}_0(\bar{H})}{\sqrt{M}} = \mathcal{O}\left(\frac{1}{\sqrt{M}}\right) \tag{6.62}$$

where again the scaling with respect to M has worsened compared to the direct measurement of the energy (equ. (6.35)). The achievable energy accuracy is now expected to scale as $\delta E \sim \frac{1}{\sqrt{M}}$ in correspondence with scaling expected from considering the statistical error at a fixed point ψ, $\sqrt{\mathrm{Var}[\bar{H}]} = \mathcal{O}\left(\frac{1}{\sqrt{M}}\right)$.

Molecular Force Errors Lastly, we again consider the molecular forces. Since molecular forces are not variational to begin with, a realistic measurement scheme does not affect the scalings:

$$\delta \boldsymbol{F}_v = \boldsymbol{\alpha}_{\boldsymbol{F}} \|\delta\psi\| + \mathcal{O}(\delta\psi^2) \qquad \mathrm{Var}[\bar{F}] = (\sigma_0(\bar{\boldsymbol{F}}))^2 + \mathcal{O}(\delta\psi) \tag{6.63}$$

We therefore find that the error at $\delta\psi_*$ (equ. (6.33)) is given by, to first order in $\delta\psi$:

$$\delta \boldsymbol{F}_v(\delta\psi_*(M)) = \boldsymbol{\alpha}_{\boldsymbol{F}} \sqrt{\frac{\tilde{\sigma}_0(\bar{H})}{\alpha_E}} \frac{1}{\sqrt[4]{M}} = \mathcal{O}\left(M^{-1/4}\right) \tag{6.64}$$

$$\delta \boldsymbol{F}_s(\delta\psi_*(M), M) = \tilde{\sigma}_0(\bar{\boldsymbol{F}}) \frac{1}{\sqrt{M}} = \mathcal{O}\left(M^{-1/2}\right) \tag{6.65}$$

$$\Rightarrow \quad \delta \boldsymbol{F}_{\mathrm{RMSE},0}(\delta\psi_*(M), M) = \sqrt{\boldsymbol{\alpha}_{\boldsymbol{F}}^2 \frac{\tilde{\sigma}_0(\bar{H})}{\alpha_E} \frac{1}{\sqrt{M}} + (\tilde{\sigma}_0(\bar{\boldsymbol{F}}))^2 \frac{1}{M}} \tag{6.66}$$

Unlike for the direct energy measurement scenario, where we found a variational-like scaling, equ. (6.40), we can see that in the current case the variational and statistical errors in the force scale differently with M. Since the variational principle for the variance of the energy is broken, resolving the energy $\delta E_v \sim \delta E_s$ no longer ensures that the molecular force is also always sufficiently resolved $\delta \boldsymbol{F}_v \not\sim \delta \boldsymbol{F}_s$. In the limit of large M, however, we do expect the molecular force to be sufficiently resolved, as $\delta \boldsymbol{F}_s$ decreases faster than $\delta \boldsymbol{F}_v$, and hence the variational-like scaling is restored $\delta E \sim \sqrt{\delta \boldsymbol{F}}$. However, this does not preclude the possibility that in practical calculations we are interested in the region of M where $\delta \boldsymbol{F}_v$ is unresolved, and hence no variational-like scaling applies, $\delta E \sim \delta \boldsymbol{F}$. This behaviour is not guaranteed to

be observed since it is not an asymptotic limit and hence our leading order approximations might break down. Again, considering a different viewpoint, these results indicate that the number of shots M_E required to resolve a systematic energy error $\delta E_v(\delta\psi)$ at $\delta\psi$ is no longer sufficient to always resolve the corresponding systematic molecular force error $\delta\boldsymbol{F}_v(\delta\psi)$. In particular in the limit that we approach ψ_0, M_E shots is expected to be more than sufficient to resolve $\delta\boldsymbol{F}_v(\delta\psi)$ with the minimum number of shots $M_{\boldsymbol{F}}$ required to resolve the molecular force error being considerably lower, $M_{\boldsymbol{F}} \ll M_E$. However, away from ψ_0, resolving $\delta\boldsymbol{F}_v(\delta\psi)$ could potentially require a considerably larger number of shots, $M_{\boldsymbol{F}} \gg M_E$. As in the previous section we note that, while from this viewpoint the applicability of a variational-like scaling might seem to be reversed – far from ψ_0 molecular forces require more shots and hence are more difficult to estimate than energies, while near ψ_0 molecular forces require fewer shots and are hence easier to estimate than energies – such a conclusion neglects to account for the systematic variational errors and their dependence on M through $\delta\psi_*(M)$.

Two Level System As in the previous section we again consider a two level system as an analytic example. To model the realistic measurement scenario we consider the Hamiltonian to be split and measured as two separate observables $\hat{A}_\pm$ given by:

$$\hat{A}_\pm = \frac{\hat{H}}{2} \pm \hat{J} \qquad \text{with} \qquad \hat{J} = J(|g\rangle\langle e| + |e\rangle\langle g|) \tag{6.67}$$

The two observables are no longer diagonal due to the inclusion of off diagonal terms via $\hat{J}$, but clearly satisfy $\hat{A}_+ + \hat{A}_- = \hat{H}$ as required by the measurement scheme. We note that this example might seem contrived, as $\hat{A}_\pm$ are more complex than $\hat{H}$ to begin with, however this is simply a result of specifying observables in the energy eigenbasis. As discussed previously, the linearity of quantum mechanical expectation values ensures that $\langle\hat{A}_+\rangle + \langle\hat{A}_-\rangle = \langle\hat{H}\rangle$ and hence that $\mathrm{E}[\bar{H}] = \langle\hat{H}\rangle$. The variance of each sub-observable is given by:

$$\mathrm{Var}_\psi(\hat{A}_\pm) = \mathrm{Var}_\psi\left(\frac{\hat{H}}{2}\right) + \mathrm{Var}_\psi(\hat{J}) \pm \mathrm{Cov}_\psi(\hat{H}, \hat{J}) \tag{6.68}$$

$$= \left(J\cos(2\theta) \;\pm\; \frac{\Delta}{4}\sin(2\theta)\right)^2 \tag{6.69}$$

and since each observable measurement is uncorrelated we find:

$$\mathrm{Var}_\psi[\bar{H}] \propto \mathrm{Var}_\psi(\hat{A}_+) + \mathrm{Var}_\psi(\hat{A}_-) = \frac{4J^2}{M}\cos^2(2\theta) + \frac{\Delta^2}{4M}\sin^2(2\theta) \tag{6.70}$$

An explicit expansion shows that the variational principle for the variance has been broken:

$$\mathrm{Var}_\psi[\bar{H}] \sim \frac{4J^2}{M} + \frac{\Delta^2}{M}\theta^2 = \mathcal{O}(1) \tag{6.71}$$

with a resulting shot noise floor given by:

$$\theta_* \sim \sqrt{\frac{2J}{\Delta}}\, M^{-1/4}. \tag{6.72}$$

and energy error:

$$\delta E_{\mathrm{RMSE},\,0}(\theta_*, M) = \sqrt{2}\frac{2J}{\sqrt{M}} \tag{6.73}$$

where again we have made identifications $\|\delta\psi\| \to \theta$, $\tilde{\sigma}_0(\bar{H}) \to 2J$ and $\alpha_E \to \Delta$. Similarly, a split measurement scheme is modelled by two operators $\hat{F}^a$ and $\hat{F}^b$ satisfying $F_{00}^{a/b} = F_{11}^{a/b} = F_d^{a/b}$, $F_{01}^{a/b} = F_{10}^{a/b} = F_c^{a/b}$ and $F_{c/d}^a + F_{c/d}^b = F_{c/d}$. The variance of this molecular force estimator is:

$$\mathrm{Var}_\psi[\bar{F}] \propto \mathrm{Var}_\psi(\hat{F}^a) + \mathrm{Var}_\psi(\hat{F}^b) = \left((F_c^a)^2 + (F_c^b)^2\right)\cos^2(\theta) \tag{6.74}$$

$$\sim \left((F_c^a)^2 + (F_c^b)^2\right)(1 - 4\theta^2) = \mathcal{O}(1) \tag{6.75}$$

which results in errors:

$$\delta F_v(\theta_*) = 2F_c\sqrt{\frac{2J}{\Delta}}\, M^{-1/4} \qquad \delta F_s(\theta_*, M) = \sqrt{(F_c^a)^2 + (F_c^b)^2}\, M^{-1/2} \tag{6.76}$$

$$\Rightarrow \quad \delta F_{\mathrm{RMSE},0}(\theta_*, M) = \sqrt{\frac{8JF_c^2}{\Delta}\, M^{-1/2} + \left((F_c^a)^2 + (F_c^b)^2\right)\, M^{-1}} \tag{6.77}$$

6.2.2 Circuit Noise

The variational principle as discussed in sec. 2.2 only applies to wavefunction errors within the manifold of pure states. Circuit noise, however, does not merely move us to a different pure state within the same variational manifold; instead, it implements a noisy quantum channel and therefore prepares, in general, a mixed state ρ even though the underlying ansatz state is pure. Given the (pure) ground state $|\psi_0\rangle$ with corresponding density matrix $\rho_0 = |\psi\rangle\langle\psi|$ the variation induced by circuit noise $\rho = \rho_0 + \delta\rho$ constitutes a variation outside the space in which the standard stationary principle is formulated and applies.

The inapplicability of the variational principle to this induced variation can be illustrated by considering a mixed state formed from a convex combination of two pure states. Let $|\psi_0\rangle$ and $|\psi_e\rangle$ be the ground state and some excited state of $\hat{H}$ with

energies E_0 and E_e and associated density matrices $\rho_0 = |\psi_0\rangle\langle\psi_0|$ and $\rho_e = |\psi_e\rangle\langle\psi_e|$ respectively. For $\lambda \in [0, 1]$ we consider a "perturbation" of the exact ground state given by:

$$\rho(\lambda) = \lambda\rho_0 + (1 - \lambda)\rho_e \tag{6.78}$$

The energy of this is linear in λ:

$$E[\rho(\lambda)] = \mathrm{Tr}\left(\rho(\lambda)\hat{H}\right) = \lambda\,\mathrm{Tr}(\rho_0\hat{H}) + (1 - \lambda)\,\mathrm{Tr}(\rho_e\hat{H}) = \lambda E_0 + (1 - \lambda)E_e \tag{6.79}$$

which can clearly be seen to not be stationary with respect to the strength of the "perturbation", λ:

$$\left.\frac{dE}{d\lambda}\right|_{\lambda=0} = E_0 - E_e \neq 0 \qquad (\text{unless } E_0 = E_e) \tag{6.80}$$

Most importantly, this shows that for a general variation of the density matrix there is no variational advantage gained from considering the energy, $E[\rho(\lambda)]$, near the exact ground state ρ_0 so that we expect:

$$\delta E = \mathrm{Tr}[\delta\rho\,\hat{H}] = \mathcal{O}(\|\delta\rho\|) \qquad \text{and} \qquad \delta\boldsymbol{F} = \mathrm{Tr}[\delta\rho\,\hat{\boldsymbol{F}}] = \mathcal{O}(\|\delta\rho\|) \tag{6.81}$$

This is the behaviour expected for incoherent circuit noise: realistic stochastic noise channels, such as depolarising or amplitude damping noise, can be viewed as inducing errors by perturbing the ideal pure state in the general space of density matrices, $\rho = \rho_0 + \delta\rho$, for which no variational advantage in the energy exists. In particular, we expect the degradation of energy and molecular force accuracies to follow a non-variational scaling of $\delta E \sim \delta F$, an unexpected behaviour from the viewpoint of standard electronic structure theory in which the variational principle applies.

Finally, it is important to also note the existence of coherent errors such as systematic over- or under-rotations of gates. Coherent errors correspond to the systematic application of an incorrect unitary within the pure-state manifold and are as a result likely to behave differently compared to general circuit noise considered above. However, a VQE optimisation is in principle able to absorb any coherent error by adjusting the variational parameters, provided the coherent offsets are stable on the timescale of the optimisation and the ansatz circuit is expressive enough. Furthermore, we note that in the present work, our device-calibrated noise model focuses only on incoherent/stochastic contributions (see sec. 6.1.2), and does not explicitly include coherent errors.

6.3 Numerical Results: Shot Noise

As previously discussed, in this section we consider the effect of the stochastic noise resulting from an expectation value estimation using a finite number of shots, M. Similarly to sec. 5.4, we are interested in determining the value of M required to reach chemical accuracy for both energies and molecular forces, or equivalently the limitations imposed by a finite M. Again, in analogy to standard convergence studies in electronic structure theory, we assess the convergence of both quantities towards the exact diagonalisation (ED) results for $\hat{H}$, however now as M is increased and hence the stochastic noise is reduced, towards the noiseless limit, $M \to \infty$. Similarly to sec. 5.4, we quantify the accuracy of our calculations by the energy error, ΔE, equ. (5.38), and molecular force error, ΔF, equ. (5.39), with respect to ED results, with our accuracy targets for chemical relevance given by ΔE_{chem} and ΔF_{chem}. We are interested in determining not only the level of M necessary to reach chemical accuracy for both energies and molecular forces, but also the extent to which the variational principle (or potential absence thereof) affects these results given the theoretical discussion presented in sec. 6.2.1.

As an additional point of reference, we also compare the achieved energy accuracies to the shot-noise floor set by the variance of the energy estimator at the exact ground state (under the same measurement protocol), by evaluating equ. (6.2) using the ED ground state. As discussed in sec. 6.2.1 the realistic measurement protocol breaks the zero-variance property of the ED ground state, and hence provides a fundamental lower bound on the attainable stochastic error at finite M.

6.3.1 Simulation Methodology

The shot-noise calculations presented in this section were implemented using the same QISKIT package versions and hardware as the ideal simulations described in sec. 5.4.1.

As discussed in sec. 6.1.1, we have opted not to simulate the finite-shot measurement process by sampling individual bit strings according to the measurement probabilities. Instead, it is approximated by sampling estimators from distributions determined by their ideal, noise-free expectation values and variances. The required exact expectation values and variances were evaluated using the QISKIT Aer estimator, `qiskit_aer.primitives.Estimator`, with `shots=None` and `approximation=True`. In the absence of a circuit-noise model, it uses a state vector simulator to evaluate exact expectation values without introducing shot or circuit noise.

In contrast to the noiseless calculations of sec. 5.4, reverse-mode state vector

Hyper-parameter	Candidate values
a	$\{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$
c	$\{10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$
A	$\{10, 10^{2}, 10^{3}\}$

Table 6.1: Discrete hyper-parameter grid used for SPSA tuning

gradients were not used for the current L-BFGS calculations. As a simulator-specific differentiation technique, reverse-mode differentiation does not account for shot noise and cannot be adapted straightforwardly to include shot noise in a manner consistent with a hardware-compatible PS gradient protocol. For L-BFGS, gradients were therefore evaluated directly using the PS rules described in sec. 5.2.4. Shot noise in the PS gradients was included using the same approach as for the energy expectation values.

For SPSA, the calculations were parallelised over independent VQE instances, corresponding to different random initialisations and different shot-noise levels. For L-BFGS, the calculations were again parallelised over independent VQE instances. Additional parallelisation over the many independent PS evaluations required for each gradient evaluation was found to be beneficial and applied by configuring the estimator using the settings `max_parallel_threads=4`, `max_parallel_shots=1`, and `max_parallel_experiments=4`.

The use of the more efficient and optimised QISKIT Aer `Estimator`, rather than the `Estimator` provided in QISKIT and used in sec. 5.4, is particularly important here since the number of circuit evaluations required – for SPSA on account of it slow convergence and for L-BFGS on account of the direct use of PS gradients – is substantially larger than for the calculations presented in sec. 5. As in sec. 5.4.1, we have again opted for an exact state vector simulator over alternative simulation methods.

6.3.2 SPSA

In our calculations we found that the accuracy achieved by the SPSA optimiser depends strongly on the choice of its hyper-parameters (HPs), the triple (a, c, A) (see sec. 4.5.3). In particular, the default SPSA implementation provided in QISKIT was unable to reach anywhere close to our chemical-accuracy targets, instead effectively sampling the parameter landscape at random. We therefore performed an extensive HP optimisation. We employed a grid search over the discrete HP sets presented in tab. 6.1. For each HP triple (a, c, A) we performed $N_{\text{seed}} = 10$ independent

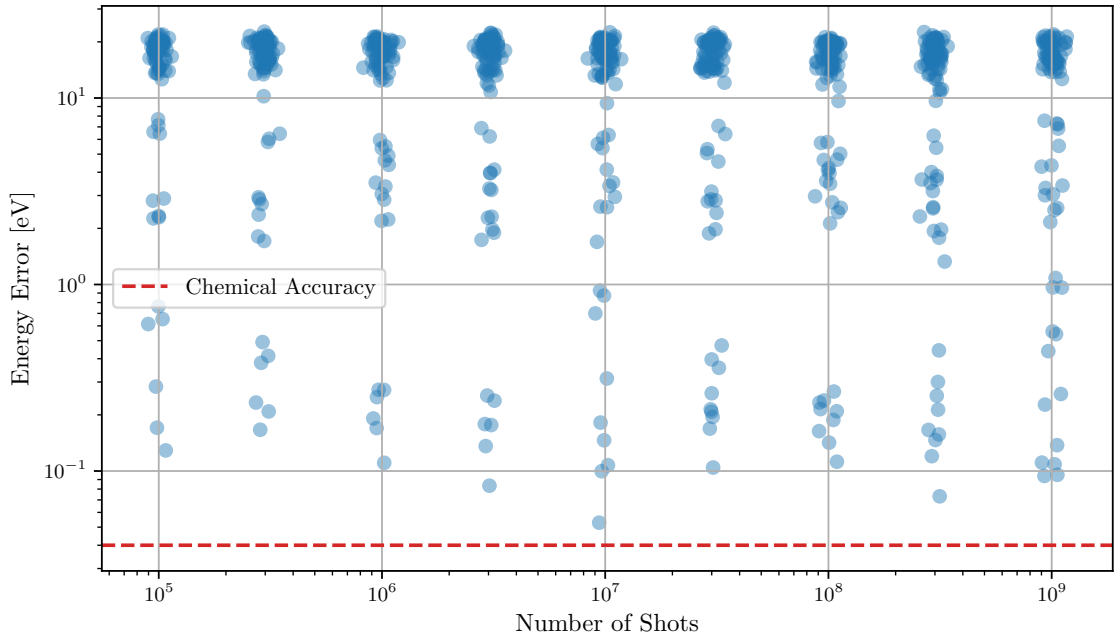


Figure 6.1: SPSA hyper-parameter grid search outcomes versus number of shots, M, showing the energy error, ΔE, for the result of each individual VQE calculation. Markers are plotted with partial transparency and a small horizontal jitter for clarity; darker regions therefore indicate higher point density due to overlap.

VQE optimisations with random initial parameters, using an SPSA iteration budget of $N_{\text{iter}} = 2000$. The HPs were optimised independently for each shot count M considered. As described, the total HP optimisation required an effective serial CPU time of $t_{\text{CPU}} \approx 4000\,\text{h}$.

The results of this grid search are depicted in fig. 6.1. As can be seen, the majority of HPs performed very poorly, effectively sampling the objective function at random. Only a very small set of HPs was able to converge towards chemical accuracy. Furthermore, even the best performing HPs were not able to reach chemical accuracy directly due to the limited iteration count. We therefore selected the optimal HP set as the one minimising the mean energy error ΔE averaged over the N_{seed} runs. The results of the HP optimisation are shown in tab. 6.2. Overall we found little variance in the optimal HPs as M is varied, indicating that further computational cost savings in the HP optimisation are likely possible in realistic calculations.

Using the SPSA optimisation algorithm with these optimised HPs, we performed $N_{\text{seed}} = 10$ individual VQE runs at each M. In the absence of an established termination condition for SPSA we fixed each calculation to an iteration budget of $N_{\text{iter}} \approx 40\,000$, set by a classical computational limit of $t_{\text{wall}} = 8\,\text{h}$ per individual

M	a	c	A
1×10^5	1×10^{-3}	1×10^{-2}	10
3×10^5	1×10^{-3}	1×10^{-2}	100
1×10^6	1×10^{-3}	1×10^{-2}	100
3×10^6	1×10^{-3}	1×10^{-3}	100
1×10^7	1×10^{-3}	1×10^{-3}	10
3×10^7	1×10^{-3}	1×10^{-2}	10
1×10^8	1×10^{-3}	1×10^{-2}	100
3×10^8	1×10^{-3}	1×10^{-3}	100
1×10^9	1×10^{-3}	1×10^{-2}	100

Table 6.2: Optimal SPSA hyper-parameters, (a, c, A), for each number of shots, M, based on a grid search and selected by minimising the mean energy error, ΔE, of the results of repeated VQE runs.

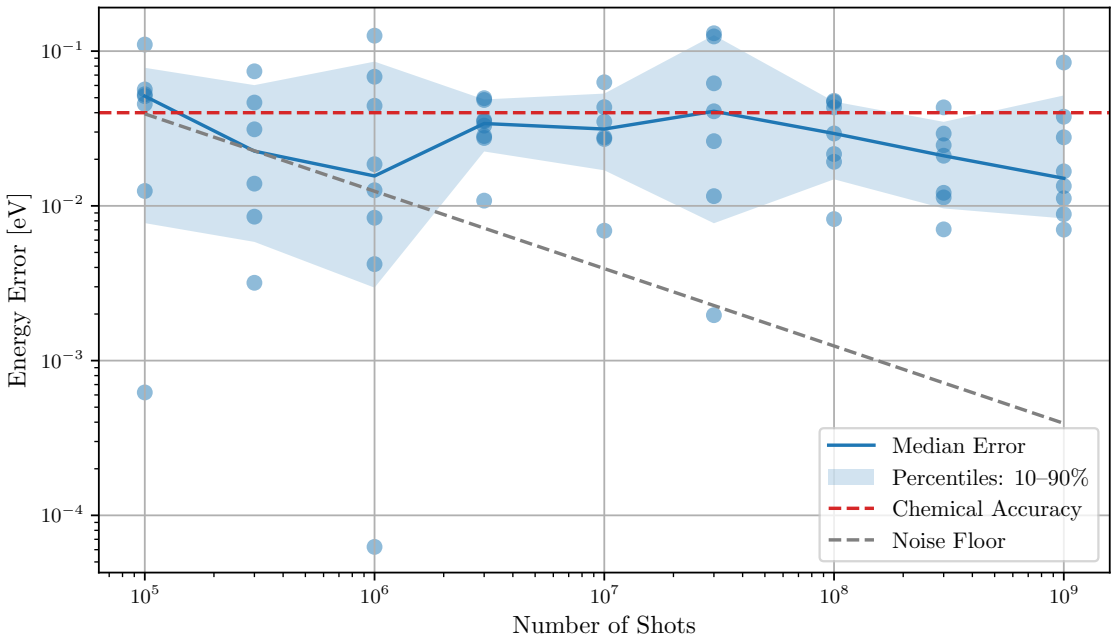


Figure 6.2: Energy error, ΔE, for the result of VQE calculations using SPSA with optimised HPs as a function of the number of shots, M. Each marker corresponds to the result of an individual VQE calculation.

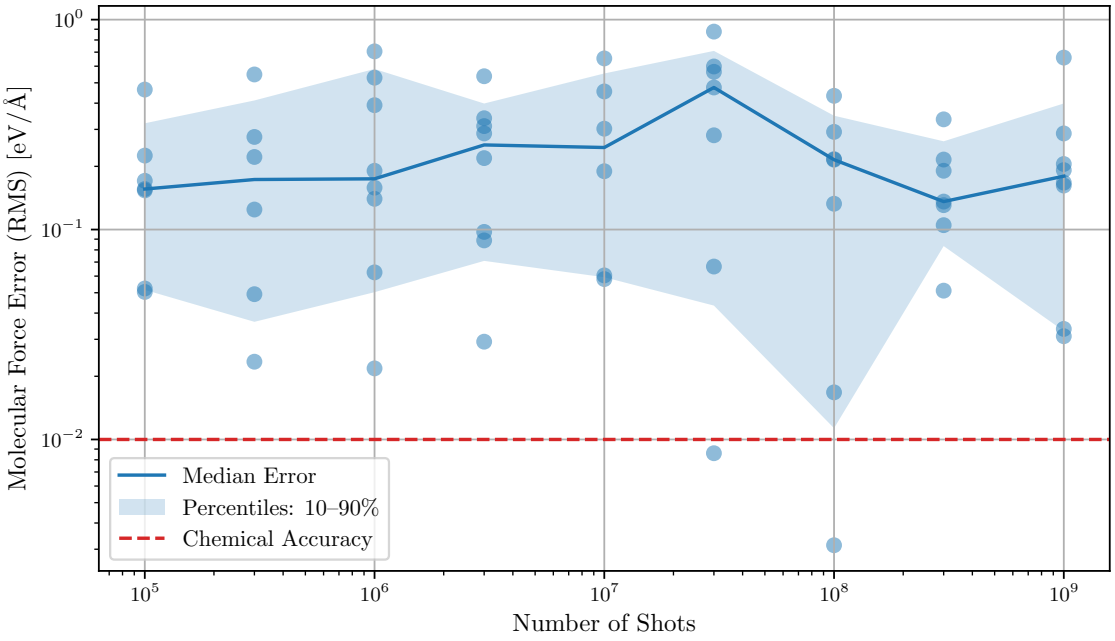


Figure 6.3: Molecular force error, ΔF, for the result of the same VQE calculations as in fig. 6.2, shown as a function of the number of shots, M. Each marker corresponds to the result of an individual VQE calculation.

VQE calculation, resulting in a total CPU time of $t_{\text{CPU}} \approx 700\,\text{h}$. We note that as discussed in sec. 6.1.1, this sets a constant N_{iter} for all M since the classical computational cost is independent of M for our simulation method.

We first consider the energy accuracy. The results of fig. 6.2 show that even with the extensive HP optimisation performed, SPSA is only just able to somewhat inconsistently reach $\Delta E < \Delta E_{\text{chem}}$ across repeated calculations, for any shot count as low as $M = 1 \times 10^5$, the lowest M considered. Even as M increases the accuracy of our SPSA calculations does not markedly improve. For the same set of VQE calculations fig. 6.3 shows the molecular force accuracy. We see that for molecular forces SPSA is completely unable to reach the chemical accuracy target required for our model, i.e. $\Delta F > \Delta F_{\text{chem}}$ for all M.

6.3.3 L-BFGS

For the L-BFGS algorithm we similarly performed $N_{\text{seed}} = 10$ VQE calculations at each M considered. Unlike SPSA, L-BFGS does not depend on any HP and does not require an HP optimisation to be performed. Each individual VQE calculation required $t_{\text{wall}} \approx 8\,\text{h}$. In total the calculations required a CPU time of $t_{\text{CPU}} \approx 2000\,\text{h}$. Fig. 6.4 shows ΔE as M is varied. As can be seen L-BFGS is able to reach $\Delta E <$

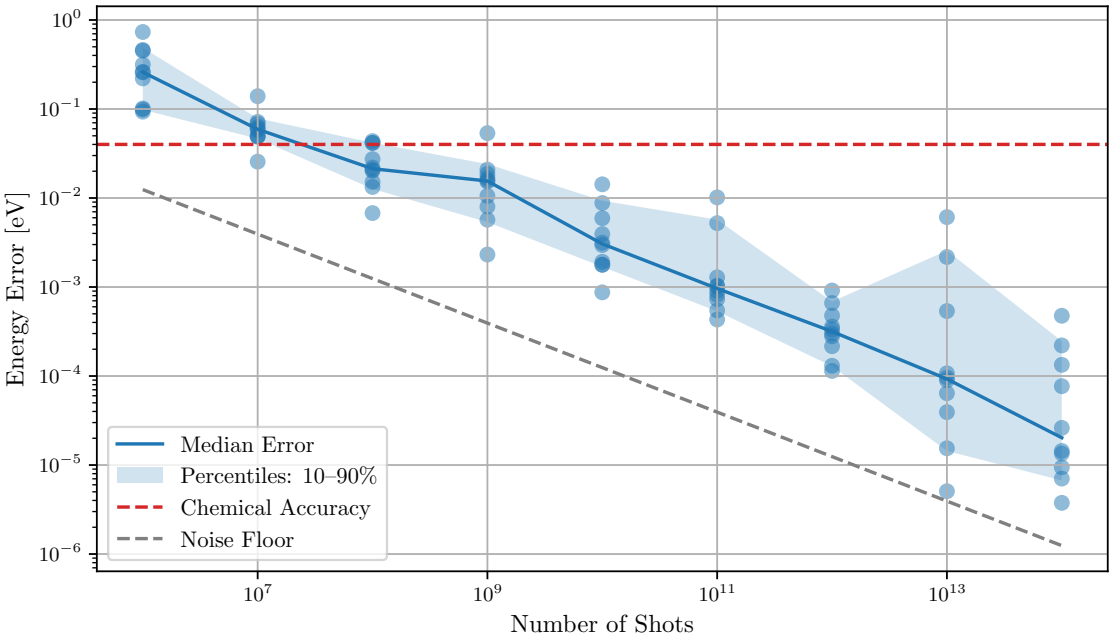


Figure 6.4: Energy error, ΔE, for VQE calculations using L-BFGS as a function of the number of shots, M. Each marker corresponds to the result of an individual VQE calculation.

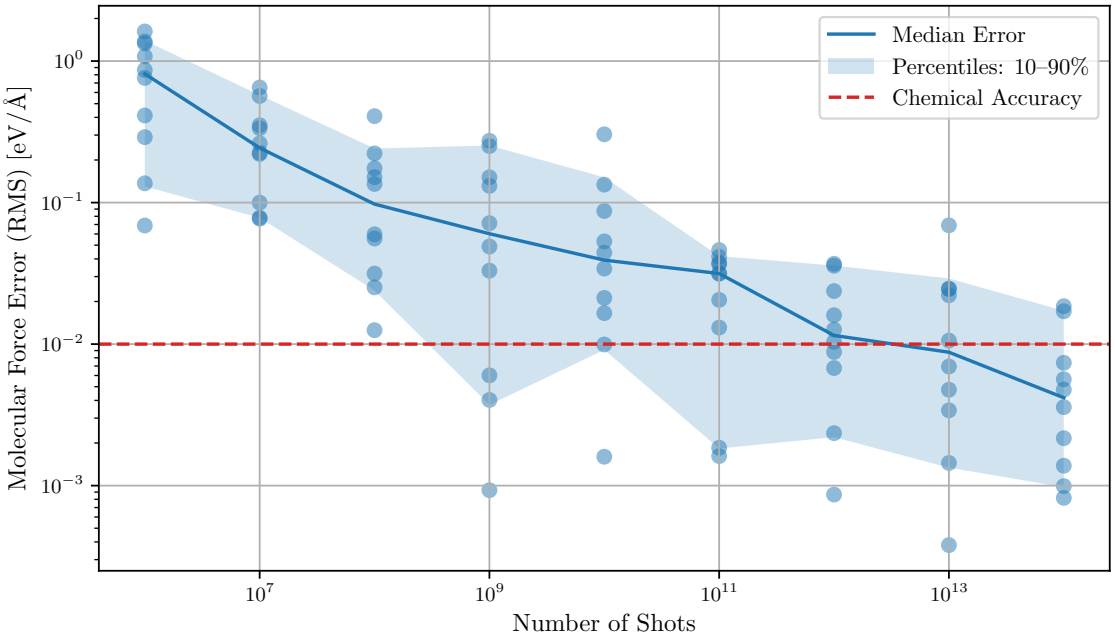
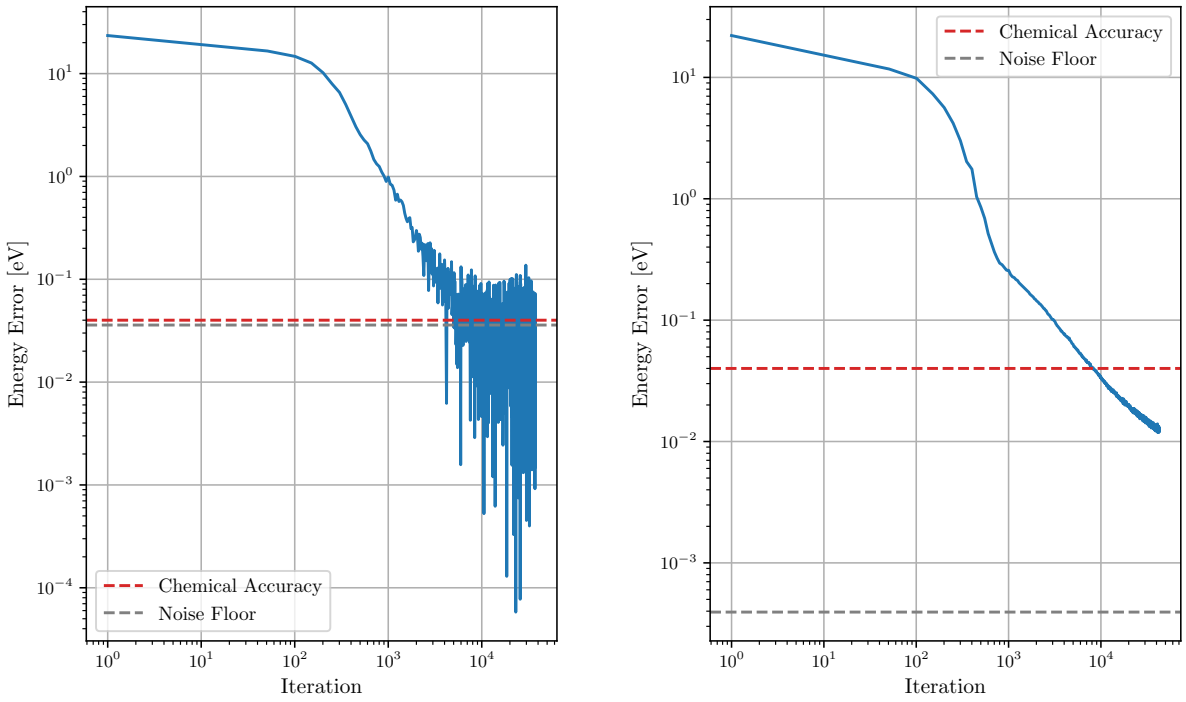


Figure 6.5: Molecular force error, ΔF, for the result of the same VQE calculations as in fig. 6.4, shown as a function of the number of shots, M. Each marker corresponds to the result of an individual VQE calculation.



(a) $M = 1 \times 10^5$ **(b)** $M = 1 \times 10^9$

Figure 6.6: Representative SPSA convergence traces of energy error, ΔE, for a selected VQE calculation at two shot budgets: (a) $M = 1 \times 10^5$ and (b) $M = 1 \times 10^9$. We note the use of log–log scales.

ΔE_{chem} consistently starting at $M = 1 \times 10^8$. Fig. 6.4 shows ΔF as M is varied. Reaching $\Delta F < \Delta F_{\text{chem}}$ requires a considerably larger M. Only starting from $M = 1 \times 10^{14}$ do the VQE calculations reach chemical accuracy in the molecular forces.

6.3.4 Discussion

SPSA

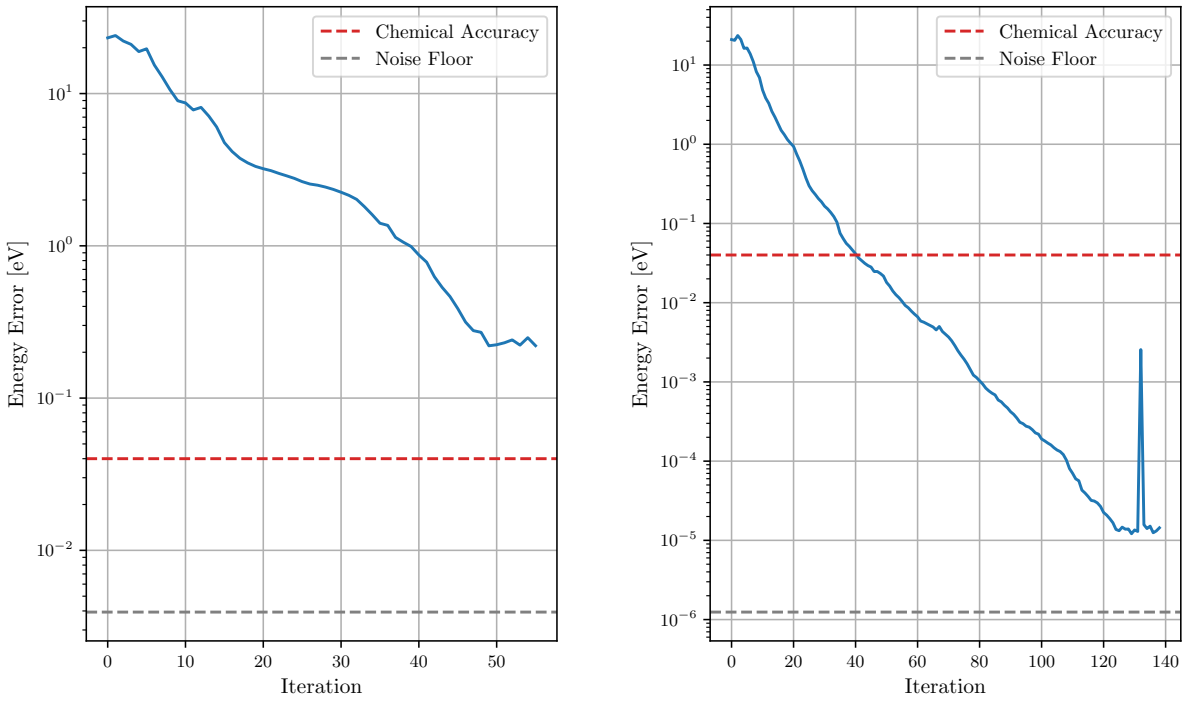
The low accuracy of the SPSA calculations for molecular forces can be explained by the slow convergence of the SPSA algorithm. Two representative SPSA convergence traces of ΔE for $M = 1 \times 10^5$ and 1×10^9 are shown in fig. 6.6; further convergence traces for other M are provided in appendix B. As can be seen from these figures, for $M = 1 \times 10^5$ SPSA is able to converge to the shot-noise floor, which for this M coincides with chemical accuracy. However, for $M = 1 \times 10^9$ the slow, sub-linear convergence, $\Delta E \sim \mathcal{O}(\frac{1}{k})$ (note the log–log scale in the convergence plots), prevents SPSA from consistently reaching chemical accuracy within the considered iteration budget. In practice, this means that even when the shot-noise floor is lowered by increasing M, SPSA fails to exploit the improved estimator precision within the

available N_{iter}, and the final error is instead set by the residual optimisation error. As discussed in sec. 2.2, this is particularly problematic for molecular forces, since the variational principle implies that a substantially higher wave-function accuracy is required. Consequently, while SPSA is occasionally able to reach chemical accuracy for energies (fig. 6.2), it does so inconsistently and remains unsuitable for the high-accuracy regime we ultimately care about – especially for molecular forces, where $\Delta F > \Delta F_{\text{chem}}$ for all M (fig. 6.3). While one might expect that, with a sufficiently large iteration budget, SPSA could eventually reach ΔF_{chem} in the high-M limit, the associated cost in N_{iter} would be prohibitive given the $\Delta E \sim \mathcal{O}(\frac{1}{k})$ convergence rate, making SPSA impractical for high-accuracy, realistic electronic-structure calculations.

Moreover, the results highlight a second drawback of SPSA for realistic electronic-structure applications: its performance is highly dependent on the choice of HPs. Even after an expensive grid-based HP optimisation for which only a small subset of (a, c, A) values yield meaningful convergence (fig. 6.1), the resulting best-case performance remains limited. While there is certainly scope for improvement – for example via more sophisticated HP selection strategies (multi-stage / coarse-to-fine tuning), shot-frugal schedules that allocate fewer measurements early in the optimisation, or adaptive schemes that adjust (a, c, A) online – these remedies further increase algorithmic and implementation complexity as well as problem dependence. It is likely that different problem instances can require widely varying sets of HPs and algorithm implementations, based not only on the molecular systems, geometries and ansatz design, but also on the target accuracies. This makes SPSA highly unsuitable for realistic electronic-structure calculations, where robustness and predictability across a wide range of problem instances is essential.

L-BFGS

In contrast to SPSA, L-BFGS exhibits fast, reliable convergence once M is large enough that stochastic fluctuations do not completely obscure descent information. Again, two representative convergence traces for L-BFGS at $M = 1 \times 10^6$ and 1×10^{14} shots are shown in fig. 6.7, with further convergence traces in appendix B. Unlike SPSA, L-BFGS can be seen to, even in the noisy setting, converge super-linearly. The fast and reliable convergence allows L-BFGS to easily exploit the increase in precision attained as the shot noise floor is lowered with increasing M for a more accurate calculation. The fast convergence makes it highly suited to the high-accuracy calculations we are targeting, where – in contrast to SPSA – we find that it is able to consistently achieve both $\Delta E > \Delta E_{\text{chem}}$ and $\Delta F > \Delta F_{\text{chem}}$ for



(a) $M = 1 \times 10^6$ **(b)** $M = 1 \times 10^{14}$

Figure 6.7: Representative L-BFGS convergence traces of energy error, ΔE, for a selected VQE calculation at two shot budgets: (a) $M = 1 \times 10^6$ and (b) $M = 1 \times 10^{14}$. We note the use of log–log scales.

large enough M. Furthermore, a performant L-BFGS optimisation does not require an expensive HP optimisation and, as a quasi-Newton algorithm, is expected to be largely problem-agnostic. This is in stark contrast to SPSA, and highly beneficial from the viewpoint of realistic electronic structure calculations.

Total Cost

The total cost of the VQE calculation is set not only by M, but also the number of iterations and cost per iteration step of the optimisation method. While in a realistic setting each L-BFGS iteration is more expensive due to the need for evaluating parameter-shift gradients (sec. 4.5.1), this cost must be weighed against the above discussed markedly reduced number of iterations required to reach a target accuracy as a result of the super-linear convergence. The total cost can be calculated as $N_{\text{eval}} = M \times N_{\text{iter}} \times C_{\text{step}}$ where N_{iter} is the number of iterations and C_{step} is the cost per optimisation step, i.e. the number of function evaluations required. For our ansatz circuit a PS gradient requires 2×1200 evaluations of the objective function, $E(\boldsymbol{\theta})$, while SPSA requires only 2 evaluations. For our L-BFGS and SPSA implementations we therefore find $C_{\text{step}} = 2400$ and $C_{\text{step}} = 2$. As can be seen from fig. 6.8, once adjusted for the relative cost, the advantage of L-BFGS over SPSA remains

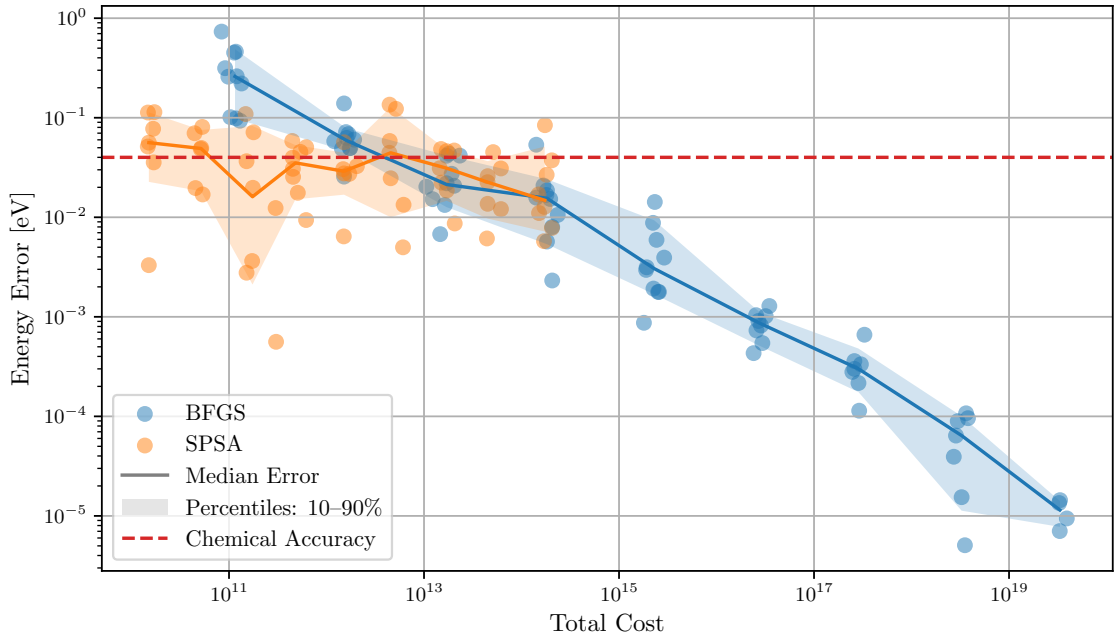


Figure 6.8: Energy error, ΔE, for the result of VQE calculations using SPSA and L-BFGS as a function of total cost, N_{eval}. Horizontal jitter results from a natural variation of N_{iter} across VQE calculations

clear. We find that for our calculations starting at $M = 1 \times 10^7$ – corresponding to a total cost of $M \approx 1 \times 10^{12}$ for both calculations – L-BFGS outperforms SPSA in terms of the achieved accuracy at a given cost as a result of its greatly increased rate of convergence.

As discussed, we are able to consistently reach chemical accuracy in energies and molecular forces, $\Delta E < \Delta E_{\text{chem}}$ and $\Delta E < \Delta E_{\text{chem}}$, in the L-BFGS calculations for $M_E > 1 \times 10^8$ and $M_F > 1 \times 10^{14}$. Based on the data presented in fig. 6.8 and fig. 6.9 we find that this translates into a total cost of approximately $N_{\text{eval},E} \approx 1 \times 10^{13}$ and $N_{\text{eval},F} \approx 1 \times 10^{20}$.

Ultimately the relevant cost is, however, the total QPU runtime required to execute a chemically accurate calculation on a real device. The ansatz circuit considered for our calculation has a depth of approximately $D \sim 1000$ assuming all-to-all connectivity and a standard basis gate set. Using a representative gate-layer time of $\tau_{\text{layer}} \sim 100\,\text{ns}$ per unit of circuit depth gives a physical circuit duration of:

$$t_{\text{circ}} \approx D \times \tau_{\text{layer}} \sim 1000 \times 100\,\text{ns} \approx 1 \times 10^{-4}\,\text{s} \tag{6.82}$$

Performing a calculation requiring N_{eval} circuit executions will result in a QPU wall-

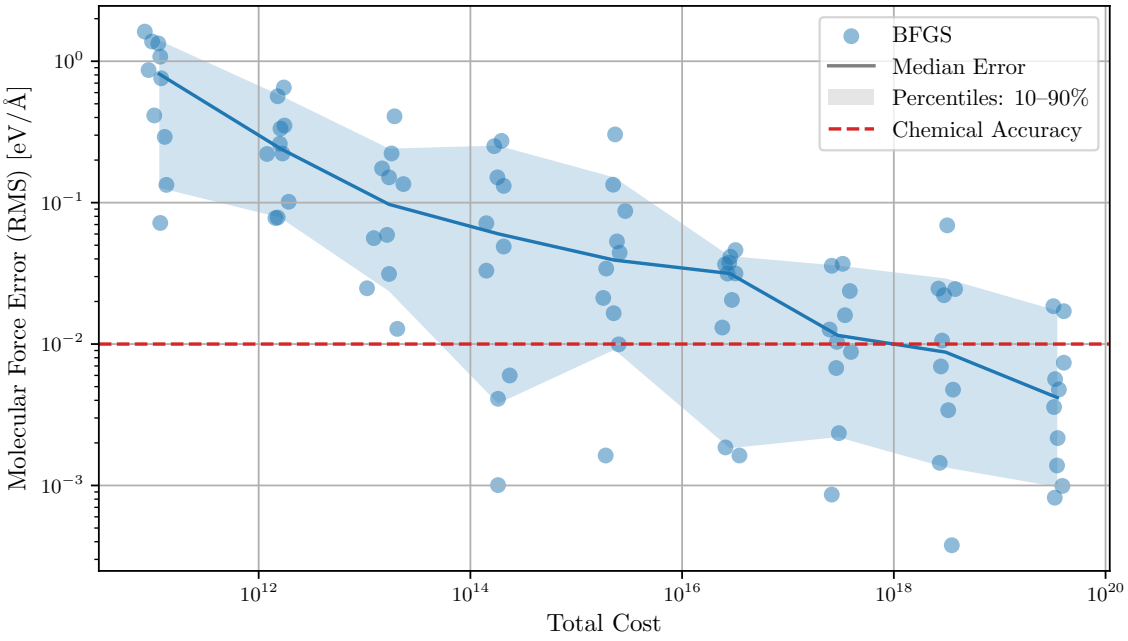


Figure 6.9: Molecular force error, ΔF, for the result of VQE calculations using L-BFGS as a function of total cost, N_{eval}. Horizontal jitter results from a natural variation of N_{iter} across VQE calculations

time of:

$$t_{\text{wall}} \approx N_{\text{eval}}\, t_{\text{circ}} \sim N_{\text{eval}} \times 1 \times 10^{-4}\,\text{s}. \tag{6.83}$$

Calculating chemically accurate energies is estimated to require a circuit execution time of:

$$t_{\text{wall},E} \sim 1 \times 10^{13} \times 1 \times 10^{-4}\,\text{s} = 1 \times 10^{9}\,\text{s} \approx 3.2 \times 10^{1}\,\text{yr}, \tag{6.84}$$

while chemically accurate molecular forces demand a wall-time:

$$t_{\text{wall},F} \gtrsim 1 \times 10^{20} \times 1 \times 10^{-4}\,\text{s} = 1 \times 10^{16}\,\text{s} \approx 3.2 \times 10^{8}\,\text{yr}. \tag{6.85}$$

We note that these estimates are deliberately optimistic. Limited connectivities in realistic devices and varying execution speeds for the different gates will further increase execution times. We have also ignored further overheads introduced by, e.g., state-preparation and measurement overheads or queueing and allocation constraints. Given these estimates, the prediction of chemically accurate energies on realistic quantum hardware can be considered just about feasible, especially since improvements in algorithmic design as well as quantum hardware will likely be able to reduce the runtime further. However, calculations targeting chemically accurate molecular forces turn out to be completely unfeasible; even under our optimistic

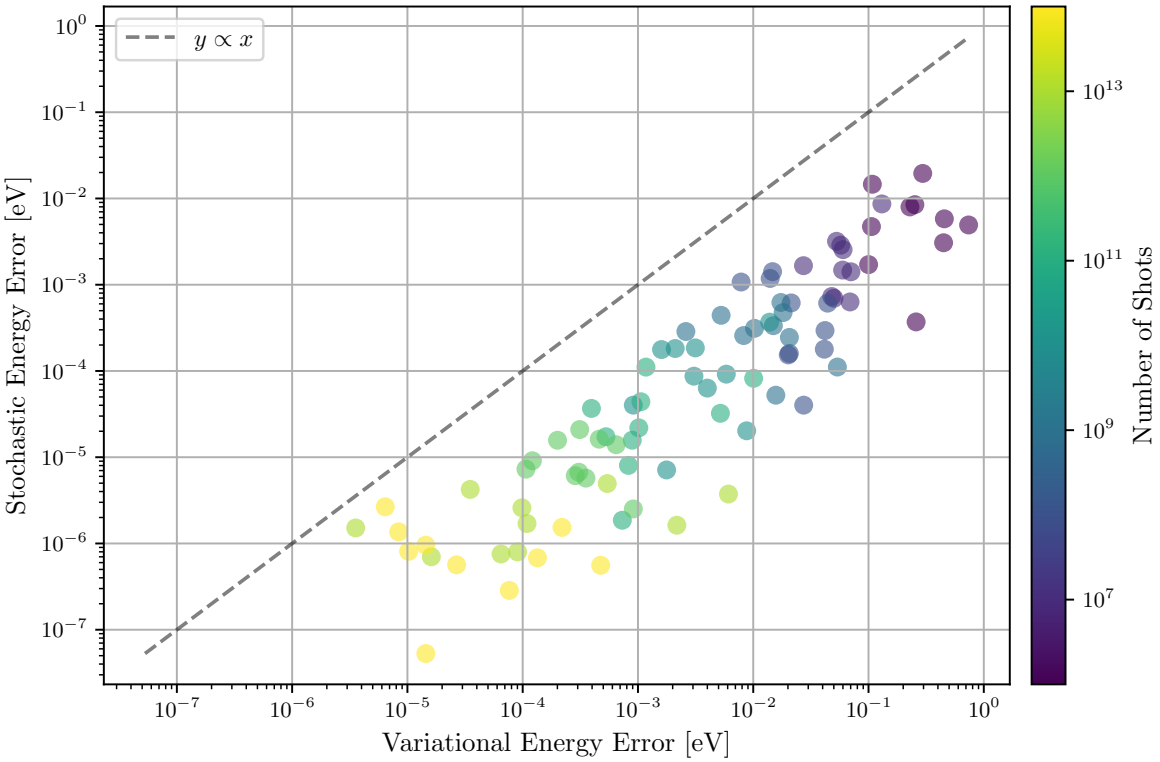


Figure 6.10: Stochastic energy error, δE_s, versus variational energy error, δE_v, for the results of VQE calculations using L-BFGS for all number of shots, M.

assumptions the required wall-time turns out to be astronomical.

Variational Principle

These results are, from an electronic-structure viewpoint, strikingly poor in terms of the sampling effort required to reach chemically meaningful accuracies. For energies in particular we require $M_E \gtrsim 1 \times 10^8$ (or equivalently $N_{\text{eval},E} \approx 1 \times 10^{13}$ for $\Delta E < \Delta E_{\text{chem}}$). The situation for molecular forces is even worse, requiring $M_F \gtrsim 1 \times 10^{14}$ (or equivalently $N_{\text{eval},F} \approx 1 \times 10^{20}$). This sampling cost is orders of magnitude beyond what one might expect by analogy with quantum Monte Carlo (QMC) simulations, where chemically useful energies – and frequently also molecular forces – can be and have been obtained with substantially fewer stochastic samples [184, 185]. The accuracy of QMC calculations is well known to benefit from the zero-variance property for the energy, equ. (2.30). However, as discussed in sec. 6.2.1, in realistic applications of VQE this property is broken by the requirement for a hardware compatible measurement protocol, and we analyse our results using the framework and arguments presented therein.

As can be seen in the results of fig. 6.4, for the L-BFGS calculations the noise floor limits the achievable accuracy to the stochastic noise at a given M, $\Delta E \sim \mathcal{O}(\frac{1}{\sqrt{M}})$,

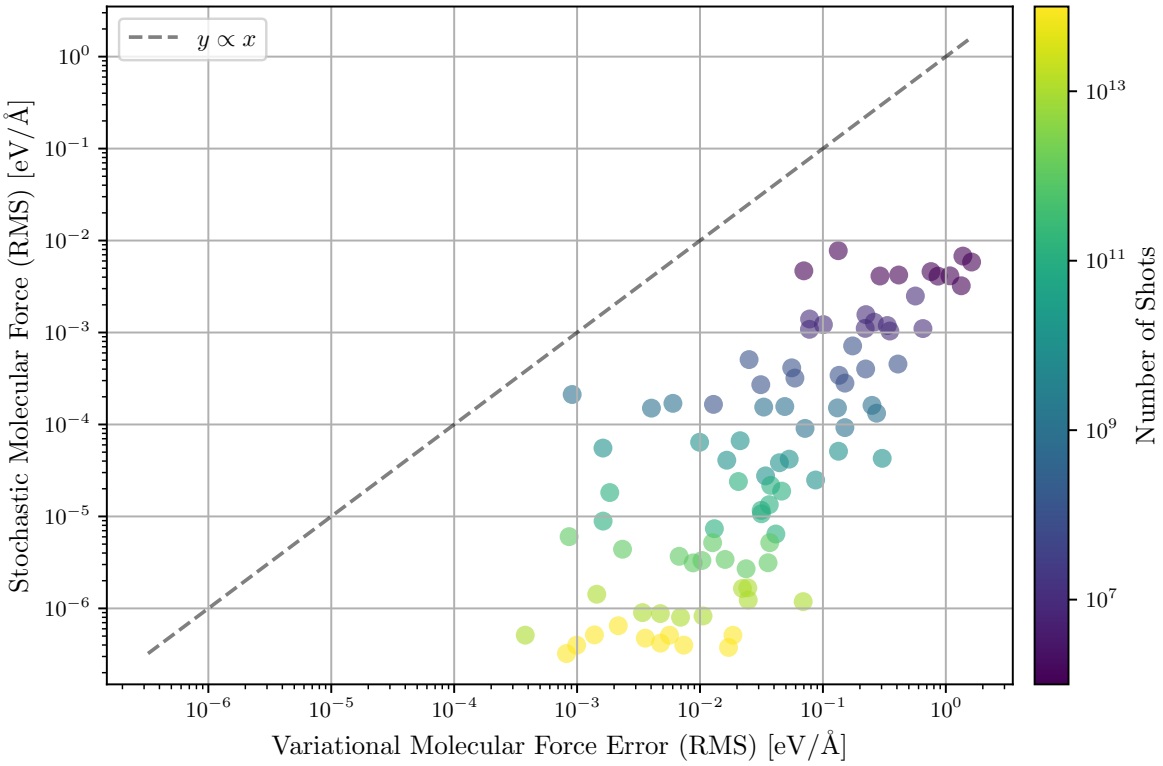


Figure 6.11: Stochastic molecular force error, δF_s, versus variational molecular force error, δF_v, for the results of VQE calculations using L-BFGS for all number of shots, M.

as argued in sec. 6.2.1 and equ. (6.61). However, an important part of the analysis, in particular with respect to the molecular forces, was based on the noise floor, equ. (6.20), limiting the calculations not only through the stochastic noise of each function evaluation, $\delta E_s(M)$, but also through the variational error that is achievable by the optimisation algorithm, $\delta E_v(\delta\psi_*(M))$. Fig. 6.10 validates this argument, showing that for the L-BFGS calculations both errors follow a linear relationship as per the noise-floor definition, equ. (6.20), and hence that they both display the scaling of $\delta E_{v/s} \sim \mathcal{O}(\frac{1}{\sqrt{M}})$, predicted by equ. (6.61) for the realistic measurement scenario. The systematic and variational energy errors are calculated based on an *ideal* noiseless energy evaluation at the optimal VQE parameters of each calculation, $E(\boldsymbol{\theta}_*)$, as $\Delta E_s = E_{\text{VQE}} - E(\boldsymbol{\theta}_*)$ and $\Delta E_v = E(\boldsymbol{\theta}_*) - E_0$, where we note that E_{VQE} is the actual result of the VQE calculation, i.e., evaluated with stochastic noise present at a finite M.

As can be seen in the results of fig. 6.5, the accuracy achievable for molecular forces at a given M instead follows an approximate relationship $\Delta F \sim \mathcal{O}(\frac{1}{\sqrt[4]{M}})$ as predicted by the results of 6.2.1 and equ. (6.64). We note that this is not a trivial result. A standard estimate based on the statistics of a point estimate $\bar{F}$ might conclude that $\Delta F \sim \mathcal{O}(\frac{1}{\sqrt{M}})$. However, as argued in equ. (6.64), the actual

dependence of ΔF is, at large M, not limited by the stochastic error on the estimator $\bar{F}$, $\delta \boldsymbol{F}_s(M)$, but rather by the accuracy with which the wave-function is resolved, and the corresponding variational error $\delta \boldsymbol{F}_v(\delta\psi_*(M))$. As such, the standard variational principle of sec. 2.2 applies, and we find that molecular forces are considerably more difficult to converge than energies. This is further exemplified by fig. 6.11, where it is clearly seen that the variational error in the molecular forces, δF_v, is the dominant contribution to ΔF, compared to δF_s.

6.4 Numerical Results: Circuit Noise

In this section, we consider the effect of the circuit noise on the accuracy of our VQE calculations. Similarly to sec. 5.4 and sec. 6.3, we are interested in determining the relative improvement in gate error probabilities, f, required to reach chemical accuracy for both energies and molecular forces, within our error model discussed in sec. 6.1.2. Again, we quantify accuracy relative to ED using the energy error ΔE, equ. (5.38), and molecular force error ΔF, equ. (5.39), with chemical-accuracy targets ΔE_{chem} and ΔF_{chem}, analysing the results both from a quantitative viewpoint as well as qualitatively based on the variational principle (or potential absence thereof) discussed in sec. 6.2.1.

6.4.1 Simulation Methodology

The circuit-noise calculations presented in this section were implemented using the same QISKIT package versions as the ideal and shot-noise simulations described in sec. 5.4.1 and sec. 6.3.1, together with `qiskit-ibm-runtime~=0.20.0`. Unlike the previous calculations, these simulations were performed using the A100 GPUs available on the Imperial HPC cluster.

As discussed in sec. 6.1.2, we chose the IBM Sherbrooke QPU as representative machine for modelling the effects of circuit noise. The ansatz circuits were transpiled using QISKIT to the native basis gates and coupling constraints of the fake IBM Sherbrooke backend, `FakeSherbrooke`. The noise model was constructed from the same `FakeSherbrooke` backend, restricted to the one- and two-qubit gate-noise contributions described in sec. 6.1.2, and rescaled according to eq. (6.5). No additional noise contributions beyond those specified in sec. 6.1.2 were included.

The VQE calculations in the current section were performed using the QISKIT Aer estimator with `shots=None` and `approximation=True`, based on a `qiskit_aer.AerSimulator` with `method="density_matrix"` and the scaled noise model described in sec. 6.1.2. With these settings the simulator directly evolves

the density matrix of the system according to the transpiled ansatz circuit and the specified gate-noise model, before evaluating exact expectation values on the final density matrix. We used this estimator to evaluate the observables for our ansatz circuit, thereby obtaining the relevant expectation values in the presence of circuit noise, but without finite-shot sampling error. This corresponds to evaluating noisy expectation values in the $M \to \infty$ measurement limit under the assumption that no noise is introduced as part of the measurement procedure. Measurement and readout errors are therefore ignored in this methodology, consistent with the restricted noise model introduced in sec. 6.1.2. As discussed in that section, this approach also ignores gate errors in the basis-transformation circuits required by the measurement protocol of sec. 5.2.3. This simplifies the simulation by avoiding the need to simulate a separate noisy circuit for each measurement group.

The calculations were performed on a GPU using the `device="GPU"` setting of the `qiskit_aer.AerSimulator`. For the $N = 6$ system considered here, each circuit evaluation in a density-matrix simulation is computationally and memory bandwidth intensive enough to benefit greatly from the highly parallel processing available on GPUs. We found that the simulations also benefited substantially from using the settings `cuStateVec_enable=True`, and `fusion_max_qubit=4`, which accelerated the circuit execution speed by a factor of up to ~ 4.

As discussed in sec. 5.4.1, alternative simulation strategies are available. In the current setting a further simulation approach is to evaluate the noisy circuit using repeated state vector simulations, sampling the stochastic noise process independently for each trajectory. However, since we are interested in exact expectation values, i.e. the $M \to \infty$ measurement limit, this would replace the exact expectation value by a finite-sample estimate, thereby introducing statistical convergence errors into the results. Since we are interested in the high-accuracy regime, we therefore again opted to use an exact method, the density-matrix simulator, which avoids the introduction of simulator-dependent approximations and errors.

6.4.2 Discussion

The purpose of the following calculations is to identify the dominant contributions in the realistic noise models described in sec. 6.1.2 and justify our assumed neglect of certain noise contributions. We consider the effect of each of the noise sources individually on the accuracy of our calculation. Specifically, we split the complete noise model into errors on gates within the ansatz circuit itself, errors on gates within the circuits used for the basis measurement (see sec. 5.2.3), readout errors, idling errors and reset and initialisation errors. As discussed in sec. 6.1.2, we consider

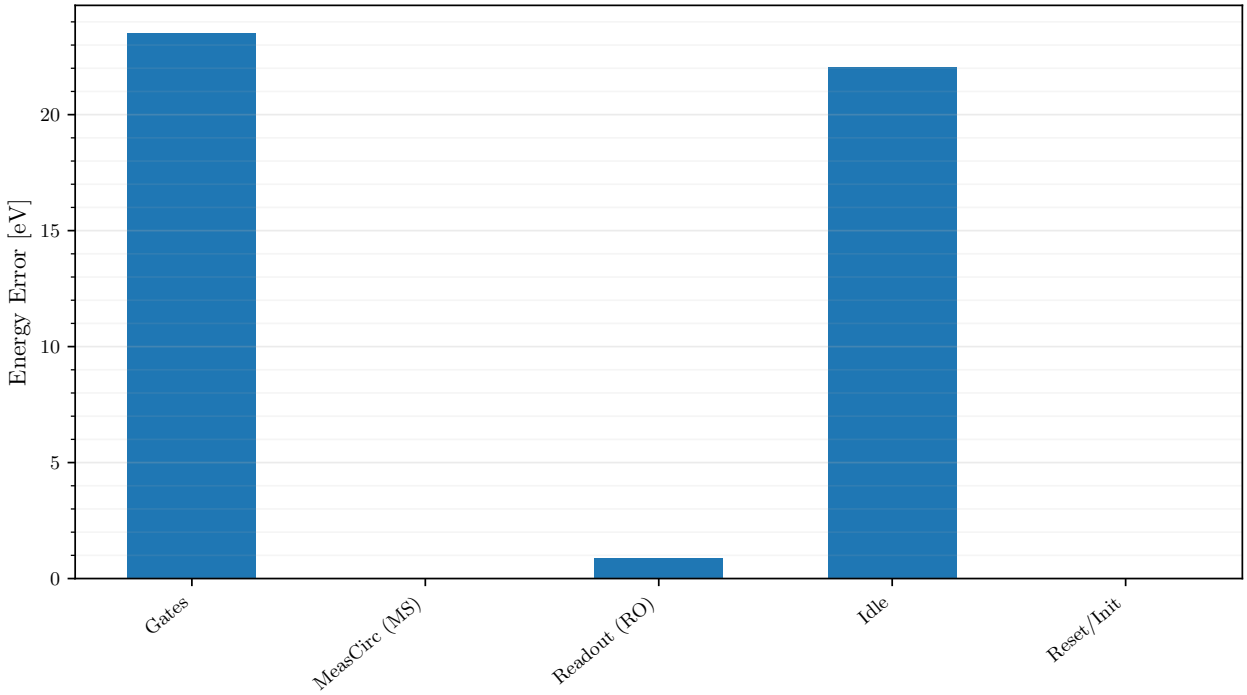


Figure 6.12: Energy error, ΔE, obtained when enabling individual components of the device-calibrated circuit-noise model with all other noise sources disabled; from left to right: ansatz-gate noise, measurement-circuit gate noise, readout error, idling noise, and reset/initialisation error,

gate noise on the ansatz circuit and on the measurement circuit separately to allow for computationally less expensive calculations. The results in fig. 6.12 show that readout, reset and idling error, and the gate noise on the measurement circuit have a negligible impact on the accuracy of the calculation and motivate our simplified noise model. While idling noise can be seen to have a non-negligible effect on the accuracy of our results – to a comparable extent as the gate noise on the ansatz circuit – we also neglect it for our calculations, since the decoherence of idling qubits can be mitigated reasonably efficiently in realistic applications [186].

Within our simplified noise model discussed in sec. 6.1.2 we first consider the effect of reducing both 1- and 2-qubit gate error probabilities independently. This corresponds to introducing the relative error probability factors f_{1Q} and f_{2Q}, which we use to independently rescale the noise channel for each of the two gate types according to equ. (6.5). We randomly selected $N_{\text{seed}} = 5$ optimal parameter sets from the VQE calculations presented in sec. 5.4. As discussed, we re-evaluated the energy at these parameters under varying levels of noise: we considered a grid of 20 logarithmically spaced values for both f_{1Q} and f_{2Q}. These simulations required a total GPU time of $t_{\text{GPU}} \approx 24\,\text{h}$. The results of fig. 6.14 and fig. 6.13 stress the importance of considering the noise on both 1- and 2-qubit gates simultaneously. For our VQE instance we find that even in the complete absence of 2-qubit gate errors

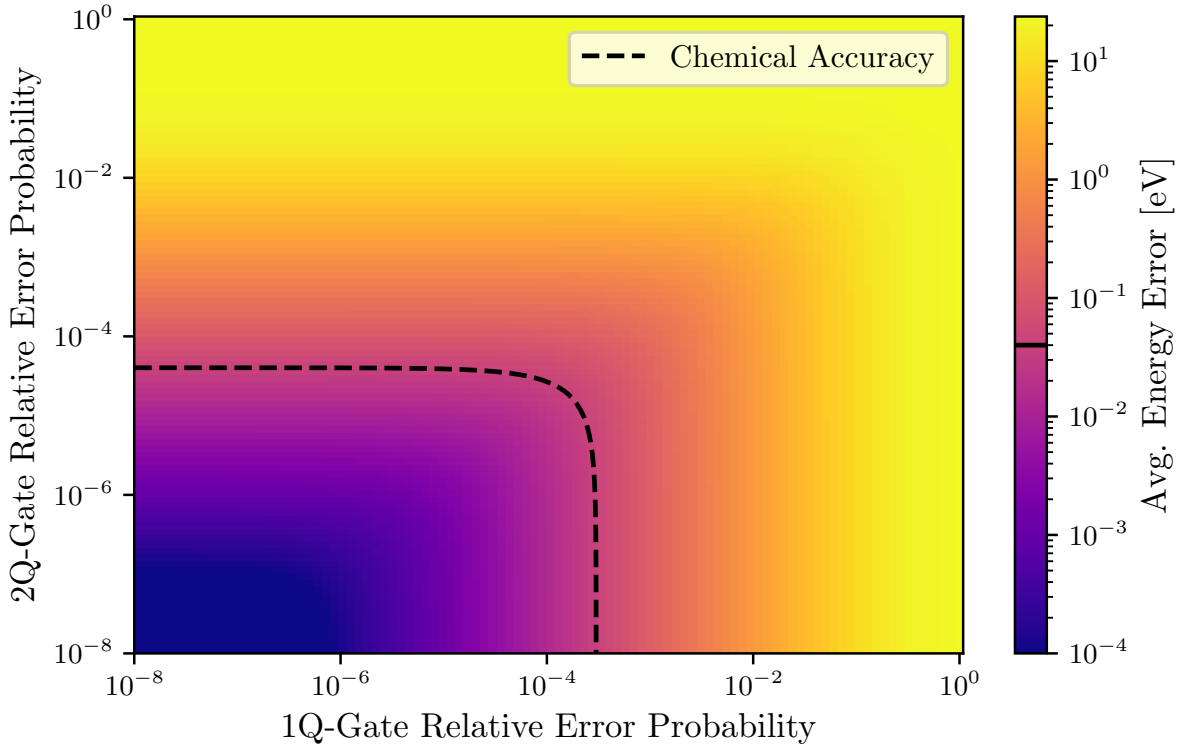


Figure 6.13: Energy error, ΔE, plotted as a heat-map with respect to the 1- and 2-qubit gate relative error probabilities, f_{1Q} and f_{2Q}, averaged over repeated VQE calculations at each (f_{1Q}, f_{2Q}).

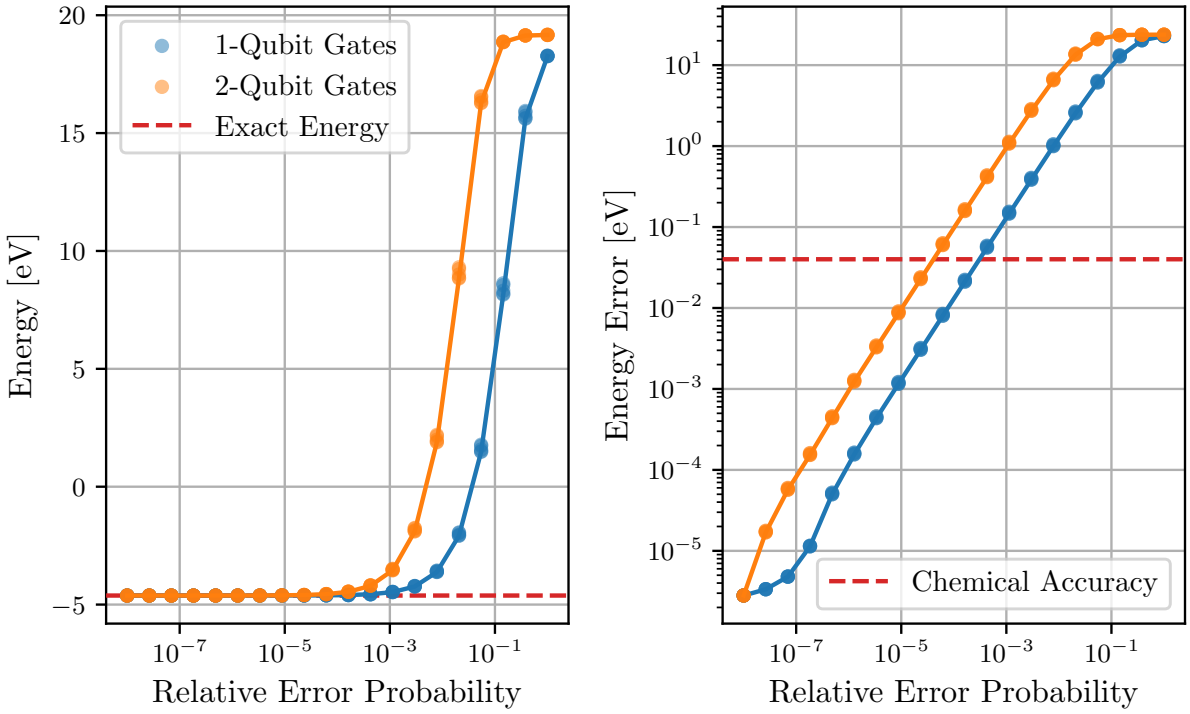


Figure 6.14: Energy error, ΔE, as a function of 1-qubit gate relative error probabilities, f_{1Q}, neglecting 2-qubit gate errors, $f_{2Q} \approx 0$, and vice versa.

the 1-qubit gate fidelities of currently available devices are not sufficient to reach anywhere close to chemical accuracy targets for the energy $\Delta E < \Delta E_{\text{chem}}$, requiring a relative reduction of error rates of $f_{1Q} \approx 1 \times 10^{-4}$. Similarly, in the absence of any 1-qubit gate errors chemical accuracy requires a reduction of $f_{2Q} \approx 1 \times 10^{-5}$. We note that this is relatively unusual from a NISQ perspective where it is generally assumed that 2-qubit gates are by far the limiting factor. The relatively "dense" nature of our ansatz circuit (see sec. 5.2.2) leads to a comparatively high 1-qubit gate count. After transpilation to the Sherbrooke native basis, the representative ansatz circuits considered here have a depth $\sim 10^4$, with $\sim 5 \times 10^3$ native two-qubit entangling gates (`ecr`) and $\sim 3 \times 10^4$ single-qubit gates in total (dominated by `rz` and `sx`). Single-qubit gates can be seen to outnumber two-qubit gates by a factor ~ 6 and errors in these gates therefore also have a relatively large impact on the accuracy of our calculations.

Given the above discussion, we instead consider a noise model in which both 1- and 2-qubit errors are simultaneously reduced, i.e. $f = f_{1Q} = f_{2Q}$. Based on the results of fig. 6.15 we find that under this noise model a simultaneous reduction of both 1- and 2-qubit errors by $f \approx 1 \times 10^{-5}$ is necessary to reach the chemical accuracy regime $\Delta E < \Delta E_{\text{chem}}$. Similarly, as shown in fig. 6.16, chemical accuracy

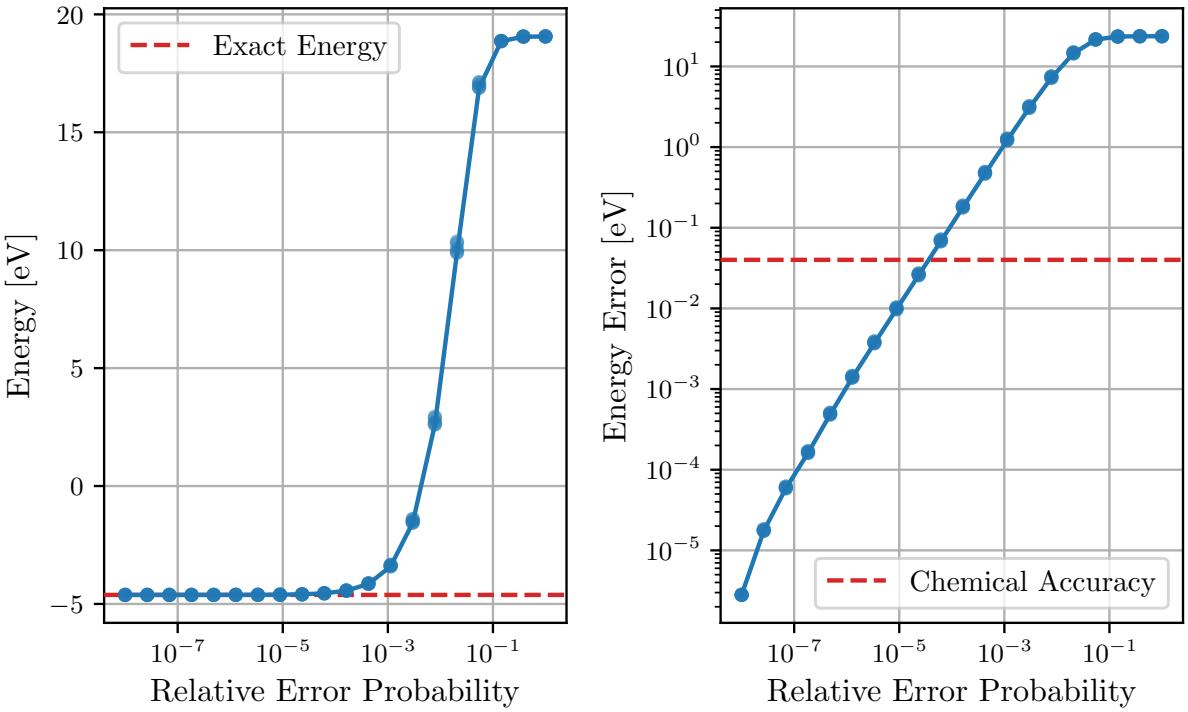


Figure 6.15: Energy error, ΔE, as a function of a combined 1- and 2-qubit gate relative error probability, $f = f_{1Q} = f_{2Q}$.

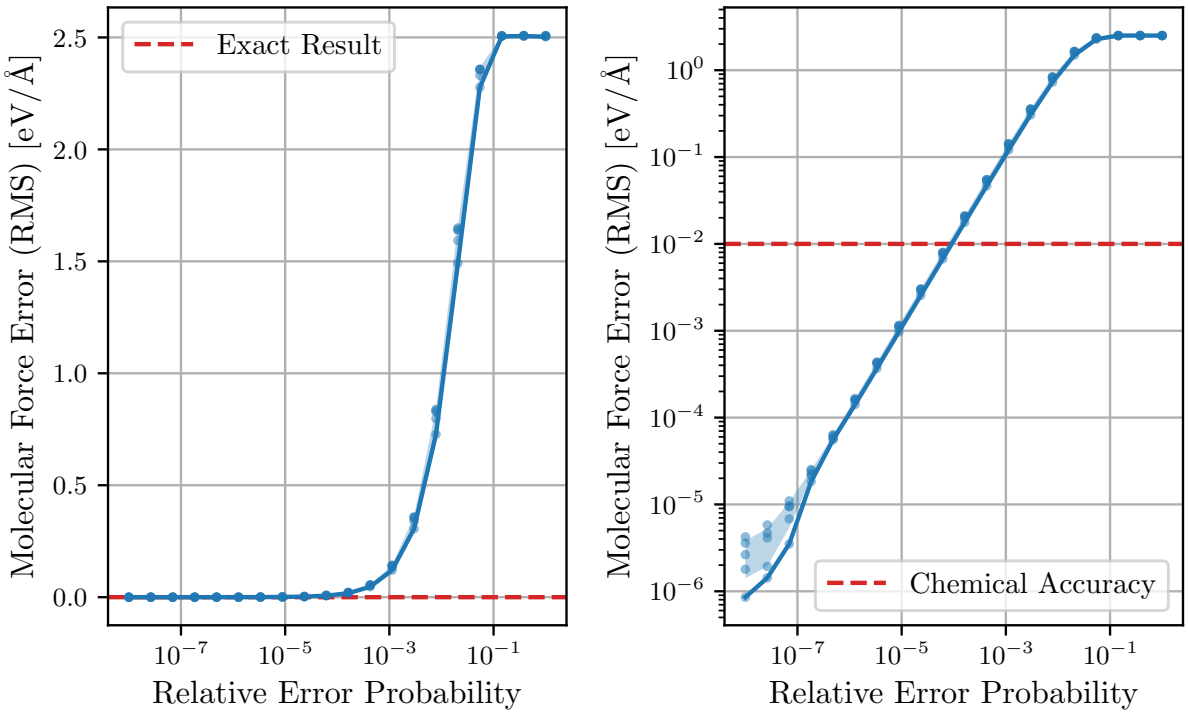


Figure 6.16: Force error, ΔF, as a function of a combined 1- and 2-qubit gate relative error probability, $f = f_{1Q} = f_{2Q}$.

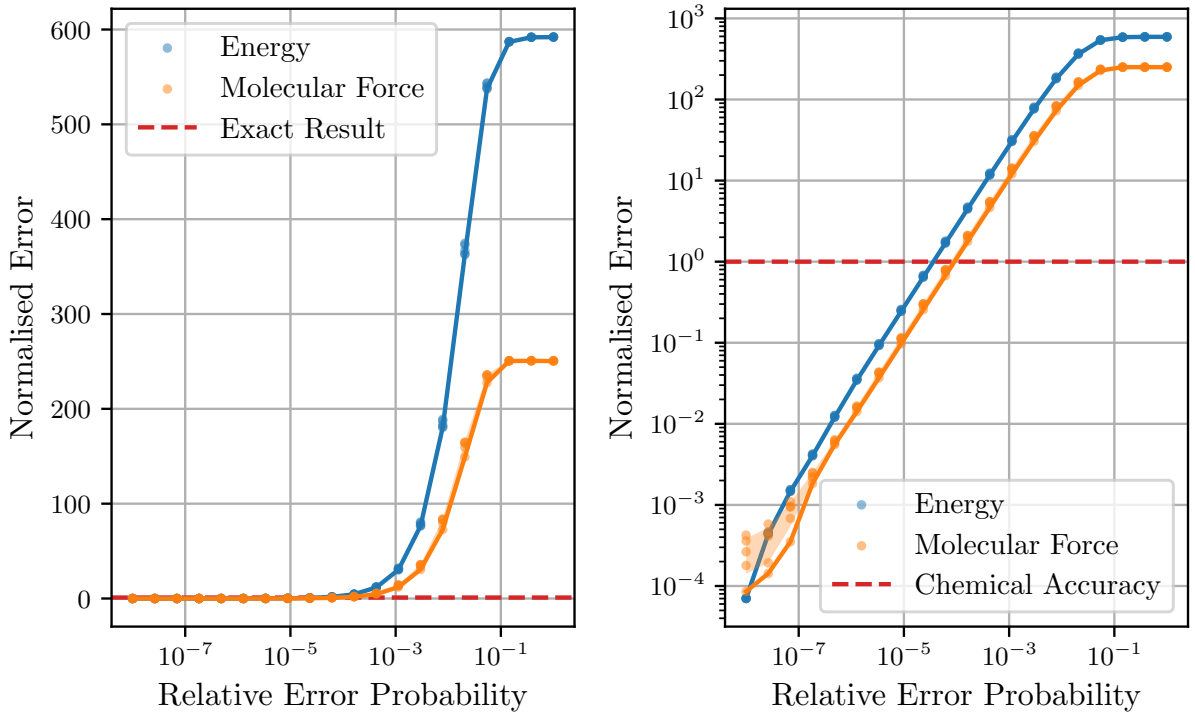


Figure 6.17: Energy error, ΔE and molecular force error, ΔF, each normalised to their respective chemical accuracy targets ΔE_{chem} and ΔF_{chem} as a function of the gate relative error probability, f.

for molecular forces, $\Delta F < \Delta F_{\text{chem}}$, requires a slightly less stringent error reduction of $f \approx 1 \times 10^{-4}$.

The above results indicate that circuit noise presents a considerable hurdle for chemically accurate simulations of our model, and that the required gate fidelities lie well outside the regime of NISQ feasibility. The current gate error rates of the IBM Sherbrooke QPU considered in our calculation are of order $\sim 10^{-4}$ and $\sim 10^{-2}$ for 1- and 2-qubit gates respectively. The required fidelity improvements of 4–5 orders of magnitude in the gate error probabilities represent physical error rates of order $\sim 10^{-8}$ and $\sim 10^{-6}$. These targets are far beyond the capabilities of NISQ devices and make realistic electronic structure calculation of our model completely unfeasible on current and near term hardware.

From the viewpoint of electronic structure calculations, however, the results are at least somewhat encouraging. In contrast to what is expected based on the variational principle of sec. 2.2, and to what was observed with respect to shot noise in sec. 6.3, in the current scenario molecular forces do not require any lower gate errors to reach chemical accuracy than energies. As a matter of fact for our particular problem instance $\Delta F < \Delta F_{\text{chem}}$ is reached at $f \sim 1 \times 10^{-4}$, a lower gate

fidelity than $\Delta E < \Delta E_{\text{chem}}$ at $f \sim 1 \times 10^{-5}$. As discussed in sec. 6.2.2, this is a result of the absence of a variational principle with respect to the error introduced by stochastic circuit noise. This is clearly demonstrated in fig. 6.17 where both energies and molecular forces can be seen to obey a linear scaling as the error probability is decreased, $\Delta E/\Delta F \sim \mathcal{O}(f)$. This in stark contrast to standard electronic structure theory, based on which it might be expected that molecular forces require greatly increased computational effort – in the current scenario corresponding to reduced gate error probabilities – compared to energies. While a variational advantage for energies is usually considered favourable in standard electronic structure theory, its absence under stochastic circuit noise implies that the limitations imposed by circuit noise and poor performance of VQE in realistic calculations do not deteriorate further when molecular forces are considered.

6.5 Conclusion

In this chapter we quantified the impact of finite-shot expectation value estimation and gate execution under realistic circuit noise, individually, on the accuracy of VQE calculations for the SSH model, in each case benchmarking both energies and molecular forces as the shot budget, M, and relative gate error probabilities, f, are increased/decreased respectively. We also analysed our results from a viewpoint of standard electronic structure theory, drawing more general conclusions about the applicability of VQE to realistic electronic structure calculations, as well as discussing the importance of considering molecular forces for realistic benchmarks of VQE calculations, focusing in particular on the (in-)applicability of the variational principle.

We found that for our implementation, chemically accurate energies lie on the verge of feasibility in the shot-noise-limited setting: the required total measurement cost corresponds to an estimated QPU run-time of order $\sim 1\,\text{yr}$. Under reasonable assumptions of improvements in both algorithmic performance and quantum hardware, this does not by itself rule out realistic energy calculations for the present model. In contrast, chemically accurate molecular forces are far more strongly constrained by a finite shot budget and require a dramatic increase in M. The resulting run-time estimate of order $\sim 1 \times 10^8\,\text{yr}$ renders the calculation of molecular forces completely infeasible within our problem instance and VQE implementation.

Although the precise numerical thresholds reported here are specific to the present model, ansatz, and implementation, we can draw several more general and qualitative conclusions. Since molecular forces constitute a substantially more strin-

gent target than energies, their consideration is important in exposing limitations of the VQE implementation for realistic electronic structure calculations. In our implementation they highlight not only the issue with optimisers such as SPSA – whose slow convergence and hyper-parameter sensitivity become prohibitive in the high-accuracy regime – but also the need for robust, systematically convergent, and largely problem-agnostic optimisation methods such as L-BFGS. Furthermore, the inclusion of molecular forces (or other response properties) is essential when assessing the feasibility of VQE for realistic electronic-structure calculations; energy-only benchmarks can substantially underestimate the true quantum resource requirements. More fundamentally, our results and discussion demonstrate that, under hardware-compatible measurement protocols, the loss of the energy's zero-variance behaviour can present a severe limitation to the accuracy achievable for molecular forces and other non-variational response quantities.

Within our simplified circuit-noise model, we found that chemically accurate simulations require substantial improvements in gate fidelities, far beyond NISQ feasibility and well within the regime of fault-tolerant quantum computing: chemically accurate energies require a reduction of gate error probabilities by $f \approx 1\times10^{-5}$, while chemically accurate molecular forces require a slightly less stringent threshold $f \approx 1 \times 10^{-4}$. Relative to current device-calibrated error rates, these thresholds correspond to 2-qubit gate error rates below 10^{-6}, placing the present calculation well outside the regime of NISQ feasibility and instead requiring fault-tolerant operation. These fidelity requirements make chemically accurate simulations of our problem instance and VQE implementation infeasible in the near term.

Again we stress that the precise numerical thresholds reported here are highly problem specific. From the viewpoint of electronic structure calculations, however, it is notable that circuit noise did not impose an additional accuracy degradation for molecular forces beyond that already present for energies: the force target is reached at a comparable, and in our instance even slightly less stringent, fidelity threshold than the energy target. Fundamentally, this reflects the absence of a variational principle for the mixed-state deviations induced by stochastic gate noise. Consequently, should hardware fidelities eventually enter a regime where chemically accurate energies become accessible for this ansatz class, force-based applications such as geometry optimisation may become feasible without incurring the extreme separations in required resources observed in the finite-shot analysis.

Within our simplified circuit-noise model, we found that chemically accurate energies require considerable simultaneous improvements in both 1- and 2-qubit gate fidelities. Varying the two error classes independently shows that neither can be

treated as parametrically negligible for the present ansatz: even in the absence of 2-qubit gate errors, the energy remains far outside chemical accuracy unless the 1-qubit error probabilities are reduced by $f_{1Q} \approx 1 \times 10^{-4}$, and conversely chemical accuracy requires $f_{2Q} \approx 1 \times 10^{-5}$ even if 1-qubit errors are neglected. Under the more realistic assumption of a uniform improvement factor, $f = f_{1Q} = f_{2Q}$, chemical accuracy is reached only at $f \approx 1 \times 10^{-5}$ for energies and at the slightly less stringent threshold $f \approx 1 \times 10^{-4}$ for molecular forces. Relative to current device-calibrated error rates, these thresholds correspond to improvements of $\sim$ 4–5 orders of magnitude in gate error probabilities, placing the present calculation well outside the regime of NISQ feasibility and instead in a regime that would require fault-tolerant operation and/or dramatic advances in hardware-level fidelities.

From the viewpoint of electronic structure calculations, however, it is notable that circuit noise does not impose an additional penalty for molecular forces beyond that already present for energies: the force target is reached at a comparable, and in our instance even less stringent, fidelity threshold than the energy target. This reflects the absence of a variational stationary principle for the mixed-state deviations induced by stochastic gate noise, for which both ΔE and ΔF exhibit an approximately linear dependence on f. Consequently, should hardware fidelities eventually enter a regime where chemically accurate energies become accessible for this ansatz class, force-based applications such as geometry optimisation may become feasible without the extreme separations in hardware requirements observed in the finite-shot analysis.

Finally, we emphasise that the two noise sources investigated here were treated in isolation to enable clear attribution of their respective limitations. In realistic NISQ calculations and experiments, finite-shot sampling and circuit noise will act simultaneously and may compound non-trivially, such as for example through the appearance of noise-induced barren plateaus [187] so that our thresholds for chemical accuracy should be interpreted as optimistic lower bounds on the resources and fidelities required for chemically accurate energy and force calculations for the present model and ansatz. Finally, we note that while fault-tolerant quantum computing may ultimately address the circuit-noise limitations by enabling effectively noiseless logical operations, it does not remove the need for finite-sample expectation-value estimation: even in a fault-tolerant setting, observable estimates would still suffer from shot noise and remain subject to severe limitations imposed on chemically accurate simulations.

Chapter 7

Conclusion

In this thesis we set out to assess the prospects of using the variational quantum eigensolver (VQE) for useful and realistic electronic structure calculations in the NISQ era, in particular from a viewpoint of an electronic structure practitioner. A central motivation for studying *trans*-polyacetylene (TPA) is that it sits in a regime that is not only industrially relevant and non-trivial from a chemistry and physics perspective, but also amenable to quantum simulations using a simple description by the Su–Schrieffer–Heeger (SSH) model. This provides a compact Hamiltonian that retains the essential interplay between electronic correlation and geometry, and – crucially for the aims of this work – admits the evaluation of molecular forces, an under-investigated component of VQE calculations and a key step towards realistic electronic structure calculations.

In Chapter 5 we detailed our VQE implementation and demonstrated the ability to reach chemical accuracy for both energies and molecular forces in ideal simulations of the SSH model. We found that from a pure ansatz expressivity viewpoint energies and molecular forces require almost the same quantum resources since the ansatz circuit used displays a unique behaviour known as a computational phase transition, which allows the preparation of an effectively exact ground state at no additional cost. From a standard electronic structure theory viewpoint this behaviour is not only highly unusual but also beneficial; it allows chemically accurate molecular forces to be calculated at negligible overhead – in spite of the applicability of the variational principle – and obviates the need for the calculation of non-variational response forces.

When finite sampling and realistic gate noise are introduced in Chapter 6, we found the practically useful calculation to lie well beyond the regime of NISQ feasibility. In the shot-noise-limited setting we found that chemically accurate energies are only marginally within reach in terms of total measurement cost, whereas chemically

accurate molecular forces become completely infeasible under the same considerations, with resource estimates increasing by many orders of magnitude. From a viewpoint of standard electronic structure theory our results are surprisingly poor, a fundamental limitation explained by the breaking of the zero-variance property in the VQE setting. In the circuit-noise-limited setting, we similarly found that chemically accurate predictions require improvements in gate fidelities by several orders of magnitude relative to current error rates, placing the calculation firmly in a regime that would require fault-tolerant operation and/or dramatic advances in hardware-level performance. Taken together, these results imply that, for the present SSH instance and VQE implementation, chemically meaningful simulations are very much outside the NISQ and near-term feasible regime.

A key contribution of this thesis is that our feasibility assessment is carried out explicitly from the viewpoint of molecular forces and the variational principle, rather than from an energy-only benchmarking perspective. From an electronic structure practitioner's viewpoint, the evaluation of molecular forces is an important part of realistic electronic structure calculations. They frequently require the inclusion of subtle non-variational response terms and a considerable increase in computational resources required to reach chemical accuracy. Our results show that molecular forces are crucial for analysing the feasibility of VQE for realistic electronic structure calculations: while in ideal simulations we found molecular forces to effectively come "for free", in realistic calculations we found that energy-only benchmarks can substantially underestimate the true quantum resources needed for chemically relevant workflows. At the same time, our results provide a more nuanced analysis of the variational principle and its applications and consequences in VQE, in particular with respect to the calculation of molecular forces and how they are affected by various ways in which realistic VQE simulations break standard electronic structure behaviours.

Finally, we conclude that based on our work, TPA, and the SSH model in particular, is a valuable benchmark system for the feasibility of VQE calculations. It occupies a productive middle ground between idealised lattice models and ab initio molecular Hamiltonians. As we have shown, molecular forces are an important consideration when assessing VQE simulations for realistic electronic structure applications. While within our implementation we do not find NISQ simulations of the SSH model to be feasible, it nonetheless provides an attractive platform on which to test the improvements in both algorithms and hardware. We therefore expect that further systematic study of the SSH model will be valuable in establishing more representative and realistic VQE calculations, as well as in guiding the development of

algorithmic and hardware advances required before quantum computers can deliver chemically meaningful simulations of realistic molecules and materials.

Bibliography

[1] P. E. M. Siegbahn, "Structures and Energetics for O2 Formation in Photosystem II", Accounts of Chemical Research **42**, 1871 (2009).

[2] G. R. Bhimanapati, Z. Lin, V. Meunier, Y. Jung, J. Cha, S. Das, D. Xiao, Y. Son, M. S. Strano, V. R. Cooper, L. Liang, S. G. Louie, E. Ringe, W. Zhou, S. S. Kim, R. R. Naik, B. G. Sumpter, H. Terrones, F. Xia, Y. Wang, J. Zhu, D. Akinwande, N. Alem, J. A. Schuller, R. E. Schaak, M. Terrones, and J. A. Robinson, "Recent Advances in Two-Dimensional Materials beyond Graphene", ACS Nano **9**, 11509 (2015).

[3] Y. Li, J. Hao, H. Liu, Y. Li, and Y. Ma, "The metallization and superconductivity of dense hydrogen sulfide", The Journal of Chemical Physics **140**, 174712 (2014).

[4] N. Marzari, A. Ferretti, and C. Wolverton, "Electronic-structure methods for materials design", Nature Materials **20**, 736 (2021).

[5] R. P. Feynman, "Simulating physics with computers", International Journal of Theoretical Physics **21**, 467 (1982).

[6] J. Preskill, "Quantum Computing in the NISQ era and beyond", Quantum **2**, 79 (2018).

[7] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, "Variational quantum algorithms", Nature Reviews Physics **3**, 625 (2021).

[8] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, "A variational eigenvalue solver on a photonic quantum processor", Nature Communications **5**, 4213 (2014).

[9] J. Tilly, H. Chen, S. Cao, D. Picozzi, K. Setia, Y. Li, E. Grant, L. Wossnig, I. Rungger, G. H. Booth, and J. Tennyson, "The Variational Quantum Eigensolver: A review of methods and best practices", Physics Reports, The

Variational Quantum Eigensolver: A Review of Methods and Best Practices **986**, 1 (2022).

[10] Y. Cao, J. Romero, and A. Aspuru-Guzik, "Potential of quantum computing for drug discovery", IBM Journal of Research and Development **62**, 6:1 (2018).

[11] S. Stanisic, J. L. Bosse, F. M. Gambetta, R. A. Santos, W. Mruczkiewicz, T. E. O'Brien, E. Ostby, and A. Montanaro, "Observing ground-state properties of the Fermi-Hubbard model using a scalable algorithm on a quantum computer", Nature Communications **13**, 5743 (2022).

[12] L. Mineh and A. Montanaro, "Solving the Hubbard model using density matrix embedding theory and the variational quantum eigensolver", Physical Review B **105**, 125117 (2022).

[13] J. Kattemölle and J. Van Wezel, "Variational quantum eigensolver for the Heisenberg antiferromagnet on the kagome lattice", Physical Review B **106**, 214429 (2022).

[14] J. L. Bosse and A. Montanaro, *Probing ground state properties of the kagome antiferromagnetic Heisenberg model using the Variational Quantum Eigensolver*, (Oct. 4, 2021) http://arxiv.org/abs/2108.08086 (visited on 06/27/2023), pre-published.

[15] D. Wu, R. Rossi, F. Vicentini, N. Astrakhantsev, F. Becca, X. Cao, J. Carrasquilla, F. Ferrari, A. Georges, M. Hibat-Allah, M. Imada, A. M. Läuchli, G. Mazzola, A. Mezzacapo, A. Millis, J. Robledo Moreno, T. Neupert, Y. Nomura, J. Nys, O. Parcollet, R. Pohle, I. Romero, M. Schmid, J. M. Silvester, S. Sorella, L. F. Tocchio, L. Wang, S. R. White, A. Wietek, Q. Yang, Y. Yang, S. Zhang, and G. Carleo, "Variational benchmarks for quantum many-body problems", Science **386**, 296 (2024).

[16] "The Hubbard model at half a century", Nature Physics **9**, 523 (2013).

[17] D. Claudino, J. Wright, A. J. McCaskey, and T. S. Humble, "Benchmarking adaptive variational quantum eigensolvers", Frontiers in Chemistry **8**, 606863 (2020).

[18] J. Wright, M. Gowrishankar, D. Claudino, P. C. Lotshaw, T. Nguyen, A. J. McCaskey, and T. S. Humble, *Numerical Simulations of Noisy Quantum Circuits for Computational Chemistry*, (May 5, 2022) http://arxiv.org/abs/2112.15540 (visited on 09/12/2023), pre-published.

[19] H. R. Grimsley, G. S. Barron, E. Barnes, S. E. Economou, and N. J. Mayhall, "ADAPT-VQE is insensitive to rough parameter landscapes and barren plateaus", Apr. 14, 2022.

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

[21] P. Pulay, "Analytical derivatives, forces, force constants, molecular geometries, and related response properties in electronic structure theory", WIREs Computational Molecular Science **4**, 169 (2014).

[22] A. Szabo and N. S. Ostlund, *Modern quantum chemistry: introduction to advanced electronic structure theory* (Dover Publications, Inc, Mineola, New York, 2012), 1 p.

[23] T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular electronic-structure theory* (Wiley, Chichester ; New York, 2000), 908 pp.

[24] P. Jørgensen, J. Olsen, M. B. Johansen, T. J. von Buchwald, A. E. Hillers-Bendtsen, K. V. Mikkelsen, and T. Helgaker, "A variational reformulation of molecular properties in electronic-structure theory", Science Advances **10**, eadn3454 (2024).

[25] C. David Sherrill and H. F. Schaefer, in *Advances in Quantum Chemistry*, Vol. 34, edited by P.-O. Löwdin, J. R. Sabin, M. C. Zerner, and E. Brändas (Academic Press, Jan. 1, 1999), pp. 143–269.

[26] M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos, "Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients", Reviews of Modern Physics **64**, 1045 (1992).

[27] N. D. M. Hine, M. Robinson, P. D. Haynes, C.-K. Skylaris, M. C. Payne, and A. A. Mostofi, "Accurate ionic forces and geometry optimization in linear-scaling density-functional theory with local orbitals", Physical Review B **83**, 195102 (2011).

[28] D. Feller, "The use of systematic sequences of wave functions for estimating the complete basis set, full configuration interaction limit in water", The Journal of Chemical Physics **98**, 7059 (1993).

[29] L. Marotta and F. Siringo, "A variational method from the variance of energy", The European Physical Journal C **44**, 293 (2005).

[30] C. J. Umrigar and C. Filippi, "Energy and Variance Optimization of Many-Body Wave Functions", Physical Review Letters **94**, 150201 (2005).

[31] D.-B. Zhang, B.-L. Chen, Z.-H. Yuan, and T. Yin, "Variational quantum eigensolvers by variance minimization", Chinese Physics B **31**, 120301 (2022).

[32] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, "Quantum Monte Carlo simulations of solids", Reviews of Modern Physics **73**, 33 (2001).

[33] J. P. Perdew and K. Schmidt, "Jacob's ladder of density functional approximations for the exchange-correlation energy", AIP Conference Proceedings **577**, 1 (2001).

[34] J. Kalikka, J. Akola, and R. O. Jones, "Simulation of crystallization in $Ge_2Sb_2Te_5$: a memory effect in the canonical phase-change material", Physical Review B **90**, 184109 (2014).

[35] A. J. Cohen, P. Mori-Sánchez, and W. Yang, "Insights into Current Limitations of Density Functional Theory", Science **321**, 792 (2008).

[36] A. J. Cohen, P. Mori-Sánchez, and W. Yang, "Fractional charge perspective on the band gap in density-functional theory", Physical Review B **77**, 115123 (2008).

[37] P. Mori-Sánchez, A. J. Cohen, and W. Yang, "Many-electron self-interaction error in approximate density functionals", The Journal of Chemical Physics **125**, 201102 (2006).

[38] K. D. Vogiatzis, D. Ma, J. Olsen, L. Gagliardi, and W. A. de Jong, "Pushing configuration-interaction to the limit: Towards massively parallel MCSCF calculations", The Journal of Chemical Physics **147**, 184111 (2017).

[39] L. Lin, J. Lu, and L. Ying, "Numerical methods for Kohn–Sham density functional theory", Acta Numerica **28**, 405 (2019).

[40] T. Helgaker, S. Coriani, P. Jørgensen, K. Kristensen, J. Olsen, and K. Ruud, "Recent Advances in Wave Function-Based Methods of Molecular-Property Calculations", Chemical Reviews **112**, 543 (2012).

[41] R. Bast, U. Ekström, B. Gao, T. Helgaker, K. Ruud, and A. J. Thorvaldsen, "The ab initio calculation of molecular electric, magnetic and geometric properties", Physical Chemistry Chemical Physics **13**, 2627 (2011).

[42] 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, "PySCF: the Python-based simulations of chemistry framework", WIREs Computational Molecular Science **8**, e1340 (2018).

[43] T. Helgaker, in *Encyclopedia of Computational Chemistry*, edited by P. Von Ragué Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer, and P. R. Schreiner, 1st ed. (Wiley, Sept. 25, 1998).

[44] P. Pulay, "Ab initio calculation of force constants and equilibrium geometries in polyatomic molecules: I. Theory", Molecular Physics **17**, 197 (1969).

[45] R. M. Martin, *Electronic Structure: Basic Theory and Practical Methods* (Cambridge University Press, Cambridge, 2004).

[46] Q. Sun, "Libcint: An efficient general integral library for Gaussian basis functions", Journal of Computational Chemistry **36**, 1664 (2015).

[47] J. P. Coe, "Analytic Gradients for Selected Configuration Interaction", Journal of Chemical Theory and Computation **19**, 874 (2023).

[48] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information*, 10th anniversary ed (Cambridge University Press, Cambridge, 2010), 1 p.

[49] H. Shirakawa, "The Discovery of Polyacetylene Film: The Dawning of an Era of Conducting Polymers (Nobel Lecture)", Angewandte Chemie International Edition **40**, 2574 (2001).

[50] A. G. MacDiarmid, ""Synthetic Metals": A Novel Role for Organic Polymers (Nobel Lecture)", Angewandte Chemie International Edition **40**, 2581 (2001).

[51] A. J. Heeger, "Semiconducting and Metallic Polymers: The Fourth Generation of Polymeric Materials (Nobel Lecture)", Angewandte Chemie International Edition **40**, 2591 (2001).

[52] N. W. Ashcroft and N. D. Mermin, *Solid state physics*, Repr (Brooks/Cole Thomson Learning, South Melbourne, 2012), 826 pp.

[53] X. Guo and A. Facchetti, "The journey of conducting polymers from discovery to application", Nature Materials **19**, 922 (2020).

[54] J. R. Reynolds, B. C. Thompson, and T. A. Skotheim, *Handbook of Conducting Polymers, Fourth Edition - 2 Volume Set* (CRC Press, Nov. 14, 2019), 1525 pp.

[55] A. J. Heeger, S. Kivelson, J. R. Schrieffer, and W. .-. Su, "Solitons in conducting polymers", Reviews of Modern Physics **60**, 781 (1988).

[56] W. P. Su, J. R. Schrieffer, and A. J. Heeger, "Solitons in Polyacetylene", Physical Review Letters **42**, 1698 (1979).

[57] W. P. Su and J. R. Schrieffer, "Soliton dynamics in polyacetylene", Proceedings of the National Academy of Sciences **77**, 5626 (1980).

[58] J. K. Asbóth, L. Oroszlány, and A. Pályi, "A Short Course on Topological Insulators: Band-structure topology and edge states in one and two dimensions", 2016.

[59] E. Jeckelmann, "Mott-Peierls transition in the extended Peierls-Hubbard model", Physical Review B **57**, 11838 (1998).

[60] M. P. Lima and G. Magela e Silva, "Dynamical evolution of polaron to bipolaron in conjugated polymers", Physical Review B **74**, 224304 (2006).

[61] S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. I. J. Probert, K. Refson, and M. C. Payne, "First principles methods using CASTEP", Zeitschrift für Kristallographie - Crystalline Materials **220**, 567 (2005).

[62] V. I. Anisimov, J. Zaanen, and O. K. Andersen, "Band theory and Mott insulators: Hubbard U instead of Stoner I", Physical Review B **44**, 943 (1991).

[63] M. Cococcioni and S. de Gironcoli, "Linear response approach to the calculation of the effective interaction parameters in the LDA + U method", Physical Review B **71**, 035105 (2005).

[64] J. W. Mintmire and C. T. White, "Local-density-functional results for the dimerization of trans-polyacetylene: Relationship to the band-gap problem", Physical Review B **35**, 4180 (1987).

[65] J. Ashkenazi, W. E. Pickett, H. Krakauer, C. S. Wang, B. M. Klein, and S. R. Chubb, "Ground state of trans-polyacetylene and the Peierls mechanism", Physical Review Letters **62**, 2016 (1989).

[66] J. Paloheimo and J. von Boehm, "Density-functional study of the dimerization of trans-polyacetylene", Physical Review B **46**, 4304 (1992).

[67] C. Ho Choi, M. Kertesz, and A. Karpfen, "The effects of electron correlation on the degree of bond alternation and electronic structure of oligomers of polyacetylene", The Journal of Chemical Physics **107**, 6712 (1997).

[68] D. Jacquemin, Eric. A. Perpète, I. Ciofini, and C. Adamo, "Assessment of recently developed density functional approaches for the evaluation of the bond length alternation in polyacetylene", Chemical Physics Letters **405**, 376 (2005).

[69] I. Ciofini, C. Adamo, and H. Chermette, "Effect of self-interaction error in the evaluation of the bond length alternation in trans-polyacetylene using density-functional theory", The Journal of Chemical Physics **123**, 121102 (2005).

[70] D. B. Nguyen, K. A. Jackson, and J. E. Peralta, "Bond length alternation of π-conjugated polymers predicted by the Fermi–Löwdin orbital self-interaction correction method", The Journal of Chemical Physics **160**, 014101 (2024).

[71] J. Nokelainen, B. Barbiellini, and A. Bansil, "Magnetic properties of polyacetylene: Exploring electronic correlation effects through first-principles modeling", Computational Materials Science **259**, 114036 (2025).

[72] G. Sun, J. Kürti, M. Kertesz, and R. H. Baughman, "Dimensional changes as a function of charge injection for trans-polyacetylene: A density functional theory study", The Journal of Chemical Physics **117**, 7691 (2002).

[73] D. Jacquemin, E. A. Perpète, G. Scalmani, M. J. Frisch, R. Kobayashi, and C. Adamo, "Assessment of the efficiency of long-range corrected functionals for some properties of large compounds", The Journal of Chemical Physics **126**, 144105 (2007).

[74] M. J. G. Peach, E. I. Tellgren, P. Sałek, T. Helgaker, and D. J. Tozer, "Structural and Electronic Properties of Polyacetylene and Polyyne from Hybrid and Coulomb-Attenuated Density Functionals", The Journal of Physical Chemistry A **111**, 11930 (2007).

[75] A. Aspuru-Guzik, A. D. Dutoi, P. J. Love, and M. Head-Gordon, "Simulated Quantum Computation of Molecular Energies", Science **309**, 1704 (2005).

[76] P. J. J. O'Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, E. Jeffrey, E. Lucero, A. Megrant, J. Y. Mutus, M. Neeley, C. Neill, C. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. C. White, P. V. Coveney, P. J. Love, H. Neven, A. Aspuru-Guzik, and J. M. Martinis, "Scalable Quantum Simulation of Molecular Energies", Physical Review X **6**, 031007 (2016).

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

[78] C. Hempel, C. Maier, J. Romero, J. McClean, T. Monz, H. Shen, P. Jurcevic, B. P. Lanyon, P. Love, R. Babbush, A. Aspuru-Guzik, R. Blatt, and C. F. Roos, "Quantum Chemistry Calculations on a Trapped-Ion Quantum Simulator", Physical Review X **8**, 031022 (2018).

[79] L. Zhao, J. Goings, K. Shin, W. Kyoung, J. I. Fuks, J.-K. Kevin Rhee, Y. M. Rhee, K. Wright, J. Nguyen, J. Kim, and S. Johri, "Orbital-optimized pair-correlated electron simulations on trapped-ion quantum computers", npj Quantum Information **9**, 60 (2023).

[80] 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 optimized unitary coupled cluster ansatz", Nature Physics **20**, 1240 (2024).

[81] P. Jordan and E. Wigner, "Über das Paulische Äquivalenzverbot", Zeitschrift für Physik **47**, 631 (1928).

[82] J. Romero, R. Babbush, J. R. McClean, C. Hempel, P. J. Love, and A. Aspuru-Guzik, "Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz", Quantum Science and Technology **4**, 014008 (2018).

[83] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information*, 10th anniversary edition (Cambridge university press, Cambridge, 2010).

[84] M. B. Hastings, D. Wecker, B. Bauer, and M. Troyer, *Improving Quantum Algorithms for Quantum Chemistry*, (Dec. 15, 2016) `http://arxiv.org/abs/1403.1539` (visited on 11/12/2025), pre-published.

[85] Y. S. Yordanov, D. R. M. Arvidsson-Shukur, and C. H. W. Barnes, "Efficient quantum circuits for quantum computational chemistry", Physical Review A **102**, 062612 (2020).

[86] D. Wecker, M. B. Hastings, N. Wiebe, B. K. Clark, C. Nayak, and M. Troyer, "Solving strongly correlated electron models on a quantum computer", Physical Review A **92**, 062318 (2015).

[87] W. J. Huggins, J. R. McClean, N. C. Rubin, Z. Jiang, N. Wiebe, K. B. Whaley, and R. Babbush, "Efficient and noise resilient measurements for quantum chemistry on near-term quantum computers", npj Quantum Information **7**, 23 (2021).

[88] I. D. Kivlichan, J. McClean, N. Wiebe, C. Gidney, A. Aspuru-Guzik, G. K.-L. Chan, and R. Babbush, "Quantum Simulation of Electronic Structure with Linear Depth and Connectivity", Physical Review Letters **120**, 110501 (2018).

[89] B. O'Gorman, W. J. Huggins, E. G. Rieffel, and K. B. Whaley, *Generalized swap networks for near-term quantum computing*, (May 13, 2019) `http://arxiv.org/abs/1905.05118` (visited on 09/22/2025), pre-published.

[90] C. Cade, L. Mineh, A. Montanaro, and S. Stanisic, "Strategies for solving the Fermi-Hubbard model on near-term quantum computers", Physical Review B **102**, 235122 (2020).

[91] B. T. Gard, L. Zhu, G. S. Barron, N. J. Mayhall, S. E. Economou, and E. Barnes, "Efficient symmetry-preserving state preparation circuits for the variational quantum eigensolver algorithm", npj Quantum Information **6**, 10 (2020).

[92] G.-L. R. Anselmetti, D. Wierichs, C. Gogolin, and R. M. Parrish, "Local, expressive, quantum-number-preserving VQE ansätze for fermionic systems", New Journal of Physics **23**, 113010 (2021).

[93] H. L. Tang, V. Shkolnikov, G. S. Barron, H. R. Grimsley, N. J. Mayhall, E. Barnes, and S. E. Economou, "Qubit-ADAPT-VQE: An Adaptive Algorithm for Constructing Hardware-Efficient Ans\"atze on a Quantum Processor", PRX Quantum **2**, 020310 (2021).

[94] Y. S. Yordanov, V. Armaos, C. H. W. Barnes, and D. R. M. Arvidsson-Shukur, "Qubit-excitation-based adaptive variational quantum eigensolver", Communications Physics **4**, 1 (2021).

[95] H. G. A. Burton, "Accurate and gate-efficient quantum Ans\"atze for electronic states without adaptive optimization", Physical Review Research **6**, 023300 (2024).

[96] J. T. Seeley, M. J. Richard, and P. J. Love, "The Bravyi-Kitaev transformation for quantum computation of electronic structure", The Journal of Chemical Physics **137**, 224109 (2012).

[97] S. Bravyi and A. Kitaev, "Fermionic quantum computation", Annals of Physics **298**, 210 (2002).

[98] Z. Jiang, A. Kalev, W. Mruczkiewicz, and H. Neven, "Optimal fermion-to-qubit mapping via ternary trees with applications to reduced quantum states learning", Quantum **4**, 276 (2020).

[99] K. Setia, S. Bravyi, A. Mezzacapo, and J. D. Whitfield, "Superfast encodings for fermionic quantum simulation", Physical Review Research **1**, 033033 (2019).

[100] C. Derby, J. Klassen, J. Bausch, and T. Cubitt, "Compact fermion to qubit mappings", Physical Review B **104**, 035118 (2021).

[101] K. Setia, R. Chen, J. E. Rice, A. Mezzacapo, M. Pistoia, and J. D. Whitfield, "Reducing Qubit Requirements for Quantum Simulations Using Molecular Point Group Symmetries", Journal of Chemical Theory and Computation **16**, 6091 (2020).

[102] D. Picozzi and J. Tennyson, "Symmetry-adapted encodings for qubit number reduction by point-group and other Boolean symmetries", Quantum Science and Technology **8**, 035026 (2023).

[103] Y. Shee, P.-K. Tsai, C.-L. Hong, H.-C. Cheng, and H.-S. Goan, "Qubit-efficient encoding scheme for quantum simulations of electronic structure", Physical Review Research **4**, 023154 (2022).

[104] D. Wecker, M. B. Hastings, and M. Troyer, "Progress towards practical quantum variational algorithms", Physical Review A **92**, 042303 (2015).

[105] A. M. Alvertis, A. Khan, T. Iadecola, P. P. Orth, and N. Tubman, "Classical Benchmarks for Variational Quantum Eigensolver Simulations of the Hubbard Model", Quantum **9**, 1748 (2025).

[106] B. Anselme Martin, P. Simon, and M. J. Rančić, "Simulating strongly interacting Hubbard chains with the variational Hamiltonian ansatz on a quantum computer", Physical Review Research **4**, 023190 (2022).

[107] Z. Cai, "Resource Estimation for Quantum Variational Simulations of the Hubbard Model", Physical Review Applied **14**, 014059 (2020).

[108] A. C. Y. Li, M. S. Alam, T. Iadecola, A. Jahin, J. Job, D. M. Kurkcuoglu, R. Li, P. P. Orth, A. B. Özgüler, G. N. Perdue, and N. M. Tubman, "Benchmarking variational quantum eigensolvers for the square-octagon-lattice Kitaev model", Physical Review Research **5**, 033071 (2023).

[109] J.-M. Reiner, F. Wilhelm-Mauch, G. Schön, and M. Marthaler, "Finding the ground state of the Hubbard model by variational methods on a quantum computer with gate errors", Quantum Science and Technology **4**, 035005 (2019).

[110] R. Wiersema, C. Zhou, Y. de Sereville, J. F. Carrasquilla, Y. B. Kim, and H. Yuen, "Exploring Entanglement and Optimization within the Hamiltonian Variational Ansatz", PRX Quantum **1**, 020319 (2020).

[111] J. Gibbs, Z. Holmes, and P. Stevenson, *Exploiting symmetries in nuclear Hamiltonians for ground state preparation*, (Feb. 15, 2024) http://arxiv.org/abs/2402.10277 (visited on 09/23/2025), pre-published.

[112] T. A. Bespalova, K. Delić, G. Pupillo, F. Tacchino, and I. Tavernelli, "Simulating the Fermi-Hubbard model with long-range hopping on a quantum computer", Physical Review A **111**, 052619 (2025).

[113] A. A. Mele, G. B. Mbeng, G. E. Santoro, M. Collura, and P. Torta, "Avoiding barren plateaus via transferability of smooth solutions in a Hamiltonian variational ansatz", Physical Review A **106**, L060401 (2022).

[114] C.-Y. Park, *Efficient ground state preparation in variational quantum eigensolver with symmetry-breaking layers*, (Jan. 30, 2024) http://arxiv.org/abs/2106.02509 (visited on 09/23/2025), pre-published.

[115] A. Mukherjee, N. F. Berthusen, J. C. Getelina, P. P. Orth, and Y.-X. Yao, "Comparative study of adaptive variational quantum eigensolvers for multi-orbital impurity models", Communications Physics **6**, 4 (2023).

[116] T. Piskor, F. G. Eich, J.-M. Reiner, S. Zanker, N. Vogt, M. Marthaler, and F. Wilhelm-Mauch, "Using gradient-based algorithms to determine ground state energies on a quantum computer", Physical Review A **105**, 062415 (2022).

[117] T. Albash and D. A. Lidar, "Adiabatic quantum computation", Reviews of Modern Physics **90**, 015002 (2018).

[118] C. Feniou, M. Hassan, D. Traoré, E. Giner, Y. Maday, and J.-P. Piquemal, "Overlap-ADAPT-VQE: practical quantum chemistry on quantum computers via overlap-guided compact Ansätze", Communications Physics **6**, 192 (2023).

[119] P. G. Anastasiou, Y. Chen, N. J. Mayhall, E. Barnes, and S. E. Economou, "TETRIS-ADAPT-VQE: An adaptive algorithm that yields shallower, denser circuit Ans\"atze", Physical Review Research **6**, 013254 (2024).

[120] M. Ramôa, P. G. Anastasiou, L. P. Santos, N. J. Mayhall, E. Barnes, and S. E. Economou, "Reducing the resources required by ADAPT-VQE using coupled exchange operators and improved subroutines", npj Quantum Information **11**, 86 (2025).

[121] C. K. Long, K. Dalton, C. H. W. Barnes, D. R. M. Arvidsson-Shukur, and N. Mertig, "Layering and subpool exploration for adaptive variational quantum eigensolvers: Reducing circuit depth, runtime, and susceptibility to noise", Physical Review A **109**, 042413 (2024).

[122] V. Verteletskyi, T.-C. Yen, and A. F. Izmaylov, "Measurement Optimization in the Variational Quantum Eigensolver Using a Minimum Clique Cover", The Journal of Chemical Physics **152**, 124114 (2020).

[123] P. Gokhale, O. Angiuli, Y. Ding, K. Gui, T. Tomesh, M. Suchara, M. Martonosi, and F. T. Chong, *Minimizing State Preparations in Variational Quantum Eigensolver by Partitioning into Commuting Families*, (July 31, 2019) http://arxiv.org/abs/1907.13623 (visited on 07/04/2023), pre-published.

[124] T.-C. Yen, V. Verteletskyi, and A. F. Izmaylov, *Measuring all compatible operators in one series of a single-qubit measurements using unitary transformations*, (Mar. 13, 2020) http://arxiv.org/abs/1907.09386 (visited on 07/05/2023), pre-published.

[125] A. F. Izmaylov, T.-C. Yen, R. A. Lang, and V. Verteletskyi, *Unitary partitioning approach to the measurement problem in the Variational Quantum Eigensolver method*, (Oct. 18, 2019) http://arxiv.org/abs/1907.09040 (visited on 07/05/2023), pre-published.

[126] O. Crawford, B. van Straaten, D. Wang, T. Parks, E. Campbell, and S. Brierley, "Efficient quantum measurement of Pauli operators in the presence of finite sampling error", Quantum **5**, 385 (2021).

[127] T.-C. Yen, A. Ganeshram, and A. F. Izmaylov, "Deterministic improvements of quantum measurements with grouping of compatible operators, non-local transformations, and covariance estimates", npj Quantum Information **9**, 1 (2023).

[128] B. Wu, J. Sun, Q. Huang, and X. Yuan, "Overlapped grouping measurement: A unified framework for measuring quantum states", Quantum **7**, 896 (2023).

[129] J. F. Gonthier, M. D. Radin, C. Buda, E. J. Doskocil, C. M. Abuan, and J. Romero, "Measurements as a roadblock to near-term practical quantum advantage in chemistry: Resource analysis", Physical Review Research **4**, 033154 (2022).

[130] J. Nocedal and S. J. Wright, *Numerical optimization*, Second edition, Springer Series in Operations Research and Financial Engineering (Springer, New York, NY, 2006), 1 p.

[131] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A Limited Memory Algorithm for Bound Constrained Optimization", SIAM Journal on Scientific Computing **16**, 1190 (1995).

[132] R. Fletcher and C. M. Reeves, "Function minimization by conjugate gradients", The Computer Journal **7**, 149 (1964).

[133] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, (Jan. 30, 2017) http://arxiv.org/abs/1412.6980 (visited on 10/08/2025), pre-published.

[134] A. Pellow-Jarman, I. Sinayskiy, A. Pillay, and F. Petruccione, "A comparison of various classical optimizers for a variational quantum linear solver", Quantum Information Processing **20**, 202 (2021).

[135] X. Bonet-Monroig, H. Wang, D. Vermetten, B. Senjean, C. Moussa, T. Bäck, V. Dunjko, and T. E. O'Brien, "Performance comparison of optimization methods on variational quantum algorithms", Physical Review A **107**, 032407 (2023).

[136] B. D. M. Jones, L. Mineh, and A. Montanaro, "Benchmarking a wide range of optimisers for solving the Fermi–Hubbard model using the variational quantum eigensolver", Quantum Science and Technology **10**, 045032 (2025).

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

[138] J. S. Kottmann, A. Anand, and A. Aspuru-Guzik, "A feasible approach for automatically differentiable unitary coupled-cluster on quantum computers", Chemical Science **12**, 3497 (2021).

[139] D. Wierichs, J. Izaac, C. Wang, and C. Y.-Y. Lin, "General parameter-shift rules for quantum gradients", Quantum **6**, 677 (2022).

[140] A. Mari, T. R. Bromley, and N. Killoran, "Estimating the gradient and higher-order derivatives on quantum hardware", Physical Review A **103**, 012405 (2021).

[141] J. Spall, "Multivariate stochastic approximation using a simultaneous perturbation gradient approximation", IEEE Transactions on Automatic Control **37**, 332 (1992).

[142] T. Q. N. Developers, *Qiskit Nature 0.6.0*, version 0.6.0, Zenodo, Apr. 14, 2023.

[143] K. Seki and S. Yunoki, "Spatial, spin, and charge symmetry projections for a Fermi-Hubbard model on a quantum computer", Physical Review A **105**, 032419 (2022).

[144] F. H. L. Essler, ed., *The one-dimensional Hubbard model* (Cambridge University Press, Cambridge New York, 2005), 1 p.

[145] L. Mineh, "Solving the Hubbard model using the variational quantum eigensolver" ().

[146] A. Choquette, A. Di Paolo, P. K. Barkoutsos, D. Sénéchal, I. Tavernelli, and A. Blais, "Quantum-optimal-control-inspired ansatz for variational quantum algorithms", Physical Review Research **3**, 023092 (2021).

[147] T. Tsuchimochi, Y. Mori, and S. L. Ten-no, "Spin-projection for quantum computation: A low-depth approach to strong correlation", Physical Review Research **2**, 043142 (2020).

[148] H. G. A. Burton, D. Marti-Dafcik, D. P. Tew, and D. J. Wales, "Exact electronic states with shallow quantum circuits from global optimisation", npj Quantum Information **9**, 75 (2023).

[149] L. W. Bertels, H. R. Grimsley, S. E. Economou, E. Barnes, and N. J. Mayhall, "Symmetry Breaking Slows Convergence of the ADAPT Variational Quantum Eigensolver", Journal of Chemical Theory and Computation **18**, 6656 (2022).

[150] T. Tsuchimochi, M. Taii, T. Nishimaki, and S. L. Ten-no, "Adaptive construction of shallower quantum circuits with quantum spin projection for fermionic systems", Physical Review Research **4**, 033100 (2022).

[151] I. Miháliková, J. Carlson, D. Neill, and I. Stetcu, *State preparation and symmetries*, (Oct. 8, 2025) http://arxiv.org/abs/2510.06702 (visited on 12/02/2025), pre-published.

[152] P. Jain, A. F. Izmaylov, and E. R. Kjellgren, *Exact Factorization of Unitary Transformations with Spin-Adapted Generators*, (Nov. 18, 2025) http://arxiv.org/abs/2511.14914 (visited on 12/01/2025), pre-published.

[153] C.-Y. Park and N. Killoran, "Hamiltonian variational ansatz without barren plateaus", Feb. 16, 2023.

[154] B. Murta, G. Catarina, and J. Fernández-Rossier, "Berry phase estimation in gate-based adiabatic quantum simulation", Physical Review A **101**, 020302 (2020).

[155] A. J. Ceulemans, *Group theory applied to chemistry*, Second edition, Theoretical Chemistry and Computational Modelling (Springer, Dordrecht, 2024), 348 pp.

[156] Z. Jiang, K. J. Sung, K. Kechedzhi, V. N. Smelyanskiy, and S. Boixo, "Quantum algorithms to simulate many-body physics of correlated fermions", Physical Review Applied **9**, 044036 (2018).

[157] C. C. J. Roothaan, "Self-Consistent Field Theory for Open Shells of Electronic Systems", Reviews of Modern Physics **32**, 179 (1960).

[158] H.-Y. Liu, T.-P. Sun, Y.-C. Wu, Y.-J. Han, and G.-P. Guo, "Mitigating barren plateaus with transfer-learning-inspired parameter initializations", New Journal of Physics **25**, 013039 (2023).

[159] M. Cerezo, M. Larocca, D. García-Martín, N. L. Diaz, P. Braccia, E. Fontana, M. S. Rudolph, P. Bermejo, A. Ijaz, S. Thanasilp, E. R. Anschuetz, and Z. Holmes, "Does provable absence of barren plateaus imply classical simulability?", Nature Communications **16**, 7907 (2025).

[160] H. H. S. Chan, N. Fitzpatrick, J. Segarra-Martí, M. J. Bearpark, and D. P. Tew, "Molecular excited state calculations with adaptive wavefunctions on a quantum eigensolver emulation: reducing circuit depth and separating spin states", Physical Chemistry Chemical Physics **23**, 26438 (2021).

[161] T. E. O'Brien, B. Senjean, R. Sagastizabal, X. Bonet-Monroig, A. Dutkiewicz, F. Buda, L. DiCarlo, and L. Visscher, "Calculating energy derivatives for quantum chemistry on a quantum computer", npj Quantum Information **5**, 1 (2019).

[162] K. Mitarai, Y. O. Nakagawa, and W. Mizukami, "Theory of analytical energy derivatives for the variational quantum eigensolver", Physical Review Research **2**, 013129 (2020).

[163] I. O. Sokolov, P. K. Barkoutsos, L. Moeller, P. Suchsland, G. Mazzola, and I. Tavernelli, "Microcanonical and finite-temperature ab initio molecular dynamics simulations on quantum computers", Physical Review Research **3**, 013125 (2021).

[164] U. Azad and H. Singh, "Quantum Chemistry Calculations using Energy Derivatives on Quantum Computers", Chemical Physics **558**, 111506 (2022).

[165] T. E. O'Brien, M. Streif, N. C. Rubin, R. Santagati, Y. Su, W. J. Huggins, J. J. Goings, N. Moll, E. Kyoseva, M. Degroote, C. S. Tautermann, J. Lee, D. W. Berry, N. Wiebe, and R. Babbush, "Efficient quantum computation of molecular forces and other energy gradients", Physical Review Research **4**, 043210 (2022).

[166] J. Lai, Y. Fan, Q. Fu, Z. Li, and J. Yang, "Accurate and efficient calculations of Hellmann–Feynman forces for quantum computation", The Journal of Chemical Physics **159**, 114113 (2023).

[167] Q. Gong, Q. Man, J. Zhao, Y. Li, M. Dou, Q. Wang, Y.-C. Wu, and G.-P. Guo, "Simulating chemical reaction dynamics on quantum computer", The Journal of Chemical Physics **160**, 124103 (2024).

[168] R. M. Parrish, E. G. Hohenstein, P. L. McMahon, and T. J. Martınez, "Hybrid Quantum/Classical Derivative Theory: Analytical Gradients and Excited-State Dynamics for the Multistate Contracted Variational Quantum Eigensolver",

[169] R. M. Parrish, G.-L. R. Anselmetti, and C. Gogolin, *Analytical Ground- and Excited-State Gradients for Molecular Electronic Structure Theory from Hybrid Quantum/Classical Methods*, (Oct. 11, 2021) `http://arxiv.org/abs/2110.05040` (visited on 04/18/2024), pre-published.

[170] K. Omiya, Y. O. Nakagawa, S. Koh, W. Mizukami, Q. Gao, and T. Kobayashi, "Analytical energy gradient for state-averaged orbital-optimized variational quantum eigensolvers and its application to a photochemical reaction", Journal of Chemical Theory and Computation **18**, 741 (2022).

[171] S. Yalouz, E. Koridon, B. Senjean, B. Lasorne, F. Buda, and L. Visscher, "Analytical Nonadiabatic Couplings and Gradients within the State-Averaged Orbital-Optimized Variational Quantum Eigensolver", Journal of Chemical Theory and Computation **18**, 776 (2022).

[172] E. G. Hohenstein, O. Oumarou, R. Al-Saadon, G.-L. R. Anselmetti, M. Scheurer, C. Gogolin, and R. M. Parrish, "Efficient Quantum Analytic Nuclear Gradients with Double Factorization", The Journal of Chemical Physics **158**, 114119 (2023).

[173] A. Javadi-Abhari, M. Treinish, K. Krsulich, C. J. Wood, J. Lishman, J. Gacon, S. Martiel, P. D. Nation, L. S. Bishop, A. W. Cross, B. R. Johnson, and J. M. Gambetta, *Quantum computing with Qiskit*, (June 19, 2024) http://arxiv.org/abs/2405.08810 (visited on 01/08/2026), pre-published.

[174] T. Jones and J. Gacon, *Efficient calculation of gradients in classical simulations of variational quantum algorithms*, (Sept. 6, 2020) http://arxiv.org/abs/2009.02823 (visited on 06/27/2023), pre-published.

[175] V. Bergholm, J. Izaac, M. Schuld, C. Gogolin, S. Ahmed, V. Ajith, M. S. Alam, G. Alonso-Linaje, B. AkashNarayanan, A. Asadi, J. M. Arrazola, U. Azad, S. Banning, C. Blank, T. R. Bromley, B. A. Cordier, J. Ceroni, A. Delgado, O. Di Matteo, A. Dusko, T. Garg, D. Guala, A. Hayes, R. Hill, A. Ijaz, T. Isacsson, D. Ittah, S. Jahangiri, P. Jain, E. Jiang, A. Khandelwal, K. Kottmann, R. A. Lang, C. Lee, T. Loke, A. Lowe, K. McKiernan, J. J. Meyer, J. A. Montañez-Barrera, R. Moyard, Z. Niu, L. J. O'Riordan, S. Oud, A. Panigrahi, C.-Y. Park, D. Polatajko, N. Quesada, C. Roberts, N. Sá, I. Schoch, B. Shi, S. Shu, S. Sim, A. Singh, I. Strandberg, J. Soni, A. Száva, S. Thabet, R. A. Vargas-Hernández, T. Vincent, N. Vitucci, M. Weber, D. Wierichs, R. Wiersema, M. Willmann, V. Wong, S. Zhang, and N. Killoran, *PennyLane: Automatic differentiation of hybrid quantum-classical computations*, arXiv.org, (Nov. 12, 2018) https://arxiv.org/abs/1811.04968v4 (visited on 01/08/2026).

[176] J. Kim, J. Kim, and D. Rosa, "Universal effectiveness of high-depth circuits in variational eigenproblems", Physical Review Research **3**, 023203 (2021).

[177] M. Larocca, N. Ju, D. García-Martín, P. J. Coles, and M. Cerezo, "Theory of overparametrization in quantum neural networks", Sept. 23, 2021.

[178] E. R. Anschuetz and B. T. Kiani, "Quantum variational algorithms are swamped with traps", Nature Communications **13**, 7760 (2022).

[179] K. Dalton, C. K. Long, Y. S. Yordanov, C. G. Smith, C. H. W. Barnes, N. Mertig, and D. R. M. Arvidsson-Shukur, "Quantifying the effect of gate errors on variational quantum eigensolvers for quantum chemistry", npj Quantum Information **10**, 18 (2024).

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

[181] J. M. Kübler, A. Arrasmith, L. Cincio, and P. J. Coles, "An Adaptive Optimizer for Measurement-Frugal Variational Algorithms", Quantum **4**, 263 (2020).

[182] E. Fontana, N. Fitzpatrick, D. M. Ramo, R. Duncan, and I. Rungger, "Evaluating the noise resilience of variational quantum algorithms", Physical Review A **104**, 022403 (2021).

[183] N. C. Rubin, R. Babbush, and J. McClean, "Application of fermionic marginal constraints to hybrid quantum algorithms", New Journal of Physics **20**, 053020 (2018).

[184] R. Assaraf and M. Caffarel, "Zero-Variance Principle for Monte Carlo Algorithms", Physical Review Letters **83**, 4682 (1999).

[185] R. Assaraf and M. Caffarel, "Zero-variance zero-bias principle for observables in quantum Monte Carlo: Application to forces", The Journal of Chemical Physics **119**, 10536 (2003).

[186] N. Ezzell, B. Pokharel, L. Tewala, G. Quiroz, and D. A. Lidar, "Dynamical decoupling for superconducting qubits: A performance survey", Physical Review Applied **20**, 064027 (2023).

[187] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, "Noise-induced barren plateaus in variational quantum algorithms", Nature Communications **12**, 6961 (2021).

[188] V. L. C. Jr and M. Cococcioni, "Extended DFT + U + V method with on-site and inter-site electronic interactions", Journal of Physics: Condensed Matter **22**, 055602 (2010).

[189] S. Kivelson, W.-P. Su, J. R. Schrieffer, and A. J. Heeger, "Missing bond-charge repulsion in the extended Hubbard model: Effects in polyacetylene", Physical Review Letters **58**, 1899 (1987).

[190] J. D. Cloizeaux, "Orthogonal Orbitals and Generalized Wannier Functions", Physical Review **129**, 554 (1963).

[191] B. Bradlyn, L. Elcoro, J. Cano, M. G. Vergniory, Z. Wang, C. Felser, M. I. Aroyo, and B. A. Bernevig, "Topological quantum chemistry", Nature **547**, 298 (2017).

[192] W. Kohn, "Analytic Properties of Bloch Waves and Wannier Functions", Physical Review **115**, 809 (1959).

[193] J. D. Cloizeaux, "Analytical Properties of n-Dimensional Energy Bands and Wannier Functions", Physical Review **135**, A698 (1964).

[194] D. Vanderbilt, *Berry Phases in Electronic Structure Theory: Electric Polarization, Orbital Magnetization and Topological Insulators* (Cambridge University Press, Cambridge, 2018).

[195] N. Marzari and D. Vanderbilt, "Maximally localized generalized Wannier functions for composite energy bands", Physical Review B **56**, 12847 (1997).

[196] I. Souza, N. Marzari, and D. Vanderbilt, "Maximally localized Wannier functions for entangled energy bands", Physical Review B **65**, 035109 (2001).

[197] D. Vanderbilt and R. D. King-Smith, "Electric polarization as a bulk quantity and its relation to surface charge", Physical Review B **48**, 4442 (1993).

[198] T. Thonhauser, D. Ceresoli, D. Vanderbilt, and R. Resta, "Orbital Magnetization in Periodic Insulators", Physical Review Letters **95**, 137205 (2005).

[199] C. Brouder, G. Panati, M. Calandra, C. Mourougane, and N. Marzari, "Exponential Localization of Wannier Functions in Insulators", Physical Review Letters **98**, 046402 (2007).

Appendix A

Bond-centred TPA Model

While the SSH model is able to explain many of the features observed in TPA, it is an independent electron model that does not include any two-body interaction terms and hence cannot account for electronic interactions beyond renormalisation of the non-interacting Hamiltonian parameters. A more accurate model of TPA has to include additional terms to further account for this interaction.

In second quantisation the electronic interaction Hamiltonian takes the general form:

$$H_{\text{e-e}} = \sum_{i,j,k,l,\sigma,\sigma'} U_{ijkl} c^{\dagger}_{i,\sigma} c^{\dagger}_{j,\sigma'} c_{k,\sigma} c_{l,\sigma'} \tag{A.1}$$

where $U_{ijkl} = \langle \psi_i(\boldsymbol{x}), \psi_j(\boldsymbol{x}') | V_{e-e}(|\boldsymbol{x} - \boldsymbol{x}'|) | \psi_k(\boldsymbol{x}) \psi_l(\boldsymbol{x}') \rangle$ is the matrix element of the (possibly screened) Coulomb repulsion and creation and annihilation operators now explicitly include spin indices. Generally, only a very small set of terms in this sum is included. In the famous Hubbard Hamiltonian the electronic interaction is greatly simplified, retaining only an on-site repulsion $U_{iiii} \equiv U$. In the extended Hubbard Hamiltonian the direct interaction $U_{ijij} \equiv V$ between nearest neighbour pairs $\langle ij \rangle$ is also included yielding:

$$H_{\text{e-e}} = \sum_{i} U n_{i,\uparrow} n_{i,\downarrow} + \sum_{\langle ij \rangle} V n_i n_j \tag{A.2}$$

where $n_{i,\sigma} = c^{\dagger}_{i,\sigma} c_{i,\sigma}$ is the occupation number operator and $n_i = \sum_{\sigma} n_{i,\sigma}$. These two terms, oftentimes referred to simply as U and V, are the terms most commonly included to account for electronic repulsion.

It would be desirable if we were able to appropriately account for the electronic interaction using only an additional U term, since by itself it already leads to an immense increase in complexity. Addition of a V term would not only increase the complexity of the model even further, but also make the parametrisation of the model

itself more difficult. While parameterising U from DFT+U calculations [63] has become a relatively standard task, parameterising V from DFT+U+V calculations [188] is a much less well established procedure. To this end we investigate whether a bond-centred model is able to capture the relevant physics in a less complex model.

In a chemical picture simultaneous occupation of neighbouring lattice sites is roughly equivalent to double occupation of a bonding orbital: a double bond. The V term which penalises simultaneous occupation of neighbouring lattice sites can therefore be expected to be somewhat equivalent to a U term penalising double occupation of a bonding orbital. This motivation can be expressed slightly more mathematically using a simple tight-binding chain, by considering the mapping of the U term between a lattice site basis and a new basis of bonding orbitals between neighbouring sites:

$$|\psi_i\rangle = \frac{1}{\sqrt{2}}(|\phi_n\rangle + |\phi_{n+1}\rangle) \tag{A.3}$$

Under this simple basis transformation the U term in the new bonding basis, $\tilde{U}$, is given by:

$$\begin{aligned}\tilde{U} = \tilde{U}_{iiii} = \langle\psi_i, \psi_i|\hat{V}_{\text{e-e}}|\psi_i, \psi_i\rangle = \\ \frac{1}{4}\big(\langle\phi_i, \phi_i| + \langle\phi_i, \phi_{i+1}| + \langle\phi_{i+1}, \phi_i| + \langle\phi_{i+1}, \phi_{i+1}|\big)\hat{V}_{\text{e-e}} \\ \big(|\phi_i, \phi_i\rangle + |\phi_i, \phi_{i+1}\rangle + |\phi_{i+1}, \phi_i\rangle + |\phi_{i+1}, \phi_{i+1}\rangle\big)\end{aligned} \tag{A.4}$$

This term can be seen to include contributions from both the $U = U_{iiii}$ and $V = V_{ii+1ii+1}$ terms of the original lattice site basis. Additionally, it also includes other contributions from exchange terms of the form U_{ijji} and "off-diagonal" terms U_{ijjj} between neighbouring sites $j = i+1$. These terms are oftentimes neglected in simple models, even though their magnitude can be comparable to the standard U and V terms [189]. A bond-centred model that includes only a U term might therefore be able to account for physics beyond a simple atom-centred model that also includes only a U term.

A.1 Independent Electron Models

An important quality that a bond-centred model should have, is that it reproduces the independent electron behaviour of the original model. Motivated by the previous discussion, we therefore first consider equivalence between an independent electron bond- and atom-centred model.

The Hamiltonian of each system is given by:

$$H_{\text{Atom}} = \sum_n \left\{ \epsilon \cdot (c^\dagger_{n,A} c_{n,A} + c^\dagger_{n,B} c_{n,B}) + (u \cdot c^\dagger_{n,A} c_{n,B} + v \cdot c^\dagger_{n,B} c_{n+1,A} + h.c.) \right\} \tag{A.5}$$

$$H_{\text{Bond}} = \sum_n \left\{ g \cdot \tilde{c}^\dagger_{n,A} \tilde{c}_{n,A} + h \cdot \tilde{c}^\dagger_{n,B} \tilde{c}_{n,B} + t \cdot (\tilde{c}^\dagger_{n,A} \tilde{c}_{n,B} + \tilde{c}^\dagger_{n,B} \tilde{c}_{n+1,A} + h.c.) \right\} \tag{A.6}$$

where $c^\dagger_{n,A/B}$ and $\tilde{c}^\dagger_{n,A/B}$ are used to distinguish the creation and annihilation operators in the two models.

The atom-centred model is the standard SSH model, equ. (3.7), where we have now explicitly included an onsite energy ϵ for each orbital. As the orbitals are considered to be situated on each carbon atom, the environment on each site is equivalent, but the separation between neighbouring sites is staggered. As such, the onsite energies are constant while the hoppings are staggered. In the bond-centred model, however, the orbitals are situated on the bonds. Therefore, the environment of the single and double bond sites are different, while the separation between neighbouring bond sites is constant. This leads to a situation in which the onsite energies are staggered, while the nearest neighbour hoppings are all equal, the reverse of the atom-centred model.

Exact equivalence between the two models would require a unitary basis transformation which, when applied to the basis states of the atom-centred model, transforms the atom-centred Hamiltonian H_{Atom} into the bond-centred Hamiltonian H_{Bond}. The existence of such a transformation immediately implies that the Hamiltonians have the same bandstructures. Similarly, equal bandstructures imply the existence of such a transformation. In matrix language, this is because if $\mathbf{U}$ and $\tilde{\mathbf{U}}$ are the matrices diagonalising $\mathbf{H}_{\text{Atom}}$ and $\mathbf{H}_{\text{Bond}}$ to the same diagonal matrix $\mathbf{\Lambda}$, then $\mathbf{U}\tilde{\mathbf{U}}^\dagger$ transforms between the two Hamiltonians:

$$(\mathbf{U}\tilde{\mathbf{U}}^\dagger)^\dagger \mathbf{H}_{\text{Atom}} (\mathbf{U}\tilde{\mathbf{U}}^\dagger) = \tilde{\mathbf{U}}\mathbf{U}^\dagger \mathbf{H}_{\text{Atom}} \mathbf{U}\tilde{\mathbf{U}}^\dagger = \tilde{\mathbf{U}}\mathbf{\Lambda}\tilde{\mathbf{U}}^\dagger = \mathbf{H}_{\text{Bond}} \tag{A.7}$$

Since both Hamiltonians are periodic, after forming the Bloch basis states this reduces to finding a basis transformation between the single particle momentum

space Hamiltonian matrices $\mathbf{H}_{\text{Atom}}(k)$ and $\mathbf{H}_{\text{Bond}}(k)$:

$$\mathbf{H}_{\text{Atom}}(k) = \begin{bmatrix} \epsilon & u + ve^{-ika} \\ u + ve^{+ika} & \epsilon \end{bmatrix} \tag{A.8}$$

$$\mathbf{H}_{\text{Bond}}(k) = \begin{bmatrix} g & t(1 + e^{-ika}) \\ t(1 + e^{+ika}) & h \end{bmatrix} \tag{A.9}$$

where a is the lattice constant of the unit cell and we have assumed $u, v, t \in \mathbb{R}$. The dispersion relations for the two models are given by:

$$E_{\text{Atom}}(k) = \epsilon \pm \sqrt{u^2 + v^2 + 2uv\cos(ka)} \tag{A.10}$$

$$E_{\text{Bond}}(k) = \frac{h+g}{2} \pm \frac{1}{2}\sqrt{(h+g)^2 - 4hg + 8t^2(1 + \cos(ka))} \tag{A.11}$$

The two band-structures can be made exactly equal when the parameters of the bond-centred model are chosen based on the atom-centred parameters as:

$$g = \epsilon + u - v \qquad h = \epsilon + v - u \qquad t = \sqrt{uv} \tag{A.12}$$

with a resulting Hamiltonian:

$$\mathbf{H}_{\text{Bond}}(k) = \begin{bmatrix} \epsilon + u - v & \sqrt{uv}(1 + e^{-ika}) \\ \sqrt{uv}(1 + e^{+ika}) & \epsilon + v - u \end{bmatrix} \tag{A.13}$$

When this is fulfilled a basis transformation between the the two momentum space Hamiltonians can be constructed easily as $\mathbf{U}(k)\tilde{\mathbf{U}}^\dagger(k)$, where $\mathbf{U}(k)$ and $\tilde{\mathbf{U}}(k)$ are the matrices diagonalising $\mathbf{H}_{\text{Atom}}(k)$ and $\mathbf{H}_{\text{Bond}}(k)$ respectively. Inverting the Bloch function construction then gives a transformation between the real space Hamiltonians, expressing the bond-centred orbitals $\tilde{c}^\dagger_{n,A/B}$ in terms of the atom-centred orbitals $c^\dagger_{n,A/B}$ as:

$$\begin{aligned} \tilde{c}^\dagger_{n,A} &= \sum_m \int_{-\pi/a}^{+\pi/a} dk e^{-ik(n-m)a} \left[\mathbf{U}(k)\tilde{\mathbf{U}}(k)^\dagger\right]_{AA} c^\dagger_{m,A} + \left[\mathbf{U}(k)\tilde{\mathbf{U}}(k)^\dagger\right]_{BA} c^\dagger_{m,B} \\ \tilde{c}^\dagger_{n,B} &= \sum_m \int_{-\pi/a}^{+\pi/a} dk e^{-ik(n-m)a} \left[\mathbf{U}(k)\tilde{\mathbf{U}}(k)^\dagger\right]_{AB} c^\dagger_{m,A} + \left[\mathbf{U}(k)\tilde{\mathbf{U}}(k)^\dagger\right]_{BB} c^\dagger_{m,B} \end{aligned} \tag{A.14}$$

This transformation is a specific example of a Wannier function construction (see sec. A.2) for the Hamiltonian H_{Atom}. The matrix $\mathbf{U}(k)$ forms the Bloch eigenfunctions of $\mathbf{H}_{\text{Atom}}(k)$, and $\tilde{\mathbf{U}}^\dagger(k)$ forms new quasi-Bloch states out of the conduction and valence band Bloch eigenstates and can be identified as being the gauge matrix.

However, unlike in the usual construction of Wannier functions the gauge matrix cannot be chosen freely. Rather, it is mostly fixed by the form of the Hamiltonian that we want our new tight-binding basis to display. A small amount of gauge freedom still exists as the eigenvectors of $\mathbf{H}_{\text{Bond}}(k)$ making up $\tilde{\mathbf{U}}(k)$ are only fixed up to an arbitrary global phase. As such, the transformation is given by:

$$\mathbf{U}_{A-B}(k) = \mathbf{U}(k)\mathbf{\Theta}(k)\tilde{\mathbf{U}}^{\dagger}(k) \tag{A.15}$$

where $\mathbf{\Theta}(k)$ is a diagonal unitary matrix, that is, containing phase factors on the diagonal.

As is usual in the construction of Wannier functions, we want our new bond-centred orbitals to be exponentially localised. As discussed in sec. A.2, the requirement for exponential localisation are smooth and periodic quasi-Bloch functions. Applied to our case above we therefore require that the momentum space transformation $\mathbf{U}_{A-B}(k)$ is smooth and periodic across the whole Brillouin zone, a condition most easily (but not only) achieved by imposing the same condition on $\mathbf{U}(k)$, $\mathbf{\Theta}(k)$, and $\tilde{\mathbf{U}}^{\dagger}(k)$ individually.

The existence of smooth and periodic diagonalising matrices $\mathbf{U}(k)$ and $\tilde{\mathbf{U}}(k)$ is equivalent to the existence of smooth and periodic Bloch eigenfunctions of $\mathbf{H}_{\text{Atom}}(k)$ and $\mathbf{H}_{\text{Bond}}(k)$. As discussed in sec. A.2, the existence of smooth and periodic quasi-Bloch functions is guaranteed for any time-reversal system. However, only in the case of completely isolated bands, where quasi-Bloch states are also eigenfunctions, does this guarantee the existence of smooth and periodic Bloch-eigenstates. Both conditions, time-reversal symmetry and isolated bands, are fulfilled in our case and we expect to be able to construct the diagonalising matrices to be smooth and periodic.

Depending on the diagonalisation procedure used, many problems can occur in the explicit construction of smooth and periodic eigenfunctions. As an example, a "naive" diagonalisation of $\mathbf{H}_{\text{Atom}}(k)$ might yield eigenvectors given by:

$$\mathbf{v}^{c/v}_{\text{Atom}}(k) = \begin{pmatrix} \pm\sqrt{u + ve^{-ika}} \\ \sqrt{u + ve^{+ika}} \end{pmatrix} \tag{A.16}$$

While both conduction and valence bands $(+/-)$ are smooth and periodic eigenfunctions when $|u| > |v|$, when $|u| < |v|$ the complex argument of the square-root encircles the branch point of $\sqrt{z}$ at $z = 0$ as k varies across the Brillouin zone. This will break the periodicity due to a discontinuity at $ka = \pi \equiv -\pi$. Similarly a

possible set of eigenvectors of $\mathbf{H}_{\text{Bond}}(k)$ is given by:

$$\mathbf{v}^{c/v}_{\text{Bond}}(k) = \begin{pmatrix} u - v \pm \sqrt{u^2 + v^2 + 2uv\cos(ka)} \\ \sqrt{uv}(1 + e^{ika}) \end{pmatrix} \tag{A.17}$$

Conduction and valence bands $(+/-)$ are again smooth and periodic. However, depending on the relative magnitude of u and v, one of them becomes the zero vector at the zone-boundaries, and therefore normalising the above eigenvectors introduces a discontinuity at that point, breaking periodicity and/or smoothness.

An appropriate set of unnormalised smooth and periodic eigenvectors for both Hamiltonians is given by:

$$\mathbf{v}^{c/v}_{\text{Atom}}(k) = \begin{pmatrix} \pm\frac{\sqrt{u^2+v^2+2uv\cos(ka)}}{u+ve^{ika}} \\ 1 \end{pmatrix} \tag{A.18}$$

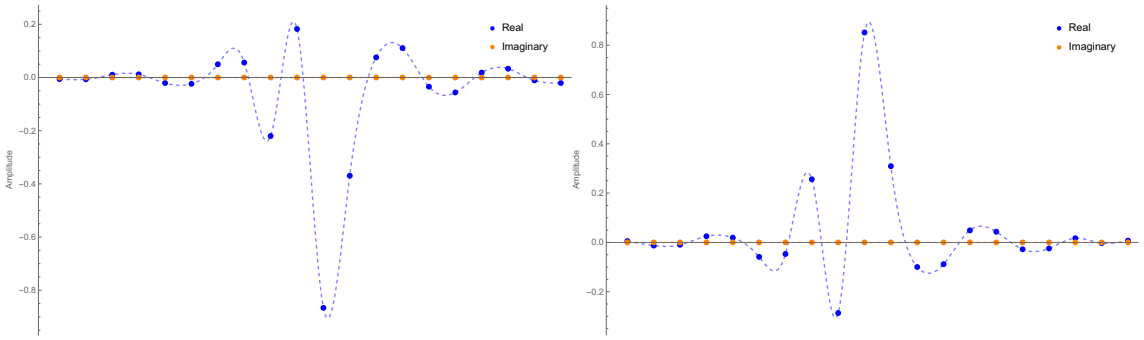
$$\mathbf{v}^{v}_{\text{Bond}}(k) = \begin{pmatrix} (u - v) - \sqrt{u^2 + v^2 + 2uv\cos(ka)} \\ \sqrt{uv}(1 + e^{ika}) \end{pmatrix} \tag{A.19}$$

$$\mathbf{v}^{c}_{\text{Bond}}(k) = \begin{pmatrix} \sqrt{uv}(1 + e^{-ika}) \\ -(u - v) + \sqrt{u^2 + v^2 + 2uv\cos(ka)} \end{pmatrix} \tag{A.20}$$

The above atom-centred eigenvectors $\mathbf{v}^{c/v}_{\text{Atom}}(k)$ can be normalised while maintaining smoothness and periodicity for any values of u and v. The bond-centred eigenvectors $\mathbf{v}^{c/v}_{\text{Bond}}(k)$, however, can only be normalised (are non-zero at all points ka) when $u < v$; therefore we will from now on restrict discussions to that case. Comparing with equ. (A.5) and assuming that both $u, v < 0$ the hopping is more negative between lattice sites in the same unit cell and as such this is the scenario in which the unit cell is centred on the double bond.

Using equ. (A.14) with $\mathbf{U} = (\mathbf{v}^{v}_{\text{Atom}}(k), \mathbf{v}^{c}_{\text{Atom}}(k))$ and $\tilde{\mathbf{U}} = (\mathbf{v}^{v}_{\text{Bond}}(k), \mathbf{v}^{c}_{\text{Bond}}(k))$ the new bond-centred eigenfunctions can be constructed numerically, examples of which are shown in fig. A.1. As constructed the Wannier function are nicely localised. They also show the oscillations typical for Wannier functions far from the central lattice sites, which ensure orthogonality between the Wannier functions centred on different sites. However, they do not show any nice symmetry properties. As discussed in sec. A.2 it is often desirable to construct Wannier functions which represent the symmetry operations of the underlying crystal. It is therefore reasonable to ask whether we can construct them to also have some additional symmetry.

The desired symmetry of the Wannier functions places symmetry conditions on the quasi-Bloch states they are constructed out of. These conditions are known as the induced representations of site-symmetry groups [190]. While for a general



(a) **(b)**

Figure A.1: Wannier functions constructed numerically according to equ. (A.14) using parameters $u = -4$ and $v = -3$. Depicted is the amplitude of the new basis states (a) $\tilde{c}_A^\dagger$ and (a) $\tilde{c}_B^\dagger$ on each of the original lattice sites $c_{A/B}^\dagger$. Interpolated lines are for visualisation purposes and do not represent any physical observable.

three dimensional crystal these conditions can become quite involved, in the case of a one dimensional model such as the SSH model they become exceedingly simple. In one dimension the only possible point symmetry beyond the identity is inversion symmetry. As such, the only two possible choices for symmetry adapted Wannier functions are either inversion symmetric pairs or inversion (anti-)symmetric functions. Constructing Wannier functions (anti-)symmetric under inversion requires:

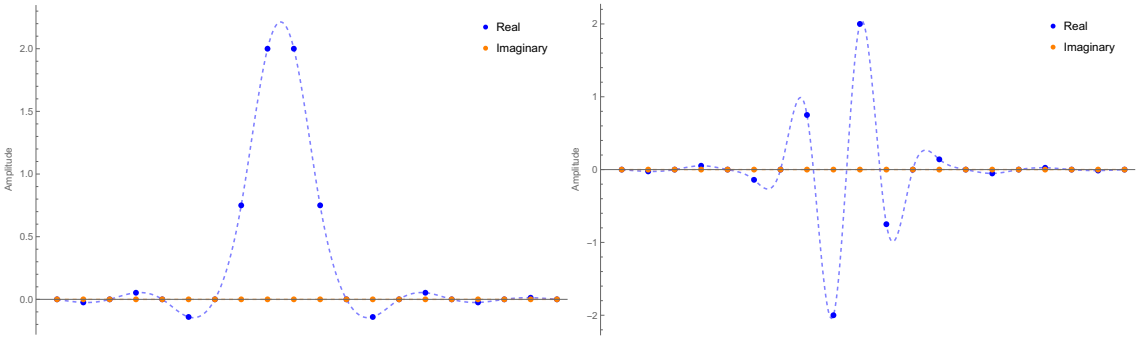
$$\hat{P}|w_{\pm,n}\rangle = \pm|w_{\pm,n}\rangle \rightarrow \hat{P}|\tilde{\psi}_{\pm,k}\rangle = \pm|\tilde{\psi}_{\pm,-k}\rangle \tag{A.21}$$

while constructing inversion symmetric Wannier function pairs instead requires:

$$\hat{P}|w_{1/2,n}\rangle = |w_{2/1,n}\rangle \rightarrow \hat{P}|\tilde{\psi}_{1/2,k}\rangle = |\tilde{\psi}_{2/1,-k}\rangle \tag{A.22}$$

While both symmetry and exponential localisation are desirable properties for Wannier functions, satisfying both is less than trivial. The existence of such Wannier functions has even recently been used as a definition for topological crystalline insulators [191]. However, in the special case of isolated bands in systems that are both time reversal symmetric and inversion symmetric, it is a well known result that the Wannier function belonging to the isolated band can be constructed to be both (anti-)symmetric and exponentially localised [192, 193]. These conditions are both fulfilled by the atom- and bond-centred Hamiltonians, equ. (A.5) and (A.6).

In the particular case of the atom-centred SSH model the inversion symmetry acts in real space as $\hat{P}c_{n,A}^\dagger = c_{-n,B}^\dagger$, or on the tight-binding Bloch basis of equ. (3.4) as $\hat{P}c_{k,A}^\dagger = c_{-k,B}^\dagger$. As a result the symmetry condition for the components of the



(a) **(b)**

Figure A.2: (a) Symmetric and (b) anti-symmetric Wannier functions constructed numerically using parameters $u = -4$ and $v = -3$. Depicted is the amplitude of the new basis states each of the original lattice sites. Interpolated lines are for visualisation purposes and do not represent any physical observable.

eigenvectors of $\mathbf{H}_{\text{Atom}}(k)$ is

$$\mathbf{v}_{\text{Atom,A}}(k) = \pm\mathbf{v}_{\text{Atom,B}}(-k) \tag{A.23}$$

For the case of $u < v$ with $u, v < 0$ an appropriate choice of (anti-)symmetric and smooth Bloch eigenfunctions is:

$$\mathbf{v}^{c/v}_{\text{Atom}}(k) = \begin{pmatrix} \pm\sqrt{-u - ve^{-ika}} \\ \sqrt{-u - ve^{+ika}} \end{pmatrix} \tag{A.24}$$

The corresponding real space Wannier functions are shown in fig. A.2.

However, in our particular Wannier function construction, equ. (A.14), the atom-centred Bloch eigenfunctions are followed by a gauge transformation determined by the form of the bond-centred Hamiltonian and its eigenfunctions. In order for a gauge transformation of already (anti-)symmetric Bloch eigenstates to construct new (anti-)symmetric quasi-Bloch functions, the gauge has to fulfil certain properties:

$$|\tilde{\psi}'_{\pm,k}\rangle = \alpha(k)|\tilde{\psi}_{+,k}\rangle + \beta(k)|\tilde{\psi}_{-,k}\rangle \tag{A.25}$$

Specifically, $\alpha(k)$ and $\beta(k)$ have to be even and odd (or odd and even) respectively to form new symmetric (anti-symmetric) wavefunctions.

Unfortunately, as can be seen from the form of the bond-centred eigenfunctions equ. (A.18) which constitute the gauge transformation $\tilde{\mathbf{U}}^\dagger = (\mathbf{v}^v_{\text{Bond}}(k), \mathbf{v}^c_{\text{Bond}}(k))^\dagger$, this conditions is not fulfilled. The identified gauge freedom in equ. (A.15) cannot be used to fulfil this condition either, as it is restricted to a diagonal matrix of phase factors. As a result it is not possible for our particular construction to yield bond-centred Wannier functions that are additionally also (anti-)symmetric under

inversion.

A.2 Wannier Functions

Wannier functions can be considered a basis transformation from momentum space Bloch functions to real space tight-binding functions, that is, back to a basis that explicitly shows the translational symmetry of the crystal $\langle \boldsymbol{r}|w_{\boldsymbol{0}}\rangle = \langle \boldsymbol{r}-\boldsymbol{R}|w_{\boldsymbol{R}}\rangle$. With the added condition that the tight-binding basis should be orthonormal, the Wannier functions associated to an isolated group of N bands, $|\psi_{n,\boldsymbol{k}}\rangle$ can be defined as [194]:

$$|w_{n,\boldsymbol{R}}\rangle = \int_{BZ} d\boldsymbol{k}\ e^{-i\boldsymbol{k}\cdot\boldsymbol{R}} U_{nm}(\boldsymbol{k})|\psi_{m,\boldsymbol{k}}\rangle = \int_{BZ} d\boldsymbol{k}\ e^{-i\boldsymbol{k}\cdot\boldsymbol{R}}|\tilde{\psi}_{m,\boldsymbol{k}}\rangle \tag{A.26}$$

where $n, m = 1, 2, \ldots, N$, $U_{nm}(k)$ is an $N \times N$ unitary matrix referred to as gauge (matrix), and $|\tilde{\psi}_{m,\boldsymbol{k}}\rangle$ are referred to as quasi-Bloch functions. For the case of a simple isolated band $U_{nm}(k) = U(k) = e^{i\phi(k)}$, which can be recognised as simply being the freedom of choosing the global phase of a wavefunction. The gauge freedom makes the Wannier function construction highly non-unique.

An important property of Wannier functions is their localisation. If the quasi-Bloch functions are chosen to be smooth and periodic functions of $\boldsymbol{k}$ across the whole Brillouin zone, then the Wannier functions calculated via the Fourier transform will be exponentially localised. However, due to the highly non-unique nature of the Wannier functions, problems concerning the existence and construction of exponentially localised ones have hindered their application for many decades. The introduction of a variational approach to Wannier functions that minimises a specific spread functional, the maximally localised Wannier function construction [195], has made wide-spread computational use possible. They have found many applications including interpolation of bandstructures [196], calculation of polarisation and magnetisation [197, 198], and allowing interpretation in terms of localised chemical bonding orbitals [195]. The question of the mathematical existence of exponentially localised Wannier functions in general bandstructures (any dimension up to three and any number of bands) has only recently been solved and has been shown to be intimately related to the topological properties of the bandstructure. The main result being that in time-reversal symmetric systems smooth and periodic quasi-Bloch states can be formed for any group of bands in any dimension $d \leq 3$ [199].

Oftentimes the symmetry properties of Wannier functions are also considered [190], specifically that they respect the symmetry of the underlying crystal, e.g. in centrosymmetric crystals they are even or odd under inversion or come in inversion

pairs. Symmetry adapted Wannier functions can be useful: for example they provide a more physical and intuitive picture, or explicitly retain the crystal symmetry during nearest neighbour approximations.

A.3 Interacting Electron Models

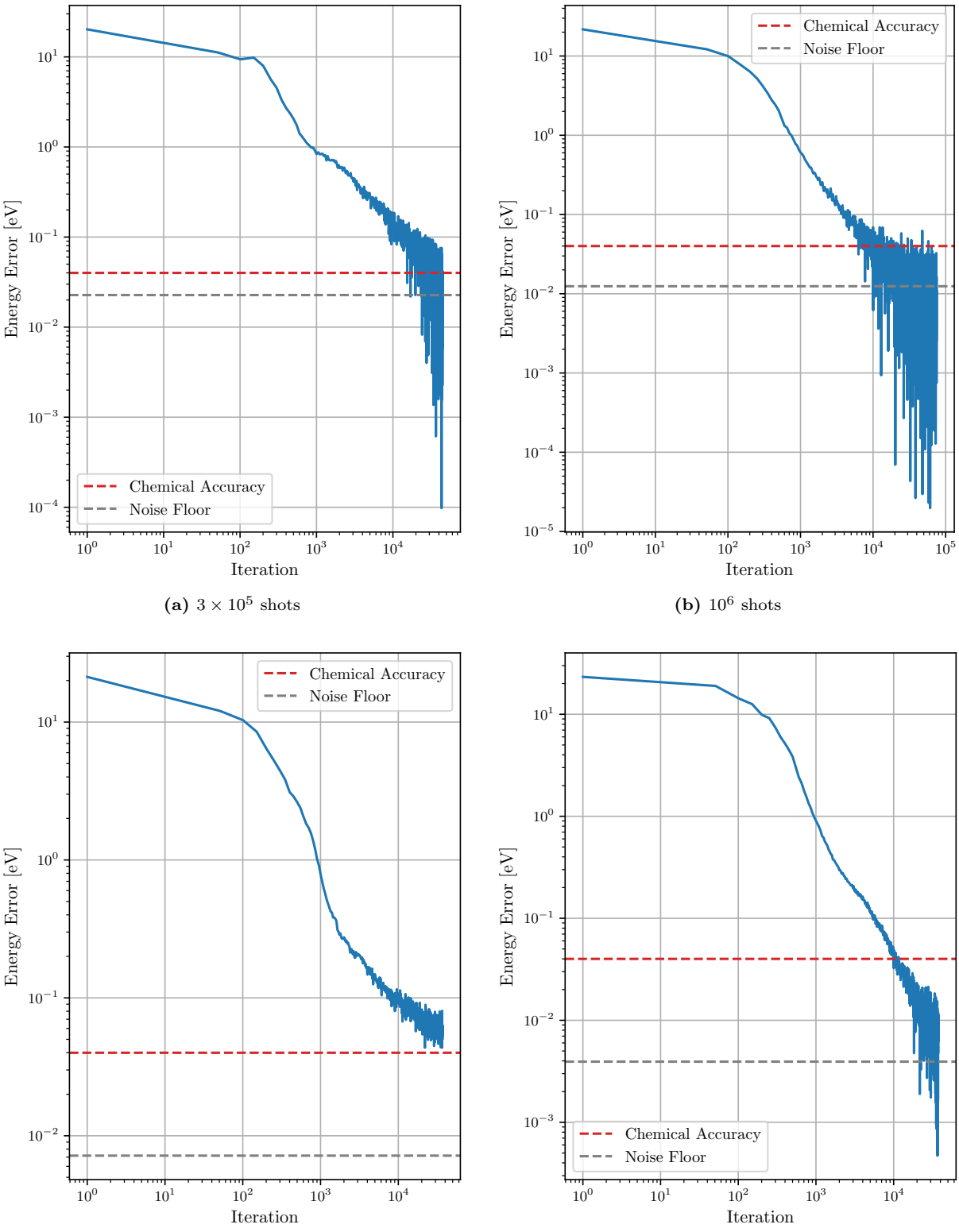
The basis transformation constructed in sec. A.1 transforms between the two independent electron Hamiltonians. If the full electronic interaction, equ. (A.1), were included in both interacting models, the basis transformation would still provide an exact equivalence between them. However, as exemplified in equ. (A.4), once the electronic interaction Hamiltonian is truncated in both models, exact equivalence is no longer guaranteed. The onsite U term in one model will can be mapped under a basis transformation $\tilde{c}^\dagger_{i,\sigma} = \sum_j C_{ij} c^\dagger_{i,\sigma}$ onto higher order electronic interaction terms in the other model:

$$H_{\text{e-e}} = \sum_i U \tilde{n}_{i,\uparrow} \tilde{n}_{i,\downarrow} = \sum_i U \tilde{c}^\dagger_{i,\uparrow} \tilde{c}^\dagger_{i,\downarrow} \tilde{c}_{i,\uparrow} \tilde{c}_{i,\downarrow} \tag{A.27}$$

$$= \sum_{i,j,k,l} U \left(\sum_p C_{pi} C_{pj} C^*_{pk} C^*_{pl} \right) c^\dagger_{i,\uparrow} c^\dagger_{j,\downarrow} c_{k,\uparrow} c_{l,\downarrow} \tag{A.28}$$

Appendix B

Shot Noise: Further Numerical Results



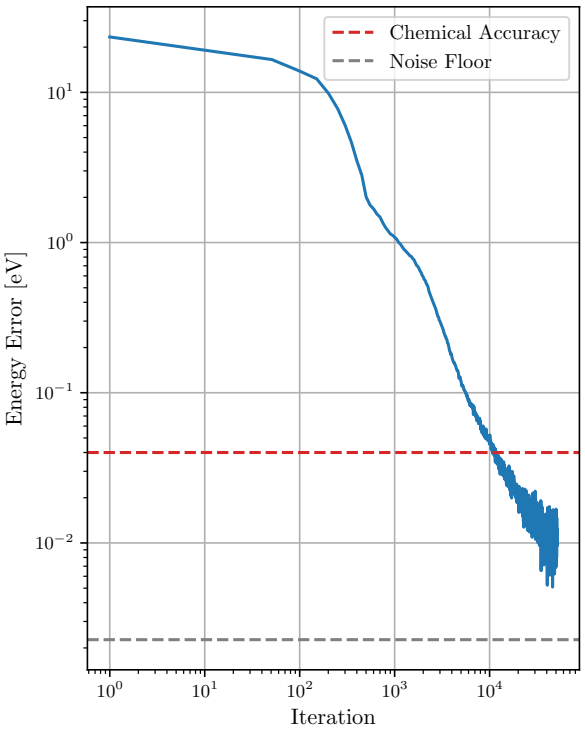
(a) 3×10^5 shots

(b) 10^6 shots

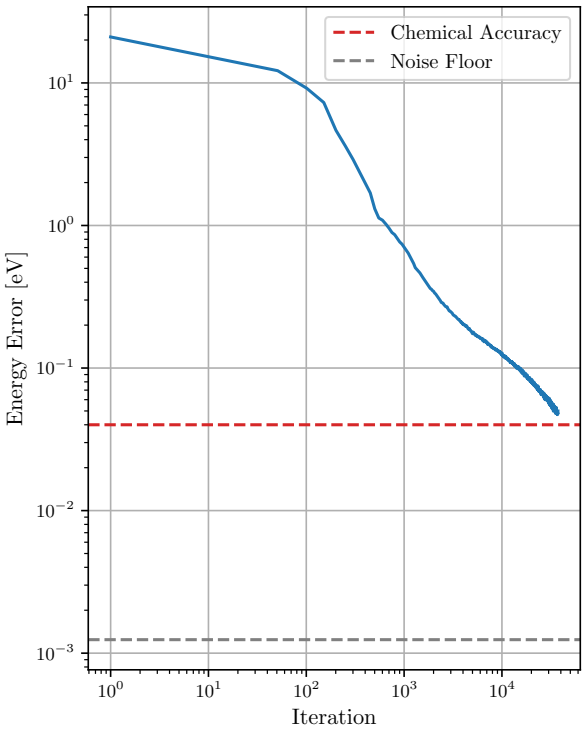
(c) 3×10^6 shots

(d) 10^7 shots

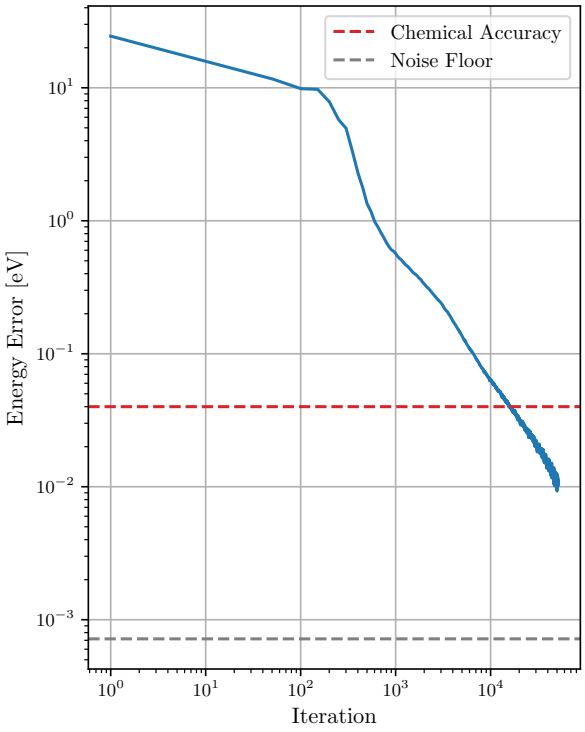
Figure B.1: Representative SPSA convergence traces of energy error, ΔE, for a selected VQE calculation at different shot budgets, M



(e) 3×10^7 shots

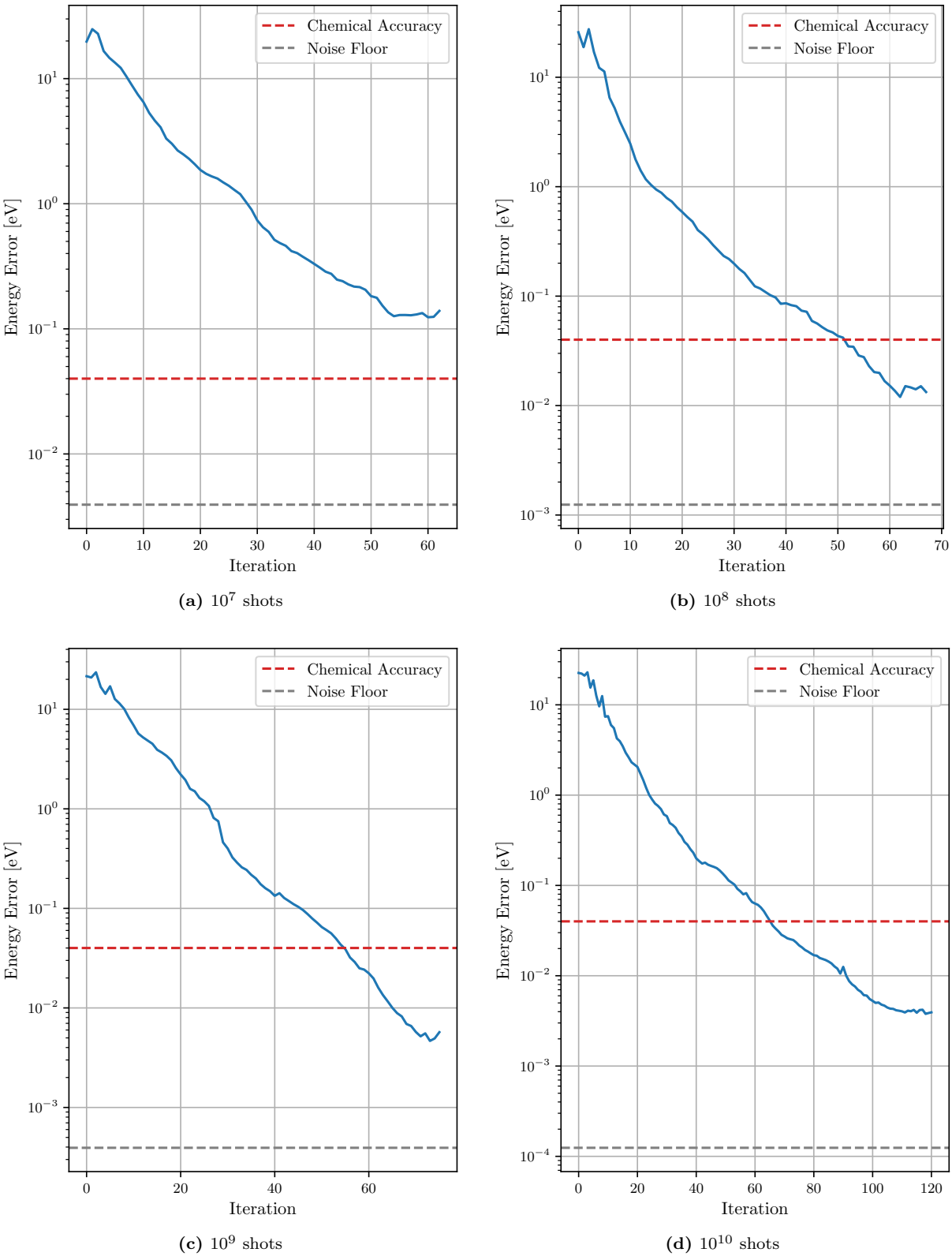


(f) 10^8 shots



(g) 3×10^8 shots

Figure B.1: Representative SPSA convergence traces of energy error, ΔE, for a selected VQE calculation at different shot budgets, M (continued).



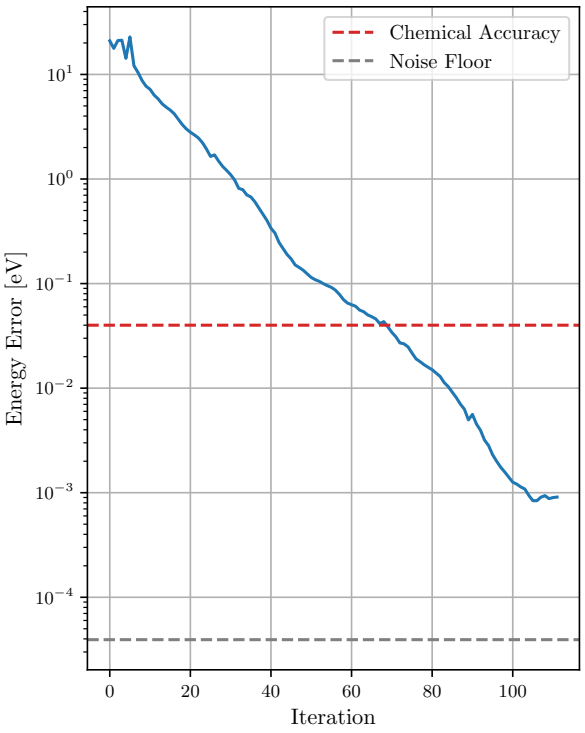
(a) 10^7 shots

(b) 10^8 shots

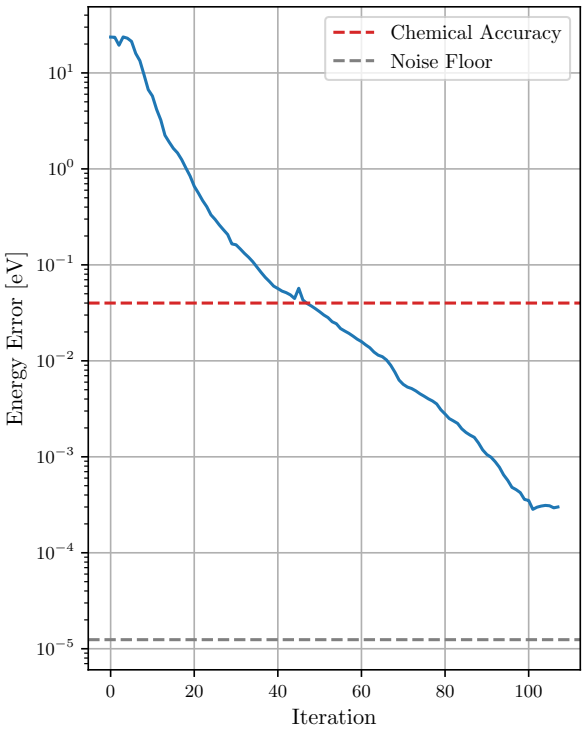
(c) 10^9 shots

(d) 10^{10} shots

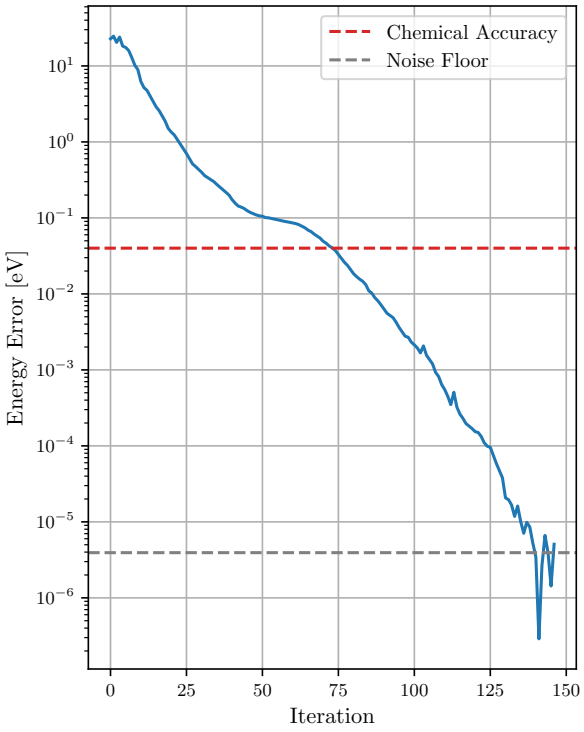
Figure B.2: Representative L-BFGS convergence traces of energy error, ΔE, for a selected VQE calculation at different shot budgets, M



(e) 10^{11} shots



(f) 10^{12} shots



(g) 10^{13} shots

Figure B.2: Representative L-BFGS convergence traces of energy error, ΔE, for a selected VQE calculation at different shot budgets, M (continued).