Methods for variational computation of molecular properties on near term quantum computers

Jules L. Tilly

A thesis presented for the degree of

Doctor of Philosophy

of

University College London

Department of Physics and Astronomy

University College London

22nd November, 2022

I, Jules Tilly confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

In this thesis we explore the near term applications of quantum computing to Quantum Chemistry problems, with a focus on electronic structure calculations. We begin by discussing the core subroutine of near-term quantum computing methods: the variational quantum eigensolver (VQE). By drawing upon the literature, we discuss the relevance of the method in computing electronic structure properties, compare it to alternative conventional or quantum methods and outline best practices. We then discuss the key limitations of this method, namely: the exploding number of measurements required, showing that parallelisation will be relevant for VQE to compete with conventional methods; the barren plateau problem; and the management of errors through error mitigation - we present a light touch error mitigation technique which is used to improve the results of experiments presented later in the thesis.

From this point, we propose three methods for near term applications of quantum computing, with a focus on limiting the requirements on quantum resources. The first two methods concern the computation of ground state energy. We adapt the conventional methods of complete active space self consistent field (CASSCF) and energy-weighted density matrix embedding theory (EwDMET) by integrating a VQE subroutine to compute the electronic wavefunctions from which reduced density matrices are sampled. These method

allow recovering additional electron correlation energy for a given number of qubits and are tested on quantum devices. The last method is focused on computing excited electronic states and uses techniques inspired from the generative adversarial machine learning literature. It is a fully variational method, which is shown to work on current quantum devices.

Acknowledgments

I want to thank foremost my supervisor Jonathan Tennyson, who has given me the freedom to pursue any direction I found interesting and promising, and has served me as a guide and mentor throughout my PhD.

I also want to thank George Booth and Ivan Rungger who have worked closely with me and have also taught me most of the quantum chemical and electronic structure theory upon which my thesis is built.

I would like to particularly thank Hongxiang Chen for the numerous insightful conversations and sharing of ideas. I further want to thank many of the key co-authors, collaborators and colleagues who have accompanied me throughout the years: Shuxiang Cao, Francois Jamet, Rachel Kerber, Glenn Jones, Cedric Weber, Robert Anderson, and P.V. Sriluckshmy.

I want to thank and acknowledge UCL for giving me the opportunity to pursue this PhD thesis, and acknowledge the support received from my industry CASE sponsor Rahko (now acquired by Odyssey Therapeutics).

Finally, I must express my very profound gratitude to my family for providing me with unfailing support and continuous encouragement throughout my years of study and the process of researching and writing this thesis.

Impact statement

The Variational Quantum Eigensolver (VQE) could significantly alter many fields of research. Condensed matter, material design drug discovery all relies on accurate and fast paced computation of quantum chemical properties. The fundamental limits of conventional computation means that numerous such problems become unsolvable beyond a certain size. From an academic research standpoint, the thesis provides a clear curriculum for study of VQE, its best practices as of 2022 and clear directions for impactful research question that will need to be addressed before VQE can be considered a viable method. In the context of the CASE sponsorship, the thesis resulted in two patents filed by my initial sponsor Rahko, contributing to its valuation and recent acquisition by Odyssey Therapeutics. A few years down the line, it is hoped that the research conducted will be used as a stepping stone for my sponsor's commercial research of novel oncology treatment.

Contents

Chapter 1

Introduction and overview

Quantum computing has undergone rapid development over recent years: from first conceptualisation in the 1980s [1, 2], and early proof of principles for hardware in the 2000s [3–14], quantum computers can now be built with hundreds of qubits [15–17]. While the technology remains in its infancy, the fast progress of quantum hardware and the massive financial investments all over the world have led many to assert that so-called Noisy-Intermediate Scale Quantum (NISQ) devices [18, 19] could outperform conventional computers in the near future [20–23]. NISQ devices are near-term quantum computers, with a limited number of qubits, and too few physical qubits to implement robust error correction schemes. Existing NISQ computers have already been shown to outperform conventional computers on a limited set of problems designed specifically to fit quantum computers' capabilities [20–22]. Algorithms running on these restricted devices may require only a small number of qubits, show some degree of noise resilience, and are often cast as hybrid algorithms, where some steps are performed on a quantum device and some on a conventional computer. In particular, the number of operations, or quantum gates, must

remain moderate, as the longer it takes to implement them, the more errors are introduced into the quantum state, and the more likely it is to decohere. Due to these restrictions, there are severe limits on the scope of algorithms that can be considered. In particular, well-known quantum algorithms such as Shor's algorithm [6, 24–30] for factoring prime numbers, or Grover's algorithm [31–36] for unstructured search problems, are not suitable for NISQ.

In this thesis, we explore the use of a NISQ method, the Variational Quantum Eigensolver (VQE), as a quantum subroutine for computation of molecular properties. The VQE was originally developed by Peruzzo *et al.* [37], and its theoritical framework was extended and formalised by McClean *et al.* in Ref. [38]. The VQE is among the most promising examples of NISQ algorithms. In its most general description, it aims to compute an upper bound for the ground-state energy of a Hamiltonian, which is generally the first step in computing the energetic properties of molecules and materials. The study of electronic structures is a critical application for quantum chemistry (for instance: [39–41]) and condensed matter physics (for instance: [42, 43]). The scope of VQE is therefore very wide-ranging, being potentially relevant for drug discovery [44, 45], material science [46] and chemical engineering [47]. Conventional computational chemistry, grounded in nearly a century of research, provides efficient methods to approximate such properties, but it becomes intractably expensive for very accurate calculations on increasingly large systems. This poses challenges in the practical application of such methods. One of the main reasons why computational chemistry methods can lack sufficient accuracy in molecular systems is an inadequate treatment of the correlations between constituent electrons. These interactions between electrons formally require computation that scales exponentially in the size of the system studied (i.e. the total time it takes to implement the computation is an

exponential function of the system size), rendering exact quantum chemistry methods in general intractable with conventional computing. This limitation is well studied in the literature addressing simulation of quantum computers on conventional computers, Ref. [48] provides an excellent example.

This bottleneck is the motivation for investigating methods such as the VQE, with the anticipation that these could one day outperform the conventional computing paradigm for these problems [49, 50]. In 1982, Richard Feynman theorised that simulating quantum systems would be most efficiently done by controlling and manipulating a different quantum system [2]. An array of qubits obey the laws of quantum mechanics, the same way an electronic wavefunction does. The superposition principle [51, 52] of quantum mechanics means that it can be exponentially costly to encode the equivalent information on conventional devices, while it only requires a linearly growing number of qubits. In the context of electronic structure theory [53–55] this is the appeal of quantum computing: it offers the possibility to model and manipulate quantum wavefunctions exactly, beyond what is possible with conventional computing. While largely dominated by electronic structure research, the VQE and its extensions have also been applied to several other quantum mechanical problems which face similar scaling issues. These notably include nuclear physics [56, 57] and nuclear structure problems [58, 59], high-energy physics [60–62], vibrational and vibronic spectroscopy [63–68], photochemical reaction properties predictions [69, 70], periodic systems [71–73], resolution of non-linear Schrödinger equations [74], and computation of quantum states of a Schwarzschild-de Sitter black hole or Kantowki-Sachs cosmology [75].

The VQE starts with an initialised qubit register. A quantum circuit is then applied to this register to model the physics and entanglement of the electronic wavefunction. Quantum circuit refers to a pre-defined series of quantum

operations that will be applied to the qubits [76]. The number of consequential operations in a circuit is referred to as depth. This circuit is defined by two parts: (1) a structure, given by a set of ordered quantum gates, often referred to as an 'ansatz'; and (2) a set of parameters that dictates the behavior of some of these quantum gates. Once the quantum circuit has been applied to the register, the state of the qubits is designed to model a trial wavefunction. The Hamiltonian of the system studied can be measured with respect to this wavefunction to estimate the energy. The VQE then works by variationally optimising the parameters of the ansatz in order to minimise this trial energy, constrained to always be higher than the exact ground state energy of the Hamiltonian by virtue of the variational principle [77–79]. For the VQE to be tractable, it is necessary that the number of quantum operations required to model the wavefunction is sufficiently low, imposing a relatively compact ansatz. The VQE admits wavefunction ansätze which cannot be efficiently simulated on conventional computers, indicating a possible advantage over conventional approaches if these quantum circuits are sufficiently accurate trial wavefunctions [37]. A first demonstration of the potential of these quantum ansätze was shown in Ref. [37], where an ansatz with polynomially-scaling depth in the size of the qubit register was constructed using principles grounded in conventional quantum chemistry (namely, the Unitary Coupled Cluster [37, 80, 81]). Since then, many alternatives have been proposed, with ansätze depth scaling as low as linearly [80] in the size of the qubit register. It must be understood however that a shallower ansatz, i.e. with fewer necessary quantum operations, will in general cover an overall smaller span of the space of all possible wavefunctions, and could result in lower accuracy of the ground state energy.

Despite an enormous amount of research being conducted on VQE, the

community remains divided on the potential of the method - several studies pointing out that the large number of measurements required to perform the method at scale renders it intractable [82–84]. These studies however do not discuss in depth how much best practices can affect the overall scaling of VQE, nor they take fully into consideration the potential for parallelisation of the method and its implications for hardware design. The aim of this thesis is to discuss the potential use of VQE as a subroutine for NISQ methods aiming at computing molecular properties. As such we focus on two research axes:

- **Research axis 1:** *Review of the current knowledge on VQE to determine best practices, potential for bringing quantum advantage and limitations of the method.*
- **Research axis 2:** *Development of methods for computation of relevant molecular properties that can be implemented and tested on the current generation of quantum computers.*

1.1 Contribution and synopsis of thesis

In Chapter 2, we provide an overview of the VQE, its relationship to alternative methods of energy computation for a molecular system. This overview is followed by a discussion of best practices drawn upon the literature (as outlined in Tilly *et al.* [85]).

In Chapter 3, we outline the key limitations of the VQE that remain to be overcome. The first one we discuss is the exploding number of measurements required to sample the wavefunction. Using the Chromium dimer as a case study, we discuss the possible runtime of a single iteration of VQE assuming best practices. From this assessment it is clear that large scale parallelisation

will be relevant for VQE to compete with conventional methods. The second limitation listed is the barren plateau problem, which implies that the gradients of parameters in the ansatz may vanish exponentially in given properties of the VQE problem. Finally, the third limitation is the ability to manage the impact of error on the optimisation and final result of VQE. After a brief introduction to error mitigation, we present a light-touch error mitigation method, that while not suitable for large scale applications, offers a convenient mean to improve results of current NISQ experiments at no computational cost.

Chapter 4 focuses on presenting two methods for recovering additional electron correlation energy at a given number of qubits, these are both adaptation of conventional quantum chemistry methods to quantum computing. These rely on VQE to resolve the energy of a correlated subspace of the wavefunction in an inner loop, while using conventional methods to update the Hamiltonian representation in an outer loop. In order to perform these method, one must sample the two-body reduced density matrices (RDM) from the resolved wavefunction on the correlated subspace. We therefore first present a study of RDM sampling on a quantum computer showing the extent to which finite sampling noise can affect the accuracy of the result. The first method is a quantum version of the complete active space self consistent field (CASSCF) which allow for improving upon the Hartree-Fock one- and two-body orbitals at each outer loop iteration, thereby providing a better representation of the Hamiltonian. The method is tested on a quantum device, shown to converge even with minimal error mitigation and used to correctly estimate a dipole moment. The second method is a quantum version of the Energy weighted Density Matrix Embedding Theory (EwDMET), and is tested on a quantum device to compute density of states for the Bethe lattice.

In Chapter 5, we present a NISQ method for computing molecular ex-

cited states. The focus of this method is to reduce the requirements on the complexity of the quantum computing operations compared to previously developed methods. Using techniques based on generative adversarial machine learning, we propose a method to generate states that would be orthogonal to the ground state, thereby allowing to progressively learn molecular excited states. The method, named discriminative VQE, is tested on several quantum devices (IBM, Honeywell, and Rigetti) and shown to work for small examples on current quantum computers.

1.2 Statement of authorship

In this thesis, I present work from the following research articles:

- Jules Tilly, Hongxiang Chen, Kanav Setia, Ying Li, Edward Grant, Leonard Wossnig, Ivan Rungger, George H. Booth, Jonathan Tennyson, The variational quantum eigensolver: a review of methods and best practices, *Physics Reports, 986, 1–128* (2022)

- Jules Tilly, PV Sriluckshmy, Akashkumar Patel, Enrico Fontana, Ivan Rungger, Edward Grant, Robert Anderson, Jonathan Tennyson, George H. Booth, Reduced density matrix sampling: Self-consistent embedding and multiscale electronic structure on current generation quantum computers, *Phys. Rev. Research 3, 033230* (2021)

- Jules Tilly, Glenn Jones, Hongxiang Chen, Leonard Wossnig, Edward Grant, Computation of molecular excited states on IBM quantum computers using a discriminative variational quantum eigensolver, *Phys. Rev. A 102, 062425* (2020)

Details of my personal contribution to these works are outlined in the document annexed at the end of the thesis: *Declaration of published work in thesis*. I have additionally contributed to the following research articles which are not included in this thesis:

- Jules Tilly, Ryan J. Marshmann, Anupam Mazumbar, Sougato Bose, Qudits for witnessing quantum-gravity-induced entanglement of masses under decoherence, *Phys. Rev. A 104, 052416* (2021)

- Hongxiang Chen, Max Nusspickel, Jules Tilly, George H. Booth, Variational quantum eigensolver for dynamic correlation functions *Phys. Rev. A 104, 032405* (2021)

- Martine Schut, Jules Tilly, Ryan J Marshman, Sougato Bose, Anupam Mazumdar, Improving resilience of quantum-gravity-induced entanglement of masses to decoherence using three superpositions *Phys. Rev. A 105, 032411* (2022)

Chapter 2

Variational quantum methods for quantum chemistry

In this Chapter, we provide a formal definition of the VQE, and an appreciation of where the algorithm is positioned compared to both conventional electronic structure, and other quantum computing methods. Later on, we suggest some best practices collected from the literature and a perspective on the overall resources that could be required for the VQE to achieve quantum advantage, as assessed in Tilly *et al.* [85].

2.1 A formal definition of the VQE

The VQE was first presented by Peruzzo *et al.* in Ref. [37] and its theoretical framework was significantly extended by McClean *et al.* in Ref. [38]. It is grounded in the variational principle (or the Rayleigh-Ritz functional [77–79]), which optimises an upper bound for the lowest possible expectation value of an observable with respect to a trial wavefunction. Namely, providing a Hamilto-

nian $\hat{H}$, and a trial wavefunction $|\psi\rangle$, the ground state energy associated with this Hamiltonian, E_0, is bounded by

$$E_0 \leqslant \frac{\langle\psi|\,\hat{H}\,|\psi\rangle}{\langle\psi|\psi\rangle}. \tag{2.1}$$

The objective of the VQE is therefore to find a parameterisation of $|\psi\rangle$, such that the expectation value of the Hamiltonian is minimised. This expectation value forms an upper bound for the ground state energy, and in an ideal case should be indistinguishable from it to the level of precision desired. In mathematical terms, we aim to find an approximation to the eigenvector $|\psi\rangle$ of the Hermitian operator $\hat{H}$ corresponding to the lowest eigenvalue, E_0.

In order to translate this minimisation task into a problem that can be executed on a quantum computer, one must start by defining a so-called ansatz wavefunction that can be implemented on a quantum device as a series of quantum gates. Given that we can only perform unitary operations or measurements on a quantum computer, we do this by using parameterised unitary operations. We hence express $|\psi\rangle$ as the application of a generic parameterised unitary $U(\boldsymbol{\theta})$ to an initial state for N qubits, with $\boldsymbol{\theta}$ denoting a set of parameters taking values in $(-\pi, \pi]$. The qubit register is generally initialised as $|0\rangle^{\otimes N}$, written as $|\mathbf{0}\rangle$ for simplicity, although low-depth operations can be performed for alternative initialisations before the unitary is applied. Noting that $|\psi\rangle$ (as well as any $U(\boldsymbol{\theta})\,|\psi\rangle$) is necessarily a normalised wavefunction, we can now write the VQE optimisation problem as

$$E_{\mathrm{VQE}} = \min_{\boldsymbol{\theta}} \langle\mathbf{0}|\,U^{\dagger}(\boldsymbol{\theta})\hat{H}U(\boldsymbol{\theta})\,|\mathbf{0}\rangle\,. \tag{2.2}$$

Eq. (2.2) is also referred to as the cost function of the VQE optimisation

problem, a terminology inherited from the machine learning and optimisation literature. We can continue this description by writing the Hamiltonian in a form that is directly measurable on a quantum computer, as a weighted sum of spin operators. Observables suitable for direct measurement on a quantum device are tensor products of spin operators (Pauli operators). We can define these as Pauli strings: $\hat{P}_a \in \{I, X, Y, Z\}^{\otimes N}$, with N the number of qubits used to model the wavefunction. The Hamiltonian can be rewritten as

$$\hat{H} = \sum_{a}^{\mathcal{P}} w_a \hat{P}_a, \tag{2.3}$$

with w_a a set of weights, and $\mathcal{P}$ the number of Pauli strings in the Hamiltonian. Eq. (2.2) becomes

$$E_{\text{VQE}} = \min_{\boldsymbol{\theta}} \sum_{a}^{\mathcal{P}} w_a \left\langle \mathbf{0} \right| U^\dagger(\boldsymbol{\theta}) \hat{P}_a U(\boldsymbol{\theta}) \left| \mathbf{0} \right\rangle, \tag{2.4}$$

where the hybrid nature of the VQE becomes clearly apparent: each term $E_{P_a} = \langle \mathbf{0} | U^\dagger(\boldsymbol{\theta}) \hat{P}_a U(\boldsymbol{\theta}) | \mathbf{0} \rangle$ corresponds to the expectation value of a Pauli string $\hat{P}_a$ and can be computed on a quantum device, while the summation and minimisation $E_{VQE} = \min_{\boldsymbol{\theta}} \sum_a w_a E_{P_a}$ is computed on a conventional computer.

2.2 The VQE pipeline

The VQE, as presented using Eq. (2.4), can be decomposed into a number of components, which all entail significant choices that impact the design and overall cost of the algorithm. We refer to the layering of these different components as the VQE pipeline. Most choices made on specific elements of this

pipeline have significant implications on the entire VQE process. We summarise the iterative process (and the main VQE loop) in Fig. 2.1 to provide a schematic description of the algorithm and its main components. We list the key components below and provide a brief introduction to each of them and how they fit together:

- **Hamiltonian construction and representation:** The first step in the VQE is to define the system for which we want to find the ground state. This can be an *ab initio* molecular Hamiltonian for electronic structure [86–89], a solid-state system [42, 90–92], a spin lattice model [93], a nuclear structure problem Hamiltonian [59], or a Hamiltonian describing any other quantum system. For each of these, one starts with a specific geometry (or conformation) of the system, specifying for example the distance between each atom, or the geometry of the lattice. Constructing the Hamiltonian involves finding specific operators and their weights between basis functions spanning the physical problem, where the basis functions represent the individual single-particle degrees of freedom. Given the Hamiltonian defines the quantum observable for the total energy associated with a wavefunction, the choice of basis is critical to define the space its spans. It can have a significant impact on the accuracy and cost of the final result, as the type of basis and number of basis functions chosen both determine the size of the computation required and the accuracy of the representation. In the case of electronic structure, these different representations could include, as examples, molecular orbitals from a prior mean-field calculation, plane-wave functions, or local atomic functions, all representing the spatial distributions (or 'orbitals') for the single-particle Fock states, from which

the many-body basis is formed [88]. Additional complexity arises when specifically looking at electrons: following the Pauli exclusion principle [94, 95] the wavefunction must be antisymmetric with respect to the exchange of two electrons. From a mathematical perspective, this means that we must decide whether we enforce this antisymmetry through the definition of the wavefunction or through the definition of the operators. These are referred to (for historical reasons) respectively as first and second quantisation [88]. In second quantisation the Hamiltonian is expressed in terms of fermionic operators, also known as creation ($\hat{a}_j^\dagger$) and annihilation ($\hat{a}_j$) operators. These correspond to the action of adding, or removing an electron from a given basis function with integer index j, respectively (e.g. an orbital or a lattice site), ensuring appropriate fermionic antisymmetry with respect to permutation of any two particles. (An introduction to Hamiltonian construction and discuss the wider implications of particular representation choices is presented in Sec. 3 of Tilly *et al.* [85]).

- **Encoding of operators:** Qubit registers on quantum computers can only measure observables expressed in a Pauli basis, due to the two-level nature of spins: $\hat{P}_a \in \{I, X, Y, Z\}^{\otimes N}$, for N qubits. In first quantisation, the operators can be directly translated into spin operators that can be measured on quantum computers [96], as they are not used to enforce antisymmetry of the wavefunction. In second quantisation the Hamiltonian is expressed as a linear combination of fermionic operators which are defined to obey this antisymmetry relationship, unlike Pauli operators. The role of a fermionic to spin encoding is therefore to construct observables, from Pauli operators, which maintain this relationship. A

transformation of fermionic operators to spin operators that meets this criterion was demonstrated a long time ago [97], and recent research has focused on improving on this initial work. The key factors determining the efficiency of an encoding are their Pauli weight (the maximum number of non-identity elements in a given spin operator), the number of qubits required, and the number of Pauli strings produced. (For a list of the most relevant encodings for second quantised Hamiltonians see Sec. 4 of Ref. [85]). It is worth noting that for certain ansatz choices, in particular those defined in terms of fermionic operators, the encoding can have significant implications on gate depth and trainability (Sec. 3.2). Cases of encoding particles others than fermions (e.g. bosons), and which do not require antisymmetry to be enforced are also possible and far simpler..

- **Measurement strategy and grouping:** The next step in the VQE pipeline is to determine how measurements are distributed and organised to efficiently extract the required expectation values from the trial wave function. In general, to achieve a precision of ϵ on the expectation value of an operator, we are required to perform $\mathcal{O}(1/\epsilon^2)$ repetitions (usually denoted as shots) of the circuit execution, each completed with a measurement at the end [38]. The objective of the measurement strategy is to make the number of repetitions as low as possible. Several techniques are available to achieve this, in particular, the use of efficient weighting of the number of measurements across the operators [98–100]. This can be further optimised by using properties of the Lie algebra in which Pauli strings are defined. *Via* processing of the encoded Pauli strings to be measured, it is possible to identify commuting groups of operators that

can be measured jointly, and subsequently find the measurement bases in which all operators of a given group can be simultaneously measured [101–103]. In order to perform this joint measurement, a short quantum circuit must therefore be designed and applied for each group, to rotate the measurement basis and to perform this joint measurement. Alternatively, because of information overlap between different Pauli strings, one can also try to reduce the number of measurements required using inference methods from fewer shots [104, 105] (See Sec. 5 of Ref. [85] for details).

- **Ansatz and state preparation:** Once the Hamiltonian has been prepared such that its expectation value can be measured on a quantum device, we can turn to the preparation of the trial wavefunction. In order to do this, one must decide on a structure for the parametrised quantum circuit, denoted as ansatz. It is used to produce the trial state, with which the Hamiltonian can be measured. Upon successful optimisation of the ansatz parameters, the trial state becomes a model for the ground state wavefunction of the system studied. A wide range of ansätze are possible, and the appropriate choice depends on the problem being addressed. The key aspects of the ansatz are its expressibility and trainability. The expressibility defines the ability of the ansatz to span a large class of states in the Hilbert space [106, 107], defining the maximum accuracy its approximation of relevant low-energy states can achieve (assuming all parameters can be perfectly optimised). Its trainability describes the practical ability of the ansatz to be optimised using techniques tractable on quantum devices [106, 108] (related to the total number of parameters, their linear dependence, the structure of the

optimisation surface, and to the concept of barren plateaus [109], which can arise when gradients almost vanish thereby preventing optimisation). A good ansatz must be sufficiently expressive to guarantee that it can appropriately approximate the ground state wave function, however, it must not be so expressive that it renders the search for the target state intractable. Another important aspect of the ansatz choice is the scaling and complexity of its circuit depth with system size. This is particularly important for near-term application of the VQE, as it determines in great part the noise resilience of the method employed. Details about ansatz selection are presented in Sec. 6 of Ref. [85].

- **Parameter optimisation:** The parameters of the ansatz used need to be updated iteratively until convergence. In general, this requires sampling the expectation value of the Hamiltonian several times for a given parameter set in the ansatz in order to define an update rule for the parameters (i.e. the updated value of the parameters is a function of the expectation value measured). The choice of optimisation algorithm is critical for at least three main reasons: (1) it directly impacts the number of measurements required to complete an optimisation step, as e.g. computing the numerical gradient of a quantum circuit can require value estimation of the Hamiltonian with respect to several slightly modified wave functions (this is also generally true for gradient-free methods) [110, 111]; (2) some optimisers have been designed to alleviate specific optimisation issues, such as the barren plateau problem [112–116]; (3) it directly impacts the number of iterations required to reach convergence (if it allows for convergence to be reached at all) [114–116].

- **Error mitigation:** Quantum noise is one of the main hurdles in the

viability of the VQE, given that the method is to be used without error correction schemes on NISQ devices. Error mitigation aims to reduce the impact of quantum noise through post-processing of the measurement data (or occasionally through post-processing of the trial wave function ahead of measurements). Error mitigation techniques vary widely in terms of cost and benefits (see Sec. 3.3), and in general, a mix of these can be implemented jointly to achieve the best balance.

It is worth briefly outlining the distinction between the VQE pipeline and that of other variational quantum algorithms (VQAs) [123]. The key distinguishing feature of VQE is that it is restricted to finding the eigenstate of a quantum observable, which is not necessarily the case of other VQAs (such as Quantum Approximation Optimisation Algorithms [124] or the Variational Quantum Linear Solver [125, 126]). As such, the process of encoding the Hamiltonian is specific to VQE. Similarly, while all VQAs would benefit from efficient measurements, the nature of the observable used in VQE (often scaling polynomially in the system size) mean that efficient grouping and measurements strategies will likely have a far greater impact on the overall scaling of the method.

A similar distinction can be drawn with the field of machine learning using Quantum Neural Networks (QNN) [127, 128]. While such methods can be considered as variational algorithms, the reverse is not necessarily true. Most VQAs, and the VQE, aim at finding the solution to a given problem from initial inputs. Machine learning on QNN aims at abstracting, and generalising [129], a pattern from already solved problems used as initial input. As such the algorithmic pipeline and challenges of both methods are largely different. In the case of QNN, one will be less concerned about the representation of the

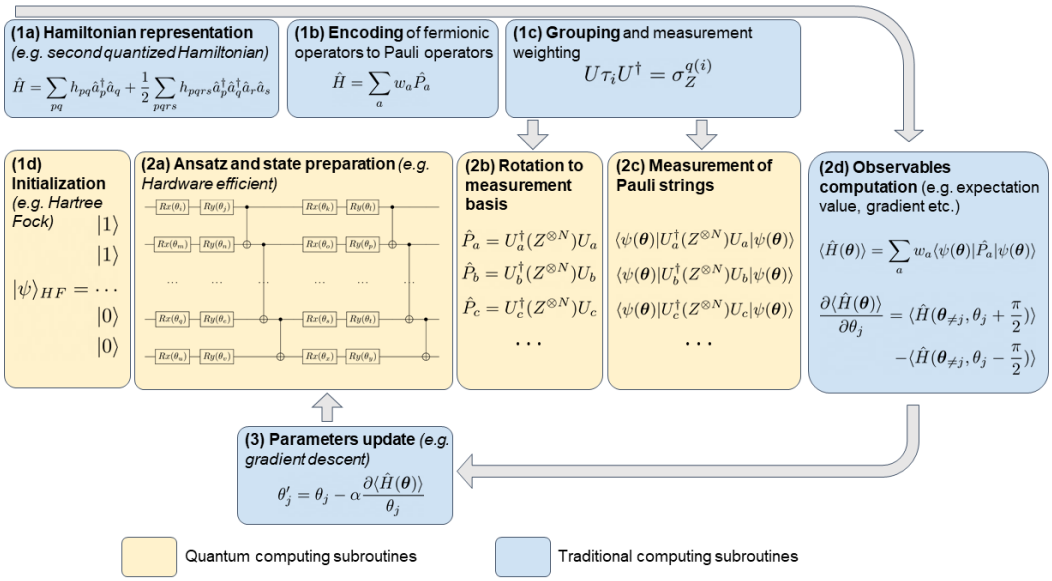


Figure 2.1: The VQE Pipeline - Formulas are illustrative and do not correspond to best practices. **(1) Pre-processing: (a) Hamiltonian representation:** A set of basis functions is defined for the Hamiltonian to be expressed as a quantum observable of the electronic wave function; **(b) Encoding:** The Hamiltonian is encoded into a set of operators that can be measured on a Quantum Computer, using the qubit register wavefunction. To do this, fermionic operators in the Hamiltonian are mapped to spin operators using an encoding; **(c) Grouping and measurement strategy:** Operators defined in (b) are grouped in order to be measured simultaneously later on, usually requiring an add-on to the quantum circuit for each group in order to rotate the measurement basis in a basis in which all operators in the group are diagonalised. It is also the step in which we decide the measurement weighting strategy; **(d) State initialisation:** Decide how the state to which the ansatz is applied is initialised. In general, the Hartree-Fock wavefunction is used [89, 117–120], but other options are also possible - **(2) The VQE loop: (a) Ansatz and trial state preparation:** Apply the ansatz to the initialised qubit register, before the first iteration of the VQE all the parameters of the ansatz also need to be initialised (randomly or using a specific method, e.g. Ref. [121, 122]); **(b) Basis rotation and measurement:** Once the trial wave function has been prepared, it must be rotated into the measurement basis of the operator of interest, or a diagonal basis of a specific group of Pauli strings **(c) Observable computation:** The expectation value to be computed depends on the optimisation strategy used, in any case however these are reconstituted by weighted summation on conventional hardware; **(d) Parameters update:** Compute and apply updates to the ansatz parameters and begin a new iteration of the VQE - **(3) Post-processing, Error mitigation:** error mitigation is a layer of additional computation on measurement output (or directly on the quantum state prior to measurement) aimed at reducing the impact of quantum noise on the results.

observable and poorly scaling measurement requirements. Instead, deciding on the process for encoding conventional data into a quantum state (often referred to as quantum feature map) [130, 131] will be critical in determining potential for quantum advantage and is completely absent from the VQE pipeline.

2.3 Advantage argument, assumptions, and limitations of the VQE

Quantum supremacy is achieved when algorithms running on quantum computers can produce results that surpass those generated on conventional computing resources in accuracy and/or resources required [132]. It was demonstrated on a tailored sampling task by several research teams [20–23], although the magnitude of the advantage has also been contested [133, 134]. Quantum advantage is in general used interchangeably with quantum supremacy, however in this work we use quantum advantage to refer to an instance of quantum supremacy where the advantage has relevant, tangible applications. This concept can rely on theoretical scaling arguments or practical demonstrations. A precise definition of the resources required and of the accuracy metrics is required for any specific demonstration of quantum advantage, which needs to include all overheads and initialisation requirements. Computing resources can be defined in many different ways, including overall absolute runtime, time scaling, memory requirements, or indeed 'indirect' metrics such as the financial cost of the computation or energy consumption (for a discussion on the energy consumption of quantum computers, we recommend Ref. [135]).

The VQE allows for the probabilistic measurement of observables over certain classes of parameterised approximate wavefunctions, which can neither

be sampled from nor have their properties computed efficiently (e.g. in polynomial time) on conventional devices as the system size gets large. Of course, this implies that the Hamiltonians studied can be written as a polynomially growing sum of independent observables [37], as is commonly found in a number of fields such as quantum chemistry, condensed matter physics, or nuclear physics. If these wavefunction forms, accessible via the VQE yet practically inaccessible via conventional means, admit sufficient accuracy in their approximation to the ground state, quantum advantage can be considered within this paradigm. The argument outlined above defines a necessary condition for the VQE to become a practically useful method for computing properties of quantum systems. It is clear from the literature, and outlined in Sec. 2.6, that under certain assumptions this condition is theoretically achievable [37, 80, 98, 136–138].

There are however many restrictions of quantum computing that this approach does not take into account, and we therefore propose two more stringent conditions. The first one is that VQE must demonstrate similar or higher accuracy than any conventional method, but with lower computational time-to-solution. This condition takes into account possible limitations due to hardware runtime, potentially resulting in a large pre-factor for VQE computations. In this thesis, the pre-factor refers to the multiplier applied to a scaling rule to obtain the actual run time of the method. If the VQE has better asymptotic scaling than conventional methods, but a large pre-factor, this means an advantage could only be achieved in the asymptotic regime of very large systems (with runtime possibly too large for VQE to be realistically usable). This would make it difficult to demonstrate quantum advantage for practical moderately sized systems. The second condition, which is also the most stringent form of quantum advantage for the VQE, is to achieve at least as good

accuracy, and with faster compute time, for a system of sufficient complexity to accurately model a real problem of physical and chemical relevance. This involves demonstrations on systems, where the approximation error in defining the specific Hamiltonian for the original problem is of smaller magnitude than its solution using the VQE. This could be as simple as ensuring that basis sets are sufficiently saturated [89], or that the complexity of the interactions with a wider system were sufficiently resolved. For instance, consider computing the energy of a series of protein-ligand complexes (for which methods extending VQE have already been proposed [139, 140]): even if the VQE achieves better accuracy in lower computation time, it is not guaranteed that these accuracy gains lead to a practical advantage. For example, the accuracy gains may still be insufficient to predict the most appropriate ligand in a physical experiment due to the approximations in the treatment of the environment in the Hamiltonian. Some researchers have attempted to estimate the tipping point for quantum computing-based quantum chemistry to overtake conventional methods. As one example, Elfving *et al.* [82] estimate the size of basis set (and hence the number of qubits) that would be required for a tangible quantum advantage of quantum computing based methods to lie somewhere between 19 and 34 molecular orbitals (or twice as many spin orbitals and hence twice as many qubits).

Despite sound theoretical arguments for the polynomial scaling of VQE [37, 38], a number of potential limitations have been identified as well, which could prevent the VQE from achieving quantum advantage:

- The VQE could be limited by a large pre-factor linked to the cost of accurate observable sampling. Several studies have analyzed the overall cost of VQE and whether it can reach a tipping point, at which it

becomes advantageous compared to conventional methods [82, 84, 98]. They have so far all concluded that given certain assumptions and the current state of research surrounding VQE functionality, the algorithm cannot outperform conventional methods within a remit of applications considered tractable. The main bottleneck identified in these studies (based on noiseless estimates) is the substantial number of measurements that are required to estimate the expectation value of the Hamiltonian using VQE (for further details see Sec. 3.1). The field of research is fast-moving however, and much research has been devoted to efficient operator sampling. Using the parallelisation potential of VQE (see Sec. 3.1.2) could also be a solution to this measurement problem but would require a paradigm shift in the way quantum hardware is conceived.

- The VQE involves solving an optimisation problem. As such, to understand the true cost of implementing VQE, one needs to assess the complexity of the optimisation process. The true cost and scaling are dependent on the optimiser and on the optimisation landscape of the specific problem studied. While some optimisers have been shown to converge in polynomial time for convex cost functions, the VQE is far from having such a favorable landscape [141, 142]. In fact, the VQE optimisation is shown to be NP-hard [141], which means that in the worst possible case, finding the optimal solution to the problem is intractable. Of course, this is to be expected as all optimisation problems can suffer from the same issue [143]. The key open question is to know whether VQE can be optimised heuristically in a polynomial number of iterations, and converge to an approximate yet accurate enough solution, for instances of practical relevance.

- Even if one could show VQE would theoretically converge in a tractable number of iterations, this would assume that expectation values and gradients are computed exactly. This assumption is not valid in the context of quantum computation, and it has been shown that the number of measurements required to accurately measure gradients could scale exponentially in certain parameterisations of systems, due to the barren plateau problem [109] (discussed in detail in Sec. 3.2). A number of mitigating methods have been proposed, such as the identity block initialisation [121] or the use of a local encoding for the Hamiltonian [108, 144]. Nevertheless, the extent to which barren plateaus can indeed be managed for VQE remains an open question.

- Related to both of the above, the extent to which VQE is resilient to quantum noise is also an open question, but actively mitigating errors will likely be unavoidable for relevant use of NISQ algorithms. Although error mitigation methods have shown great success in improving the accuracy of VQE on the current generations of quantum computers (for instance [145–150]), it can significantly increase the resource requirements of the algorithm. It is unclear whether this increase in resources is an acceptable cost or likely to be a critical limitation in larger-scale application of the VQE. A recent paper [151] is rather pessimistic on this point, showing that the increase in cost is exponential when the ansatz circuit grows deeper. Conversely, it was suggested in the early days of VQE that variational algorithms possess inherent noise resilience since the optimisation can effectively adapt to the noise [38]. This resilience has helped VQE to be more successful than other algorithms on the current generation of quantum devices, and has been numerically demonstrated

in small qubit simulations [152]. However, it remains unclear whether this resilience from noise can be retained in larger quantum experiments, where one is confronted with a more complex ansatz, with more noise coming from the difficulty of controlling large numbers of qubits with precision.

An important additional point to stress here is that while in theory the exact state could in principle be spanned by a number of qubits that scales linearly with system size, this exactness is in general forgone in VQE via the imposed parameterisation. At this point, a strictly exact limit within a defined Hamiltonian is only expected to be recovered with an exponential number of variational parameters (and hence circuit depth) [153]. Therefore, to achieve advantage, the classes of states accessible within the VQE framework must admit superior approximations to quantum many-body systems of interest compared to accessible conventional descriptions of quantum states, as well as their scaling with system size. The key question regarding the realm of current applicability of the VQE is therefore whether it can achieve higher accuracy on at least some representative systems, with some appropriate resource metric, compared to conventional computational chemistry methods.

2.4 VQE and conventional computational chemistry

The first step in any application of VQE to *ab initio* electronic structure is to define the basis functions determining the resolution and representation of the system. A common (but not required) approach to this would be to use the molecular orbitals obtained from a prior mean-field Hartree-Fock (for a

description of this method, see Ref. [89]) or density functional theory (DFT) calculation [154–157] (for comprehensive reviews of DFT, see Refs. [158, 159]). These orbitals are used to define the representation of the Hamiltonian, and thus compute the operator weights of the resulting Pauli strings. In this way, the VQE already relies on the techniques of conventional quantum chemistry for its use. Furthermore, in order to clarify the challenge for quantum advantage, as well as the expected scope and applicability of the VQE in the context of computational chemistry, we provide a very brief review of existing methods in this domain.

Although exceptions exist, it should be noted that most conventional approaches for high accuracy *ab initio* ground-state energetic properties of molecular systems rely on wavefunction approximations, in keeping with the wavefunction approximation inherent in the VQE approach [160]. Other quantum variables (such as densities, density matrices, or Green's functions) can be used, but are in general unable to reach state-of-the-art accuracy for ground state energies [161]. As such, methods like DFT which is widely used in material sciences, and offers a competitive cost-accuracy trade-off for large systems are not direct competitors to VQE, due to the lack of systematic improvability of their results. Despite some quantum algorithms for electronic structure presenting algorithms with scalings competitive or even lower than DFT (for example, Ref. [162]), the most likely competitors for short to medium term applications of VQE are accurate wavefunction approaches, which can scale as high polynomial or even exponential, but which still are able to access comparatively large system sizes.

2.4.1 Full Configuration Interaction

Full configuration interaction (FCI) provides the benchmark for exact representation of a quantum state for a given Hamiltonian and basis set [88, 163]. This results in the approach being in most cases intractable, with practical limits for *ab initio* systems currently being 18 orbitals [164]. However, its numerically exact treatment of the correlated physics ensures that it occupies a unique and important position in quantum chemistry and electronic structure. FCI builds the variationally optimal wavefunction as a linear superposition of all possible configurations of electrons within the available degrees of freedom. Whilst the inclusion of all possible configurations ensures that the final result is invariant to the precise single-particle representation of the orbitals considered, it is common to perform FCI in a basis of Hartree-Fock molecular orbitals to improve the convergence rate. The Hartree-Fock method provides the variationally optimised single Slater determinant, as appropriate for closed-shell systems [89], approximating the ground state wavefunction at the mean-field level. In this basis, the orbitals have individual single-particle energies associated with them, since they diagonalise the single-particle Fock matrix. The structure of the FCI wavefunction then takes the following form, where the configurations can be classed by the number of particle-hole excitations they create in the reference Hartree-Fock configuration, as

$$|\psi\rangle_{\mathrm{FCI}} = c_0 |\psi\rangle_{\mathrm{HF}} + \sum_{ia} c_{ia} \hat{a}_a^\dagger \hat{a}_i |\psi\rangle_{\mathrm{HF}} + \sum_{ijab} c_{ij,ab} \hat{a}_a^\dagger \hat{a}_b^\dagger \hat{a}_j \hat{a}_i |\psi\rangle_{\mathrm{HF}} + \ldots, \quad (2.5)$$

where the first sum represents 'singly-excited' configurations where an occupied spin-orbital, denoted by the indices $i, j, \ldots$ is depopulated, and a vir-

tual spin-orbital, denoted by the indices $a, b, \ldots$ is populated (here we use terminology corresponding to electronic structure theory, however these considerations are valid for any quantum system expressed in a finite basis). This (de)population is achieved while preserving antisymmetry of the overall wavefunction, via the use of the fermionic second quantised operators, $\hat{a}^{(\dagger)}$. These number preserving excitations from the reference Hartree-Fock determinant can be extended to double excitations (second sum) all the way up to m-fold excitations, where m is the number of electrons. This then spans the full space of configurations, and due to the linear parameterisation, ensures that the minimisation of the Ritz functional (Eq. 2.1) can be written as a diagonalisation of the full Hamiltonian in this basis [165, 166]. Exact excited states (within the defined basis and resulting Hamiltonian) can then also be computed as successively higher-lying eigenvalues of the Hamiltonian matrix in this basis.

While the FCI represents the 'ground truth' solution for the defined combination of Hamiltonian and basis set, the core aim of much of electronic structure is to truncate the complexity of this FCI solution (ideally to polynomial scaling with system size), while minimising the loss in accuracy resulting from this truncation [88]. It is also advantageous to have the ability to systematically relax this truncation of the approximate ansatz, allowing for improvable results when the situation demands (for instance, see the methods described in Ref. [160]). To this end, a large number of approximate parameterisations of the FCI wavefunction have been explored, which differ in their accuracy, scaling, functional form, and method of optimisation. Many of these approaches have enabled chemical accuracy and beyond to be routinely reached in systems far larger than those accessible by FCI [167–172]. The technical definition of chemical accuracy is specifically the accuracy required to compute accurate enthalpies (heats) of reactions, which numerically corresponds to a method

achieving an output within 1.6 milli-Hartree (mE_H) [173] of experimental results. It is widely used as a benchmark for numerical methods, although it is worth noting that other chemical properties need higher accuracy (sometimes by 2 to 3 orders of magnitude for instance in the case of spectroscopic properties). It is in general considered extremely difficult (or impossible) to reach for large systems due to the approximations which are made when constructing the model. Oftentimes computational methods aim at a correct qualitative description of the chemical properties instead. We refer to chemical precision as the benchmark for the precision at which the model is solved, irrelevant of the uncertainties and approximation made when constructing the model (see Ref. [82] for a thorough discussion of chemical accuracy vs. precision in the context of computational chemistry).

The considerations described in devising an effective parameterisation also largely echo the developments of ansätze for the VQE, although the functional forms of ansätze which admit efficient evaluation on quantum devices are different. In the next section, we explore a few of these parameterisations which are used on conventional devices, and how these considerations have influenced and transferred over to the choice of ansatz developed in the context of the VQE.

2.4.2 Efficient approximate wavefunction parameterisations for conventional computation

While the complexity of the exact FCI ansatz (Eq. 2.5) is clearly combinatorial with the number of degrees of freedom, many accurate and more compact wavefunction forms have been established. It results in more efficient approaches than FCI which can access larger system sizes, with only small tradeoffs in

accuracy. As an illustration of the capabilities of state-of-the-art methods, a recent study presents a blind test comparison of nine different methods applied to benzene on an active space of 30 electrons and 108 molecular orbitals [160]. The root mean square deviation between the results produced by these methods was only 1.3mE_h, demonstrating a consistent and reliable level of accuracy between these methods, expected to be close to FCI accuracy. A similar (albeit not blind) study was conducted in Ref. [161], with applications to transition metal systems, again showing excellent agreement between the most accurate wavefunction methodologies in these systems.

To rationalise some of these parameterisations, an obvious first approximation to Eq. (2.5) can be made via truncation in the total number of excitations from the reference configuration, allowing retention of the efficient linear form. The most common of these is the configuration interaction with singles and doubles ansatz (CISD), where only up to double excitations are retained [88]. More recent adaptive, selective or stochastic inclusion of desired configurations exploit the sparsity in the optimised amplitudes, and can extend the ansatz further in accuracy, resulting in methods such as Adaptive Sampling CI (ASCI) [169, 174–176], Semistochastic Heat-Bath CI (SHCI) [172, 177–182], or Full Configuration Interaction Quantum Monte Carlo [170, 183–186]. However, these truncated linear approximations can suffer from size-intensive total energies, where the energy does not scale appropriately with respect to the number of electrons, ensuring that the energy error per particle becomes increasingly large as systems grow in size [88]. Nevertheless, they can result in excellent variational approximations to FCI for small systems. An alternative approach is to construct a multi-linear approximation to the FCI wavefunction, which results in the matrix product state functional form. This form can be efficiently optimised within the density matrix renormalisation

group (DMRG), and can also yield accurate and systematically improvable approximations to FCI both in the case of *ab initio* molecular Hamiltonians and for lattice models [167, 168, 187–196]. More broadly, the development of tensor network theory [197, 198] and the use of matrix product states has resulted in significant improvements of methods for the resolution of lattice models (for instance [199, 200]). Finally, it is worth mentioning that a larger class of approximate wavefunction ansätze are able to be optimised within the framework of 'Variational Monte Carlo'. In these approaches, an approximate ansatz is chosen whose probability amplitude can be efficiently sampled at arbitrary electron configurations, but where a closed-form polynomial expression for the energy of the state is not accessible [201–203]. Within this criterion, the parameters of the ansatz can be optimised via Monte Carlo integration of the energy functional in a very general framework, albeit with the necessity of controlling for stochastic errors and care in optimisation of the parameters. Many of these considerations transfer through to the VQE.

Largely to correct for the size inconsistency in linear ansätze, the popular coupled-cluster ansatz truncates and then exponentiates the form of Eq. (2.5). This results in an appropriately sized consistent method, with an excellent accuracy vs. cost balance [204–206]. The coupled-cluster with single, double and perturbative triple excitations retained in the ansatz (known as CCSD(T)) is often referred to as the 'gold standard' of quantum chemistry where the correlations are not too strong [207], while other approximate coupled-cluster forms suitable for stronger correlation effects have also been developed (for recent examples, see Ref. [208, 209]). The coupled cluster is also the motivation of the use of the *unitary* coupled cluster ansatz of VQE [153], where a similar structure of exponentiated excitations based around a reference configuration is constructed [37, 80, 81, 210–212], with modifications to ensure efficient use

on a unitary set of quantum circuits. Similar considerations of dynamic inclusion of additional excitations are also possible with the ADAPT-VQE ansatz [213–215]. Furthermore, the Efficient Symmetry Preserving ansatz [216] looks to ensure the ability to systematically improve its span of the FCI description of Eq. (2.5) ensuring the preservation of symmetries inherent in its form. However, ensuring this systematic coverage of the FCI ansatz means that this form remains exponential in the system size in a number of realistic cases, meaning that true FCI may also be out of reach for the VQE.

An important consideration in the application of these approximate conventional parameterisations is that the size of the errors is different for different systems. Over time and use, an understanding has emerged from both theoretical and numerical analysis for the domain of applicability of these approaches and physical properties of the state which enable their accuracy, e.g. low-rank excitations (coupled cluster), locality (DMRG) or sparsity in the state (selected CI, FCIQMC). This understanding has promoted their effective use in appropriate circumstances, and stimulated further developments to improve their accuracy and scope. The analysis of the errors in different systems using VQE ansatz for quantum simulation, as well as the reasons underpinning or limiting their accuracy, is only starting to be performed, with more work necessary to fully classify and numerically investigate the approximations made in their form [153, 217].

Overall, these established wavefunction methods based on conventional computing (some of which are briefly described in this section) constitute the state of the art in high-accuracy quantum chemistry, at least for the ground state energetics. It should be stressed again that these approaches, as opposed to FCI, constitute the ultimate benchmark on which the success of VQE should be measured, as they represent approaches to systematically achieving chem-

ical accuracy but with a greater efficiency than exact FCI. This constitutes a demanding target for VQE to meet, with many decades of research in this area. Furthermore, continuing research for other parameterisations suitable for conventional devices, such as the rapidly emerging field of machine-learning inspired ansätze [203, 218–224], will continue to push the boundaries of accuracy on conventional devices to challenge the criteria for VQE superiority.

2.5 VQE and Quantum Phase Estimation

The Quantum Phase Estimation algorithm (QPE) [225–229] provides a method to find a given eigenvalue of a Hamiltonian from an approximated eigenstate (ground or excited). QPE can compute an eigenvalue to a desired level of precision with a probability proportional to how close the approximated eigenstate is to the true eigenstate [230]. It does so however using quantum circuits of depths that are far beyond what is achievable in the NISQ era of quantum computing [231]. As part of our discussion on the VQE we briefly outline QPE and how the two compare.

2.5.1 Overview of the quantum phase estimation

A representation of the quantum circuit used to implement QPE is presented in Fig. 2.2, and the process can be described as follows (adapting the descriptions in Refs. [76, 96]):

- The objective of QPE, like the objective of VQE, is essentially to compute an eigenvalue of a Hamiltonian. In the case of QPE however, the problem presented in Eq. (2.2) is slightly reformulated. For a given Hamiltonian $\hat{H}$, and a given eigenstate $|\lambda_j\rangle$ (usually the ground state: $|\lambda_0\rangle$), one tries

to find a eigenvalue E_j such that:

$$e^{i\hat{H}} |\lambda_j\rangle = e^{iE_j} |\lambda_j\rangle . \tag{2.6}$$

The Hamiltonian is exponentiated to obtain a unitary operator, and without loss of generality we can write $e^{iE_j} = e^{2\pi i\theta_j}$, with θ_j the 'phase' QPE aims at discovering.

- The only inputs available however are the Hamiltonian, and an approximation of the ground state $|\psi_0\rangle \sim |\lambda_0\rangle$, which can be generally expressed in the eigenbasis of the Hamiltonian as

$$|\psi_0\rangle = \sum_{j=0}^{2^{N_q}-1} \alpha_j |\lambda_j\rangle , \tag{2.7}$$

where N_q is the number of qubits used to represent the electronic wavefunction of the Hamiltonian (which therefore has a total of 2^{N_q} eigenstates).

- A register of ancilla qubits is used to map the eigenvalue sought-after, in general to a binary number. The number of ancillas required therefore depends on the type of implementation and desired precision (more ancillas mean a longer binary string, and therefore a higher precision [76]). This ancilla register is initialised as an equally weighted superposition of all possible state in the computational basis (all possible binary strings). If we have a total of N_a ancilla qubits, there are 2^{N_a} such basis elements. Starting from a register in state $|0\rangle^{\otimes N_a}$, a Hadamard gate (Had) is applied to each qubit. We recall that $\text{Had}\,|0\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$, and $\text{Had}\,|1\rangle = (|0\rangle - |1\rangle)/\sqrt{2}$. After these operations, the state of the ancilla

register is

$$|\psi_{\text{anc}}\rangle = \frac{1}{\left(\sqrt{2}\right)^{N_a}} \sum_{x=0}^{2^{N_a}-1} |\text{bin}(x)\rangle\,, \tag{2.8}$$

where x represents integers from 0 to $2_a^N - 1$ and $\text{bin}(x)$ the binary representation of of x. When both the ground state approximation and the ancilla register are considered together we get the total state of the qubit register $|\psi_{\text{tot}}\rangle = |\psi_{\text{anc}}\rangle \otimes |\psi_0\rangle$ such that

$$\begin{aligned} |\psi_{\text{tot}}\rangle &= \left(\frac{1}{\left(\sqrt{2}\right)^{N_a}} \sum_{x=0}^{2^{N_a}-1} |\text{bin}(x)\rangle \right) \otimes \left(\sum_{j=0}^{2^{N_q}-1} \alpha_j \, |\lambda_j\rangle \right) \\ &= \frac{1}{\left(\sqrt{2}\right)^{N_a}} \sum_{j=0}^{2^{N_q}-1} \sum_{x=0}^{2^{N_a}-1} \alpha_j \, |\text{bin}(x)\rangle \otimes |\lambda_j\rangle\,. \end{aligned} \tag{2.9}$$

- In the superposition state above, there is no clear relation between an ancilla state $|\text{bin}(x)\rangle$ and the j-th eigenstate $|\lambda_j\rangle$. That is, if one were to measure the ancillas resulting in a binary number $\text{bin}(x)$, no information could be gained about the state of the wavefunction register which encodes $|\lambda_j\rangle$, and therefore no information can be gained about the associated eigenvalues. In the following, we will apply unitary gates to this superposition state such that there is a clear one to one correspondence between a measured binary number $\text{bin}(x)$ and the eigenvector $|\lambda_j\rangle$. Consider the following unitary

$$U^{(k)} = \left(e^{i\hat{H}}\right)^{2^k}, \tag{2.10}$$

with k an arbitrary number for the time being. Following Eq. (2.6), if this unitary is applied to eigenstate $|\lambda_j\rangle$, it effectively results in a phase $e^{(2\pi i\theta_j 2^k)}$. Now suppose that k is the index of the ancilla qubits, i.e.

$k \in [0, N_a - 1]$ and that, for each k, $U^{(k)}$ is applied to the ground state approximation only if the ancilla qubit of index k is in state $|1\rangle$. This operation can be performed by mean of a controlled unitary operation, which applies a unitary operation subject to the value of a control qubit [76]. For a superposition instance $|\text{bin}(x)\rangle$ of the ancilla register, this means that the unitary U is applied x times in total to the ground state (consider for example the ancilla superposition $|\text{bin}(5)\rangle = |101\rangle$, here qubits are indexed from right to left to correspond to binary strings. The unitary is applied for $k = 0$, and for $k = 2$, hence following Eq. (2.10) it is applied $5 = 1 \cdot 2^2 + 1 \cdot 2^0$ times). We obtain the state

$$|\psi_{\text{tot}}\rangle = \frac{1}{\left(\sqrt{2}\right)^{N_a}} \sum_{j=0}^{2^{N_q}-1} \sum_{x=0}^{2^{N_a}-1} e^{2\pi i \theta_j x} \alpha_j \, |\text{bin}(x)\rangle \, |\lambda_j\rangle \, . \tag{2.11}$$

- The next step is to transfer information from each superposition instance's phase to the basis state by applying an inverse quantum Fourier transform (QFT) to the ancilla register. QFT is a transformation from the computation basis to the Fourier basis, mapping a single computational basis element $|\text{bin}(y)\rangle$ to a superposition of all computational basis elements each with different relative phase (due to the periodicity of the phase, each relative phase is a different point on the 2π period, with a total of 2^{N_a} different points)

$$\text{QFT} \, |\text{bin}(y)\rangle = \sum_{y=0}^{2^{N_a}-1} e^{2\pi i (xy/2^{N_a})} \, |\text{bin}(x)\rangle \, . \tag{2.12}$$

If we set $y = 2^{N_a}\theta_i$, we can observe that applying the inverse QFT to

the ancilla register in Eq.(2.11) results in

$$(\mathrm{QFT}^{-1} \otimes I^{\otimes N_q}) \left|\psi_{\mathrm{tot}}\right\rangle = \sum_{j=0}^{2^{N_q}-1} \alpha_j \left|\mathrm{bin}(2^{N_a}\theta_j)\right\rangle \left|\lambda_j\right\rangle , \tag{2.13}$$

where for simplicity we have assumed $2^{N_a}\theta_i \in \mathbb{N}$.

- Measuring the ancilla register in the Z basis returns the binary string $\mathrm{bin}(2^{N_a}\theta_i)$ with probability $|\alpha_j|^2$, from which θ_i, and E_i can be recovered easily. The complete qubit register then collapses to the state $|\mathrm{bin}(2^{N_a}\theta_i\rangle \left|\lambda_j\right\rangle$.

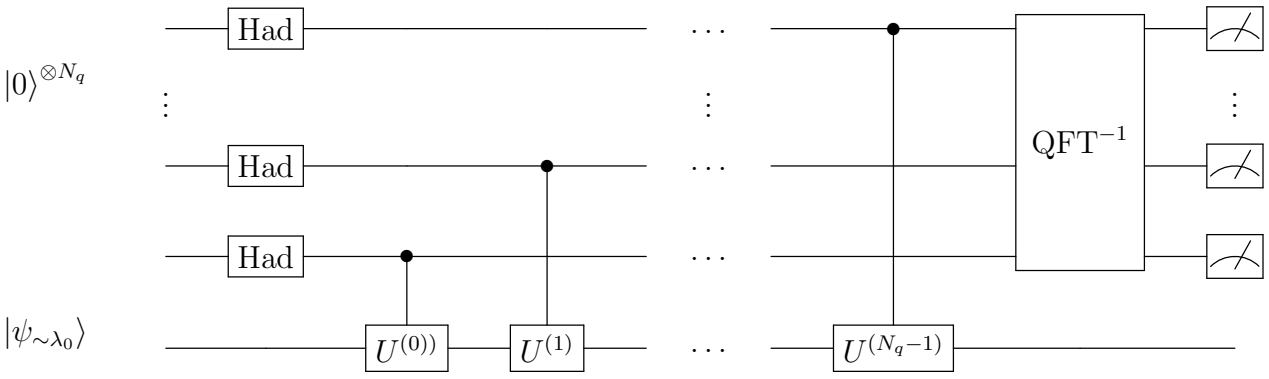


Figure 2.2: Quantum circuit for Quantum Phase Estimation

2.5.2 Discussion and comparison

Due to the ground state being measured directly in binary form in QPE, the number of ancilla qubits required is directly related to the precision ϵ targeted. Ancilla qubits provide one bit of information each, and as such, their number scales $\mathcal{O}(1/\log_2(\epsilon))$ in precision. The number of controlled unitaries is doubled for each ancilla qubit, and therefore scales $\mathcal{O}(1/\epsilon)$. These unitaries are effectively representing the action of the Hamiltonian on a state. The core

component of QPE is therefore efficient Hamiltonian simulation (for an overview of relevant methods see Ref. [47, 232]). Provided a non-restricted pool of qubits, Babbush *et al.* [233] show that by engineering the correct Hamiltonian representation (namely, a plane wave basis in first quantisation), one can achieve sub-linear scaling in the number of basis elements for Hamiltonian simulation.

As mentioned above, QPE is only likely to succeed if the fidelity of the input state to the unknown target eigenstate approaches one. In turn, this implies that using a randomised state as input is not an option as its expected fidelity to the target eigenstate approaches zero exponentially in the system size, resulting in QPE becoming exponentially costly with imperfect input state preparation [230]. Numerous methods have indeed been proposed to prepare a good enough approximation of the target eigenstate in a tractable manner, often grounded in conventional quantum chemistry (for example [234–238]) or in adiabatic quantum computation (for example [229, 239, 240]).

There have been a number of successful implementations of QPE on quantum devices [137, 241–246]. These have only been on small systems, as large scale implementations require quantum resources which are not currently available. In particular, large scale controlled unitaries, required for QPE, cannot be reliably implemented on NISQ devices. This is also the case for the inverse QFT. Several numerical studies have been performed to assess the complete cost of implementing QPE on relevant systems, and estimate runtime on a fault-tolerant quantum computer. The problem of nitrogen fixation has become a *de facto* benchmark for this algorithm [247]. Reiher *et al.* [231] estimate that the 54 electrons, 108 spin orbitals of FeMo-co would require over $\mathcal{O}(10^{15})$ T gates, 200 millions qubits and would need to run for over a month to obtain quantitatively accurate results (assuming 100 ns gate times and error threshold of 10^{-3}).

Berry *et al.* rely on qubitisation (a method which aims at transforming the evolution operator into a quantum walk) [248] to reduce the gate requirements to $\mathcal{O}(10^{11})$ Toffoli gates, despite an extended active space [249]. Lee *et al.* [250] further improve on these results and estimate they could perform this energy computation with four million physical qubits and under four days of runtime, with a similar $\mathcal{O}(10^{-3})$ error threshold. There have been many resource estimates performed for condensed matter models (for instance [251, 252]), with estimates as low as $\sim 500,000$ physical qubits running for a few hours to solve a 100 site version of the Fermi-Hubbard model. Finally Elfving *et al.* [82] estimate that with similar error rates, the chromium dimer (Cr_2) with an active space of 52 spin orbitals and 26 electrons, would require $\mathcal{O}(10^7)$ physical qubits running for about 110 hours. Research has progressed rapidly, and despite estimated runtimes and hardware requirements which remain daunting, offers a promising outlook for QPE, at least on targeted quantum chemistry tasks (examples of which are suggested in Ref. [82]).

The VQE trades off the depth and number of qubits required under QPE with a higher number of measurements and repetitions of the circuit, as well as the constraints of an approximate ansatz for the state. As presented in Ref. [253], QPE requires $\mathcal{O}(1)$ repetitions with circuit depth scaling $\mathcal{O}(1/\epsilon)$ in precision ϵ, VQE requires $\mathcal{O}(1/\epsilon^2)$ shots with circuit depth scaling $\mathcal{O}(1)$ in precision. While many other factors affect the overall time scaling of both methods, this point illustrates the asymptotic efficiency of QPE compared to VQE assuming access to fault-tolerant quantum computers, but also the resource efficiency of VQE over QPE for NISQ-era devices. The frontier between NISQ and fault-tolerant quantum computation is blurry, and as pointed out by Wang *et al.* [253] so is the frontier between VQE and QPE. They present an interpolation between the two algorithms, labeled Accelerated VQE (or α-

VQE), which uses smaller scale QPE calculations as sub-routines for the VQE. This method introduces a parameter $\alpha \in [0, 1]$ which allows tuning of the circuit depth $\mathcal{O}(1/\epsilon^{\alpha})$ and number of samples $\mathcal{O}(1/\epsilon^{2(1-\alpha)})$ (one recovers the QPE scaling if $\alpha = 1$, and the VQE scaling if $\alpha = 0$). In general, rather than being mutually exclusive methods for solving an electronic structure problem, VQE and QPE are likely to provide the most benefit when combined as complementary approaches, offering algorithmic flexibility that can be adjusted depending on the progress of quantum hardware.

It is worth noting that there exist other methods which attempt to reduce the computing requirements of QPE and bring quantum eigensolvers closer to being suitable to near-term devices. For example, Somma [254] proposes a method that identifies eigenvalues of a given Hamiltonian in a time polynomial in ϵ but only requiring a single ancilla qubit and without needing to implement the inverse QFT.

2.6 Some suggested best practices for VQE and their scaling assessment

In this section, we focus on combining compatible methods throughout the VQE pipeline, which offer the most promising scaling without compromising excessively on accuracy. The definition of a series of best practices for the VQE may suffer from many pitfalls since there remain many open research questions that affect the choice of optimal methods. It is also worth noting that it is likely that a method that is optimal for one system is not for another, and that this optimal compromise will change as quantum hardware improves. With this in mind, we provide some suggestions for best practices on current

devices for two broad families of systems. In particular, we can distinguish between lattice models [42, 90, 92, 93] and *ab initio* molecular systems [86–89]. These two categories usually require different encodings, measurement strategies, and ansätze. Table 2.1 summarises the most promising VQE methods that we have identified, together with their scaling. The key distinctive factor separating *ab initio* molecular systems and lattice models is that the former makes no assumption on the range and type of interaction between the fermionic modes (beyond it being a two-body interaction), while the latter usually has a simplified and parameterised form which often only connects fermionic modes following a nearest-neighbor lattice structure and/or features a lower effective rank of interactions.

As noted in the introduction, while the majority of the literature on VQE relates to electronic structure computation and lattice models, other applications have been proposed. Proposing best practices for these alternative applications is challenging as the research is sparse and therefore we avoid discussing these in this section.

Table 2.1: Summary of state of the art methods identified for the VQE for both *ab initio* molecular systems and lattice models. These methods and scalings are indicative only, as there remain a number of uncertainties with respect to their behavior on large scale systems and in noisy environments.

Task	***ab initio* Molecular systems**	**Lattice models**
Hamiltonian construction	**Second quantisation**	**Second quantisation**
	$\mathcal{O}(n^4)$ Hamiltonian terms and $N = \mathcal{O}(n)$ qubits, with n number of basis functions	*idem*
	First quantisation could be advantageous on some systems, but further research is needed	*idem*

Continued on next page

Table 2.1 – continued from the previous page

Task	*ab initio* Molecular system	Lattice models
Fermion to spin encoding	**Ternary tree Encoding** [255] Number of operators: $\mathcal{O}(N^4)$, Pauli weight $\mathcal{O}(\log_3(2N))$, Low weight encodings could result in more resilience to barren plateau [108, 144, 257], and more compact ansätze, though there is at least some suggestions that it may increase impact of quantum noise [258]	**Generalised Superfast Encoding** [256] Qubit number: $N = \mathcal{O}(nd/2)$; for a regular lattice, number of operators scales $\mathcal{O}(ND)$, where D is the lattice dimension; Pauli Weight: $\mathcal{O}(\log_2(d))$, with d the fermionic-interaction graph maximum degree *idem*, the Compact encoding [259, 260] results in a lower number of qubits for $D = 2, 3$ but has not yet been generalised to higher dimensions.
Grouping and measurement strategy	**Decomposed interactions** [103, 261] Operators to measure reduced to $\mathcal{O}(N)$; additional basis rotation circuit depth $\mathcal{O}(N/2)$ Full rank optimisation (in particular its extensions) [261] seem to achieve better overall measurement reduction for a given precision ϵ than the basis rotation method [103], but cost scaling remains unclear. Classical shadows [105] have been shown in at least one case [267] to outperform the scaling of decomposed interactions, though further numerical studies will be required to demonstrate dominance.	**Qubit-wise commutation** [38, 99, 101, 102, 147, 262–266] Operators to measure reduced by a scalar, Additional basis rotation circuit depth $\mathcal{O}(1)$ QWC grouping benefits from low Pauli weight encoding and comes at virtually no cost
Ansatz	**k-UpCCGSD** [80]	**Hamiltonian variational ansatz (HVA)** [98, 138]

Continued on next page

Table 2.1 – continued from the previous page		
Task	***ab initio* Molecular system**	**Lattice models**
	Circuit depth of $\mathcal{O}(kN)$, number of parameters $\mathcal{O}(kN^2)$	Circuit depth and number of parameters: $\mathcal{O}(k\tilde{C})$, with $\tilde{C}$ the number of commutative groups in the Hamiltonian (at most $\mathcal{O}(ND)$ for a regular lattice)
	Promising scaling, and good accuracy [80, 217] but uncertainty remains for applications on large highly correlated systems. Uncertainty around k, the number of repetitions required. Adaptive ansätze [213, 268, 269] may perform better, but their scaling requires more investigation.	HVA has shown resilience to barren plateau and efficacy on lattice models [138]. k is the number of repetition of the anstaz required to reach the desired accuracy.
Optimiser	**Trigonometric computation of optimal parameters** (e.g. independently discovered under different names: Sequential optimisation, coordinate descent, RotoSolve [114, 115, 270, 271] or Fraxis [272, 273])	**Simultaneous perturbation stochastic approximation (SPSA) [274, 275]**
	Requires sampling three values for each parameter at each step	SPSA has the advantage of requiring significantly less measurements to perform an optimisation step while likely requiring more iterations.
	Some indication of faster convergence [115] including against SPSA [114], but does not allow for full potential for parallelisation of VQE, and requires more values to sample than most optimisers.	SPSA has been shown to outperform trigonometric resolution methods on instances of the Hubbard model [276], and other gradient-based optimisers on Hubbard and hydrogen chains [277, 278].
Continued on next page		

Table 2.1 – continued from the previous page		
Task	***ab initio* Molecular system**	**Lattice models**
Error mitigation strategy	Despite a number of error mitigation methods having been proposed and tested [279], none has been shown to be outperforming others so far (see Sec. 3.3.2 for a brief discussion on the main methods). A fair comparison of the performance and cost between different error mitigation methods requires further research.	*idem*

2.6.1 Best practices for *ab initio* electronic structure of molecular systems

Hamiltonian construction: In the case of an *ab initio* molecular system, the Hamiltonian representing its electronic energy landscape is initially defined by a series of atoms and the spatial coordinates of their nuclei. The first choice to make is the basis in which the Hamiltonian is expressed. This directly impacts the number of qubits required for the implementation of VQE, which is proportional to the number of basis functions. Since the number of qubits is a limited resource in NISQ, we recommend using a molecular orbital basis, as it is in general more compact for a given target accuracy (compared to, for example, atomic or plane wave bases). Once a basis is decided upon, we must choose whether the Hamiltonian is prepared in first quantisation (antisymmetry maintained by the wavefunction) or second quantisation (antisymmetry maintained by the operators). The number of qubits in first quantisation scales

as $\mathcal{O}(m \log_2(n))$ [226, 280], with m the number of electrons and n the number of basis functions, against $\mathcal{O}(n)$ in second quantisation [97]. The former also requires additional depth to enforce the antisymmetry of the wavefunction. There has not been, to the best of my knowledge, a rigorous study of the efficiency of using first quantisation in VQE. While the scaling for first quantisation could be advantageous in systems with few electrons and large basis sets (e.g. if a plane wave basis is used [233]), second quantisation is generally preferred.

Encoding: The next decision to take is that of the mapping used to transform the fermionic, second quantised Hamiltonian into a weighted sum of Pauli operators. The most relevant encodings for *ab initio* molecular system include Jordan-Wigner [97], the Parity [281], Bravyi-Kitaev [281–283], and ternary tree mappings [255]. Out of these, the most promising encoding is the ternary tree mapping [255], since asymptotically it has the lowest Pauli weight (maximum number of non-identity Pauli operators in the string), resulting in lower circuit depth and possibly higher resilience to the barren plateau problem [108, 144]. It is however still unclear whether this lower circuit depth does indeed result in more noise resilience, as pointed out in Ref. [258]. The resulting qubit Hamiltonian can also be further reduced using tapering off methods based on symmetries [284–286].

Measurement strategy: The large number of measurements required to sample the numerous terms in the Hamiltonian is often cited as the most detrimental bottleneck of VQE [82, 84, 98]. Deciding on an efficient strategy for grouping and measuring Hamiltonian terms can go a long way in reducing this bottleneck. The decomposed interactions methods [103, 261] provide on

balance the most promising means to jointly measure the Hamiltonian. They allow measuring an entire molecular Hamiltonian with $\mathcal{O}(N)$ groups and require additional circuit depth of $\mathcal{O}(N)$ to perform the necessary basis rotation, which is usually acceptable since this scaling is equivalent or better than for most ansätze. While it was shown that methods using grouping based on general commutativity of Pauli strings (e.g. [102, 287]) require fewer shots to achieve a given accuracy [261] in some numerical studies (in particular when using the Sorted Insertion heuristic [288]), this reduction will likely not be worth the additional circuit depth scaling $\mathcal{O}(N^2)$ [101] required to perform the joint measurements. It is also worth noting that among the decomposed interactions methods, the Variance-estimate Greedy Full Rank Optimisation [261] appears to perform best, although it requires minimisation search of decomposition parameters. While this cost could be tractable there has been, to the best of my knowledge, no thorough research on how it would scale on large systems. For this reason, the Basis Rotation Group methods [103], which have a predictable cost, is a more cautious choice currently. For additional efficiencies and variance reduction, one can distribute shots according to the weights of each group in the Hamiltonian [98–100]. It is worth noting however that the $\mathcal{O}(N)$ scaling in number of groups is not an ideal proxy for the scaling in number of measurements required to achieve a given precision on observable estimation. This is due to possible covariances arising from the joint measurement of different operators [38]. Classical shadows [105] is also a promising method for reducing measurement count in VQE and has been shown in one study to have a better asymptotic scaling than decomposed interactions [267]. Further numerical studies will be required to establish to true performance of classical shadows compared to grouping methods.

Ansatz: We now have to decide on an ansatz to model the electronic wavefunction on the qubit register. Deciding on an ansatz remains challenging because it is often unclear which is expressive and efficient enough to allow for a good approximation of the ground state. The ansatz with the best scaling, and some evidence for appropriate accuracy for the ground state representation [217], is the Unitary paired Generalised Coupled Cluster Singles and Doubles (UpCCGSD) ansatz [80]. This ansatz scales linearly with the number of qubit $\mathcal{O}(kN)$, but may require to be repeated k times to reach the desired accuracy. The scaling of required repetition of the ansatz k has been partially studied [80] but remains uncertain for large systems. This ansatz also has the advantage of only needing a fairly low number of parameters ($\mathcal{O}(kN^2)$). Adaptive ansätze (such as ADAPT-VQE [213], iterative Qubit Coupled Cluster [289] and Cluster VQE [290]) are also promising, as they may provide resilience against barren plateaus. Their main drawback is that these adaptative methods come at the cost of selecting an operator to grow the ansatz (or Hamiltonian) and the need to fully re-optimise the ansatz at each iteration. Numerical studies have suggested that additional measurements may be required compared to fixed structure ansätze [291], although further research is required to provide an exhaustive costs and benefits analysis. That said, these methods have been shown to remain resilient to barren plateaus [215].

Optimiser: It is challenging to systematically compare different optimisers since no thorough large scale studies of their convergence rate have been conducted. For the time being, the sequential optimisation / coordinate descent / Rotosolve optimiser (for simplicity, the last is used in the remainder of the thesis) has been shown to converge significantly faster than several gradient-based optimisers [114, 115, 270, 271]. It offers the advantage of not relying

on any meta-parameters (such as a learning rate), which makes it a very easy optimiser to implement. However, Rotosolve presents two significant caveats: firstly, each iteration requires sampling three different values instead of two for most gradient-based methods (one can avoid this overhead by finding one of the values from the optimisation of the previous parameter, but this could result in correlated noise); secondly, parameters must be updated sequentially, thereby restricting the scope for parallelisation of the VQE. The Fraxis method [272, 273] works in a similar manner and has been shown in some numerical studies to perform at least as good as Rotosolve or even outperform it. For the time being, given there are currently no optimisers that have been shown to have superior convergence rates, and given it is unlikely that there will be a sufficient number of quantum computers to fully exploit the parallelisation potential of the VQE in the NISQ era (see Sec. 3.1.2), we propose the Rotosolve / Fraxis optimisers over other alternatives (although it is worth noting that the Quantum Natural Gradient [116] has been shown to perform well and to benefit from resilience to barren plateaus [292, 293], albeit at a significant cost [116]).

Scaling: Based on the discussion above, we can now construct a scaling estimate for a single iteration of the state-of-the-art VQE for *ab initio* molecular systems. The overall scaling is expressed in terms of the number of quantum gate time steps that must be performed (i.e. several gates applied on disjoint sets of qubits can be implemented within the same time step) over all the circuit repetitions required to perform a complete iteration. The computation of the expectation value of a single operator at a precision ϵ requires $\mathcal{O}(1/\epsilon^2)$ repetitions of the ansatz. In principle, ϵ should aim for chemical precision, generally accepted as 1.6 $\mathrm{mE_H} \approx 10^{-3}\mathrm{E_H}$. However, it is worth noting that, in practice

gradients may become lower than chemical precision (due to the barren plateau problem for instance, described in Sec. 3.2). In this situation, estimating gradients may require a more precise ϵ and more measurements, but it also means that optimisation may rapidly become impossible. If the k-UpCCGSD ansatz is chosen, this scales as $\mathcal{O}(kN)$, while choosing to use the decomposed interactions requires $\mathcal{O}(N)$ different operators to be measured (and therefore a gate depth of $\mathcal{O}(N)$ for rotation to the joint measurement basis) resulting in a total scaling for a single estimation of the entire Hamiltonian of $\mathcal{O}(kN^2/\epsilon^2)$.

There are $\mathcal{O}(kN^2)$ parameters in the k-UpCCGSD ansatz, hence this represents the cost scaling of updating each parameter using the Rotosolve optimiser. As this optimiser is not parallelizable, one may prefer to use a different method if sufficient sets of qubits are available. Overall, this gives us a total scaling for one iteration of the VQE of $\mathcal{O}(k^2N^4/\epsilon^2)$ without parallelisation, and $\mathcal{O}(kN)$ with full parallelisation (the circuit depth). This perfect parallelisation would require $\mathcal{O}(kN^3/\epsilon^2)$ sets of $\mathcal{O}(N)$ qubits. Note that while qubits within one set need to be entangled for the course of a single measurement, there is no requirement for entanglement between qubits of different sets of parallel quantum compute nodes. The sets of qubits can therefore be either all within on one quantum computer, or else also distributed across different separated quantum computers (see Sec. 3.1.2). It should be noted that the precision ϵ is generally required to achieve chemical precision. However, if a barren plateau occurs, ϵ may need to be reduced by orders of magnitude to compute gradients accurately enough to achieve a satisfactory optimisation.

So far we have only considered the scaling of one iteration. It is still an open research question how the number of iterations required to achieve convergence scales with system size for the VQE. This depends on numerous factors, including the ansatz, the optimiser used, and the system studied.

One important point to note is that convergence tends to be rapid at the beginning of the optimisation process, with large gradients that require only a low number of shots to be computed accurately enough to progress. It becomes more challenging close to the optimum, where gradients are smaller, requiring a larger number of shots to continue the optimisation. As such the last few iterations of the VQE are likely orders of magnitude more expensive than the rest of the optimisation, if the algorithm is implemented efficiently.

There are other overheads that may be worth consideration in the initial setup of the system Hamiltonian. The computation of the Hamiltonian matrix elements generally has a polynomial scaling, while naive implementations of Hartree-Fock scale $\mathcal{O}(n^4)$ [294], with n the number of basis functions, and it can be reasonably assumed that $n = N$ for *ab initio* molecular systems. Similarly, applying a decomposed interactions method to diagonalise operators and reduce measurements requires rewriting the Hamiltonian in a different basis [103, 261]. However, these costs only occur once at the beginning of the VQE process, and are unlikely to be a bottleneck. Despite possibly higher scaling than that of a VQE iteration, they are likely to have a significantly lower pre-factor (as implemented on conventional hardware), and as such are not likely to constrain the application of the algorithm except far in the asymptotic realm. However, less investigated is that these joint measurement bases may result in covariance between measurements of different Hamiltonian terms, thereby requiring additional measurements [38, 103, 261, 288] which could significantly affect overall cost for the VQE.

2.6.2 Best practices for lattice models

Our suggestions for lattice models differ from their *ab initio* counterparts. Lattice models for the most part only include terms in the Hamiltonian between nearest-neighbors on their respective lattice, with interactions between more distant sites significantly truncated in range. In particular, this limited degree of connectivity of lattice models provides the option to construct mappings with much lower Pauli weight, enabling more compact ansätze to be efficient, though this can come at the cost of additional qubits.

Hamiltonian construction: In the case of a lattice model, the Hamiltonian is given as a small number of empirical parameters, requiring no prior computation of matrix elements. These models are generally most naturally and efficiently expressed in terms of fermionic operators in the site representation, where the locality of the interaction can be exploited to reduce the number of measurements.

Encoding: There is significant literature on the fermion-to-spin mapping for lattice models. These mappings are in general designed to minimise the required Pauli weight of the operators for a given lattice structure. The most important property of a lattice is the maximum degree of connectivity (coordination) of the sites, denoted d. For instance, a square lattice has $d = 4$, and an equivalent hypercubic lattice of dimension D has $d = 2D$. If one is limited by the number of qubits available, the most appropriate mapping for a lattice is an adaptation of the Bravyi-Kitaev mapping (based on Fenwick trees) [295], and which has a Pauli weight scaling as $\mathcal{O}(\log(v))$, where v is the minimum number of sites in any one dimension for a $D = 2$ lattice. It has the advantage of reducing the Pauli weight of the operators produced, compared

to a naive implementation of Bravyi-Kitaev on a lattice, while maintaining the number of qubits required equal to the number of sites, n. If however the number of qubits is not a hard constraint, the Generalised Superfast Encoding method [256] provides a lower Pauli weight scaling of $\mathcal{O}(\log_2(d))$ at the cost of requiring an increased number of qubits for each site, with an overall scaling of $\mathcal{O}(nd/2)$ qubits. The Compact encoding [259, 260] requires a lower number of qubits ($\sim 1.5n$), but has not yet been generalised for regular lattices of more than three dimensions. Beyond their relationship to the resilience to barren plateaus in the optimisation [108, 144, 257], the relevance of the Pauli weight in the context of VQE is also in how it affects the choice of ansatz, and in particular whether the ansatz is initially expressed in terms of fermionic operators. If the chosen ansatz is not dependent on fermionic terms, then Bravyi-Kiteav or Jordan-Wigner mappings are preferred. Furthermore, the number of qubits required to represent the Hamiltonian can be reduced using the tapering off methods based on symmetries as described Refs. [284, 285].

Grouping and measurement strategy: The number of operators in lattice models scales, in general, with the number of edges of the lattice graph, E. For example, for a hypercubic lattice of dimension D, the number of edges, and therefore, the number of operators will scale $\mathcal{O}(nD)$ (though it is worth noting that the pre-factor to this scaling may change significantly depending on the encoding used). Because it is in general possible to reach low Pauli-weight encodings for lattice models, qubit-wise commutativity (QWC) grouping [38, 99, 101, 102, 147, 262–266] could possibly offer significant potential for reductions in the number of terms. In particular, regular grid-based lattice models such as the Hubbard model can use QWC to reduce scaling of the number of operators to measure to $\mathcal{O}(D)$, and remain independent to the

number of lattice sites [276, 296]. It is also worth considering the fact that ansätze for lattice models tend to be shallower, and as such, the cost of basis rotations in methods based on general commutativity of Pauli operators could be too significant to justify its use. As such we would therefore recommend QWC grouping until further research is conducted.

Ansatz: Direct application of ansätze suited for *ab initio* molecular systems (such as Unitary Coupled Cluster, UCCSD and its extensions) have been shown to work in practice using generalised encodings such as Jordan-Wigner (see for instance, Ref. [297]). However, one can see that the underlying physics motivating the ansatz is not ideally suited to strongly correlated lattice models, requiring care to ensure that they are efficient ansatz for these systems [153]. Since these ansätze are formulated in a basis of Hartree-Fock or other mean-field orbitals, they do not allow using some of the low-weight encodings easily and do not enable exploitation of the low degree of connectivity of the model. Instead, one ansatz that has been shown to be very suitable for correlated lattice problems is the Hamiltonian variational ansatz (HVA) [98, 138]. The ansatz leverages the more compact structure of the lattice model Hamiltonian and is built using fermionic operators, thereby making the most of low Pauli weight lattice encodings [256, 259, 260, 298]. HVA was also shown in [138] to be particularly resilient to the appearance of barren plateaus in the optimisation problem. The ansatz has a depth and number of parameters scaling with the number of commutative Pauli groups in the Hamiltonian (though it may need to be repeated several times to account for lower expressibility compared to UCCSD). For a regular lattice, this can result in an overall scaling that is lower than $\mathcal{O}(nD)$. Extensions of HVA could also be considered for specific systems. For instance, Fourier Transform-HVA [233] could be very efficient

on certain models (in particular jellium in Ref. [233], which is a continuous model, but with the Hamiltonian defined by a single parameter). Symmetry breaking HVAs [299, 300] are also a promising avenue, and though numerical tests from Ref. [300] show excellent fidelities of the state produced, results in Ref. [299] show some instabilities of the ansatz. Overall, until further research is conducted, is seems that HVA is safest and most general option for lattice models. Finally, we have neglected consideration of adaptive ansätze, as it is difficult to make a scaling argument for their efficiency in this domain, where the ability to justify the inclusion of some terms over others in the ansatz is likely to be diminished.

Optimiser: While we find that Rotosolve [114, 115, 270, 271] and Fraxis [272, 273] are also generally supported by strong empirical evidence, the SPSA algorithm has been shown to outperform these gradient-free methods on the Hubbard model [276, 277] and could therefore be considered preferable. SPSA has the great benefit of only requiring a fixed number of measurements, irrelevant of system size, significantly reducing the computational burden of completing an iteration.

Scaling: Based on the above, one can make an argument for the scaling of VQE as implemented for lattice models. The number of shots required to achieve a precision of ϵ when computing an expectation value scales $\mathcal{O}(1/\epsilon^2)$, with ϵ the target precision. For lattice models however, the target is usually not chemical accuracy and instead depends on the aim of the calculation. Generally, the aim is to resolve some correlated features of the electronic structure (e.g. predicting the parameter regimes for different phases, or response properties Ref. [301–304]). The HVA ansatz has a scaling capped by the number

of terms in the lattice model Hamiltonian and the number of repetitions of the ansatz k, hence $\mathcal{O}(knD)$. However, there may be a few caveats to this. Since some encodings require additional qubits to reduce the Pauli weight, they would increase the depth and the number of terms in the ansatz, while at the same time reducing the depth per term of the ansatz due to the reduced Pauli weight. Overall these two effects are likely to only affect the pre-factor of the VQE, noting that this is an area of open research. This results in single energy evaluation scaling as $\mathcal{O}(k(nD/\epsilon)^2)$.

For a regular hypercubic lattice with only at most nearest-neighbor terms in the Hamiltonian, there are $\mathcal{O}(knD)$ parameters in the HVA, bringing the total scaling of computing gradients for a lattice model VQE to $\mathcal{O}((k)^2(nD)^3/(\epsilon)^2)$. This can however be considered a rather pessimistic scaling as lattice models will oftentimes offer opportunities to exploit structural features to reduce the number of measurements required. As mentioned above, in the case of the Hubbard model, one can use the grid like structure to measure sets of non-overlapping (or more broadly QWC) pauli strings concurrently reducing the number of shots for a single energy calculation to $\mathcal{O}(D/\epsilon^2)$. Similarly, as SPSA has $\mathcal{O}(1)$ scaling and has been shown to perform well on Hubbard lattices this could reduce the overall scaling for a complete iteration of this model to $\mathcal{O}((kn(D)^2/(\epsilon)^2)$ (it is worth noting however that the use of SPSA would likely also require additional iterations to reach convergence).

In either case, if the full parallelisation potential is exploited, this would give a scaling of $\mathcal{O}(knD)$ for one iteration of the VQE. Discussions regarding the number of iteration are identical to the considerations raised in the previous section. Similarly, while the depth of HVA scales with the number of commutative groups in the Hamiltonian, it is not yet clear whether this would reduce scaling below that of the number of terms $\mathcal{O}(nD)$.

Chapter 3

Limits of near-term variational quantum computing

As outlined in the previous chapter, the potential advantage of VQE remains uncertain due to three main impediments: (1) the pre-factor to scaling, resulting from the large sampling requirements of the method, (2) the barren plateau problem, which could be worsening the number of measurements required, and (3) the mitigation and control of errors. These are discussed in the three following sections.

3.1 Runtime estimates for VQE and the measurement problem

I first present a runtime estimate for a single iteration of VQE, constructing on the best practices outlined in the previous chapter. We then discuss the large potential for paralellisation of measurements and discuss requirements on hardware and outstanding research questions.

3.1.1 Cost and runtime estimates for VQE

There have been several studies estimating the resources required to perform VQE on a system that is too large to be accurately treated using conventional methods. Wecker et al. [98] developped the Hamiltonian Variational Ansatz (HVA), and presented a numerical study of the accuracy of some ansätze (the HVA and various UCC based ansätze) and the number of repetitions required. They also possibly unveiled the existence of the barren plateau problem ahead of it being characterised in Ref. [109], by numerically showing that more expressive forms of UCC cannot reach the same accuracy as less expressive forms on larger systems. They estimate that the total number of samples required to compute the ground state energy of Fe_2S_2 to chemical precision (using the STO-3G basis with $n = 112$ spin orbitals) is of the order $\mathcal{O}(10^{19})$, which is far beyond what could be considered tractable. Of course, conventional methods aiming at resolving exactly, and directly, the ground state of a $n = 112$ spin orbitals Hamiltonian would also be intractable (this would equate to finding the lowest eigenvalue of a 2^{112} matrix). However this does not exclude the possibility that more refined conventional methods, accepting some approximation, could an acceptable level of accuracy on Fe_2S_2 (as an example, please see Ref. [305]). The literature has progressed significantly since this 2015 study, and there are now more efficient ansätze and grouping methods that may change this conclusion.

Kühn *et al.* [83] numerically assess the number of qubits and circuit depths required for UCC based ansätze. They show that to model a medium-sized organic molecule such as naphthalene ($C_{10}H_8$, with 68 electrons) would roughly require about 800 to 1500 qubits, and a number of two-qubit gates of about $\mathcal{O}(10^8)$ using UCCSD. This latter number may be significantly lowered if the k-

UpCCGSD ansatz is used (assuming it can achieve the desired accuracy). They also claim that the runtime for a VQE implementation would be impractical, even using full parallelisation potential, without unfortunately providing more details about how this conclusion is reached.

Gonthier *et al.* [84] provide what probably constitutes the most comprehensive study of the VQE resource requirements to date by estimating the cost of combustion energy computation for nine organic molecules (including methane, ethanol, and propane). They provide a detailed estimate of the runtime for one energy estimation ranging from 1.9 days for methane (CH_4), which requires 104 qubits for accurate treatment, to 71 days for ethanol (C_2H_6O), which requires 260 qubits (this estimate uses a frozen natural orbital basis, with 13 functions, i.e. 13 qubits for each electron). The analysis is rather exhaustive since it takes into consideration the joint measurements of Hamiltonian terms, and different optimisation methods.

It is worth noting that the studies mentioned above do not take into account three obstacles we listed at the beginning of this section (namely, the complexity of the optimisation, the barren plateau problem, and the impact of quantum noise). At the same time, Refs. [82–84] do not discuss the potential for parallelisation (with the exception of [82] which touches upon it briefly). For instance, the runtime estimates of 1.9 to 71 days presented in [84] can be parallelised efficiently, although this would require a significant quantity of qubits arranged in sets on which parallel computation can be performed, possibly resulting in a variety of new problems such as overhead communication cost and additional quantum noise (see Sec. 3.1.2 for a discussion).

We provide our estimated runtimes for the steps in the VQE for a representative example system in Table 3.1, including the general scaling estimate for such types of systems. The example system considered corresponds to the

one proposed in Ref. [82], and is the *ab initio* computation of the chromium dimer (Cr_2) with an active space of 26 electrons in 26 molecular orbitals (52 spin orbitals and 52 qubits).

It is very difficult to estimate the pre-factor of the VQE, which would very much be dependent on the hardware, and a detailed numerical analysis is not within the scope of this work. To estimate the depth we compiled a 52 qubit version of k-UpCCGD, using qiskit, assuming $k = 1$ and full connectivity of the qubit register. We founnd a depth of $\sim 27,000$ timesteps, denoted L, and 170 parameters, denoted p (to illustrate the impact of the connectivity of the qubit register, we compiled the same ansatz assuming a linear qubit register, and found that the depth required is increased by more than tenfold). Note that choosing $k = 1$ is likely insufficient [80, 217]. Gate time is assumed to be $T = 100$ ns (similar to what is presented in [84], which itself refers to Table 1 in [306]), which is better than what can currently be achieved for superconducting qubits (~ 500 ns), but is probably achievable over the next few years. We assume readout and reset times are negligible compared to the circuit runtime. The pre-factor for the number of operators ($\mathcal{P}$) to measure can easily be assumed to be 16 as each fermionic operator result in two Pauli strings under generalised mappings, and there are four fermionic operators in each two-body term in the Hamiltonian. Using a form of decomposed interactions we estimate that the number of operators is $\sim 16N = 832$ (we do not consider the impact of covariances may have on the noise of the estimates, though point out that 16 is a conservative pre-factor), and assume that the required circuit depth for basis rotation is negligible. Finally, we set the target precision ϵ to 10^{-3} $\mathrm{mE_H}$, which is close to chemical precision (1.6 $\mathrm{mE_H}$) [173] and roughly assume that $S = 1/\epsilon^2 = 1,000,000$ shots are used for the estimation. A much lower number of shots would be sufficient to progress the initial part

Table 3.1: Indicative estimates of the runtime of one iteration of the VQE making the following assumptions: gate time: 100 ns. This assumes an active space of 26 molecular orbitals for Cr_2 spanned over $N = 52$ qubits, and a gate runtime of $T = 100$ ns. The model uses k-UpCCG SD, k = 1, targets $\epsilon = 10^{-3}$ mE_H (using decomposed interactions methods).

Operation	Scaling	Formula	Runtime
Single shot	$\mathcal{O}(kN)$	$L \times T$	3 ms
One expectation	$\mathcal{O}(\frac{kN^2}{\epsilon^2})$	$L \times T \times \mathcal{P} \times S$	25 days
Full iteration (Rotosolve)	$\mathcal{O}(\frac{k^2N^4}{\epsilon^2})$	$L \times T \times \mathcal{P} \times S \times 3p$	35 years
Full iteration (gradient-based)	$\mathcal{O}(\frac{k^2N^4}{\epsilon^2})$	$L \times T \times \mathcal{P} \times S \times 2p$	24 years

of the optimisation, and this high number of shots is only required in the last iterations of a VQE close to convergence to reach chemical precision (note however that this number of shots may need to be much higher in case of barren plateaus). It is therefore likely that the last few iterations before convergence are the most costly and time-consuming, largely dominating the cost. However, despite some optimistic assumptions listed above, it is clear that the time cost of VQE implemented on a single set of qubits remains orders of magnitude too large to be realistically viable, pointing to the dependence of the method on parallelisation.

3.1.2 Parallelisation potential of the VQE

The potential for parallelisation of the VQE was already identified in the initial paper by Peruzzo *et al.* [37] and subsequently mentioned in many VQE papers, although in-depth studies are still sparse. Parallelism could however be relevant for the performance of the method. Parallelism of the VQE offers a direct way to convert runtime cost into hardware cost by splitting the shots required onto different sets of qubits (which can be arranged in different threads on a single quantum computer, or multiple, disconnected quantum computers). To illustrate this point, we adapt the estimates presented in Table 3.1 assuming that perfect parallelisation is possible, and present the results in Table 3.2.

From this analysis, it appears that parallelisation will be a critical part of any future success of the VQE method when applied to *ab initio* molecular systems. The scale of benefits that can be gained from parallelisation is however less clear with regards to lattice models, and in particular when applied to the Hubbard model. Cai [296] provides an assessment of the possible runtime of VQE on a Hubbard model of 50 qubits (which cannot be solved accurately on current conventional devices) and shows that an iteration could take approximately 1.7 days, or merely 10 minutes with parallelisation over 200 QPUs. Further improvements to the implementation of HVA (proposed for instance by Cade *et al.* [276]) could reduce this runtime significantly. As such, while lattice models would likely also benefit from parallelisation, it is less likely to be a requirement than for *ab initio* molecular systems. It is worth noting that the use of parallelisation has already been shown to be less advantageous on an implementation of the Hubbard model in a case where the SPSA optimiser is used [307].

Considering the case of *ab initio* molecular systems, broad availability of

Table 3.2: Indicative estimates of the runtime and number of quantum computers used for one iteration of the VQE assuming perfect parallelisation of the method can be achieved and neglecting any communication overheads - using the same assumptions stated in Table 3.1

Operation	Time scaling	Runtime	Scaling sets of qubits	Sets of 52 qubits
Single shot	$\mathcal{O}(kN)$	3 ms	$\mathcal{O}(1)$	1
One expectation	$\mathcal{O}(kN)$	3 ms	$\mathcal{O}(\frac{N}{\epsilon^2})$	$\sim 800 \times 10^6$
Full iteration (Rotosolve)	$\mathcal{O}(k^2N^3)$	0.5 s	$\mathcal{O}(\frac{N}{\epsilon^2})$	$\sim 2,500 \times 10^6$
Full iteration (gradient-based)	$\mathcal{O}(kN)$	3 ms	$\mathcal{O}(\frac{kN^3}{\epsilon^2})$	$\sim 280 \times 10^9$

quantum computers with increasing number of qubits could significantly speed-up the VQE process, however there are significant caveats to that. One key observation is that full parallelisation would require a number of quantum computers cores (or threads) that scales $O(pN^4/\epsilon^2)$, with p the number of parameters in the ansatz. This could clearly be a prohibitive number for large computation given the current state of hardware technology, and it is possible that fault-tolerant technology could arrive before we are able to produce such large quantities of devices.

Even if it was possible to build large quantities of quantum computers, there are many caveats to the potential of parallelisation for the VQE. First, as it is the case for conventional parallel computing, parallel quantum computing will suffer from communication overheads. These overheads are the computational cost of coordinating the parallel tasks, which can include the likes of synchronisation cost, data aggregation and communication (possibly latency if the different sets of qubits are connected through the cloud). Second, parallelisation could result in higher noise levels. We note two possible sources of additional noise: (1) if parallelisation is done on multi-threaded quantum computers, there is higher chance of cross-talk between qubits; (2) variational algorithms are considered to be somewhat noise resilient as they can learn the systematic biases of a given hardware [38, 152] - if the algorithm is run on multiple quantum computers these noise resilient effects may disappear, as systematic biases on one set of qubits, which differs on another, may no longer be learned through the variational process.

3.2 The barren plateau problem

Ansatz selection is a central part of the VQE pipeline. The right choice of ansatz is critical to obtain a final solution that is close to the true state of interest. To achieve this, it is essential to maximise the span of the ansatz in parts of the Hilbert space that contain the solution (i.e., a state that is sufficiently close to the desired state which globally minimises the expectation value of the Hamiltonian). The span of possible states an ansatz can reach is referred to as its expressibility. However, optimising a highly expressible state could easily become intractable due to the number of parameters, the number of iterations required for convergence, or the number of shots required to achieve sufficient gradient accuracy to continue the optimisation. Whether an ansatz can be optimised in a tractable manner is referred to as its trainability. In practice, it is better to choose an ansatz spanning a smaller subspace, but remaining trainable. Designing an efficient ansatz for a given number of qubits hence involves finding an optimal trade-off between expressibility and trainability.

Expressibility: The expressibility of an ansatz describes its span across the unitary space of accessible states [106, 107, 308]. One can quantify the expressibility of an ansatz by assessing the distance between the distributions of the unitaries that can be generated by the ansatz, and the uniform distribution of unitaries in the corresponding Hilbert space [308], also known as the Haar measure. A given ansatz is called a t-design if it is indistinguishable from the Haar measure up to the t^{th} moment. A 2-design ansatz can produce any possible state in the Hilbert space considered, from any input state: it is maximally expressive. As a side note, Hubregtsen et al. [309] also study the

relationship between the expressibility of quantum neural networks and their accuracy in a classification task.

More formally, one can define as $\mathbb{U}$ the set of unitaries accessible by an ansatz, and $\mathcal{U}(N)$ the complete unitary group in which the ansatz is expressed (with N the number of qubits it spans), such that $\mathbb{U} \subseteq \mathcal{U}(N)$ [106–108, 308]. The following super operator, representing the second order difference between the Haar measure on $\mathcal{U}(N)$ and the uniform distribution of $\mathbb{U}$ can be constructed (we follow the formalism in [106]):

$$\mathcal{A}_{\mathbb{U}}(\cdot) := \int_{\mathcal{U}(n)} d_\mu(V) V^{\otimes 2}(\cdot)(V^\dagger)^{\otimes 2} - \int_{\mathbb{U}} dU U^{\otimes 2}(\cdot)(U^\dagger)^{\otimes 2}, \tag{3.1}$$

with $d_\mu(V)$ the volume element of the Haar measure, and dU the uniform distribution over $\mathbb{U}$, $V \in \mathcal{U}_N$ and $U \in \mathbb{U}$. If $\mathcal{A}_{\mathbb{U}}(\hat{O}) \to 0$, then the ansatz producing $\mathbb{U}$ approaches a 2-design and therefore offers maximal expressibility. From this super-operator, one can compute a metric for expressibility of an ansatz as

$$\varepsilon^{\rho}_{\mathbb{U}} := ||\mathcal{A}_{\mathbb{U}}(\rho^{\otimes 2})||_2 \tag{3.2}$$

$$\varepsilon^{\hat{P}}_{\mathbb{U}} := ||\mathcal{A}_{\mathbb{U}}(\hat{P}^{\otimes 2})||_2. \tag{3.3}$$

Consequently, the expressibility of an ansatz can be expressed with respect to an initial input state (ρ), or with respect to a measurement operator ($\hat{P}$). Following the equations above, one can interpret that if $\varepsilon = 0$ the ansatz is maximally expressive, while expressibility decreases as ε increases. Expressibility has also been shown to be a convenient metric for assessment of parameterised quantum circuits more generally [107]. In addition, several methods

have been proposed to remove redundant parameters from quantum circuits without decreasing expressibility [310] or reducing the set of states that can be generated through the circuit [311].

Trainability: The trainability of an ansatz refers to the ability to find the best set of parameters of the ansatz by (iteratively) optimising the ansatz with respect to expectation values of the Hamiltonian in a tractable time [106, 108]. More specifically, an ansatz is considered trainable if its expected gradient vanishes at most polynomially as a function of the various metrics of the problem (e.g. system size, circuit depth). On the other hand, if the gradient vanishes exponentially, it is said to suffer from the barren plateau problem.

A key issue that is inherent to all types of variational quantum algorithms is the risk of vanishing gradients, either during training or as a result of a random initialisation [109]. This refers to the risk of the cost function gradients vanishing exponentially as a function of specific properties of the optimisation for a problem. McClean *et al.* [109] provide the first formal characterisation of this barren plateau problem (some early numerical evidence of this problem are outlined in Ref. [98], without a characterisation being provided), and show that cost function gradients are vanishing exponentially in the number of qubits in the quantum register when provided with random initialisation of the circuit parameters. Even though this problem is akin to the vanishing gradient problem in machine learning, it has two striking differences that make it significantly more impactful on the prospects of variational quantum algorithms [109]:

- The estimation of the gradients on a quantum device is essentially stochastic. Any observable can only be measured to a certain precision, increasing

as the inverse square root of the number of shots. If gradients are exponentially approaching zero, it means that distinguishing between a positive and a negative gradient becomes increasingly difficult. Failing to establish the sign of the gradient reliably transforms the optimisation into a random walk, overall requiring an exponential number of shots to continue optimisation.

- The barren plateau problem is dependent on the number of qubits (while the problem is dependent on the number of layers for the vanishing gradient problem). Additional research also shows that it can be linked to other factors specific to quantum circuits, including expressibility of the ansatz [106], degree of entanglement of the wavefunction [122, 312], non-locality of the wavefunction [108, 144, 313], or quantum noise [314].

Before describing key drivers of the barren plateau problem in more detail, and potential methods to address it, it is worth briefly discussing the typical cost function landscape for single parameters in the variational quantum eigensolver. Another problem that affects this landscape is that of 'narrow gorges' (initially characterised in [108]). It refers to the fact that the local minimum (well defined by the region starting from the end of a barren plateau and going towards a local minimum) contracts exponentially in the number of qubits. Interestingly, it was shown that narrow gorges and barren plateaus always occur concurrently and can therefore be considered equivalent [315]. An alternative way to present the barren plateau problem is that it implies the expectation value of an observable with respect to a random state concentrates exponentially around the mean value of that observable [109], rendering intractable optimisation away from the mean.

In the context of the VQE, the barren plateau problem can be formally

characterised as follows. Consider a VQE optimisation problem with cost function:

$$E(\boldsymbol{\theta}) = \langle \psi(\boldsymbol{\theta}) | \hat{H} | \psi(\boldsymbol{\theta}) \rangle , \tag{3.4}$$

with $\hat{H}$ the molecular Hamiltonian operator, and $|\psi(\boldsymbol{\theta})\rangle$ the parameterised wave function with a vector $\boldsymbol{\theta}$ of parameters. This cost function exhibits a barren plateau if, for any $\theta_i \in \boldsymbol{\theta}$ and for any $\epsilon > 0$ there is $b > 1$ such that:

$$Pr(|\partial_{\theta_i} E(\boldsymbol{\theta})| \geqslant \epsilon) \quad \leqslant \quad \mathcal{O}(\frac{1}{b^N}), \tag{3.5}$$

which is an immediate consequence of Chebyshev's inequality and the result from above (for the expectation value and variance) [108]. This means that the probability of a gradient being above a certain threshold (which could be arbitrarily small), can always be upper-bound by a number that decreases exponentially in the system size N. It is however important to note that while defined with respect to a cost gradient, the barren plateau problem also affects gradient-free optimisers [312, 316] (e.g. COBYLA, Powel, Nelder-Mead, RotoSolve). It is easy to understand, as gradient-free optimisers usually rely on sampling the cost landscape of specific parameters. If the variance across the landscape is too small, then it becomes impossible to accurately progress through the optimisation step.

3.2.1 Drivers of the barren plateau problem

System size and random initialisation [109]: The barren plateau problem refers to the fact that the gradient of a cost function incorporating a layered ansatz has an exponentially vanishing variance, and values approaching zero in the number of qubits, provided ansatz parameters are initialised

randomly. A layered ansatz for a random parameterised quantum circuit can be described as [109]

$$U(\boldsymbol{\theta}) = \prod_{l=1}^{L} U_l(\theta_l)\mathcal{W}_l, \tag{3.6}$$

where $U(\theta)_l = e^{-i\theta_l \hat{V}_l}$, with $\hat{V}_l$ a hermitian operator, and $\mathcal{W}_l$ a generic non-parameterised unitary. The cost function is as described in Eq. (3.4), taking $|\psi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})\,|0\rangle$. The gradient of this cost function with respect to any given parameter θ_i can be conveniently computed as

$$\partial_{\theta_i} E = i\,\langle 0|\, U^{\dagger}_{1\to(i-1)} \left[\hat{V}_i, U^{\dagger}_{i\to L}\hat{H}U_{i\to L}\right] U_{1\to(i-1)}\,|0\rangle\,. \tag{3.7}$$

where $1 \to (i-1)$ represent the ansatz layers from layer index 1 to layer index $(i-1)$, and $i \to L$ represent the ansatz layers from layer index i to layer index L. From the computation of the gradient, McClean et al. [109] show that if both $U_{1\to(i-1)}$ and $U_{i\to L}$ are 2-designs, the variance of the gradient is clearly vanishing exponentially in the system size:

$$\mathrm{Var}[\partial_{\theta_i} E] \approx \frac{1}{2^{(3N-1)}} Tr\left[\hat{H}^2\right] Tr\left[\rho^2\right] Tr\left[\hat{V}^2\right] \tag{3.8}$$

Cases in which either of $U_{1\to(i-1)}$ or $U_{i\to L}$ is not a 2-design are also addressed in Ref. [109], with similar outcomes (I direct readers to this reference for a full demonstration, as well as detailing of the rules needed to compute the expected value of a variance over an ansatz). Further analysis conducted by Napp in Ref. [317] shows additional analytical bounds for unstructured variational ansätze, moving away from the layered ansatz described above.

Expressibility [106]: Holmes et al. show that trainability and expressibility of the ansatz are inversely related. In other words, the more expressive an

ansatz is, the more prone it is to barren plateaus. This does not mean that low-expressivity ansätze are not affected by barren plateaus, as other drivers can otherwise trigger the problem (for instance, system size and random initialisation, as above, or a very non-local cost function [108, 144, 313]). This observation implies that one cannot lower-bound gradients as a function of expressibility, but they can be upper-bounded.

This is shown in Ref. [106] by extending the expression of the barren plateau problem as explained in Ref. [109], and setting an upper bound for the variance of the cost gradient as a function of the ansatz' distance to a 2-design. As such, they use the same layered ansatz template (Eq. 3.6) and resulting gradients (Eq. 3.7). As an illustration, Ref. [106] find a generalised bound for gradient variance as a function of expressibility (we encourage the reader to read the finer details directly from the source material), as

$$\mathrm{Var}[\partial_{\theta_i}, E] \leqslant \frac{g(\rho, \hat{P}, U)}{2^{2N}-1} + f(\varepsilon_L^{\hat{P}}, \varepsilon_R^{\rho}), \tag{3.9}$$

where the first half of the bound corresponds to the maximally expressive ansatz (and $g(.)$ is a function defining a pre-factor on this expressibility, defined in detail in Appendix E of Ref. [106]). The function $f(.)$ is the extra bound resulting from the expressibility (or lack thereof) of the ansatz, defined as

$$f(\varepsilon_x, \varepsilon_y) := 4\varepsilon_x\varepsilon_y + \frac{2^{N+2}(\varepsilon_x||\hat{P}||_2^2 + \varepsilon_y||\rho||_2^2)}{2^{2N}-1}, \tag{3.10}$$

and in which $\varepsilon_L^{\hat{P}}$ (the expressibility metric for the part of the ansatz to the left of the parameter i, where the gradient is taken with respect to the measurement operator), and ε_R^{ρ} (the expressibility metric for the part of the ansatz to the right of the parameter i, where the gradient is taken with respect to the state

density matrix), have been used as arguments. From this equation, one can see that the gradient variance admits an upper-bound approaching $\mathcal{O}(\varepsilon_L^{\hat{P}}\varepsilon_R^{\rho})$ as $N \to \infty$. As a result of the definitions in Eqns. (3.1-3.3), it shows that high expressibility (low ε) lowers the gradient variance bound and therefore limits the trainability of the ansatz. For further information on expressibility of ansätze, Nakaji et al. [107] provide a study of the expressibility of the shallow alternating layer ansatz.

Cost function non-locality: [108, 144, 313]: In Ref. [108], Cerezo et al. show that an ansatz trained on local cost functions are more resilient to the barren plateau problem than those trained on global cost functions. They illustrate this point by comparing a cost function constructed around the expectation value of a global observable: $\hat{O}_G = \mathbb{1} - |0\rangle\langle 0|^{\otimes N}$ to a cost function constructed around the expectation value of a local observable $\hat{O}_L = \mathbb{1} - \frac{1}{N}\sum_i^N |0\rangle\langle 0|_i \otimes \mathbb{1}_{\neq i}$, the latter being local as each component of the observable only applies to a single qubit. It is shown in particular that while alternative layered ansätze trained on the global cost function are challenging to train (this is not necessarily true on other types of ansätze), ansätze trained on the local cost function are considered trainable if their depth scales logarithmically with the circuit width (i.e. $\mathcal{O}(\log(N))$ or below). Cerezo et al. [108] also show that ansätze with a scaling $\mathcal{O}(\text{poly}(\log(N)))$ could also be either trainable or not. Ref. [144] extend these findings to a wider range of cost functions, and Ref. [313] demonstrated the occurrence of this phenomenon in the case of Dissipative Perceptron-Based Quantum Neural Networks.

An important consequence for the VQE, as pointed out in [108] is that local encodings such as Bravyi-Kitaev [281–283], ternary tree encoding [255] or Generalised Superfast Encoding [256] with lower Pauli weight, would offer more

resilience than encodings such as Jordan-Wigner [97] which has Pauli weights scaling of $\mathcal{O}(N)$), therefore resulting in a very non-local VQE cost function. It is also worth noting that the best known scaling for a VQE ansatz is linear (e.g. k-UpCCGSD [80], Fourier Transform-HVA [233] on some systems). Uvarov *et al.* [257] compares numerically the impact of using Jordan-Wigner compared to Bravyi-Kitaev on a Hubbard-like model and find that in this case the latter results in gradient variance nearly one order of magnitude larger than the former. This should be caveated by the fact that the numerical results in Ref. [257] also show that the number of layers used in the ansatz (in this case, a symmetry preserving ansatz is used, similar to the one presented in Ref. [136]) ultimately dominates and reduces gradient variance to a negligible number in either case.

Noise induced barren plateau (NIBP): Wang *et al.* [314] show that incorporating quantum noise in a variational optimisation can accelerate the occurrence of barren plateaus, and additionally result in vanishing of the amplitude of the expectation value. Their main result is a bound on the value of a parameters' gradient (with notation slightly changed from Ref. [314]), as

$$\left| \frac{\partial \langle \hat{H}(\boldsymbol{\theta}) \rangle}{\partial \theta_i} \right| \leq G(N) q^{L+1}, \tag{3.11}$$

where $q < 1$ is a parameter representing the strength of the noise model (the lower it is, the more noise there is), $G(N) \sim \mathcal{O}(2^{-\alpha N})$ with α an arbitrary, positive constant, and L represents the number of layers in the ansatz. A few important points can be raised as a consequence of this bound [314]. First, the noise-induced barren plateau is independent of parameter initialisation, or locality of the cost function, meaning that some of the strategies listed in

Sec. 3.2.2 will not work in a noisy setting. Second, this bound is conceptually different from the previously described drivers of barren plateaus as it is a bound on the gradient, rather than on the variance of the gradient. Rather than the flattening optimisation landscape described previously, NIBP results in the vanishing of the amplitude of the expectation value function and a bias away from the minimum.

Large degrees of entanglement [122, 312]: The degree of entanglement of the trial wavefunction has also been shown to be associated with the barren plateau problem. In particular, Patti *et al.* show that one can link vanishing gradients to the entanglement entropy [76] of the trial state wavefunction even at low circuit depth. Ortiz Marrero *et al.* reach a similar conclusion by first showing that entanglement between visible and hidden units in a Quantum Neural Network reduces trainability. The result is then extended to unitary networks (very much similar to UCC based ansätze) and quantum Boltzmann machines. A final point to note is that it was also shown that higher-order derivatives of the cost function are also affected by the barren plateau problem, and therefore cannot be used as a means to circumvent it [318].

3.2.2 Methods to address barren plateau problem

It follows from Refs. [109] and [106] that addressing the barren plateau problem can be done through modification of the ansatz. In particular, techniques focusing on selectively reducing the expressibility of the ansatz, or in other words, avoiding a 2-design (which would be the maximally expressive unitary on a given Hilbert space) are expected to be more resilient to barren plateaus. In the context of the VQE, this can be done by restricting the span of the

ansatz to a section of the Hilbert space of interest. In particular, it was shown that adaptive ansätze exhibit some resilience to the barren plateau problems. In addition, there are optimisation methods which aim at tempering this problem (for example Ref. [113]). Another means available to contain the barren plateau problem is to select a local encoding with low Pauli weights, as discussed above and explained in Ref. [108].

Some methods have also been developed specifically to address barren plateaus. A first example consists of initialising ansatz parameters such that subsections of the ansatz (as split when computing the gradient) do not form a 2-design, at the very least avoiding to start the optimisation process in a barren plateau [121]. Starting from the layered ansatz in Eq. (3.6), we can divide the ansatz into K blocks of depth D (such that the total depth $L = KD$). The depth D of each block considered in isolation needs to be shallow enough to ensure that the block does not approach a 2-design. Each block $U_k(\boldsymbol{\theta}_k)$, parameterised by a vector $\boldsymbol{\theta}_k$ can then be split into two parts of equal depth, such that

$$U_k(\boldsymbol{\theta_k}) = \prod_{d=1}^{D/2-1} U_d(\theta_{d,1}^k) W_d \prod_{d=D/2}^{D} U_d(\theta_{d,2}^k) W_d, \tag{3.12}$$

where $\theta_{d,1}^k$ can be initialised at random, but where $\theta_{d,2}^k$ are initialised such that $U_d(\theta_{d,2}^k)W_d = (U_d(\theta_{d,1}^k)W_d)^\dagger$. The result is that, for all k, before any optimisation, we have

$$U_k(\boldsymbol{\theta}_k) = I_k, \quad \text{and} \quad U(\boldsymbol{\theta}_{init}) = I. \tag{3.13}$$

Grant et al. [121] show that this parameter initialisation strategy could slow down the optimisation process of the VQE as the initial state produced by the circuit would have no entanglement. They propose to initialise the

qubit register with a random entangled state, using a shallow random unitary which remains constant throughout the optimisation process. While showcasing promising results on small systems, the method is however quite challenging to implement in practice. Identifying a block initialisation is not directly possible for all ansatz structures: for example it is not straightforward for unitary coupled cluster based ansatz [37, 81, 319] without repeating some operators, and it is in general not possible exactly with hardware efficient ansätze.

Nonetheless, the idea of adjusting parameter initialisation to improve resilience to the barren plateau problem has been extended to alternative mitigation techniques. In Ref. [320], Sauvage *et al.* propose to select optimal initial parameters with the help of a machine learning model (FLexible Initialiser for arbitrarily-sised Parameterised quantum circuits, or FLIP). The model is trained to identify structures of parameters that best suit specific families of quantum circuit optimisation problems, and is numerically shown to provide significant improvements. Similarly, Kulshrestha and Safro show that initialising ansatz parameters by picking them from a beta distribution reduces the impact of the barren plateau problem compared to picking them from a uniform distribution [321]. They also show that adding perturbations to the parameters between each optimisation step also helps in mitigating vanishing gradients.

Several additional methods have been developed for the barren plateau in the general case of parameterised quantum circuits. In Ref. [322], Volkoff et al. show that one can reduce the dimensionality of the parameter space by using spatially and temporally correlated parameterised quantum gates, resulting in higher resilience to barren plateaus. The ansatz can also be trained layer by layer to the same effect [323], though limitations of this method were shown in Ref. [324]. This is somewhat akin to adaptive ansätze, but generalised to

any quantum neural network optimisation. Patti *et al.* [122] also propose several additional mitigating methods including an alternative initialisation strategy in which two qubit registers are initially not entangled, regularisation on the entanglement, the addition of Langevin noise, or rotation into preferential cost function eigenbases. Sack *et al.* [325] showed that barren plateaus can be partially mitigated as part of a classical shadow measurement scheme. Wu *et al.* [326] propose to mitigate the impact of NIBP by defining an alternative cost function with the same optimal state but without sensitivity to vanishing gradients, by identifying and eliminating the dominant term in the Pauli representation of the observable measured. Finally, though not directly relevant to VQE, Pesah et al. [327] show numerically that Quantum Convolutional Neural Networks exhibit natural resilience to the barren plateau problem. Similarly, Zhang *et al.* [290] show that Quantum Neural Networks with tree tensor structure and step-controlled architectures have gradients that vanish at most polynomially in the system size.

3.2.3 Comments on barren plateau in the context of the VQE

In Ref. [106], Holmes et al. point out that the problem structure of VQE can be used to limit the impact of barren plateaus. The features in question include symmetries of the problem, which can be used to reduce the portion of the Hilbert space, or physically-motivated ansätze targeting a restricted part of the Hilbert space in which good approximations of the ground state are expected to be. Such relevant structures are in general more difficult to find when considering the wider field of quantum neural networks. Adopting an initialisation strategy such as the ones presented in Refs. [121, 122, 320]

would allow an ansatz to begin optimisation away from a barren plateau, and as such, away the target operator mean. One could argue that given the optimisation problem aims at finding a minimum, a reliable optimiser should always move a state away from the mean expectation value and therefore away from the barren plateaus regions. This point however could be invalidated by several aspects of the optimisation process. These include the existence of local minima, the increase in entanglement of the trial wavefunction as the optimisation progresses, or the presence of noise which results in vanishing of the value function amplitude.

Focusing on local encodings has been shown to provide some resilience to barren plateaus [108, 144, 257], suggesting that using VQE on lattice models with limited dimensions could be performed relatively better than on a molecular Hamiltonian in that respect. As pointed out in Ref. [108], anstze scaling logarithmically in the system size, and measured on local observable are resilient to barren plateaus. At this stage, however, there is no known VQE ansatz scaling as such that also guarantees an accurate description of the ground state.

3.3 Error mitigation and error control

NISQ and near-term quantum computation assumes that hardware does not meet the requirements to implement error correction schemes. State of the art error correction would require $\mathcal{O}(10^6)$ physical qubits assuming an error rate of $\epsilon \sim 10^{-3}$ [231]. Both the number of qubits and expected error rates are far too restrictive for near term quantum devices. As such, error mitigation places itself as a computationally more efficient alternative to error correction, possibly accepting a cost in terms of the accuracy of the final result.

3.3.1 Review of main error mitigation methods

In order to provide context to the use and discussion of error mitigation, we present here a few promising techniques that have been developed in the literature.

In the context of quantum chemical computation, a first example of error mitigation is filtration of symmetry breaking measurements [328, 329]. This technique was demonstrated experimentally on an implementation of VQE [146]. This technique of course guarantees that all accepted measurements will be in the desired symmetry subspace, preventing a significant proportion of errors, however resulting in a potential large increase in the number of measurements required [330]. It is worth noting that symmetry error mitigation can also be implemented through mid-circuit measurements [330].

The next example of error mitigation is the family of extrapolation based techniques. The idea is simply to infer the value of a noiseless measurements from different noisy measurements. These method therefore rely on the ability to artificially increase the noise of the quantum device to gather further datapoints on which the inference is based. It can be achieved for example by re-scaling the intensity of the pulse producing quantum gates [331], by randomly adding Pauli gates into the circuit (referred to as Pauli Twirling) [332], or by adding blocks of CNOT gates [333, 334]. From the data gathered, inference of noiseless measurement can be made using a linear, polynomial, [331] or exponential extrapolation models [335].

A final example of error mitigation that we select to present here is the exponential error suppression [336, 337], also referred to as virtual distillation. The idea behind this technique is that the mixed state resulting from application of a noise channel will likely remain dominated by the noise-free state.

By applying operations that increase the amplitude of the dominant eigenstate in the noisy, mixed state, one can exponentially increase the likelihood of measuring the noiseless state.

Of course, many other methods have been proposed, but scalability of error mitigation remains under question. In Ref. [151], Takagi *et al.* provide a bound regarding the capability of error mitigation schemes to reduce computation errors as a function of the sampling overhead required for their implementation. In their analysis, they suggest that this bound scales exponentially in circuit depth.

3.3.2 Light touch error mitigation

To improve results from near-term experiments, we present here a light touch error mitigation technique. The reasoning employed is similar to the method proposed in Ref. [338]. It does not aim to provide an alternative to more advanced error mitigation methods, but rather to provide a tool to assess the lower bound of what error mitigation can achieve at no additional computational cost to that of a VQE. This method was initially presented as an appendix to Ref. [149].

This light-touch technique is a simplified version of extrapolation, based on the assumption that the impact of the circuit errors is evenly distributed on the output state. The aim is to recover an approximation of the true expectation value of a quantum circuit with respect to an operator, assuming a specific bias on the output results. It is worth noting that this method primarily focus on correcting gate and readout errors, and as such does not attempt to maintain or restore the purity of the quantum state produced. It is not expected to be an effective approach (when used in isolation) for mitigating errors on deep

quantum circuits where the primary source of noise is decoherence.

We define the true (desired) expectation value of the circuit with respect to an operator, $\hat{O}$, as $\langle\hat{O}\rangle$, while the measured expectation value is denoted as $\langle\hat{O}_{\text{mes}}\rangle$ (and where it is assume to the operator studied only has two eigenvalues). The outcome of a circuit can be associated with measurement eigenvalues ± 1, with the probability of measuring 1 equal to $P_1 = \Pr(O_m = 1)$, and with O_m referring to a single measurement of operator $\hat{O}$ at the end of the quantum circuit. Similarly, we have $P_1 = 1 - P_{-1} = \Pr(O_m = -1)$. We can associate the true expectation value with:

$$\langle\hat{O}\rangle = P_1 - P_{-1} = 2P_1 - 1. \tag{3.14}$$

We now assume that there is a certain probability, P_{err}, that at least one gate error occurs during propagation and measurement of the quantum circuit. Any gate error changes the balance of probabilities between $\Pr(O_m = 1)$ and $\Pr(O_m = -1)$. One can assume that there exists a value between ± 1, representing the expectation value of $\hat{O}$ given the error rate, which we denote $\langle\hat{O}_{\text{err}}\rangle$. This value, as well as the probability distribution of the operator measurements O_m are unknown and cannot be recovered easily.

We can approximate the expectation value of the measured operator as $\langle\hat{O}_{\text{mes}}\rangle = (1 - P_{err})\langle\hat{O}\rangle + P_{err}\langle\hat{O}_{err}\rangle$, assuming a linear relationship between the gate errors and effect on the expectation value. We can rewrite this as

$$\langle\hat{O}\rangle = \frac{\langle\hat{O}_{meas}\rangle - P_{err}\langle\hat{O}_{err}\rangle}{(1 - P_{err})}. \tag{3.15}$$

We make the further assumption on the value of $\langle\hat{O}_{err}\rangle$ that, given a sufficiently large number of a random circuit errors, the probability of getting either eigen-

value when an error occurs is equal (or $\langle \hat{O}_{err} \rangle = 0$). This assumption is based on two observations: (1) there is no way to tell exactly what the impact of a gate error will be on measurement probability except that it will bias measurement averages towards 0 (since the dominant probability will on average be affected by more errors), and (2) given that the gate errors are random, these will not result in always measuring 1 or -1, ensuring that these are the least likely values for $\langle \hat{O}_{err} \rangle$. It is however likely that these observations will not always hold true in practice as some sources of error can bias measurements towards a certain result. For instance, readout errors on some QPUs can be biased towards a lower energy outcome [339]. Likewise, amplitude damping noise channel can also favour a measurement outcome. While some of these systematic biases can be learned away through the variational process [152] it is likely that $\langle \hat{O}_{err} \rangle$ may be significantly different from 0, in particular as the system studied and the complexity of the ansatz grow.

This method is largely sufficient for the purpose of our experiments, and can act as a lower bound for the benefits error mitigation can achieve with no additional computing cost. With this approximation, and for eigenvalues ± 1, we can ignore this final term, and the expression simplifies to

$$\langle \hat{O} \rangle = \frac{\langle \hat{O}_{\mathrm{meas}} \rangle}{(1 - P_{err})}, \tag{3.16}$$

while for binary eigenvalues of 1 and 0, we get

$$\langle \hat{O} \rangle = \frac{\langle \hat{O}_{\mathrm{meas}} \rangle - 0.5 P_{err}}{(1 - P_{err})}. \tag{3.17}$$

In order to estimate the final bias on the true value of the operator, we have considered the electron number operator (trace of the one-body RDM). The

deviation from the set number of electrons in the system gives us an estimate for the bias induced by quantum noise P_{err}. Given our assumptions, the bias factor can be recovered as follows:

$$\frac{1}{1 - P_{err}} = \frac{N_{elec}}{N_{\mathrm{meas}}}, \tag{3.18}$$

with N_{elec} the target number of electrons, and N_{meas} the number of electron measured:

$$N_{\mathrm{meas}} = \sum_i \gamma_{ii} = \sum_i \langle \psi | \, \hat{a}_i^\dagger \hat{a}_i \, | \psi \rangle \tag{3.19}$$

This comes at no extra computational cost as the one-body RDM terms used are necessarily computed as part of the VQE process.

An alternative method would be to estimate P_{err} directly by computing it through the reported gate calibration data from the QPU provider (compounding the gate fidelities), but we found that in general this approach is less reliable. This is most likely due to the fact that using this latter method treats the bias resulting from circuit errors completely classically: it ignores any part of the bias that could be due to the reduced purity of the quantum state produced, which can otherwise be captured by the former method.

Results from using this method are presented and discussed throughout Sec. 4 and Sec. 5, along with the novel methods presented in this thesis.

Chapter 4

VQE as a solver of correlated subspaces in multiscale methods

The VQE has been applied as a sub-routine to resolve the low-energy electronic structure in a number of existing approaches, thereby adapting many hybrid methods of conventional quantum chemistry methods to quantum computing. These include a number of 'quantum embedding' methods, where the full space of the problem is partitioned, with each solved at a differing level of theory. In these, it is generally the strongly correlated low-energy partition of orbitals that are amenable to use within a VQE solver which are then, in various ways, coupled back to the rest of the system (potentially self-consistently) at a lower level of theory on a classical device. These multi-resolution methods can substantially extend the scope and applicability of the VQE, under additional constraints arising from this choice of partitioning and coupling of the spaces. Below, we demonstrate the feasibility of two of these self-consistent algorithms in a unified approach applicable to both quantum chemistry and condensed matter physics.

In Sec. 4.1 we present a review of multi-scale embedding methods on both conventional and quantum computers. In Sec. 4.2 we review reduced density matrices and their sampling within QPUs as the self-consistent quantum variables in these multi-scale methods. We show that judicious grouping of commuting terms allows even large active space RDMs to be realistically sampled, with the proposed groupings opening the prospect for higher-rank RDMs and perturbative couplings to active spaces. In Sec. 4.3 we consider the complete active space self-consistent field (CASSCF) [340, 341], an approach for the simulation of molecular systems with strong quantum effects (such as those encountered routinely in inorganic chemistry, systems with competing spin states, excited states, and systems at bond-breaking geometries [342–344]). We show application to carbon monoxide and implementation on IBM Quantum services (IBMQ) machines, where active space wave function are optimised on the QPU within the variational quantum eigensolver (VQE) [37]. Finally, in section 4.4 we focus on extended bulk systems, with the strongly correlated Bethe-Hubbard lattice considered. Specifically, we observe the QPU description of a local region to allow for the opening of Hubbard bands in the material within the QPU-coupled energy-weighted density matrix embedding theory (EwDMET) [345–347] approach. For both of these approaches, we show the fidelity of the QPU sampling of the active space RDMs required for a fully QPU-coupled self-consistent algorithm, and consider the scaling of sampling operations as the active space increases in size in future applications. In particular, using a light-touch error mitigation strategy (as described in Sec. 3.3.2), noise resulting from gate infidelities does not prevent stable convergence of the algorithms presented in this paper.

4.1 Embedding methods in conventional and quantum computing

Complete active space approaches: The simplest and most widespread approach in quantum chemistry for isolating and treating a correlated set of low-energy degrees of freedom at a higher level of theory are the Complete Active Space (CAS) approaches. In these, a subset of high-energy occupied and low-energy unoccupied Hartree–Fock orbitals are considered to span the dominant strongly correlated quantum fluctuations, and treated with an accurate correlated treatment within this subspace (often full configuration interaction, see Sec. 2.4.1). This subspace Hamiltonian includes the presence of a Coulomb and exchange mean-field potential from the remaining electrons outside this space. In this way, the active space electrons are fully correlated within that manifold, leading to the Complete Active Space Configuration Interaction (CAS-CI) approach [89, 348]. Furthermore, the CAS-CI can be variationally optimised, by updating the choice of molecular orbitals defining the low-energy CAS space via single-particle unitary rotations among the entire set of orbitals in the system. This method is generally referred to as Complete Active Space Self-Consistent Field (CASSCF) [340, 341], or the related Multi-Configurational Self-Consistent Field (MCSCF) where the active space is not solved at the level of full configuration interaction. The CASSCF wavefunction can therefore be written as follows:

$$|\Psi_{\text{CASSCF}}\rangle = |\mathbf{R}, \mathbf{c}\rangle = e^{-\mathbf{R}} \sum_{\mu} c_{\mu} |\mu\rangle \,, \tag{4.1}$$

where $\mathbf{R}$ parameterises the single-particle unitary operator defining the rotation of the active space, $|\mu\rangle$ the complete set of Slater determinants in the

active space, and $\mathbf{c}$ defines the coefficients of the configurations indexed by μ. In implementation on a quantum device, the rotation operator defining the active space, $\mathbf{R}$, can be optimised on a classical device, while the parameterised description of the active space wavefunction can be sought via the VQE. These approaches constitute the bedrock for simulation of molecular systems with strong correlation, in particular in systems with competing spin states, excited states, systems at bond-breaking geometries, and inorganic chemistry [342–344]. These CAS-based approaches were initially proposed in combination with VQE as a solver for the active space in Ref. [231] and were subsequently successfully demonstrated practically on quantum computers (an example of which is presented in the following section) in Refs. [149, 349, 350], including self-consistent optimisation of the active space.

It should be noted that in order to achieve this optimisation of the active space, the two-body reduced density matrix of the active space is required, which can have ramifications on the number of measurements required by the VQE [149]. However, in strongly correlated quantum chemistry, it is generally also important to include a description of the correlation within the orbitals external to the active space, generally via low-order perturbation theory, resulting in methods such as complete active space second-order perturbation theory (CASPT2) [351]. These however require computation of the 3-body reduced density matrix (and potentially higher) in order to couple the active space correlations to this perturbative treatment and are therefore considered a daunting proposition for VQE. There is also a wider range of extensions to the CASCI approach, including extensions to embedding with density functional theory (DFT) description of the environment, which has also been explored by Rossmannek et al. [352] within a VQE description of a correlated active space. Shade *et al.* [353] also extend these ideas to the reduced density mat-

rix function theory (RDMFT) and demonstrate an implementation of their method to a Hubbard-like system on a quantum device.

Density matrix embedding theory (DMET): Similar to the active space methods mentioned above, DMET [354, 355] aims at embedding an accurately correlated subspace in a mean-field environment. In contrast to CAS-CI, this 'active space' is chosen through locality criteria, starting from a local fragment space and augmenting it with the minimal number of additional orbitals (denoted the bath space) to ensure that the active space recovers the Hartree–Fock description, and explicitly captures quantum entanglement between the fragment and its environment. In this way, the DMET approach can be considered as having a similar ambition to dynamical mean-field theory [356], but cast as a static wavefunction theory (see below). In order to optimise the mean-field state of the system, the one-body reduced density matrix is matched between the individual fragment spaces between the correlated and mean-field descriptions.

Integration of DMET with a VQE for the correlated subspace solver has been the subject of several publications [357–361], and has been implemented on quantum computers with proof of principles for relevant applications such as protein-ligand interactions for drug design [140] (with an alternative method based on perturbation theory proposed in Ref. [139]). Energy weighted DMET (EwDMET) which builds on DMET to improve its description of dynamical fluctuations for small fragment sizes (thereby moving systematically towards a DMFT description described below) [345–347] was also tested and implemented on a quantum device [149] (see developments in Sec. 4.4), allowing quantum phase transitions to be captured which were out of the scope of DMET. A wide range of possible alternative formulations exist for embedding

correlated subspaces in (static) mean-field environments - especially when that subspace is only weakly coupled to the environment, and the explicit entanglement between the subspace and the environment can be neglected.

Dynamical Mean-Field Theory (DMFT): DMFT again relies on a similar embedding of a (local) correlated subspace in a mean-field environment. However, this environment allows for local quantum fluctuations in its description, thereby including the effects of correlation in the local propagation of particles through the environment. This effect is captured by a local self-energy, which is the self-consistent quantum object in DMFT [356, 362]. This necessitates a formalism built around the single-particle Green's function (a specific dynamical correlation function), which is the object which must be sampled within DMFT on a quantum device. At the heart of DMFT is a mapping from the system of interest to an impurity model, which describes a local correlated fragment coupled to a wider non-interacting set of degrees of freedom, denoted the 'bath'. This impurity model can be represented in a Hamiltonian formulation, from which the single-particle Green's function must be sampled, with various approaches to solve for this Green's function known as 'impurity solvers'. The techniques presented earlier in this section can be used to sample this Green's function in either a time or frequency domain at each iteration in the self-consistent loop. The use of quantum computers as an impurity solver was proposed initially in Ref. [363] in the time domain, but frequency domain solvers have often been more amenable to the low-depth NISQ era. These were explored in the context of DMFT impurity solvers in physical realisations of correlated material systems via VQE-type parameterised algorithms in Refs. [364–367]. An alternative method to compute the Green's function over the whole energy range is based on the quantum

subspace expansion [368].

Overall, embedding methods using the VQE as a high accuracy and scalable solver to describe the correlations within a subspace self-consistently coupled to a wider environment are a promising avenue to extend the applicability of quantum computation towards practical applications. In general, they allow for recovery of significant parts of the electron correlation energy, while avoiding treatment of the full system, thereby reducing qubits number in exchange for additional classical resources in defining the embedding, as well as a self-consistent loop. It is worth noting that the possibilities for embedding the VQE and more general quantum algorithms within wider multi-method and multi-resolution hybrid schemes extends far beyond just the quantum embedding methodologies presented above, and are likely to be of central importance in the utility of quantum algorithms in molecular modeling in all contexts in the future.

4.2 Sampling Reduced Density Matrices on a Quantum Computer

The reduced density matrices (RDM) used in this work are not defined by tracing out a subsystem, but rather tracing out the entire phase space of many electrons from the full N-particle density matrix of a pure state. For a m-body reduced density matrix, Γ_m, this integration over $N - m$ electronic

variables can be written as

$$\Gamma_m(\mathbf{x}_1, \ldots, \mathbf{x}_m; \mathbf{x}'_1, \ldots, \mathbf{x}'_m) = m! \binom{N}{m} \times \int \Gamma_N(\{\mathbf{x}\}_N; \{\mathbf{x}'\}_N) \Delta^N_{m+1} d\mathbf{x}_{m+1} \ldots d\mathbf{x}_N d\mathbf{x}'_{m+1} d\mathbf{x}'_N, \tag{4.2}$$

where $\mathbf{x}_i$ represents the combined spatial and spin coordinate for electron i, and $\Delta^N_{m+1} = \prod_{i=m+1}^N \delta(\mathbf{x}_i - \mathbf{x}'_i)$. Fortunately, these reduced-body density matrices can be directly computed, rather than requiring tracing from higher-rank density matrices. By projecting the electronic coordinates into a basis set, we can define the two-body RDM as

$$\Gamma_{ijkl} \equiv \langle \psi | \, \hat{a}^\dagger_i \hat{a}^\dagger_j \hat{a}_l \hat{a}_k \, | \psi \rangle \,, \tag{4.3}$$

with other rank RDMs defined equivalently, and where the indices $i, j, \ldots$ label spin-orbital degrees of freedom, $\hat{a}^{(\dagger)}_i$ are the fermionic annihilation (creation) operators, and we have omitted the explicit subscript denoting the rank of the RDM where it is obvious from the number of indices. In this example, the partial trace down to the one-body RDM can then be written as

$$\gamma_{ik} = \frac{1}{N-1} \sum_j \Gamma_{ij,kj}. \tag{4.4}$$

Despite tracing out large numbers of degrees of freedom, these two-body RDMs still contain all the information about a quantum system required for physical observables of interest which depend on (up to) pairwise operators, including the total energy. The rank of an operator defining a given observable determines the rank of the RDM required to compute its corresponding expectation value. For example, the electric dipole moment is a one-body quant-

ity, requiring the one-body RDM, while the Hamiltonian defining the energy is a two-body expectation value, requiring the two-body RDM to evaluate. Non-observable quantities of interest, such as entanglement entropies or mutual information, can also in general be computed from reduced-body density matrices [369].

Furthermore, using RDMs one can compute the probability of a given m-electron distribution, as the diagonal of the m-RDM. The sum over this distribution then gives the number of m-tuples of particles in the system, which can be used as a normalisation condition, e.g. the trace of the 2-electron distribution giving the number of pairs of electrons, as

$$\sum_{ij} \Gamma_{ij,ij} = \frac{N(N-1)}{2}. \tag{4.5}$$

Overall, these m-RDMs have all the information about the distribution and entanglement of m particles in a given state of an N particle system, which rationalises their use as method-agnostic, low-rank quantities in order to couple quantum systems described at different levels of theory.

In this work, we consider second-quantised Hamiltonians where spin symmetry is preserved, allowing further tracing out of spin degrees of freedom, defining the central spin-free two-body RDM of interest as

$$\overline{\Gamma}_{pqrs} \equiv \sum_{\sigma\tau} \Gamma_{p_\sigma q_\tau r_\sigma s_\tau}, \tag{4.6}$$

where $p, q, \ldots$ denote spatial degrees of freedom and σ, τ denote spin labels. Further permutational symmetries can be used which reduce the number of

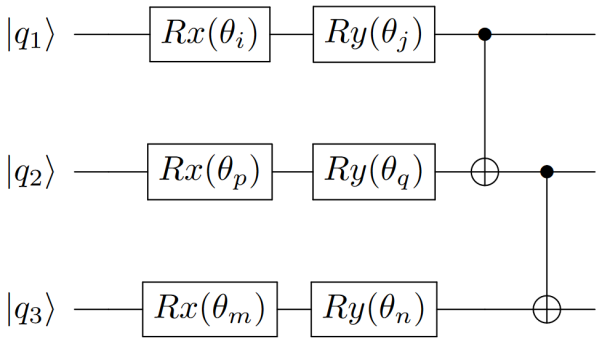


Figure 4.1: Illustration of one layer of HEA as used in this work. Each layer is composed of two rotation gates on each qubit and a ladder of entangling gates.

independent quantities to evaluate, as

$$\Gamma_{p_\sigma q_\tau r_\sigma s_\tau} = -\Gamma_{p_\sigma q_\tau s_\tau r_\sigma} = \Gamma_{q_\tau p_\sigma s_\tau r_\sigma} = \Gamma_{r_\sigma s_\tau p_\sigma q_\tau}, \tag{4.7}$$

with time-reversal symmetry ensuring

$$\Gamma_{p_\sigma q_\tau r_\sigma s_\tau} = \Gamma_{p_{\bar{\tau}} q_{\bar{\sigma}} r_{\bar{\tau}} s_{\bar{\sigma}}}. \tag{4.8}$$

The resulting set of fermionic operators must be mapped to spin operators for sampling of the state on a QPU. For this, we use the Jordan-Wigner mapping [97], though other mappings (e.g. Bravyi-Kitaev mapping [281, 282]) could be used, as long as the mapping is consistent. While each fermionic operator will in general be mapped to several spin operators, one can find several efficiencies to reduce this overall number of terms. We discuss scaling of the number of terms for RDM sampling using efficient grouping methods in Appendix A, which will be essential for scaling to larger numbers of qubits or for the extraction of beyond two-body properties.

In order to test RDM sampling on a quantum computer, we computed

the one- and two-body RDMs of magnesium porphyrin after optimisation of a hardware efficient ansatz (HEA) [262, 370] wave function using the gradient-free RotoSolve optimiser [114, 115, 270]. Unless specified otherwise, the HEA used in this thesis is composed of repeated layers of two rotation gates on each qubit followed by a ladder of entangling gates, as presented in Fig. 4.1. We use an active space restricted to 2 orbitals and 2 electrons (a 4 qubit Hamiltonian) for this test. It is worth noting that while the HEA is convenient for studying small systems due to its relatively small pre-factor, it is expected to have difficulties scaling to larger active spaces [109], where other VQE ansätze are expected to be preferred such as the unitary coupled-cluster [37, 80, 81, 319]. The resulting RDMs are then compared to the RDMs obtained via exact methods for the same active space, and the distribution of relative errors in the elements are shown in Fig. 4.2. We conducted this test on a simulated QPU with different number of shots, as well as current quantum hardware (IBMQ Athens QPU, details about each QPU used in this experiment can be found in Appendix B)).

In our investigation of RDM observable sampling errors, we also considered approaches to reliably mitigate for these errors via physically justified extrapolation techniques. It is clear from the results of Fig. 4.2 that these can substantially ameliorate quantum noise and sampling errors. The simple extrapolation technique used in this work for error mitigation is presented previously in Sec. 3.3.2; the technique relies on a binomial distribution of independent errors in the quantum circuit, and has the benefit of requiring no additional measurements (for a more involved approach to extrapolation, we recommend Ref. [335]).

This simple error mitigation technique significantly improves the overall

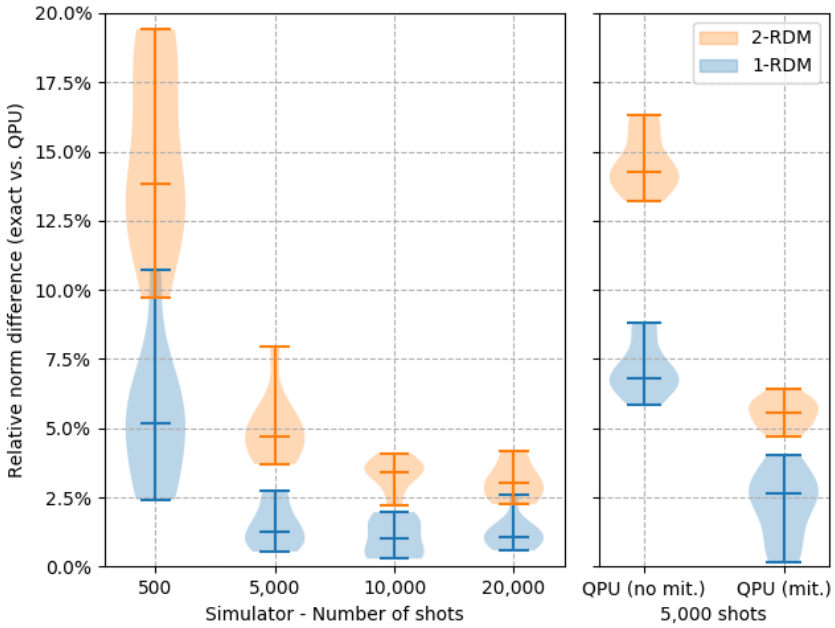


Figure 4.2: Distribution of relative errors in the Frobenius norm for both QPU and quantum emulated sampling of the RDMs, compared to exact classical calculation. To obtain the distributions, we repeated the computation of the Frobenius norms differences over 20 realisations of the RDMs for each number of measurements considered. On the left, results are presented for a QPU simulator (assuming perfect qubits) and therefore displays the impact of finite sampling noise. On the right, the results computed on IBMQ Athens (4-qubits, depth 3 HEA), with and without error mitigation.

accuracy of the QPU estimates, bringing it almost on par with the results of the simulator at an equivalent number of measurements. A key question that remains is whether the norm error presented above has a significant impact on the ability to use these RDMs reliably within subsequent quantum chemical calculations, where manipulation of these noisy RDMs may prevent convergence or lead to unacceptable bias in desired quantities. In order to test this we apply this sampling to a QPU-solved complete active space self-consistent field method (computing both energetics and molecular dipole moments following the optimisation), as well as a QPU version of the energy-weighted density matrix embedding theory, as examples of multi-scale approaches to allow quantum resources to be applied to realistic systems in electronic structure calculations.

4.3 Quantum CASSCF

The complete active space self-consistent field (CASSCF) approach is generally the starting point in quantum chemistry for molecular systems exhibiting stronger correlation effects, and therefore a key step in the development of electronic structure methods suitable for quantum computation [340, 344, 371–385]. As stated above, CASSCF is a post Hartree-Fock method in that it builds on an already computed mean-field Hamiltonian and ground energy to recover some of the electron correlation energy. As such it can be used as solver for any molecular energy problems but lands itself in particular to systems with a strongly correlated group of electrons in a limited range of orbitals. Applications are wide ranging, a few examples are presented in Ref. [386] which relate to describing surface hopping problems and include cis-trans isomerization, surface crossing with electron-transfer, and surface crossing with electron pair

transfer. As such the method can find numerous applications across quantum chemistry and condensed matter physics.

The central tenet of CASSCF is that the dominant strong quantum fluctuations required to qualitatively describe an electronic system are spanned by a small number of low-energy degrees of freedom about the chemical potential. The changes caused by explicitly considering interaction-driven virtual excitations in this space can change the occupation and induce entanglement of these orbitals, giving rise to correlated physics far from a mean-field description. The first step of CASSCF is therefore to partition the orbitals into three subspaces, denoted core, active and virtual. Core orbitals are deep-lying orbitals, which are considered to be chemically-inert and fully occupied, while conversely, the virtual orbitals are considered high-energy states which remain unoccupied. Together, these denote the 'external' space. The active space denotes the degrees of freedom which are considered to span the dominant electron correlations corresponding to low-energy virtual excitations of the N_{act} electrons within it, with the full set of quantum fluctuations among this set to be considered. No entanglement or particle/spin fluctuations are considered between the external and active spaces. The overall CASSCF wave function at any point can therefore be written as

$$|\Psi_{\text{CASSCF}}\rangle = |\psi_{\text{active}}\rangle \otimes \det[\phi_c], \tag{4.9}$$

where $|\psi_{\text{active}}\rangle$ denotes an N_{act}-electron wave function spanning the active degrees of freedom, while $\det[\phi_c]$ is a single product state over the core orbitals, accounting for the $N - N_{\text{act}}$ remaining electrons.

A key initial step for CASSCF is therefore to choose the orbitals in each set. These are selected from an initial mean-field calculation, where to a first

approximation, the highest-energy occupied and lowest-energy virtual orbitals about the chemical potential are chosen as the active space. However, this choice is often augmented with other criteria for selection of the active space, including symmetry, locality and/or 'chemical intuition', with approaches for automatic selection of this space, e.g., from quantum information arguments, a source of recent developments [387–389]. However, it is clear that selecting these orbitals from an initial mean-field calculation has an inherent flaw. The active space, designed to capture the strong correlations and dominant entanglement between single-particle states, is chosen from a theory with no correlations or entanglement via simple mean-field orbital energetics, which can change substantially in the presence of electron correlation. To account for this, a self-consistency in the choice of the active space orbitals is required for meaningful and qualitatively accurate results in the presence of strong correlation. This involves a variational optimisation of the state given in Eq. 4.9, to account for an arbitrary mixing between all three classes of orbitals, defined by the exponential of an anti-hermitian one-particle operator. This allows the character of the orbitals to change, by rotating core and virtual components into the active space in a variationally optimal way.

The CASSCF method from another perspective can be considered as an embedding of the correlated effects of the active space into a mean-field description of its 'environment' (as given by the electrons in core orbitals), as presented in Ref. [352] for quantum emulation. However, this active 'embedding region' is chosen largely on energetic criteria, with a strictly separable form and no entanglement with the core electrons. We contrast this with an alternative criteria based on locality in Sec. 4.4. The limitations of the approach come from the size of the active space, which for an exact treatment is often accepted to be 16 electrons in 16 orbitals [344], with some instances

of computation up to 20 electrons in 20 orbitals [164]. This is due to the exponential scaling of classical resources with respect to this size in order to represent and optimise $|\psi_{\text{active}}\rangle$. Beyond this, approximate descriptions of the active space wave function are increasingly being investigated, although all have their limitations [176, 181, 390–393]. This active size constraint hinders the application of CASSCF to systems with larger valence spaces, where a small active space is not sufficient and convergence of desired properties with respect to active size is not reached.

This limitation is a potential opportunity for NISQ computers to exhibit a quantum advantage in this keystone method in quantum chemistry, with the active space paradigm often being touted as a near-term prospect for quantum computers [349, 387] (see e.g. Ref. [82] for a recent review of the limits of classical computers in this field and the requirements for quantum advantage). However, beyond simple analysis of gate depth and qubit number required is the question of the practical feasibility of a robust and convergent algorithm for the self-consistency of the full CASSCF method, which has been demonstrated for a single orbital optimisation step without full self-consistency in the work of Takeshita *et al.* [349]. In the algorithm which we use, the coupling of the active space correlations to the orbital rotations required for self-consistency is provided by the two-body RDM within the active space. Therefore, the faithful sampling of this two-body RDM with sufficient fidelity is critical for a well-behaved algorithm. This is especially important as the orbital rotation updates involve non-linear functionals of the sampled two-body RDM elements, meaning that we expect noise from the QPU sampling to manifest as systematic error in the final results, even in the case that the sampling of the underlying RDM elements is unbiased. We investigate the two-body RDM active space sampling for this purpose on QPUs as well as the

importance of error mitigation, by using a parameterised gate circuit as the active space wave function optimised via the variational quantum eigensolver (VQE). However, the use of VQE in this work could be replaced by quantum Krylov or imaginary-time solvers suitable for NISQ devices [293, 394], as well as quantum phase estimation algorithms when suitable devices are available.

4.3.1 Fully self-consistent algorithm

We briefly summarise the key steps of the (two-step) CASSCF approach (sometimes also described as the related multi-configurational self-consistent field method), with more details available in Ref. [341]. We start with the second quantised electronic Hamiltonian in a basis, as

$$\hat{H} = \sum_{ij} h_{ij} \hat{a}_i^\dagger \hat{a}_j + \frac{1}{2} \sum_{ijkl} g_{ijkl} \hat{a}_i^\dagger \hat{a}_j^\dagger \hat{a}_l \hat{a}_k + E_{\text{nuc}}, \tag{4.10}$$

where h_{ij} and g_{ijkl} are the one and two-body integrals respectively, with E_{nuc} the scalar nuclear repulsion. We can parameterise the orbitally-optimised CASSCF wave function of Eq. 4.9 as

$$|\Psi_{\text{CASSCF}}\rangle = |\mathbf{R}, \mathbf{c}\rangle = e^{-\sum_{ij} R_{ij} \hat{a}_j^\dagger \hat{a}_i} \sum_n c_n |n\rangle \,, \tag{4.11}$$

where the one-body matrix operator $\mathbf{R}$ parameterises the single-particle unitary rotation operator of the molecular orbital basis ($R_{ij} = -R_{ji}$), $|n\rangle$ the complete set of Slater determinants spanning the active space, and $\mathbf{c}$ defines the coefficients of the configurations indexed by n, spanning the selected active

space. The full optimisation problem can then be written as

$$E = \min_{\mathbf{R},\mathbf{c}} \frac{\langle \mathbf{R}, \mathbf{c}| \, H \, |\mathbf{R}, \mathbf{c}\rangle}{\langle \mathbf{R}, \mathbf{c}|\mathbf{R}, \mathbf{c}\rangle}. \tag{4.12}$$

Within the two-step algorithm, the optimisation of $\mathbf{R}$ and $\mathbf{c}$ are treated separately and alternated, as the optimisation of $\mathbf{R}$ can be efficiently performed for one-body unitary rotations on classical computers, given the knowledge of the active space two-body RDM,

$$\Gamma_2 = \langle \psi_{\text{active}} | \hat{a}_i^\dagger \hat{a}_j^\dagger \hat{a}_l \hat{a}_k | \psi_{\text{active}} \rangle. \tag{4.13}$$

The optimisation of $\mathbf{R}$ then proceeds via construction of the gradient and Hessian of the energy with respect to these parameters, which can then be updated at modest computational expense via a quasi-second order step, accelerated with iterative subspace methods as implemented in the `PySCF` package [341, 395, 396].

For a given rotation matrix parameterised by $\mathbf{R}$, the Hamiltonian can then be transformed into the new basis, and the Coulomb and exchange contribution from the static core electrons integrated out, resulting in an active space Hamiltonian, $H_{\text{act}}(\mathbf{R})$, which only spans the active space degrees of freedom. The optimisation of this active space wave function is then amenable to implementation within a VQE minimisation, as

$$E|_{\mathbf{R}} = \min_{\theta} \langle \psi_{\text{active}}(\theta)| \, H_{\text{act}}(\mathbf{R}) \, |\psi_{\text{active}}(\theta)\rangle \, , \tag{4.14}$$

where θ denote the angles to optimise within the chosen quantum circuit parameters [37]. Once optimised, the 2RDM elements of the active space of Eq. 4.13 can be sampled, in order to update $\mathbf{R}$ in the full space, until convergence. In

practice, convergence can be triggered in a black-box fashion over a number of variables, such as the energy, orbital gradients, or density matrices themselves. However in this work we run for a fixed number (ten) of orbital updates, which is sufficient to gauge convergence and to subsequently observe the fluctuations in the quantities of interest at this point.

4.3.2 Results

While the computational procedures for this coupled orbital optimisation are well developed for exact or near-exact solvers in the domain of quantum chemistry, their utility in a fully self-consistent algorithm with a noisy quantum computer is far from clear (although there is some relevant recent work on noisy Monte Carlo solvers for active spaces [390, 391, 397, 398]). We therefore consider the CASSCF algorithm with an active space NISQ device solver, to determine the stability of the algorithm in the presence of sampling, gate noise, decoherence, and a parameterised quantum circuit for the state. This allows us to understand the feasibility of this multiscale approach, and develop practical strategies to ameliorate potential shortcomings from the noisy active space sampling.

We apply the method to a carbon monoxide (CO) molecule in a cc-pVDZ basis set, and at a stretched bond length of 1.54 Å. This stretching of the multiple bond enhances the strong correlation in the electronic structure, as the atomic-like character of the constituent atoms is increased. An active space of two orbitals and two electrons, corresponding to the highest occupied and lowest unoccupied molecular orbitals, is selected to capture the dominant many-body entanglement in these lowest-energy quantum fluctuations. To ensure a significant level of orbital relaxation from the self-consistent proced-

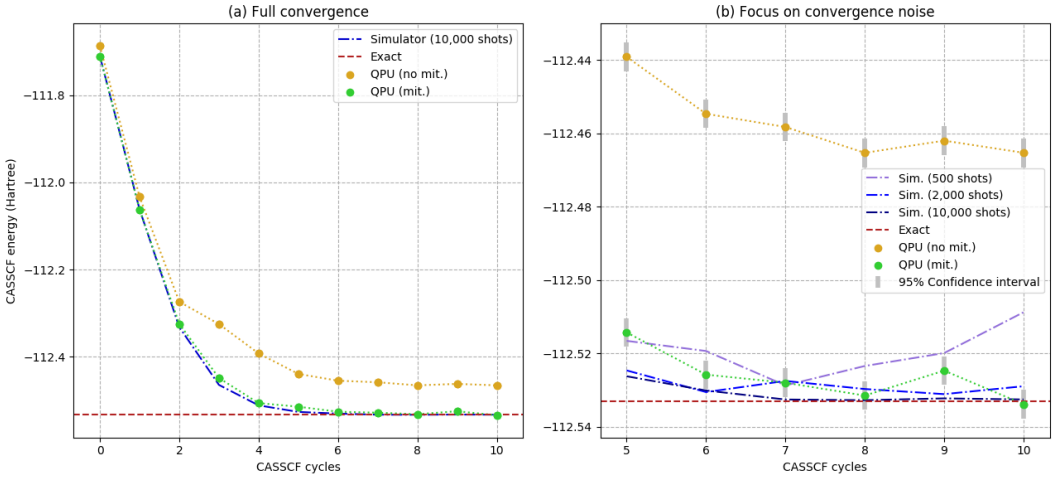


Figure 4.3: Convergence of the energy of the CASSCF state for a 2-electron, 2-orbital active space of Carbon Monoxide for each orbital update step. Results are shown for a quantum simulator, on IBMQ Bogota (without error mitigation) and on IBMQ Santiago (with error mitigation). Plot (b) focuses on the final five orbital update cycles, showing the variation and bias in converged results, with additional simulated results for 500 and 2,000 shots to illustrate the impact of finite sampling noise (results for 5,000 shots are indistinguishable from those obtained with 10,000 shots and as such were not included). Error bars for QPU results on plot (b) represent 1.96 times the standard error spread of the measurement data, or the 95 % confidence interval for the values estimated.

ure, and to test the stability of this noisy optimisation in the case of a poor initial choice of orbitals, we can select initial orbitals (and active space) from only a partially converged Hartree-Fock calculation. This was achieved by an early stopping of the mean-field self-consistent field procedure after only two updates of the Fock matrix prior to the CASSCF.

We first implemented the method on a quantum simulator with 500, 2,000, 5,000 and 10,000 shots to sample each mapped two-body RDM operator required, but in the absence of any additional noise model for the gates. For the quantum hardware experiments, we use IBMQ Bogota and IBMQ Santiago, which are both 5-qubit QPUs available through the the IBMQ platform, with equivalent levels of gate fidelity (details about each QPU used in this paper

can be found in Appendix B). The initial calculations on IBMQ Bogota were performed without accounting for any error mitigation, before applying the light-touch error mitigation strategy presented in Sec. 3.3.2 on the IBMQ Santiago hardware to assess any improvements from this. For the QPU runs, we use a measurement ramp-up schedule whereby the number of measurements is increased if the output energy at a given iteration is higher than for the previous one (which should not be the case during the optimisation). It is capped at 8,000 shots, which is also the number of measurements used for RDM sampling after the state is optimised.

We used the same ansatz for all experiments, built on a four-qubit, three-layer version of the HEA [262, 370], and the same optimiser: the gradient-free RotoSolve method [114, 115, 270, 271] . This resulted in a total of 24 variational parameters in the model. We found it unnecessary to fully converge the ansatz each iteration, and therefore investigated varying the level of ansatz optimisation each CASSCF step to improve efficiency. Five iterations of the VQE were in general sufficient on the first cycle, and we then used the parameters obtained to initialise the ansatz for the next cycle. A single iteration of the VQE for subsequent CASSCF steps after performing this warm start was sufficient to fully converge in a reasonable time, and to reach good accuracy.

The results of these CASSCF optimisations are presented in Fig. 4.3. Without error mitigation, the QPU results show significant systematic error at convergence of ~ 60 $\mathrm{mE_h}$, but nevertheless allow for a stable optimisation. Including the error mitigation allows for significantly better results, with fluctuations of less that 10 $\mathrm{mE_h}$ from the exact CASSCF value from exact 2-step optimisation of the same initial active space. As expected, the variance from the QPU experiment is significantly more than the corresponding quantum simulated results, even with error mitigation. This reflects the fact that the

error mitigation effectively removes the bias in the sampled measurements, but does not materially improve on the variance resulting from quantum noise. In our quantum simulated results, we find strong convergence for any simulation without gate noise or decoherence above 5,000 shots. Below that number, finite sampling noise prevents the algorithm reaching the sought after solution. At 500 shots, it fails to reach under 10 mE_h difference to the target state energy on average.

We can distinguish and isolate the effects of certain errors arising from the quantum solver on these results. Firstly, we have the systematic error in the VQE at each iteration, including the optimisation, gate errors and ansatz choice, which lead to a non-exact energy and state for a given set of active orbitals. Secondly, we can consider the effect of stochastic noise in the RDM due to a finite number of samples. This second error will lead to incorrect orbital updates in the CASSCF macroiterations, and a loss of precision in the final CASSCF energy due to an inability to propagate to the optimal orbitals defining the active space and its Hamiltonian. Furthermore, since the orbital choices in CASSCF are not linear functionals of the sampled density matrix elements, even if the RDM elements are entirely unbiased and correct on average, this does not preclude a systematic error entering the orbital updates at any finite sampling.

To separate these sources of error, we can consider the exact energy of each CASSCF iteration, but using the active space orbitals obtained at each iteration from the noisy VQE update from the quantum solver. This eliminates errors due to the VQE optimisation of a given active space, isolating the error due to convergence of the non-optimal orbital rotations being found at each step, primarily due to the inherent sampling noise of the RDMs. These results are shown in Fig. 4.4, and show that the overwhelming majority of the error is

arising from the bias in the VQE, while the convergence of the orbitals is highly robust to the errors in the active space VQE description and sampling errors of the RDMs. Even without error mitigation of the RDMs, orbital optimisation is accurate to within 10 mE_h, while the results of the simple error mitigation in the RDMs rendered an almost exact CASSCF energy. This demonstrates that the orbital optimisation procedure is less susceptible to the errors in the VQE and RDM sampling than the inherent errors in the energy and wave function optimisation for a given active space. This relative insensitivity to the fidelty of the RDMs bodes well for larger active space calculations on QPUs and the practicality of orbital optimisation through RDM sampling, as well as the improvements which would transfer to this approach from improved active space quantum algorithms [293, 394].

A key question remains as to whether this robustness is a property just of orbital optimisation, or whether this also extends to a broader set of expectation values which can be derived from the RDM sampling (other than the energy), as these also relax due to a more faithful description of the correlated wave function. We consider here the effect of orbital optimisation and a correlated VQE wave function on the magnitude of the dipole moment of the carbon monoxide molecule in the same active space, which can be extracted from the sampled one-body reduced density matrix as a one-particle expectation value. This quantity characterises the net charge distribution in the molecule, and from symmetry constraints can be described by a vectorial quantity which must be coincident with the carbon-oxygen bond. The magnitude of this vector is shown in Fig. 4.5, as the active space orbitals are optimised in the presence of the correlated VQE state.

Since the dipole moments are linear functionals of the one-body RDM, we would expect an unbiased sampling of the RDM to give an unbiased estimate

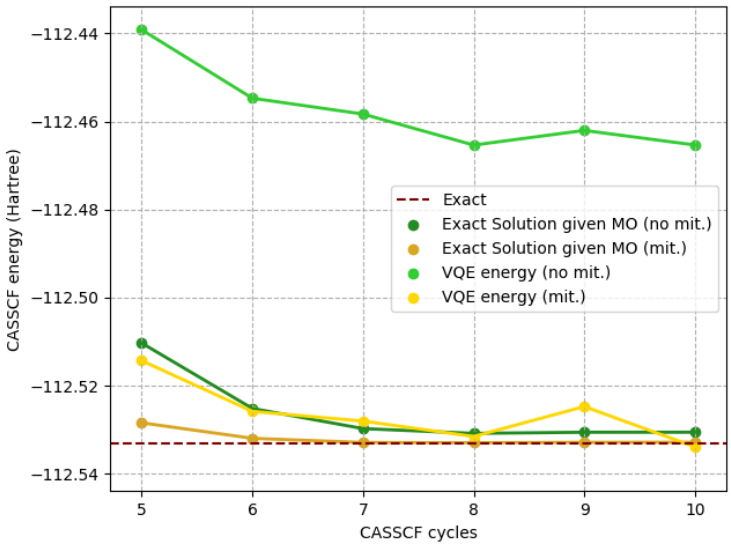


Figure 4.4: Convergence of the final six orbital update cycles of the CASSCF energy, obtained with the IBMQ Bogota (with no error mitigation) and IBMQ Santiago (with error mitigation). Two series are presented for each QPU calculation: 'VQE energy' results are equivalent to Fig. 4.3, while 'Exact solution' represent the exact energy from the current active space, that would have been obtained if the VQE was solved perfectly given the molecular orbitals found from the previous quantum VQE update step.

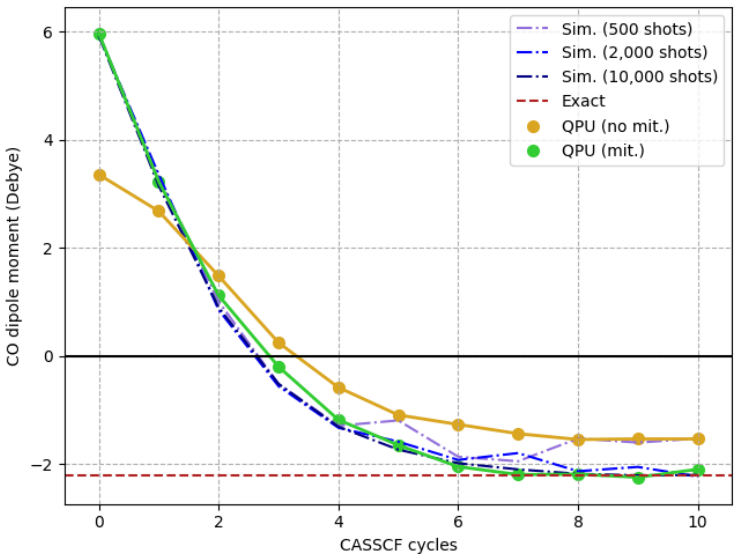


Figure 4.5: Convergence of the dipole moment in Debye from QPU-CASSCF as the orbitals are optimised, both with (IBMQ Santiago) and without (IBMQ Bogota) error mitigation, as well as quantum simulated results from an RDM sampling of 500, 2,000 and 10,000 shots. Positive dipole moments mean that the dipole moment points towards the oxygen (i.e. the oxygen atom has net negative charge), while the converged results flip the orientation of the dipole moment.

of the dipole moment from the optimised VQE-CASSCF state. We find that without error mitigation, there is still an error of $\sim 0.6D$, however the error mitigated results can effectively reduce the systematic error in the final dipole moment completely, with fluctuations in each cycle of a similar magnitude to the emulated values without quantum noise or decoherence with 5,000 shots. It is also worth noting that the dipole moment of carbon monoxide is notoriously difficult to converge theoretically [399], hence the error with respect to experiment of 2 Debye is to be expected for CASSCF with a minimal active space. At this point, the fluctuations in the dipole moment agree with the magnitude of the fluctuations expected from the original density matrix sampling experiments of Fig. 4.2, and an unbiased estimate of the exact CASSCF dipole moment is obtained. We also note that the correlation and orbital optimisation reverses the direction of the charge imbalance in this system from the starting description.

The overall runtime of the full QPU-CASSCF calculations on IBMQ Bogota was ∼14 QPU-hours, including 10 orbital updates (or cycles), VQE optimisation and RDM sampling. Each energy evaluation was done measuring 49 operators (see Appendix A) with 8,000 shots. Each of the 24 parameters is sampled 3 times to implement the RotoSolve algorithm hence each inner loop takes $8,000 \times 49 \times 3 \times 24 \sim 3 \times 10^7$ shots. We performed 5 iterations in the first cycle and 1 iteration in the subsequent 9 cycles, at the end of each cycle the full two-body RDM are sampled to compute the final values ($8,000 \times 49$ shots). All together, this results in a total of $\sim 4 \times 10^8$ shots. Of course this number of shots can be optimised further, through more efficient operator grouping, measurement weightings and other methods. Likewise, the overall QPU runtime is also worsened by latency from accessing the device through the cloud, and regular device calibrations which can take a significant amount

of time.

However, one must consider the potential for parallelisation. In this case, all 8,000 measurements conducted on the 49 operators could have been conducted in parallel, possibly reducing the overall runtime up to well under a second. While this is not a good indication for scaling and long term viability of the method (we encourage the reader to refer to Ref. [82] for an interesting assessment of the scalability of VQE), it does illustrate the potential for strong parallelisation, and corresponding error mitigation techniques, for the viability of NISQ algorithms.

Finally, it is worth discussing the viability of extensions to CASSCF on quantum devices. In quantum chemistry, CASSCF is rarely the end of the story, as it neglects the contributions to expectation values arising from interactions between the active space electrons and the external degrees of freedom. These can generally be treated at a perturbative level of theory[400], cumulant or energy-moment expansions [401] or via subspace expansions[349], and are required for quantitative accuracy for predictive calculations. These perturbative couplings between the spaces can be computed by considering the higher-rank RDMs in the active space. This approaches will significantly increase the number of terms which must be sampled. However, large reductions can be found with the appropriate groupings, and this is demonstrated in Appendix A, where 440,154 Pauli strings for the sampling of the four-body RDM within 6 qubits can be reduced to only 3,182 commuting groups.

4.4 QPU-Enhanced Energy-weighted Density Matrix Embedding

The CASSCF method exploits the locality of correlation in the energy domain, choosing and optimising a low-energy subspace for the correlated treatment. In this section, we demonstrate the utility of a faithful QPU sampling of RDMs in order to correlate and optimise a different subspace, which instead relies on spatial locality. This perspective is often more useful for strongly-correlated extended systems, where the atomic-level correlated degrees of freedom can be isolated, and where widely used methods such as density functional theory fail to provide accurate results [402]. These approaches fall under the umbrella of quantum embedding or quantum cluster methods, and are amongst the most promising for QPU-enhanced materials modelling [362, 403, 404]. We investigate the recently-developed 'Energy-weighted Density Matrix Embedding Theory' (EwDMET) as a promising candidate in this direction [345–347].

The EwDMET method connects the density matrix embedding theory (DMET) and dynamical mean-field theory (DMFT), two established approaches in quantum embedding [356, 405, 406]. Both of these 'parent' approaches have recently been adapted for use with a quantum hardware solver, as well as related embedding techniques [357–359, 363, 365, 407–410]. However, the EwDMET avoids a number of difficulties. Similar to DMET, it avoids any necessity to compute the single-particle Green's function of the resulting quantum cluster problem on the QPU, which is challenging for quantum hardware, although important progress is being made [411]. Instead, the method requires a desired number of one-particle spectral moments from the subspace problem, which can be obtained directly from the reduced density matrices of the ground

state. The number of self-consistent spectral moments can then be systematically enlarged, to approach the complete dynamical character of DMFT as an orthogonal polynomial expansion. The method also removes all explicit numerical fitting steps, and constructs a rigorous self-consistency on these spectral moments, systematically extending the DMET formulation and connecting it to its fully dynamical limit. This rigorous and algebraic self-consistency enables non-trivial results to be obtained even at the lowest truncation of the spectral moment expansion. This requirement of only computing ground-state RDMs, while at the same time benefiting from a rigorous and algebraic self-consistency for non-trivial emergent physics makes it an ideal candidate for combination with QPU-derived RDMs in the NISQ era. We briefly review the salient features of EwDMET for this QPU formulation, with more details in Ref. [347].

While EwDMET remains a fairly recent method, it focuses on similar applications to DMET which has been applied and tested in numerous studies of condensed matter systems. For example, DMET has been used to study the one-dimensional Hubbard model [405, 412–414], the Hubbard–Holstein model [415], the Hubbard–Anderson model [416], the two-dimensional spin-J1–J2 model [417]. DMET has also been, in some occasions, tested on molecular energy applications such as the study of carbon polymers, boron–nitride sheets, crystalline diamond [418] and hydrogen rings and sheets [354]. In general, the family of methods, from DMET to DMFT, has widely been used for the study of strongly correlated materials and metals, in particular when computationally more efficient methods such as density functional theory fail.

As with all quantum cluster approaches, the algorithm begins with the choice of a local correlated space. This could be the $d-$shell of a transition metal atom, or a cluster of sites for a discrete lattice model. The EwDMET

method then allows for an improvable and self-consistent description of the one-particle quantum fluctuations between this fragment and its environment [345–347]. This information is contained within the self-consistently optimised (hole and particle) spectral moments of the fragment, defined as

$$T^{(n)}_{h,\alpha\beta} = \langle\Psi|\, \hat{a}^\dagger_\alpha (\hat{H} - E_0)^n \hat{a}_\beta \,|\Psi\rangle\,, \tag{4.15}$$

$$T^{(n)}_{p,\alpha\beta} = \langle\Psi|\, \hat{a}_\alpha (\hat{H} - E_0)^n \hat{a}^\dagger_\beta \,|\Psi\rangle\,, \tag{4.16}$$

where α, β index the degrees of freedom of this local fragment, $n \geq 0$ denotes the order of these moments, optimised up to a maximum desired value n_{mom}, and $|\Psi\rangle$ is the ground state of the constructed correlated subspace. As $n_{\text{mom}} \to \infty$, the method exactly reproduces the effective dynamics of DMFT, recast as a ground-state wave function theory, while systematic truncation to low n_{mom}will still faithfully describe the dominant low-energy fluctuations from the fragment into its environment. The EwDMET method rigorously maps the full system to a subspace consisting of the chosen fragment coupled to a 'bath' space . The size of the bath is determined solely by the size of the fragment and the desired number of spectral moments to capture (and the correlated subspace is hence independent of the size of the full system). It is this correlated subspace problem which must then be solved on the QPU at each iteration, and the spectral moments of Eqs. 4.15 and 4.16 computed. With these computed spectral moments from the correlated fragment space, the one-particle description of the full system can be algebraically updated via the addition of non-interacting auxiliary states, to ensure that the fragment moments at the mean-field level over the full system exactly reproduce the

correlated subspace ones. The procedure is iterated, updating the auxiliary space and bath space of the quantum cluster problem, until convergence.

4.4.1 Infinitely coordinated Bethe-Hubbard Lattice

We apply this method to the paradigmatic Hubbard model of condensed matter physics, which describes a range of quantum phases and correlation-driven transitions. Specifically, the limit of an infinitely-coordinated extended Bethe-Hubbard lattice with local interactions defines our model of interest, which has the particular feature that correlation-driven changes to all one-particle properties are site local. This property was used to great effect to motivate the development of DMFT, by providing a non-trivial model for which it describes an exact limit [419, 420]. The EwDMET has the same exact limit for this model as $n_{\text{mom}} \to \infty$.

The Bethe lattice can be used for numerous models, and is defined as a graph, without loops, of infinitely many points each connected to a fixed number of neighbours. This number is usually labeled z and called the coordination number. In this experiment, we consider a model where $z \to \infty$ akin to the model presented in Ref. [347]. A version of the Bethe lattice for $z = 2$, which represents a simple chain, is described in Fig. 2 of Ref. [347]. An example of a Bethe lattice for $z = 4$ can be found in Ref. [421].

The model can be equivalently defined in this infinite-dimensional limit via its metallic non-interacting density of states [421], which is defined to have the form

$$A(\omega) = \frac{1}{2\pi}\sqrt{4 - \omega^2}, \tag{4.17}$$

for a bandwidth of $|\omega| < 2$. This non-interacting spectrum was fit to a single central fragment site with 200 additional degrees of freedom, to approxim-

ate this full spectrum to a high energy resolution [422, 423]. The interacting Hamiltonian is then defined as resulting from the additional on-site Hubbard interaction term, $U\hat{n}_{i\uparrow}\hat{n}_{i\downarrow}$, which is included on the fragment in the correlated subspace Hamiltonian at each iteration. The spectral moments of this central fragment site are then self consistently optimised, where we define the projection of the non-interacting system Hamiltonian into this cluster subspace as h^{clust}.

In this work, we truncate the spectral moment expansion at order $n_{\text{mom}}= 1$, defining the set of self-consistent fragment quantities. This simplifies their computation from the VQE solution for the ground state of the cluster Hamiltonian at each iteration, since these $n = 0$ and $n = 1$ moments over the fragment can be constructed from (parts of) the one- and two-body RDMs, for which we have efficient sampling as detailed previously. For instance, the $n = 1$ hole moment reduces to

$$T^{(1)}_{h,00} = \sum_{j\in\text{clust}} h^{\text{clust}}_{0j}\gamma_{0j} + U\Gamma_{0000}, \tag{4.18}$$

where 0 denotes the fragment site index. Physically, the restriction of $n_{\text{mom}}= 1$ means that the center of mass of the particle and hole spectral distributions can be self-consistently optimised on the lattice in the presence of the local correlation effects. This contrasts with DMET, where single site self-consistency cannot change the physics of the full system from the non-interacting picture in translationally symmetric systems, and so QPU emulations of this method with a single fragment site are restricted to single-shot computation without any self-consistency [357, 358, 410]. By using VQE as the solver on quantum hardware, we can identify the ground state of the cluster Hamiltonian, and subsequently sample the relevant RDMs to construct the required fragment

spectral moments. We iterate this procedure until self-consistency, which we define to be when the sum of the squared update to the (four) parameters defining the auxiliary states varies by an energy of less than 10^{-4}. These self-consistent auxiliary states consist of individual poles in a self-energy which modifies the spectral function of the system to match the correlated local moments from the VQE.

At this choice of spectral moment truncation, the cluster Hamiltonian consists of the single fragment site and a single bath orbital, resulting in a four-qubit system to solve at each iteration of the EwDMET method. This cluster is solved via VQE on the QPU with the same three-layer HEA as applied in the CASSCF section, with a Jordan-Wigner mapping to the qubit representation. Additionally, the same error mitigation is used to control the noise inherent in the sampling of the RDMs required to construct the fragment spectral moments. Emulated QPU simulations without noise models were also performed for comparison to the QPU experiments of this algorithm.

4.4.2 Results

Fig. 4.6 presents results for the single-particle spectrum for the model at self-consistency for VQE-EwDMET with $n_{\text{mom}}= 1$, at a strongly correlated limit with an on-site interaction of $U = 8$, which is twice the non-interacting bandwidth of the material. Self-consistency achieves the matching of the first two local spectral moments (Eqs. 4.15 and 4.16 for $n = 0$ and 1 over the fragment site) for both the mean-field state of the whole system, and the VQE results over the correlated cluster. At this point a spectrum can be obtained which is consistent through these local moments, via diagonalisation of the resulting (dynamical) mean-field, with specifics found in Ref. [347]. In brief,

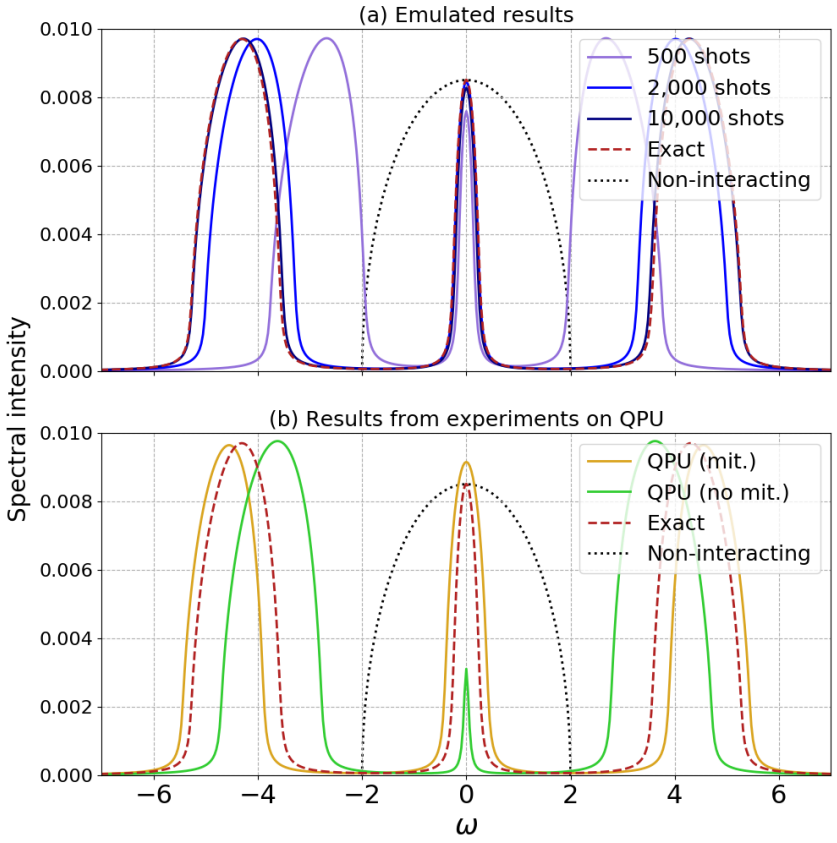


Figure 4.6: The density of states for the Hubbard model on the Bethe lattice with infinite coordination for the EwDMET method with $n_{\text{mom}}= 1$. (a) Upper results are performed on a classical QPU simulator with finite sampling noise and varying numbers of shots in the VQE solution to the cluster Hamiltonian at each iteration. (b) Lower panel shows results from two QPU experiments, on IBMQ Santiago and IBMQ Bogota, with and without error mitigation respectively, with 5,000 shots for the sampling of the required RDMs. Grey dotted lines show the original non-interacting spectrum of the model, while the red dotted line shows the EwDMET($n_{\text{mom}}= 1$) results with an exact solution of the cluster Hamiltonian each iteration.

if we note ϵ_p the eigenvalues of the dynamical mean-field the density of state presented in Fig. 4.6 can be computed using the following spectral function:

$$A(\omega) = -\frac{1}{\pi}\mathrm{Im}\left[\sum_p \frac{1}{\omega - \epsilon_p + i\eta}\right], \tag{4.19}$$

with η an arbitrary number close to 0 (for broadening of the spectra).

The final converged spectra are presented for an exact solver, an emulated quantum simulation without noise but with different numbers of shots for the sampling of the RDMs (500, 2000 and 10,000), and quantum hardware results on IBMQ machines with and without error mitigation, with 5,000 shots for the sampling. All calculations converge within 8 iterations as the auxiliary and bath spaces adapt to the correlations described over the fragment site in each VQE cluster solution, demonstrating the robustness of the self-consistency in the presence of noise. At this level of theory, the spectrum shows significant qualitative changes from the non-interacting spectrum. Upper and lower Hubbard bands develop, splitting the original density of states, with a qualitatively correct charge gap between these bands shown. However, a small quasiparticle peak remains at the Fermi level, showing that the metallic character of the system is not entirely removed by the correlations, as is to be expected from numerically exact calculations on this system such as NRG+DMFT [424]. Consistency in higher orders of the spectral moments are required to get to a true Mott insulating state [347], which can be achieved at the expense of an increasing size of bath space and sampling higher spectral moments (which requires higher-body RDMs, described in Appendix A). Nevertheless, even at this low truncation, much of the true correlated spectral density is reproduced with significant physical correlation-driven redistribution of spectral weight observed.

At 10,000 shots, the emulated results without further simulated quantum noise are are almost indistinguishable from the exact benchmark at all energies. At lower numbers of shots, the gap between the Hubbard bands is too small, as the variance in the sampled RDMs increases. As with the CASSCF method, the updated auxiliary space at each iteration is a non-linear transformation of the spectral moments (which are themselves linear functionals of the RDM elements, as can be seen in Eqs. 4.18). The consequence of this is a systematic error in the resulting spectral functions at convergence due to the increasing RDM variance, rather than simply a manifestation in a noisy but unbiased spectrum. This behavior of an underestimated gap between the Hubbard bands is also present in the QPU results, where unmitigated results feature unrepresentative Hubbard bands. However, the performance is once again considerably improved with the error mitigation, with the Hubbard bands and low-energy peak resolved to higher accuracy, suggesting the method fits well with a QPU cluster solver, and removing the necessity for the full solution of the fragment Green's function at each point within a DMFT framework.

Properties of the system can also be observed from the effective self-energy of the system, which is obtained directly from the self-consistent auxiliary states, and allows access to quantities such as Fermi liquid parameters [347]. For the same system, Fig. 4.7 shows the imaginary part of the self energy on the Matsubara frequency axis. The finite sampling results are seen to approach the exact results with increasing shots, with 10,000 shots reaching a comparable performance to the exact results, with the discrepancy far more visible in the self-energy than the original single-particle spectrum of Fig. 4.6. The quantum hardware unmitigated results correspondingly demonstrate significant overestimation of the resulting self-energy. Despite the fact that the Hubbard bands are closer, these unmitigated results show a larger effective

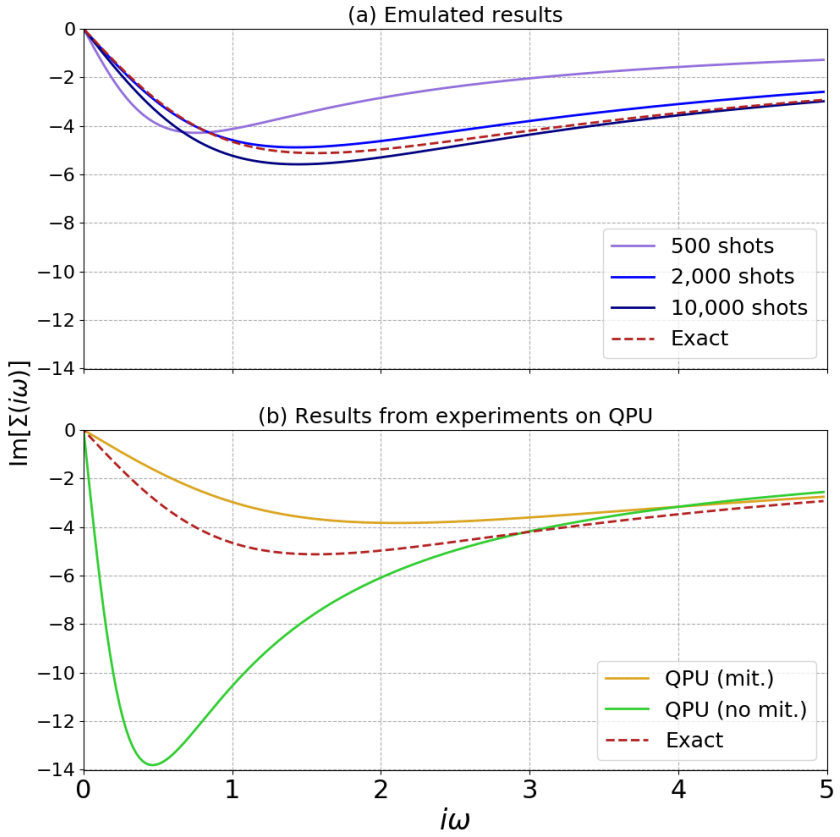


Figure 4.7: Converged effective on-site Matsubara self-energy for the Bethe lattice Hubbard model with EwDMET at $n_{\text{mom}}= 1$. (a) Upper plot shows results from the quantum simulator with finite sampling of the RDM elements, compared to exact results (infinite sampling). (b) Lower plot shows results from quantum hardware on IBMQ Santiago and IBMQ Bogota and 5,000 shots in the RDM sampling, with and without error mitigation in the RDM sampling respectively.

mass and quasi-particle renormalisation at the Fermi surface from the self-energy (a larger derivative at $i\omega \to 0$), which manifests in the smaller peak in Fig. 4.6 at that point. Error mitigated QPU self-energies are however more in line with exact results for the method, albeit now slightly underestimated at low-frequencies. Further improvement of the results can be obtained by increasing the moment order (n_{mom}) to which the dynamical quantities are all resolved. Similar to the perturbative corrections to CASSCF however, these will require sampling of higher-body RDMs, and is an avenue of continuing research.

4.5 Discussion on multiscale methods

We have presented a unified approach to self-consistent coupling of quantum and classical computational resources in quantum chemistry and condensed matter electronic structure problems. This coupling relies on the faithful and efficient sampling of reduced density matrices on quantum resources, where these objects span the correlated physics of an iteratively optimised subspace of the full system. We considered the required fidelity and sampling quality of these density matrices for robust optimisation on current generation quantum hardware, developing a simple but effective approach to mitigate for gate errors and allow this full convergence. As well as converged energetics, we also analysed the viability of the sampled density matrices for non-energetic quantities, including the dipole moment for *ab initio* simulation of chemical systems, and the self-energy and mass renormalisation of strongly correlated extended models.

Overall, the picture is encouraging, with the self-consistent optimisation found to be particularly robust to the presence of sampling noise on current

generation quantum hardware. This self-consistency is found to be more reliable than the uncertainties resulting from the state optimisation and energy obtained from the VQE at any single iteration. This points to a significant transfer from continuing improvements in both hardware and quantum algorithms for state preparation on quantum devices for self-consistent approaches. These conclusions however are restricted to relatively small correlated subspaces, and further work is required to understand the generality of these conclusions as we access QPU with larger qubit capacity. Furthermore, quantitative rather than qualitative accuracy in these application areas will require an efficient and compact description of higher-body density matrices, which will be the focus of future directions.

Chapter 5

Computation of molecular excited states

The computation of excited states is key to many processes in quantum chemistry and materials science, governing the dominant optical, transport and reactive properties [425, 426]. However, it is in general a significantly more challenging task than ground state computation, owing to the state generally being further away from a mean-field description, as well as less straightforward optimisation to avoid the variational collapse to the ground state. Conventional correlated quantum chemical approaches [427] include Equation of Motion (EOM) coupled-cluster [428], linear response theory [429], as well as multi-reference approaches for stronger correlation [209, 430].

5.1 Literature review of specific methods and experiments

Quantum computing methods can be broadly divided into two main types of methods, those that rely on computing excited states within a subspace, and fully variational methods relying on modification of the VQE cost function. We briefly review the core aspects of some of these approaches below.

Quantum Subspace Expansion: The quantum subspace expansion relies on finding an approximate Hamiltonian that spans a subspace of the full Hilbert space, but whose dimension is small and grows as only a low-order polynomial of the system size. The matrix elements of these Hamiltonians are sampled on quantum computers, but can then be tractably diagonalised on classical resources, with the higher-lying eigenvalues of these subspace Hamiltonians approximating true eigenvalues of the system. In practice, this approach starts with a ground state VQE calculation. From this ground state, it is then necessary to add additional states in order to define the span of a subspace into which the Hamiltonian can be computed. For reliable excited states, it is necessary to ensure that this space spans the dominant low-energy excitations of interest, as the whole spectrum will not be reproduced by construction. There are different approaches to choose these low energy states to span these relevant excitations, including approaches based on Krylov (or Lanczos) subspaces [237, 411, 431], and low-rank excitations of the ground state motivated by an equation-of-motion formalism [432, 433]. These approaches can also be used to yield improved ground state estimates [237, 434].

In the quantum subspace expansion based around the equation-of-motion expansion, the resulting Hamiltonian (and overlap) matrix between these states

can be found via high-order reduced density matrices evaluated from the ground state, as initially proposed in Ref. [432], and subsequently implemented on a quantum device [435]. The advantage of these methods is that they do not require particularly deep circuits to evaluate the relevant matrix elements of this subspace Hamiltonian. However, the quantum subspace expansion approaches can be quite sensitive to noise, while high-order density matrices can be expensive to sample and accumulate. Furthermore, noisy (yet unbiased) matrix elements can lead to systematic biases in eigenvalues [436, 437].

Variational approaches: An alternative approach relies on directly optimising an ansatz for specific excited states, using a modified cost function, which affords a fully variational flexibility, while maintaining orthogonality to lower-energy states. These have the advantage of not suffering from the limitations and biases of subspace expansion methods, but usually come at a higher cost in terms of quantum resources, and a restriction to a specific ansatz chosen. The simplest approach is to simply enforce symmetry constraints on the ansatz to a different symmetry sector to the ground state, in which case orthogonality to the ground state is guaranteed for the lowest-lying excitations in each symmetry [438]. This is however restricted to only specific excited states and limited by the symmetry of the system studied. Another approach which was proposed early on in the development of variational quantum algorithms (initially suggested for quantum computation in Ref. [38]), is to use the folded spectrum Hamiltonian [439]: $\hat{H}' = (\hat{H} - \gamma\mathbb{1})^2$, for which the ground state is now the eigenstate of $\hat{H}$ which has an eigenvalue closest to γ. It was applied by Liu *et al.* [440] as a mean to probe many-body localisation on a quantum computer. This method however implies squaring the Hamiltonian, which can result in a significant increase in measurements required if the operator is

dense, and requires prior knowledge of the eigenspectrum (which is somewhat less of a problem in the case of vibrational spectroscopy than in the case of electronic structure computation [68]).

The subspace search VQE (SSVQE) [441] leverages the fact that a unitary transformation between states cannot change the orthogonality of the states it is applied to. Therefore by preparing different orthogonal input states and training a VQE ansatz to minimise the energy of all these states at the same time (for instance by modifying the VQE cost function to include the sum of expectation values of the Hamiltonian with respect to each of the states, or by creating a mixed state using ancilla qubits), one can simultaneously learn the ground state and any number of subsequent excited states. It is likely however that this simultaneous optimisation of the ansatz becomes increasingly more constrained with the number of excited states desired.

Higgott *et al.* [442] proposed using a deflated Hamiltonian to iteratively compute successive excited states (oftentimes referred to as Variational Quantum Deflation, VQD). The algorithm works by first computing the ground state with VQE. Once discovered, the cost function is modified to add a penalty term, which corresponds to the overlap between the ground state and a new trial wavefunction. This new trial wavefunction is then trained to minimise both the expectation value of the Hamiltonian and maximise the overlap with the lower energy states. This process can be repeated iteratively for any number of excited states. The key challenge of this method is the computation of the overlap term which may require a quantum cost that could be significant for a NISQ device (i.e. a large number of SWAP gates), or possibly be subject to additional noise (by implementing as a circuit the complex conjugate of the ansatz used to prepare previous excited states, a method also applied in Ref. [80]), though improvements have been proposed. For instance, Jones *et al.*

[443] propose to compute the overlap term with a low depth SWAP test, and uses variational time evolution [293]. Chan *et al.* [444] extend this excited state method by merging it with ADAPT-VQE [213]. Kottmann *et al.* [445] independently also proposed an adaptation of VQD to an adaptive method which benefits from efficiency gained from gradient evaluation process presented in the same work. Wakaura and Suksomo [446] propose an adaptation of the VQE cost function to minimise the norm of the tangent vector to the energy rather than just the energy, dubbed Tangent-Vector VQE (TVVQE). While this can be used for ground state energies, it is also combined with VQD to compute excited states. While the method is shown to provide improved accuracy compared to a UCC based VQE on simple models (Hubbard, H_2, LiH), it is reported to require a run time on average five times longer than VQE [446].

An alternative approach is the Variance VQE method [447], which replaces the usual cost function of VQE by minimising the variance of a Hamiltonian with respect to a state, rather than its expectation value. The idea behind this method is that the variance of the expectation value of a Hamiltonian must be equal to zero if the state used to perform the measurement is an eigenstate of that Hamiltonian (on the zero-energy variance principle, we direct readers to Ref. [448–453]). Because all eigenstates have zero-energy variance, a simple approach will not guarantee convergence to a low-energy state. This problem is addressed in Ref. [447] by combining both energy and variance minimisation in order to allow for computation of low lying excited states. Zhang *et al.* [454] propose an adaptative variant of this method to computed highly excited states of Hamiltonians. The ansatz is grown by choosing operators from a pool of Pauli operators, akin to the method presented in Ref. [213].

5.2 Discriminative VQE

In this section, we present an alternative variational method to compute molecular excited states (the method is published in Ref. [455]). It is closest to the VQD proposed in Ref. [442] but uses a technique inspired from the generative machine learning literature to bypass the need to compute state overlap. It results in computation of molecular excited states without the use of SWAP gates, SWAP tests or circuit inversions, allowing for an implementation on current NISQ devices.

Description of the model

The Discriminative VQE (DVQE) relies on combining an orthogonality objective with an energy minimisation objective (also named VQE objective). At a high level, it aims at finding a state orthogonal to the ground state which at the same time is at a minimum of the Hamiltonian energy landscape. This will correspond to an approximation of the first excited state: the Hylleraas-Undheim and MacDonald [456, 457] theorem implies that the energy of a state orthogonal to the ground state (or any number of lower excitation states) acts as an upper bound for the next eigenvalue. Rather than directly minimising the overlap of the excited state of interest with the previous excited states and/or the ground state (as is done for instance in [442]), this method uses a combination of two quantum circuits working in tandem to learn parameterisation angles and reproduce unknown excited states. Our technique takes inspiration from Quantum Generative Adversarial Networks (QGAN). In a classical Generative Adversarial Network, an initial Generator network (denoted by G) is trained to fake an unknown data structure by learning how to fool a Discriminator network (denoted by D). The Discriminator is trained to

distinguish between the generated data structure and the unknown data structure. The QGAN is an adaptation of this algorithm where the data structure is replaced by a pure quantum state. The parameterised quantum circuit is trained to generate an approximation of an unknown pure state [458, 459].

In this case however, the logic of the QGAN is reversed. Instead of trying to fool the Discriminator, the Generator learns to create a state which makes it as easy as possible for the Discriminator to distinguish between a known quantum state (for instance, a simulated ground state) and the generated state. In effect, the Generator is identical to the ansatz circuit used for the VQE, although with different parameters. Borrowing from the QGAN logic, one can see that this change would result in producing a state which is as easily distinguishable from the known state as possible. In classical problems, this approach rarely makes sense. In quantum problems however, a state which has no overlap with a given reference state will be in the latter's orthogonality space. There are an infinite number of physically meaningful orthogonal states to a given quantum state. The VQE objective is used to guide the learning of the Generator towards a single orthogonal state. A state which is orthogonal to the ground state and at the same time minimises the energy of the entire orthogonal subspace must be the first excited state.

With this in mind, we believe the method proposed here offers the following advantages:

1. It is decisively NISQ friendly, requiring only rotation gates, entangling gates, and only one additional qubit compared to a VQE.

2. The orthogonality objectives rely on single qubit measurements, reducing exposure to read-out errors, and does not require computation of overlap terms which have been recognised as challenging for NISQ devices ([442]).

3. The method used to enforce orthogonality does not require perfect optimisation and therefore offers some resilience to quantum noise.

4. The excited state is directly and variationally minimised, rather than being inferred through non-linear postprocessing (as it is the case for example in analytic continuation of imaginary time or in subspace diagonalisation). This in turn reduces exposure to systematic bias in the estimation result.

5. Unlike some of methods outline above (in particular methods based on subspace expansion and its extention), the classical overhead is minimal and scales identically to the classical overhead of the VQE.

First consider a series of pure states $\rho_{s_i}=|s_i\rangle\langle s_i|$, with $i\in[0,n-1]$ representing adequate approximations of the ground and first $n-1$ excited states of a Hamiltonian H. It is assumed that we have a pre-trained quantum circuit that can produce these states using indexed parameters θ_i (which can be obtained using the VQE and previous iterations of this algorithm). We are now looking for a way to determine the n state: $\rho_{s_n}=|s_n\rangle\langle s_n|$.

For this, consider a state ρ_g generated through a parameterised quantum circuit applied to an initial state $|0\rangle\langle 0|^{\otimes d}$, and which is initiated as state ρ_{n-1}. We denote this Generator circuit as $G(\theta)$, with parameters θ. We have $\rho_g=G(\theta)|0\rangle\langle 0|G^\dagger(\theta)$.

Consider a Discriminator quantum circuit labelled $D(\phi)$, which is tasked with distinguishing between any of the known states and the output of the Generator. In order to accomplish this task, it takes as input either any of the known states, or ρ_g, randomly but with equal probability. It is followed by a Positive Operator Value Measurement (POVM). Because of the discriminative objective of the circuit, we can limit the required POVM outcome to only two

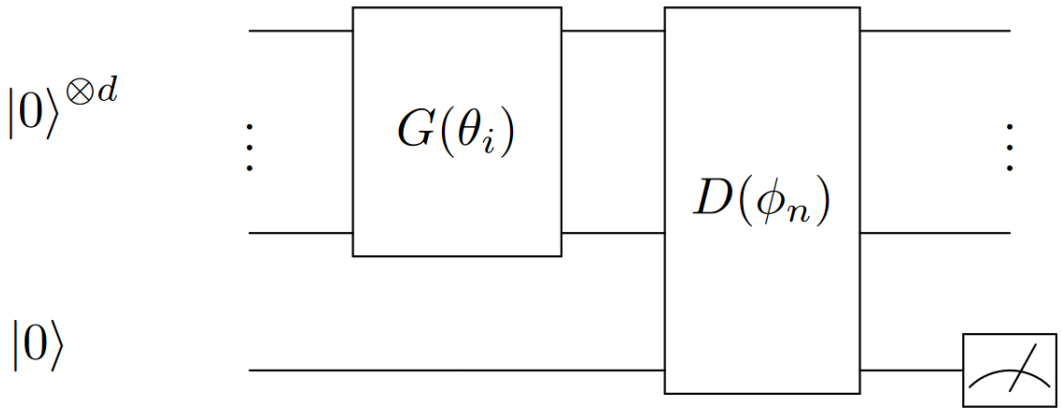


Figure 5.1: Discriminative VQE (DVQE) quantum circuit for computation of state n. The HEA is built using repeated layers composed of two rotation gates followed by entangling gates between nearest neighbour qubits, as described in [370] and illustrated in Fig. 4.1

.

elements: 0 if the Discriminator identifies one of the known states, and 1 if it identifies the generated state. We can therefore map the POVM to a single ancilla qubit, also input to the Discriminator (see quantum circuit: Fig. 5.1). We define P_0 as the projector of the circuit output state onto the zero state of the ancilla qubit.

Based on this, and considering that the Generator cost function must also take into consideration the energy minimisation objective we can define two subsequent cost functions that need to be minimised iteratively, for the Generator first and the Discriminator second. At optimum, the Generator cost function converges to the energy of the n^{th} excited state:

$$\begin{aligned} C^{(n)}_{gen}(\theta) = &\langle 0| G^\dagger(\theta) H G(\theta) |0\rangle \\ &+ \gamma \mathrm{Tr}[P_0 D(\phi)(\rho_g \otimes |0\rangle \langle 0|) D^\dagger(\phi)], \end{aligned} \tag{5.1}$$

$$C_{disc}^{(n)}(\phi) = \mathrm{Tr}[P_0 D(\phi)(\rho_g \otimes |0\rangle \langle 0|)D^\dagger(\phi)] - \sum_n \mathrm{Tr}[P_0 D(\phi)(\rho_{s_i} \otimes |0\rangle \langle 0|)D^\dagger(\phi)]. \quad (5.2)$$

We added a weighting factor γ to the Generator cost function. This is to guarantee that the minimum of the optimisation problem is indeed the state of index n. For this, we must have $\gamma > (n+1)(E_n - E_0)$. The derivation for the cost functions, convergence demonstration and explanation for the γ factor can be found in the following sections (Sections 5.2 and 5.2.1, respectively). We can find a suitable γ for all states by computing the maximum energy, running a VQE on the inverse Hamiltonian (for which the lowest eigenvalue is the highest eigenvalue of the original Hamiltonian) and taking the difference between the lowest energy state and the highest energy state.

It is worth noting that the Generator cost function is identical for any excitation level, while a term is added to the Discriminator at each new level of excitation calculated (one for each level of excitation). Therefore, there is a linear increase in the number of terms to be calculated with the number of excited states.

Derivation of the value functions

Consider an application of the DVQE circuit in which only the ancilla qubit is measured. The methodology to derive the DVQE value function is analogous to that developed for the QGAN in [459]. We note ρ_g the output state of the Generator and ρ_{s_i} the excited state of index i. Similarly, we note the state of all qubits after the DVQE circuit ρ_{Dg} for the generated state, and ρ_{Ds_i} for any state $i \in [0, n-1]$. Recalling that $D(\phi)$ is the operator resulting from the Discriminator circuit, $G(\theta)$ is the operator resulting from the Generator

circuit (we omit the ϕ and θ in the notation), that $\rho_g = G\left|0\right\rangle\left\langle 0\right|^{\otimes d} G^\dagger$, and that ρ_{s_i} represent any known energy state of the molecule, we have:

$$\rho_{Dg} = D(\rho_g \otimes \left|0\right\rangle\left\langle 0\right|)D^\dagger, \tag{5.3}$$

$$\rho_{Ds_i} = D(\rho_{s_i} \otimes \left|0\right\rangle\left\langle 0\right|)D^\dagger, \tag{5.4}$$

where one can observe that an ancilla qubit has been added, the necessity of which is explained later on.

The Discriminator therefore outputs a mixture $\rho_{D_{mix}} = p(g)\rho_{D_g} + \sum_i p(s_i)\rho_{D_{s_i}}$, with $p(g)$ and $p(s_i)$ the probabilities of presenting the generated or any state s_i to the Discriminator. We conduct a POVM on the output state, with projectors P_b, with b indexing the possible measurement outcomes such that $\sum_b P_b = \mathbb{1}$. Each possible measurement outcome P_b, can occur with a probability $p(b) = \mathrm{Tr}[P_b \rho_{D_{mix}}]$, following Born's rule. The Discriminator can either be right and the POVM identifies correctly the input state, or the Discriminator can be wrong and the POVM identifies the incorrect input state. The process through which the POVM identifies the input state is referred to as the decision rule.

Following Bayes' theorem, this decision rule should select the index b which maximises the posterior probability, or $\mathrm{argmax}_{x\in\{g,s_i\}} p(x|b)$. It has been shown that this decision function (Bayes' decision function) has the lowest probability of error of any possible decision function [460].

Our value function is built in order for the Discriminator to minimise the probability of error on a given measurement outcome. The probability of the measurement resulting in a correct decision is $\max_{x\in\{g,s_i\}} p(x|b)$. Therefore, using Bayes' decision function, the probability of error when observing any

element of the set $\{P_b\}$, can be written as:

$$\begin{aligned} p_{err}(\{P_b\}) &= \sum_b (1 - \max_x p(x|b))p(b) \\ &= \sum_b \min_x p(x|b)p(b). \end{aligned}$$

This equality is verified as the classification decision is done only over two possible categories: the Discriminator identifies a generated state g or the Discriminator identifies any of the known states s_i. We therefore have $1 - \max_x p(x|b) = \min_x p(x|b)$. Given that by Bayes' formula $p(x|b)p(b) = p(b|x)p(x)$:

$$\begin{aligned} p_{err}(\{P_b\}) &= \sum_b \min_x p(b|x)p(x) \\ &= \sum_b \min_x \mathrm{Tr}[P_b\rho_x]p(x). \end{aligned} \tag{5.5}$$

The objective function for the Discriminator being to minimise the probability of error for any given outcome obtained, it can be described by

$$p^*_{err} = \min_{\{b\}} p_{err}(\{P_b\}), \tag{5.6}$$

where $\{P_b\}$ represents the set of projectors corresponding to all possible measurement outcomes.

In this algorithm, we want the Discriminator to distinguish a generated state from any known state ρ_{s_i}. Therefore, the outcome of the POVM corresponds to the following: 0 is mapped to all the known states (ρ_{s_i}); 1 is mapped to the generated state (ρ_g).

Noting $p(g)$ and $p(s_i)$ the probabilities of the generated state and of any

known state being presented to the Discriminator, the objective function is given by:

$$
\begin{aligned}
p^*_{err} &= \min_{\{P_0,P_1\}} (p(0|g)p(g) + \sum_i p(1|s_i)p(s_i)) \\
&= \min_{\{P_0,P_1\}} (\mathrm{Tr}[P_0\rho_{Dg}]p(g) + \sum_i \mathrm{Tr}[P_1\rho_{Ds_i}]p(s_i)) \\
&= \min_{\{P_0\}} (\mathrm{Tr}[P_0\rho_{Dg}]p(g) + \sum_i \mathrm{Tr}[(\mathbb{1} - P_0)\rho_{Ds_i}]p(s_i)) \\
&= \min_{\{P_0\}} (\mathrm{Tr}[P_0\rho_{Dg}]p(g) - \sum_i \mathrm{Tr}[P_0\rho_{Ds_i}]p(s_i)) + \sum_i p(s_i). \qquad (5.7)
\end{aligned}
$$

However this is also dependent on the action of the Generator. The objective of the Generator is $\min P^*_{err}$ w.r.t. ρ_g. Incorporating this objective in the equation above we get the following shared objective function:

$$
\min_{\{\rho_g\}} \min_{\{P_0\}} (\mathrm{Tr}[P_0\rho_{Dg}]p(g) - \sum_i \mathrm{Tr}[P_0\rho_{Ds_i}]p(s_i)) + \sum_i p(s_i). \qquad (5.8)
$$

Due to the discriminative objective of the circuit, we can limit the required POVM outcome to only two elements: 0 if the Discriminator identifies the original state, and 1 if it identifies the generated state. We can map the POVM to a single ancilla qubit also input to the Discriminator. In the case we have $P_b = \mathbb{1}^{\otimes d} \otimes |b\rangle \langle b|, b \in [0, 1]$. Re-writing the state as the output of the quantum circuit we obtain the value function $\min_{\{\theta\}} \min_{\{\phi\}} V(\theta, \phi)$. Discarding

the parameterisation indices θ and ϕ we therefore aim to minimise

$$\begin{aligned} V(\theta,\phi) = &\operatorname{Tr}\left[P_0 D(\rho_g \otimes |0\rangle\langle 0|) D^\dagger\right] p(g) \\ &- \sum_{i=0}^{n-1} \operatorname{Tr}\left[P_0 D(\rho_{s_i} \otimes |0\rangle\langle 0|) D^\dagger\right] p(s_i) \\ &+ \sum_{i=0}^{n-1} p(s_i). \end{aligned} \tag{5.9}$$

The above value function is sufficient for the Generator to find at least one state belonging to the space orthogonal to all known states. However it does not guarantee that the state generated is ρ_{s_n}. In order to do so, we can add a VQE objective to the value function, whereby the Generator will also aim at finding a state which then minimises the expectation value of the Hamiltonian. Preemptively, we note that the weighting between both objectives is important in making sure the value function does converge to the desired excited state. In order to parameterise this weighting, we introduce a factor γ the value of which is discussed in the following section. Re-writing the value function accordingly, we get

$$\begin{aligned} V(\theta,\phi) = &\langle 0| G^\dagger H G |0\rangle \\ &+ \gamma \left[\operatorname{Tr}[P_0 D(\rho_g \otimes |0\rangle\langle 0|) D^\dagger\right] p(g) \\ &- \gamma \sum_{i=0}^{n-1} \operatorname{Tr}\left[P_0 D(\rho_{s_i} \otimes |0\rangle\langle 0|) D^\dagger\right] p(s_i) \\ &+ \gamma \sum_{i=0}^{n-1} p(s_i). \end{aligned} \tag{5.10}$$

All together, by grouping the terms of the value function dependent on θ and the terms of the value function dependent on ϕ, we find the cost functions

of the Generator and of the Discriminator which have already been outlined in Section 5.2:

$$C^{(n)}_{gen}(\theta) = \langle 0| G^\dagger(\theta) H G(\theta) |0\rangle + \gamma \mathrm{Tr}[P_0 D(\phi)(\rho_g \otimes |0\rangle \langle 0|) D^\dagger(\phi)], \tag{5.11}$$

$$C^{(n)}_{disc}(\phi) = \mathrm{Tr}[P_0 D(\phi)(\rho_g \otimes |0\rangle \langle 0|) D^\dagger(\phi)] - \sum_n \mathrm{Tr}[P_0 D(\phi)(\rho_{s_i} \otimes |0\rangle \langle 0|) D^\dagger(\phi))]. \tag{5.12}$$

5.2.1 Convergence demonstration

In this section, we demonstrate that one can find the correct eigenstate and eigenenergy using the value functions above, and we show how the γ factor is derived. The following demonstration assumes that the ansatz for the Generator can approximate any eigenstate of the Hamiltonian studied and that the ansatz for the Discriminator can approximate general unitaries, at least to the extent that it needs to distinguish between the known eigenstates and any unknown states. It is also assumed that both ansätze can be trained to their respective optima.

Consider a generic state $|\psi\rangle = \sum_{i=0}^{d-1} \alpha_i |s_i\rangle$ such that $|\psi\rangle = G(\theta)|0\rangle$ (recalling that d refers to the dimension of the system, and n refers to the excited state searched). We use this state in the value function derived in equation 5.10 (discarding θ and ϕ for readability):

$$
\begin{aligned}
V = & \langle\psi| H |\psi\rangle \\
& + \gamma \mathrm{Tr}\left[P_0 D(|\psi\rangle\langle\psi| \otimes |0\rangle\langle 0|) D^\dagger\right] p(g) \\
& - \gamma \sum_{i=0}^{n-1} \mathrm{Tr}\left[P_0 [D(|s_i\rangle\langle s_i| \otimes |0\rangle\langle 0|) D^\dagger\right] p(s_i) \\
& + \gamma \sum_{i=0}^{n-1} p(s_i).
\end{aligned}
\tag{5.13}
$$

The energy states $|s_i\rangle$ form an eigenbasis for the molecular Hamiltonian which can be written in the form $H = \sum_i E_i |s_i\rangle\langle s_i|$. We have $\langle s_i| H |s_i\rangle = E_i$, and we can re-write the above equation as

$$
\begin{aligned}
V = & \sum_{i=0}^{d-1} |\alpha_i|^2 E_i \\
& + \gamma \mathrm{Tr}\left[P_0 D(\sum_{i=0}^{d-1}\sum_{j=0}^{d-1} \alpha_i \alpha_j^* |s_i\rangle\langle s_j| \otimes |0\rangle\langle 0|) D^\dagger\right] p(g) \\
& - \gamma \sum_{i=0}^{n-1} \mathrm{Tr}\left[P_0 D(|s_i\rangle\langle s_i| \otimes |0\rangle\langle 0|) D^\dagger\right] p(s_i) \\
& + \gamma \sum_{i=0}^{n-1} p(s_i).
\end{aligned}
\tag{5.14}
$$

To simplify the writing, we set $p(g)$ and all $p(s_i)$ to be equiprobable, such that $p(g) = p(s_i) = \frac{1}{n+1}$ (as we use n known states $|s_i\rangle$ plus the generated state) and use the following short-hands:

$$
K_i = \frac{1}{n+1} \mathrm{Tr}\left[P_0 D(|s_i\rangle\langle s_i| \otimes |0\rangle\langle 0|) D^\dagger\right] \tag{5.15}
$$

and

$$k_i = \frac{1}{n+1}\mathrm{Tr}\left[P_0 D(\sum_{j\neq i}^{d-1} \alpha_i \alpha_j^* \left|s_i\right\rangle \left\langle s_j\right| \otimes \left|0\right\rangle \left\langle 0\right|) D^\dagger\right], \tag{5.16}$$

such that we now have:

$$\begin{aligned} V = &\sum_{i=0}^{d-1} |\alpha_i|^2 (E_i + \gamma K_i) + \sum_{i=0}^{d-1} \gamma k_i \\ &- \gamma \sum_{i=0}^{n-1} K_i + \frac{n\gamma}{n+1}. \end{aligned} \tag{5.17}$$

From here, we can see that the choice of set of parameters θ, for the Generator, affect the values of the terms α_i (and therefore, also the terms k_i) while the choice of set of parameters ϕ, for the Discriminator, affect the values of the terms K_i and k_i. Both Generator and Discriminator are trained to minimise this value function, and it is clear that, as a result of the terms k_i, both need to be trained for a meaningful minimum to be found.

It is important that the Discriminator is deep enough to be able to perform the classification between generated state and known states, we assume thereafter that it is the case. One can note that while some of the K_i have both positive and negative factors in the value function (namely for $i \in [0, n-1]$), the k_i all have positive factors. The terms k_i should go to 0 when the Discriminator is optimised. A similar argument can be made for the terms K_i such that $i \in [n, d-1]$.

Here it is worth noting that these terms are in general not accessible to the user given the states $|s_i\rangle$ for $i \in [n, d-1]$ are not known. However this does not prevent the convergence described above to occur during optimisation.

When the Generator is subsequently optimised, the value of the terms

k_i may increase as the α_i terms are updated. Subsequent updates of the Discriminator will bring these values back to 0. This implies that Generator and Discriminator will need to be updated iteratively for the DVQE to work. To simplify the demonstration, we assume that the terms k_i are sufficiently close to 0 so that we can ignore them in the following. We have

$$V = \sum_{i=0}^{d-1} |\alpha_i|^2 (E_i + \gamma K_i) - \gamma \sum_{i=0}^{n-1} K_i + \frac{n\gamma}{n+1}. \tag{5.18}$$

We now consider the case of optimising the Generator in the context of Eq. 5.18, that is finding a minimum for this equation by only modifying the α_i terms and recalling that $\sum_i |\alpha_i|^2 = 1$. Because the terms $E_i + \gamma K_i$ can be ordered from smallest to largest, optimising the Generator is equivalent to finding an index $p \in [0, d-1]$ such that $E_p + \gamma K_p < E_i + \gamma K_i$ for all $i \in [0, d-1] \setminus p$. In this case, α_p converges to 1.

In order to see that this index p should equate to n consider the ideal case in which the Discriminator is fully optimised and in which all K_i with $i \in [0, n-1]$ are equal to $\frac{1}{n+1}$. The last two terms in the Eq. 5.18 cancel each other and we obtain a simplified value function

$$V = \sum_{i=0}^{d-1} |\alpha_i|^2 (E_i + \gamma K_i), \tag{5.19}$$

which can be re-written as

$$V = \sum_{i=n}^{d-1} |\alpha_i|^2 (E_i + \gamma K_i) + \sum_{i=0}^{n-1} |\alpha_i|^2 (E_i + \gamma K_i). \tag{5.20}$$

Eq. (5.20) is important to understand how the algorithm behaves in a noisy environment, where the Discriminator cannot be fully optimised. How-

ever before discussing this, let us consider the case where the Discriminator perfectly succeeds at its task rendering $K_i = \frac{1}{n+1}$ for $i \in [0, n-1]$ and $K_i = 0$ for $i \in [n, d-1]$. We now have

$$V = \sum_{i=n}^{d-1} |\alpha_i|^2 (E_i) + \sum_{i=0}^{n-1} |\alpha_i|^2 (E_i + \frac{\gamma}{n+1}). \tag{5.21}$$

Once again, the action of optimising the Generator will result in one of the α_i being equal to 1, and the others to 0. To make sure that it is α_n we must have $E_n < E_0 + \frac{\gamma}{n+1}$ or, the γ factor, weighting the VQE and orthogonality objectives in the value function must obey

$$\gamma > (n+1)(E_n - E_0). \tag{5.22}$$

In a more general case, considering equation 5.21, for the state n to be the lowest energy of the value function, it must be that $(E_n + \gamma K_{n+1})$ is lower than $(E_i + \gamma K_i)$ for any i between 0 and $d-1$ except n. Therefore, given that together the Discriminator and the Generator push K_i towards 0 for i greater or equal to n and towards 1 for i lower than n then it is possible for the algorithm to converge to the right state given a large enough γ factor even if the Discriminator is not fully optimised. This is a particular advantage for NISQ computers where full optimisation of the Discriminator and Generator may be impossible due to circuit and read-out errors creating an optimisation barrier.

I noticed however that in the case of a noisy QPU, using a γ factor that is too high may result in the algorithm converging to the wrong value. That is because noise can prevent convergence to 0 of the k_i terms. If the Discriminator fails to bring close to 0 the term k_n, it may be that the minimum of the value

function is reached when more than one α term is non-zero.

It is worth noting that the term $\frac{n\gamma}{n+1}$ at the end of the value function has no impact on the optimisation (as it has a null gradient in all parameters of the function). We could discard it and find the same optimal point. The value function at optimal point would be different but we would still find the eigenstate and eigenenergy.

5.2.2 Simulation and error propagation analysis

Simulations: In order to test our algorithm, we first simulated the excitation levels of the 2-qubit H_2 Hamiltonian obtained using the Bravyi-Kitaev transformation in the STO-3G basis (results presented in FIG. 5.2). We used an optimisation cycle of three iterations for the Discriminator followed by three iterations of the Generator, repeated iteratively until convergence. For this test, we use successive layers of the HEA, as illustrated in Fig. 4.1, each layer being composed of two rotations (one on the Y axis and one on the X axis) on each qubit, followed by a ladder of entangling gates. This results in a total of 8 parameters. The Discriminator is composed of three such layers (applied on 3 qubits and hence 18 parameters) for the first excited state and four such layers for the second and third excited state (hence 24 parameters). The algorithm first computes the ground state using the VQE and continues to determine the first excited state. Each subsequent excited state is computed iteratively once convergence has been reached on the previous one. Typically, a precision of 10^{-3} Hartree is achieved within 20 iterations of the model using the Rprop optimiser [461]. I tested the algorithm on a 4-qubit version of the LiH Hamiltoninan, using the process detailed in Ref. [262] to build the Hamiltonians, and computing excitations until the 6^{th} excited state.

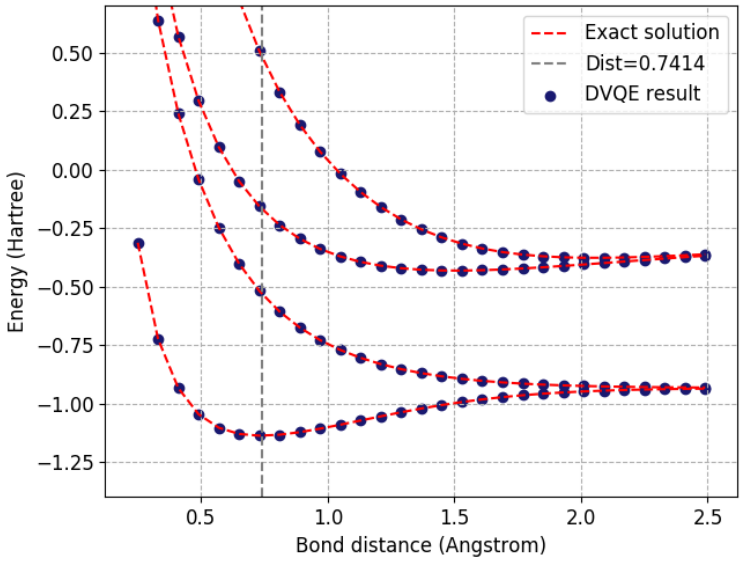


Figure 5.2: Dissociation curves for H_2 Hamiltonian using DVQE simulation and exact solver. Dotted lines represent ground and three excited states - all error to targets for this test are under 1 milliHartree

We initially used a depth of four for the generator and of six to eight for the discriminator to model the ground state and the first three excited states of LiH. Unlike the ansatz we used for H_2, we added rotations on the Z axis for each layers of the HEA as it resulted in overall significantly better accuracy. We achieved a precision of at least 1 mE_hon average across bond distances for all excited states with maximum single error of 2.5 mE_h. This is offering an initial test of the scalability of the method, showing precision is maintained on a larger system. To increase the expressiveness of the ansatz we added two layers to the generator to each subsequent energy state following the third state. Similarly, we increased depth of the discriminator by two layers for each subsequent energy state. While the initial depth is not sufficient for computation of higher excited states, further research will be necessary to determine the optimal ansatz both for the generator and the discriminator.

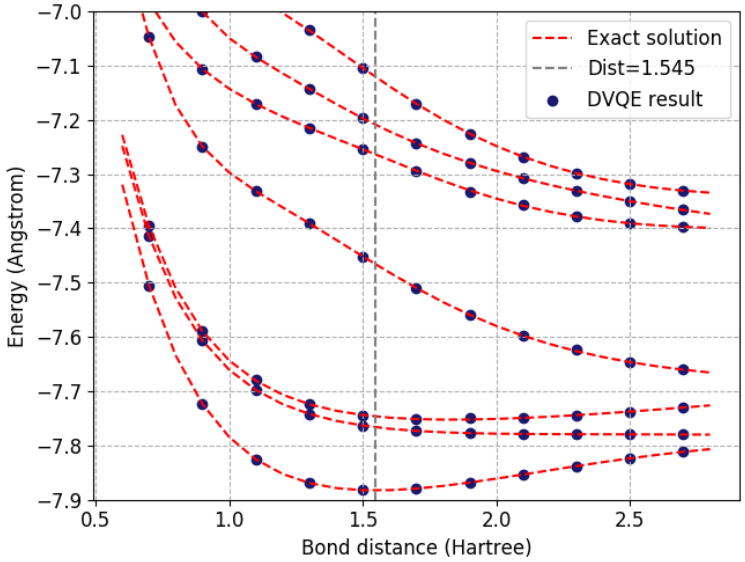


Figure 5.3: Dissociation curves for LiH Hamiltonian using DVQE simulation and exact solver. Dotted lines represent ground and six excited states. Errors are on average below 1 mE_h, with a few exceptions up to 2.5 mE_h

Error propagation analysis: In this section, we discuss the accuracy of the DVQE simulation when applied to LiH. On average, we find that convergence is reached within a 10^{-3} accuracy for all excitation levels with some outliers in higher excited states.

Figure 5.4 shows an increase in the magnitude of errors after ground and first excitation. It is worth nothing that we present average absolute error in this figure. In some cases, in particular for higher excited states, the DVQE converges slightly below the target value as a result of previous states not being perfectly orthogonal. These 'overshoots' errors however tend to be lower than 'undershoot' errors, resulting in higher excited states having better accuracy than some of the previous ones (e.g. second or fourth excited states). Additionally, it appears that most of the reduction in accuracy in higher excited states is driven by a higher frequency of outliers (instances where accuracy

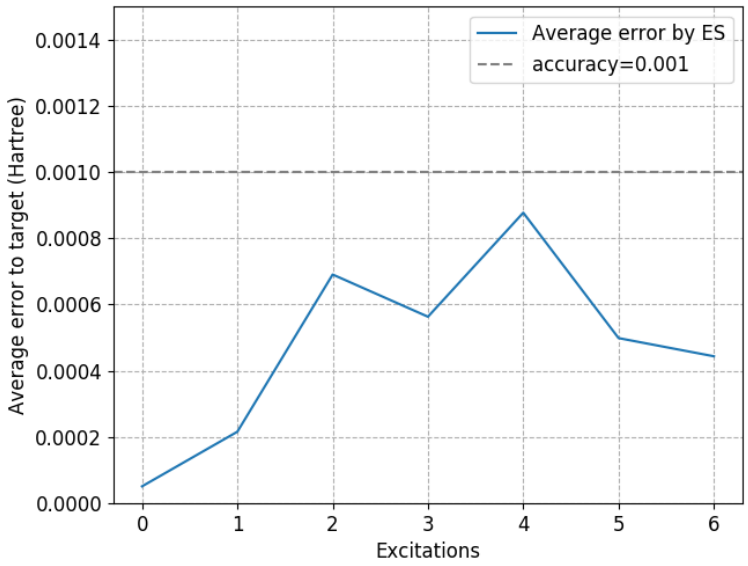


Figure 5.4: Average absolute error by excitation level over bond distances 0.7 to 2.7 (step 0.2)

is below 10^{-3}). This is particularly visible when considering the third and fourth excited states in Fig. 5.5. Large outliers are followed by low error overshoots in the following excited state. There are a number of reasons that could explains these outliers. First, this could be a result of our parameter initialisation strategy: we perform a warm start using the parameters of the nearest bond distance. While this reduces the number of iterations required, it could in some instances initialise the modeled wavefunction close to a local minimum, preventing convergence to the target value. Second, it could be that the ansatz is not expressive enough for certain bond distances (intuitively one can think that molecules with relatively higher bond distances have more entangled electrons). One factor that supports this second point is that we were able to increase average accuracy from the order of 10 to 1 mE_hat bond distance 2.3 Angstrom by increasing the ansatz for the generator by one

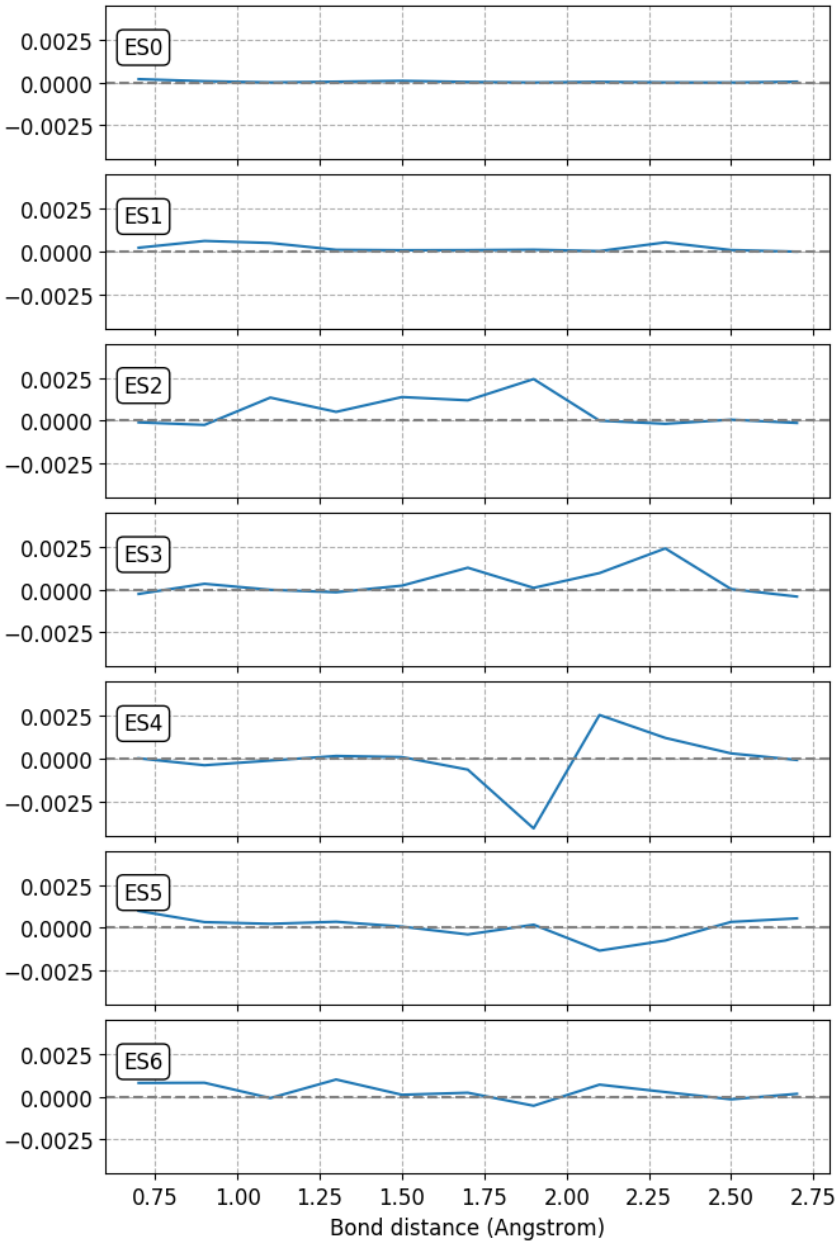


Figure 5.5: Complete set of errors by level of excitation and bond distance - Errors to target are in Hartree

layer.

5.2.3 Implementation on NISQ devices

Implementation on IBMQ: In order to test the algorithm's resilience to errors, we implemented our algorithm on IBMQ London and Vigo Quantum Processing Units (QPUs) for the H_2, two-qubit Hamiltonian (results presented in FIG. 5.6). Instead of using Rprop, we used the Rotosolve algorithm for which convergence is reached significantly faster [115] at the expense of not being parallelizable. Read-out errors are mitigated using the IBMQ Qiskit Ignis tool (described later on in this section). We computed both ground state through VQE and first state using DVQE. We found that both achieved about 10 mE_haccuracy.

Computing the second excited state would have required an additional layer for the Discriminator and as a result more involved error mitigation to obtain an accurate result. Similarly, a higher accuracy would require stronger error mitigation methods or lower circuit error rates. In particular we estimated that, given the depth of circuits used and based on the data provided by IBMQ, our circuit error on runs of the Generator was about 2% on all QPUs and of roughly 8% on runs of the full DVQE (Generator plus Discriminator).

Running an algorithm on a QPU remains computationally costly. We focused on minimising the number of single instruction requests to the QPU required to run the algorithm to an appropriate level of convergence. Each of our instruction requests covers the Rotosolve optimisation of one angle for either the Generator or the Discriminator. It includes requests to conduct estimation (through a given number of measurements, or shots) of the three

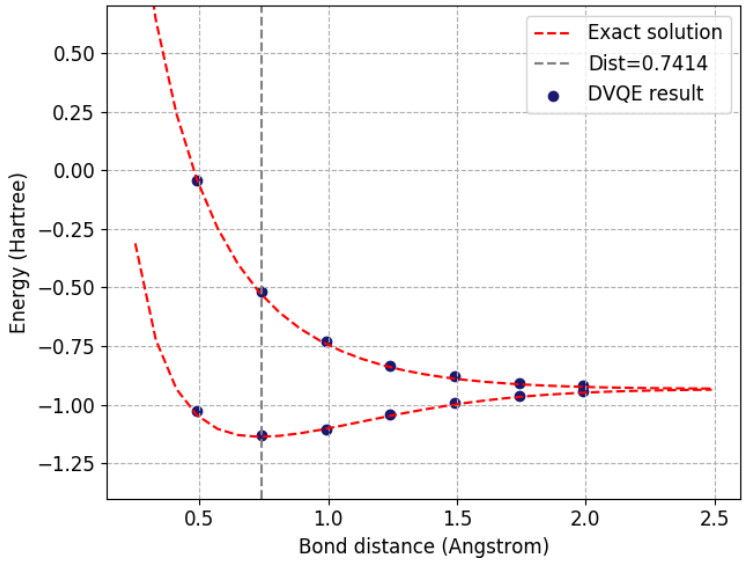


Figure 5.6: Dissociation curves for H_2 Hamiltonian using DVQE on IBMQ London, Ourense and Vigo and exact solver. Dotted lines represent ground and first excited state. Errors are within an average of 6 $mE_h$for the ground state, and 8 $mE_h$for the first excited state. Errors from ground and first excited states are mitigated using the technique presented in Sec. 3.3.2

expectation value terms required to complete a Rotosolve iteration.

Given the H_2 Hamiltonian on two qubits, we used a circuit of depth 2 for the Generator and of depth 3 for the Discriminator, with each layer composed of two rotation gates (R_X and R_Y) and an entangling gate. Hence we had to optimise 8 parameters for the Generator, and 18 for the Discriminator. The benefits of further depth could be studied but given our objective of minimising the number of calls to the QPUs we have not attempted anything further outside of simulation. For each bond distance, we use 2 iterations of the Generator and 2 iterations of the Discriminator for each iteration of the DVQE, and a total of 4 iterations of the DVQE resulting in a total of 208 separate calls to the QPU for each point (in addition to what was required

to compute the ground state, usually 2 iterations of the VQE, which has the same depth as the Generator, hence 18 calls).

This optimisation schedule was used only for calculating the energy values at a bond distance of 0.741. For other bond distances, we performed a warm start by using the θ and ϕ parameters learnt at distance 0.741 as a starting point for our optimisation process. In all cases, one iteration of the VQE and one iteration of the DVQE was sufficient to reach convergence (although more were required to show convergence). In addition, it is worth noting that as the efficacy of the Discriminator is resilient to noise, it is also resilient to small changes in the bond distance. In particular, we noticed that we did not need to re-train the Discriminator in most cases in order to reach convergence. This however may not be true when studying more complex systems and when attempting to achieve higher accuracy (for instance by increasing the number of measurements beyond $8,000$).

In order to reduce the number of shots conducted, we used a ramping-up schedule for the circuit estimate. The first few iterations of the circuit are done with a low numbers of shots, and the final iteration of the DVQE was done using $8,000$ measurements. Energies are then calculated using the final θ obtained and using repeated $8,000$ shots run to obtain an average.

It is worth noting that while we used Rotosolve for the implementation on a QPU, we used Rprop for the simulation. There are good reasons to think that this algorithm will be more relevant on a multi-core QPU than on a single QPU with a large number of qubits. Multi-core QPUs could offer tremendous opportunities for parallelisation. Because calculations of angles under Rotosolve are co-dependent on each other, it offers less parallelisation than gradient-based methods such as Rprop where all angle gradients can be calculated in parallel. Whether Rotosolve or gradient-based methods will be

Bond distance	QPU	Energy	Exact	DVQE
0.491	Ourense	Ground	-1.047	-1.025
		First	-0.046	-0.045
0.741	London	Ground	-1.137	-1.129
		First	-0.532	-0.519
0.991	Vigo	Ground	-1.103	-1.108
		First	-0.741	-0.728
1.241	Vigo	Ground	-1.048	-1.040
		First	-0.840	-0.832
1.491	Vigo	Ground	-0.999	-0.991
		First	-0.889	-0.880
1.741	Vigo	Ground	-0.967	-0.964
		First	-0.913	-0.908
1.991	Ourense	Ground	-0.949	-0.943
		First	-0.924	-0.919

Table 5.1: Detailed results of DVQE runs on IBMQ - values given are average of the last round of Rotosolve iteration (all in Hartree)

more efficient remains to be seen, however as long as QPUs are single core, Rotosolve will likely perform better for actual QPU runs, while Rprop (and other gradient-based methods) will be significantly more efficient for simulations.

Errors on the measurement results were mitigated using the IBM Qiskit Ignis error mitigation tool. The process is described here briefly. One first measures the quantum computer prepared in each of the 2^n computational basis, where n is the number of qubits. This could be easily achieved with quantum circuits using Pauli X gates and measurements. Using the measurement outcomes of the 2^n circuits, one could construct an estimate of the matrix M defined element-wise as:

$$M_{i,j} = \text{Probability}\{\text{measured state } i | \text{prepared in state j}\}$$
$$i, j \in \{0, 1, \cdots 2^n - 1\}$$

Then, one would like to apply the inverse of M to the measurement outcomes in the experiments. This is achieved by solving the following optimisation problem:

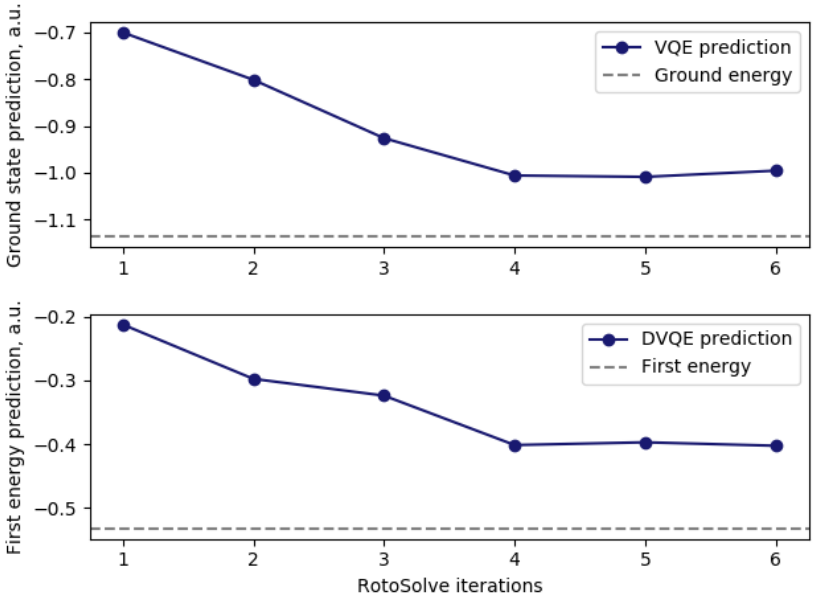
$$x = \text{argmin}_X |Y - MX|, \text{ subject to} \sum_i X_i = \sum_i Y_i$$

where Y is the vector of raw measurement outcome and x is the vector of error mitigated measurement outcome. In the ith position of each vector is the number of occurrence of the measurement outcome in state i. The vector norm is defined as $|v| = v \cdot v$.

The detailed results obtained are presented in table 5.1. The simulations were conducted using a TensorFlow backend simulator, while the actual tests on QPUs used a Qiskit backend linked to IBMQ.

Implementation on Rigetti's Aspen-4 QPU: In order the implement the DVQE onto Rigetti's Aspen-4, we ported the code to Pyquil which is Rigetti's development package and API to access their machines. The DVQE was implemented on the same molecular problem as for the IBMQ implementation: H_2 using the Bravyi-Kitaev symmetry conserving mapping and thus fitting on 2-qubits. Aspen-4 being a transmon qubits QPU it has similar speed as IBMQ QPUs, albeit with slightly lower gate fidelities. The results of the implementation, without and with error mitigation are presented in figures 5.7 and 5.8 respectively.

In addition to a simple DVQE implementation test, we also tried to run half the measurements on Aspen-4 and half the measurements on Aspen-7. Parallelisation is key to long term viability of variational algorithms such as



8

Figure 5.7: DVQE results on Rigetti's Aspen-4 QPU, Ground state energy at the top and first state energy at the bottom as a function of RotoSolve iterations. Without error mitigation

VQE and therefore it is interesting to assess the implication of running an algorithms on two different QPU cores in parallel.

While there is a clear improvement in algorithm runtime (See Figure 5.9) we found that the overall level of errors was higher than on either of the two QPU separately. While one could expect that running systematic errors of both QPUs would cancel each others, we can provide a provisional hypothesis as to why that is not the case. It is known that variational algorithms are particularly good at 'learning out' systematic biases [313]: the parameters found as optimal will necessarily take into account the systematic biases of QPU on which it is implemented. Parallel implementation fails at learning out these systematic biases and as such results in higher overall error than

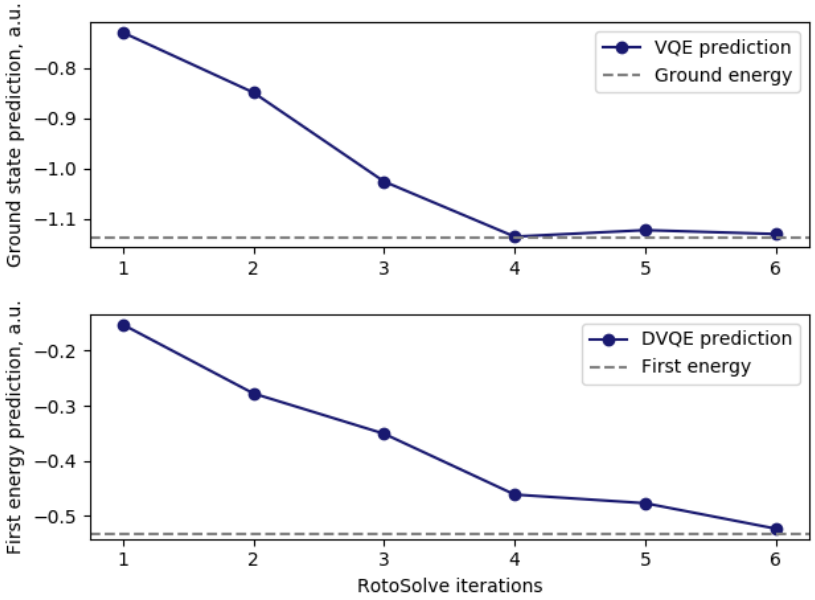


Figure 5.8: DVQE results on Rigetti's Aspen-4 QPU, Ground state energy at the top and first state energy at the bottom as a function of RotoSolve iterations. With error mitigation

on either QPU. Further tests, across different machines, would be required to test this hypothesis, however it points out to the fact that error mitigation techniques specific to parallelisation are likely to be required before VQE could be shown to exhibit a quantum advantage.

Implementation on Honeywell's HS0 QPU: Honeywell's latest QPU is an ion-trap based core with 6 qubits, and a reported quantum volume of 64 [16] (making it the QPU with highest quantum volume at time of writing, although IBM recently announced a similar performance on one of their superconducting QPU. For a definition of Quantum Volume see Ref. [462]). That said, with ion-trap QPUs being several orders of magnitude slower than superconducting qubits, implementing DVQE fully on this device would not have been possible

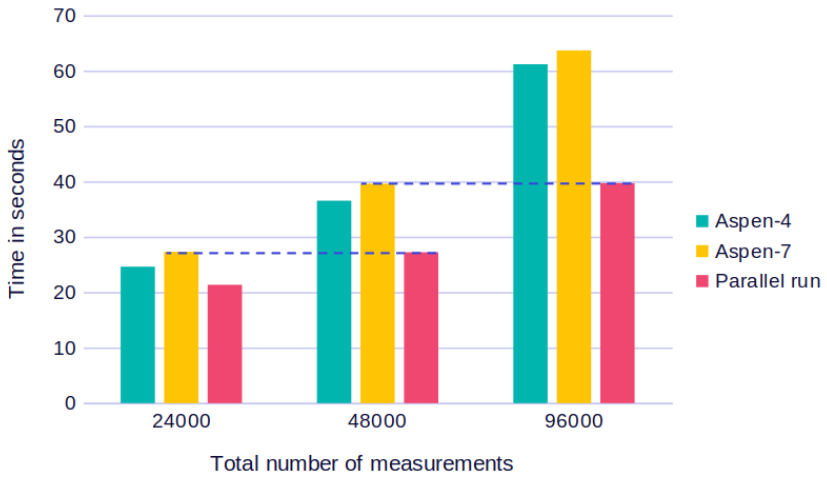


Figure 5.9: Run time for one VQE-RotoSolver iteration on Rigetti's Aspen-4, Aspen-7 and with parallel run. As expected, overall time of the parallel run is identical that that of the slowest QPU on half the measurements

in the time that was allocated.

Instead, the model was pre-trained to the optimal parameters classically, and then implemented on machine for one iteration at ground state level and one iteration at first excited state level to verify if the machine can maintain the optimal parameters.

As the QPU showcases much higher fidelities than those of IBM or Rigetti, we used both a 2-qubit and a 4-qubit version of the H_2 Hamiltonian. The larger number of qubits also required increased circuit depth. Due to the slow gate time of ion-trap QPUs, we were only able to compute one data point for each model, with and without error mitigation.

For the 2-qubit model, we obtained an accuracy to target of 98.6% for the ground state and 96.6% for the first excited state, without error mitigation (While the ground state is comparable with the results obtained on IBMQ, the first excited state is slightly better, as IBMQ's equivalent test return 93.2%

accuracy to target). With error mitigation (as presented in Sec. 3.3.2), these results are improved to 99.9% for the ground state and 99.6% for the first excited state.

For the 4-qubit model, we obtained an accuracy of 94.6% for the ground state and 89.4% for the first excited state, without error mitigation. This illustrates clearly the impact of only a few additional entangling gates on the overall level of errors. With error mitigation, these results are significantly improved to 99.6% for the ground state and 93.3% for the first excited state.

Conclusion

In the introduction we presented two research axes for this thesis:

- **Research axis 1:** *Review of the current knowledge on VQE to determine best practices, potential for bringing quantum advantage and limitations of the method.*
- **Research axis 2:** *Development of methods for computation of relevant molecular properties that can be implemented and tested on the current generation of quantum computers.*

On research axis 1: While the VQE remains among the most promising NISQ methods, we have seen that several significant hurdles remain to be addressed before a quantum advantage can even be considered (Sec. 3, Ref. [85]). These relate both to theoretical, algorithmic research and the way quantum hardware is developing. The three most significant obstacles that will need to be overcome are the exploding number of measurements, barren plateaus, and mitigation of errors.

As previously discussed, the number of measurements required to conduct VQE optimisation is astonishingly large, even when taking into account the significant scaling benefits brought by Pauli terms grouping or measurement inference methods such as classical shadows[82–85]. Large scale parallelisation

so far appears as the most suitable to overcome this large pre-factor. However there are three areas of future research that will need to be investigated before knowing whether this approach is reliable. Firstly, the benefits of parallelisation may be limited by the communication overhead of gathering and processing the large amount of data produced by the quantum devices. These hidden computing costs could significantly reduce the scaling potential of parallel quantum computation. Secondly, variational algorithms have been shown be reliably learn the systemic biases (e.g. rotation gates overshoot) from the quantum devices they are executed on, thereby offering some noise resilience [38, 152]. While executing a variational algorithm on several devices could result in biases from random errors, the resilience to systematic bias could be significantly reduced. Finally, quantum hardware providers have been largely focused on targeting developing systems for fault-tolerant quantum computation (FTQC). While FTQC methods benefit from obtaining the largest number of qubits subject to a given noise model, NISQ methods benefit from obtaining the lowest possible error rate for a given number of qubits. Similarly, while one could parallelise computation on a single chip (multi-thread), it is unlikely to perform as well as on different chips (multi-core) due to the impact of noise resulting from cross-talk between qubits. These three points will likely be required areas of research to determine the future applicability of VQE.

Regarding the barren plateau problem, many mitigations strategies have been proposed but none has been thoroughly tested at large scales. This is understandable given the challenge of simulating a VQE for more than a dozen qubits, however this raises the question of whether the barren plateau problem has been solved. This claim has been made in several studies, and most notably in the literature surrounding adaptive ansätze [213, 214, 268], where it has been numerically shown that gradients do not vanish in the system

size as long as the ansatz is built progressively [215].

Finally, no error mitigation technique has, so far, been shown to be scalable. In Sec. 3.3.2 we presented a simple error mitigation technique for small experiments on quantum devices. While this method will not be suitable to achieve accurate results on realistic applications of VQE, it illustrates the degree of predictability of the bias induced by quantum errors. Extrapolation methods directly target this predictability but are thought to be unscalable [151]. In that respect one could suppose that a machine learning system aiming to learn the specific bias for a given quantum chip, and any circuit execution could be trained based a wide range of random circuit fragments and measurements.

On research axis 2: As part of this thesis, we presented three methods for computation of molecular properties and demonstrated their implementation on current quantum devices. The first two methods use multiscale embedding to reduce the number of qubits (and thereby the depth of the circuit) required to compute ground energy at a given level of accuracy (Sec. 4, Ref. [149]). This is achieved through forcing optimisation of the Hamiltonian itself through conventional methods. The last method aims to bypass the need for swap gates and circuit inversion to variationally compute molecular excited states (Sec. 5, Ref. [455]).

A first point to note is that all these methods were successful in reaching appropriate accuracy in small systems, but all require at least some light touch error mitigation, echoing the previously raised comments about the fact that further research on error mitigation is a requirement for NISQ methods to be viable. A second, and final point to note is that current techniques used as part of VQE largely ignores the potential to efficiently remove some of the burden placed on quantum resources. A final remark on the overall computational cost of VQE is worth raising. The vast majority of the research in the field

is focused on the simple divide between sampling observables on the quantum computer. It is very likely that a future relevant application of VQE will require a more insightful split between quantum and conventional resources. Several avenues have already been proposed in this respect. For instance, Stenger *et al.* [463] propose to partially solve a many-body Hamiltonian on a conventional computer before performing a quantum based VQE. Okada *et al.* [464] alternatively show how one can perform a classical optimisation of the VQE for local-interaction states, leaving quantum sampling only for measurement of global quantities. Methods of Hamiltonian dressing such as ClusterVQE [465] work differently but aim at the same objective by absorbing in the Hamiltonian parts of the ansatz that are prone to quantum noise and worse gate scaling.

Appendix A

Operator groupings for RDM element sampling

At most, sixteen Pauli strings result from each two-body fermionic operator. While this implies polynomial scaling ($\mathcal{O}(n^4)$) in the number of terms that need to be measured on the quantum computer, the total number of measurements can be further improved upon by using commutative features of the Pauli strings. In particular, Pauli strings that commute can be measured jointly. There are two main ways to approach commutativity: qubit-wise commutation (QWC) and General Commutation (GC) of the operators. For QWC, we group two Pauli strings together if each operator in the first Pauli string commutes with the operator of corresponding index in the second Pauli string. GC is more general, and allows grouping of Pauli strings as long as they generally commute (for a review of Pauli strings commutation rules, we recommend: [38, 101, 102]). In this appendix, we study the scaling of measuring RDM using GC grouping for use in future research. These grouping strategies tend to be computationally expensive. However the final set of non-commuting

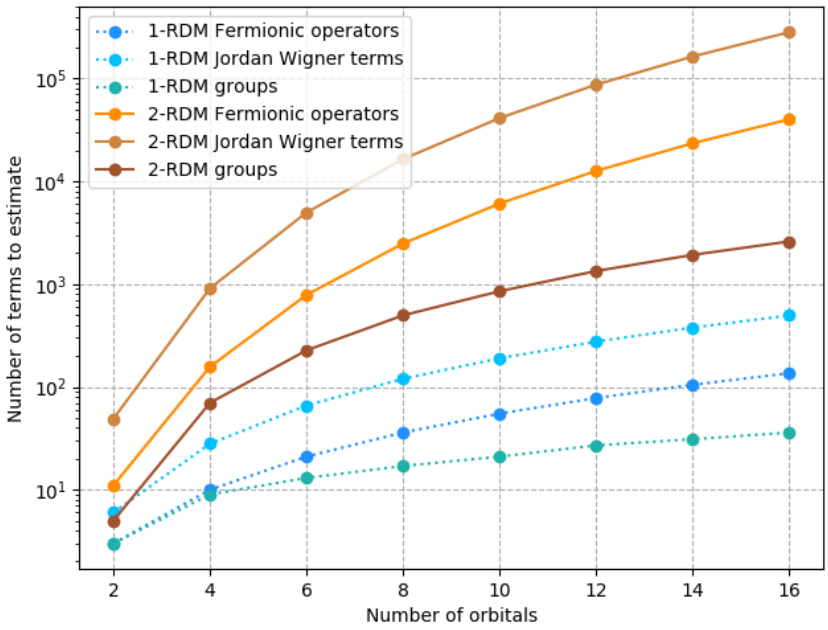


Figure A.1: Number of unique fermionic operators, corresponding set of unique Pauli strings (under Jordan-Wigner mapping), and commutative groups to be measured in order to compute all elements of the one-body (dotted line) and two-body RDM (solid line), for up to 16 orbitals (32 qubits). The groups were found via the Largest-degree First Coloring (LDFC) algorithm.

Pauli strings required to measure a given rank of RDM will be identical and agnostic to the details of the Hamiltonian for a given number of orbitals.

As a result, once an optimal set of terms is established, it can be used for all systems, in a similar fashion to the energy measurement for a VQE problem of a given size (which is equivalent to the two-body RDM).

We present in Fig. A.1 the number of commutative groups constructed for the one- and two-body RDMs as a function of the number of molecular orbitals in the active space (where each molecular orbital is mapped to two qubits). Our results are similar to those found previously in the literature (see for instance Ref. [401]), showing a significant reduction in dimensionality and

scaling of independant observable measurements as the active space increases in size. This results in a reduction of over two orders of magnitude for the number of terms to be sampled in a 16 orbital active space, with this factor increasing for larger active spaces. We present these numbers in Tab. A.1. In addition, we also present the groups for three- and four-body RDMs, up to 6 orbitals (Tab. A.2), as an investigation into the future feasibility of extended coupling schemes between the quantum region and environment such as multi-reference perturbative or subspace expansion approaches [349, 391, 400]. Finding these groups for the higher-body RDMs for larger numbers of orbitals became too computationally demanding for the current algorithm given our resources at time of writing. One point of note is that the symmetries used in Eqs. 4.6-4.8, combined with the Jordan-Wigner mapping, ensure that the one-body RDM only relies on half of the wave function (the same half of all Pauli strings required are identities, rendering half of the qubits used obsolete in the sampling). This feature can be used to easily sample the energy of any tensor product state or one-body RDM functional [148].

It is worth noting that grouping of terms may entail additional costs. Firstly, the joint measurements of Pauli operators results in a covariance between terms, potentially increasing the overall variance of the observable expectation values. In exceptional cases, this can even increase the total number of samples required for a given fidelity [38]. In general however, we should expect a reduction in number of measurements necessary for a given precision [101]. Secondly, joint measurements of Pauli strings groups require additional circuit depth to rotate the measurement basis appropriately. This additional circuit scales $O(N^2)$, with N the number of qubits [101], and therefore should be considered small on larger gate depth circuits (by comparison, the Generalised Unitary Coupled Cluster Ansatz scales $O(N^3)$ in depth. See for instance

Ref. [80]). In the case of current generation QPUs however, this additional circuit length (largely composed of entangling gates) results in quantum noise that would arguably out-weigh the benefits obtained from reduction of finite sampling noise from operator joint measurements. For this reason, while this reduction in terms to sample is a promising feature for longer-term viability of density matrix sampling, we are unlikely to benefit from this for small qubit arrays on current generation QPUs, and therefore leave the use of operator grouping in actual experiments for future work.

We present below further details on the approach we have used to group operators, as presented in Fig. A.1. In order to find these groupings, we require a graph of commutative relationships between all the Pauli strings required to measure the elements of the RDMs. To find a low number of groups of fully connected sub-graphs, we employed the Largest-Degree First Coloring (LDFC) algorithm (similar to what is proposed in [102]), a graph coloring heuristic. As an example for an alternative to the LDFC algorithm, one can start by grouping Pauli strings according to the frequency of identity operators in the string (as done, for instance in [401]).

The steps required to complete grouping of Pauli terms (using LDFC), joint measurements and measurement results aggregation are outlined at a high level below. For a more detailed description, we recommend Ref. [101].

- **Initialisation:** From the list of Pauli terms that require grouping, define a graph $G(V, e)$, with V the vertices, corresponding to each Pauli operator, and e the edges representing anti-commuting relationship between Pauli operators.

- **LDFC step 1:** Rank the elements of V according to their degree, i.e. number of edges they are connected to. Colors are represented by in-

	Active orbitals	2	4	6	8
One-body RDM	Fermionic operators	3	10	21	36
	JW	6	28	66	120
	JW-Groups	3	9	13	17
Two-body RDM	Fermionic operators	11	157	786	2,486
	JW	49	910	4,983	16,468
	JW-Groups	5	70	227	497
	Active orbitals	10	12	14	16
One-body RDM	Fermionic operators	55	78	105	136
	JW	190	276	378	496
	JW-Groups	21	27	31	36
Two-body RDM	Fermionic operators	6,085	12,651	23,492	40,156
	JW	41,325	87,354	164,115	282,968
	JW-Groups	853	1,342	1,928	2,601

Table A.1: Number of unique Pauli strings to be measured once grouped in order to sample all elements of the one- and two-body RDMs, as the number of (spatial) orbitals in the active space is enlarged. These numbers represent the terms in a direct Jordan-Wigner mapping (JW) and grouping of commuting terms from a Largest-degree First Coloring algorithm (JW-Groups).

	Active orbitals	4	6
Three-body RDM	Fermionic operators	610	8,400
	JW	4,928	71,742
	JW-Groups	189	2,049
Four-body RDM	Fermionic operators	939	40,065
	JW	11,425	440,154
	JW-Groups	163	3,182

Table A.2: Number of unique Pauli strings to be measured once grouped in order to sample all elements of the three- and four-body RDMs, as the number of (spatial) orbitals in the active space is enlarged. These numbers represent the terms in a direct Jordan-Wigner mapping (JW) and grouping of commuting terms from a Largest-degree First Coloring algorithm (JW-Groups).

tegers, the color of each vertex is initialised to 0 (unallocated).

- **LDFC step 2:** First allocate the color 1 to the element of V with the highest degree. Continue by allocating to the next element of V the lowest color that is not already attributed to one of its neighbours. Iterate likewise until all vertices have a color.

- **Joint-measurement basis identification:** The groups have now been defined. From each group, identify a basis (multiplicative), from which all the other elements of the groups can be computed. The basis size should be N, with N the number of qubits.

- **Joint-measurement circuit construction:** Once a basis is identified, construct the circuit required for joint measurements of the operators (following for instance the instructions set in Ref. [101], aiming to map

each of the operators in the basis to a single qubit Z-operator measurement.

- **Reconciliation:** From the results of the measurement, reconstruct the expectation value of each element in each group that can then be used to compute the one- and two-body RDMs.

Appendix B

IBM QPU lattice structures and additional information

In this appendix, we present additional information about the IBM QPU used during the experiment. The information provided below is sourced from IBMQ Experience reported calibration of the machines at time of running the experiment and may change slightly over time.

B.1 Lattice structures

IBMQ Bogota, Santiago and Athens are all 5-qubit QPUs, following IBM's *Canary* r3 processor type, with reported quantum volume of 32 [466]. The lattice structure, as well as the qubits used are presented in Fig. B.1, B.2, B.3 respectively.

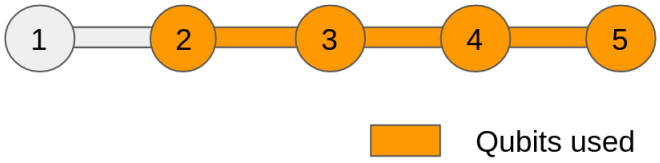


Figure B.1: IBMQ Bogota lattice structure and qubits used for experiments. This QPU was used to compute CASSCF and EwDMET with error mitigation.

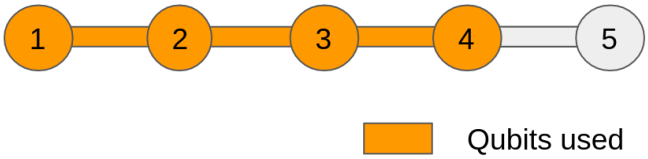


Figure B.2: IBMQ Santiago lattice structure and qubits used for experiments. This QPU was used to compute CASSCF and EwDMET without error mitigation.

B.2 Calibration information

The information presented in Table B.1 summarises the calibration data of the QPU used. It is directly taken from the IBMQ Experience Portal and may change over time as IBM re-calibrates the processors.

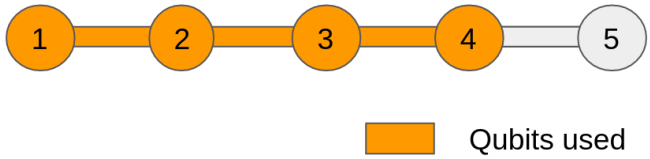


Figure B.3: IBMQ Athens lattice structure and qubits used for experiments. This QPU was used to compute the RDM sampling studies presented in Fig. 4.2, with and without error mitigation

QPU	Bogota	Santiago	Athens
Single-qubit Pauli-X error	2.95e-4	4.55e-4	4.16e-4
Qubit frequency (GHz)	4.89	4.78	5.094
Two-qubit gate error	1.40e-2	1.22e-2	1.043e-2
Two-qubit Gate time (ns)	536.89	408.89	346.67
Read-out length (ns)	5048.89	4017.78	3022.22
Read-out error	3.77e-2	1.82e-2	1.82e-2

Table B.1: Selected calibration metrics from IBMQ experience. These values are averaged for all qubits / connections and taken at a point in time near the experiment was run. They may change over time.

Bibliography

[1] Benioff, P.. The computer as a physical system: A microscopic quantum mechanical Hamiltonian model of computers as represented by turing machines. Journal of Statistical Physics 1980;22(5):563–591. doi:10.1007/bf01011339.

[2] Feynman, R.P.. Simulating physics with computers. International Journal of Theoretical Physics 1982;21(6-7):467–488. doi:10.1007/bf02650179.

[3] Chuang, I.L., Gershenfeld, N., Kubinec, M.. Experimental implementation of fast quantum searching. Physical Review Letters 1998;80(15):3408–3411. doi:10.1103/physrevlett.80.3408.

[4] Jones, J.A., Mosca, M.. Approximate quantum counting on an NMR ensemble quantum computer. Physical Review Letters 1999;83(5):1050–1053. doi:10.1103/physrevlett.83.1050.

[5] Leung, D.W., Chuang, I.L., Yamaguchi, F., Yamamoto, Y.. Efficient implementation of coupled logic gates for quantum computation. Physical Review A 2000;61(4). doi:10.1103/physreva.61.042310.

[6] Vandersypen, L.M.K., Steffen, M., Breyta, G., Yannoni, C.S.,

Sherwood, M.H., Chuang, I.L.. Experimental realization of Shor's quantum factoring algorithm using nuclear magnetic resonance. Nature 2001;414(6866):883–887. doi:10.1038/414883a.

[7] Häffner, H., Hänsel, W., Roos, C.F., Benhelm, J., Chek-al kar, D., Chwalla, M., et al. Scalable multiparticle entanglement of trapped ions. Nature 2005;438(7068):643–646. doi:10.1038/nature04279.

[8] Negrevergne, C., Mahesh, T.S., Ryan, C.A., Ditty, M., Cyr-Racine, F., Power, W., et al. Benchmarking quantum control methods on a 12-qubit system. Physical Review Letters 2006;96(17). doi:10.1103/physrevlett.96.170501.

[9] Plantenberg, J.H., de Groot, P.C., Harmans, C.J.P.M., Mooij, J.E.. Demonstration of controlled-NOT quantum gates on a pair of superconducting quantum bits. Nature 2007;447(7146):836–839. doi:10.1038/nature05896.

[10] Hanneke, D., Home, J.P., Jost, J.D., Amini, J.M., Leibfried, D., Wineland, D.J.. Realization of a programmable two-qubit quantum processor. Nature Physics 2009;6(1):13–16. doi:10.1038/nphys1453.

[11] Monz, T., Schindler, P., Barreiro, J.T., Chwalla, M., Nigg, D., Coish, W.A., et al. 14-qubit entanglement: Creation and coherence. Physical Review Letters 2011;106(13). doi:10.1103/physrevlett.106.130506.

[12] Devitt, S.J., Stephens, A.M., Munro, W.J., Nemoto, K.. Requirements for fault-tolerant factoring on an atom-optics quantum computer. Nature Communications 2013;4(1). doi:10.1038/ncomms3524.

[13] Devitt, S.J.. Performing quantum computing experiments in the cloud. Physical Review A 2016;94(3). doi:10.1103/physreva.94.032329.

[14] Monz, T., Nigg, D., Martinez, E.A., Brandl, M.F., Schindler, P., Rines, R., et al. Realization of a scalable Shor algorithm. Science 2016;351(6277):1068–1070. doi:10.1126/science.aad9480.

[15] Jurcevic, P., Javadi-Abhari, A., Bishop, L.S., Lauer, I., Bogorin, D.F., Brink, M., et al. Demonstration of quantum volume 64 on a superconducting quantum computing system. Quantum Science and Technology 2021;6(2):025020. doi:10.1088/2058-9565/abe519.

[16] Pino, J.M., Dreiling, J.M., Figgatt, C., Gaebler, J.P., Moses, S.A., Allman, M.S., et al. Demonstration of the trapped-ion quantum CCD computer architecture. Nature 2021;592(7853):209–213. URL: https://doi.org/10.1038/s41586-021-03318-4. doi:10.1038/s41586-021-03318-4.

[17] Ebadi, S., Wang, T.T., Levine, H., Keesling, A., Semeghini, G., Omran, A., et al. Quantum phases of matter on a 256-atom programmable quantum simulator. Nature 2021;595(7866):227–232. doi:10.1038/s41586-021-03582-4.

[18] Preskill, J.. Quantum computing in the NISQ era and beyond. Quantum 2018;2:79. doi:10.22331/q-2018-08-06-79.

[19] Brooks, M.. Beyond quantum supremacy: the hunt for useful quantum computers. Nature 2019;574(7776):19–21. doi:10.1038/d41586-019-02936-3.

[20] Arute, F., Arya, K., Babbush, R., Bacon, D., Bardin, J.C., Barends, R., et al. Quantum supremacy using a programmable superconducting processor. Nature 2019;574(7779):505–510. doi:10.1038/s41586-019-1666-5.

[21] Zhong, H.S., Wang, H., Deng, Y.H., Chen, M.C., Peng, L.C., Luo, Y.H., et al. Quantum computational advantage using photons. Science 2020;370(6523):1460–1463. doi:10.1126/science.abe8770.

[22] Wu, Y., Bao, W.S., Cao, S., Chen, F., Chen, M.C., Chen, X., et al. Strong quantum computational advantage using a superconducting quantum processor. 2021. arXiv:2106.14734v1.

[23] Madsen, L.S., Laudenbach, F., Askarani, M.F., Rortais, F., Vincent, T., Bulmer, J.F.F., et al. Quantum computational advantage with a programmable photonic processor. Nature 2022;606(7912):75–81. URL: https://doi.org/10.1038/s41586-022-04725-x. doi:10.1038/s41586-022-04725-x.

[24] Shor, P.. Algorithms for quantum computation: discrete logarithms and factoring. In: Proceedings 35th Annual Symposium on Foundations of Computer Science. IEEE Comput. Soc. Press. ISBN 0-8186-6580-7; 1994, p. 124–134. doi:10.1109/sfcs.1994.365700.

[25] Lu, C.Y., Browne, D.E., Yang, T., Pan, J.W.. Demonstration of a compiled version of Shor's quantum factoring algorithm using photonic qubits. Physical Review Letters 2007;99(25). doi:10.1103/physrevlett.99.250504.

[26] Lanyon, B.P., Weinhold, T.J., Langford, N.K., Barbieri, M., James,

D.F.V., Gilchrist, A., et al. Experimental demonstration of a compiled version of Shor's algorithm with quantum entanglement. Physical Review Letters 2007;99(25). doi:10.1103/physrevlett.99.250505.

[27] Lucero, E., Barends, R., Chen, Y., Kelly, J., Mariantoni, M., Megrant, A., et al. Computing prime factors with a Josephson phase qubit quantum processor. Nature Physics 2012;8(10):719–723. doi:10.1038/nphys2385.

[28] Martín-López, E., Laing, A., Lawson, T., Alvarez, R., Zhou, X.Q., O'Brien, J.L.. Experimental realization of Shor's quantum factoring algorithm using qubit recycling. Nature Photonics 2012;6(11):773–776. doi:10.1038/nphoton.2012.259.

[29] Markov, I.L., Saeedi, M.. Faster quantum number factoring via circuit synthesis. Physical Review A 2013;87(1). doi:10.1103/physreva.87.012310.

[30] Amico, M., Saleem, Z.H., Kumph, M.. Experimental study of Shor's factoring algorithm using the IBM Q experience. Physical Review A 2019;100(1). doi:10.1103/physreva.100.012305.

[31] Grover, L.K.. A fast quantum mechanical algorithm for database search. In: Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing. STOC '96; New York, NY, USA: Association for Computing Machinery. ISBN 0897917855; 1996, p. 212–219. doi:10.1145/237814.237866.

[32] Bennett, C.H., Bernstein, E., Brassard, G., Vazirani, U.. Strengths

and weaknesses of quantum computing. SIAM Journal on Computing 1997;26(5):1510–1523. doi:10.1137/s0097539796300933.

[33] Cerf, N.J., Grover, L.K., Williams, C.P.. Nested quantum search and NP-hard problems. Applicable Algebra in Engineering, Communication and Computing 2000;10(4-5):311–338. doi:10.1007/s002000050134.

[34] Ambainis, A.. Quantum search algorithms. ACM SIGACT News 2004;35(2):22–35. doi:10.1145/992287.992296.

[35] Ambainis, A.. Quantum walk algorithm for element distinctness. SIAM Journal on Computing 2007;37(1):210–239. doi:10.1137/s0097539705447311.

[36] Bernstein, D.J.. Grover vs. McEliece. In: Lecture Notes in Computer Science; vol. 6061. Springer Berlin Heidelberg; 2010, p. 73–80. doi:10.1007/978-3-642-12929-2_6.

[37] Peruzzo, A., McClean, J., Shadbolt, P., Yung, M.H., Zhou, X.Q., Love, P.J., et al. A variational eigenvalue solver on a photonic quantum processor. Nature Communications 2014;5(1). doi:10.1038/ncomms5213.

[38] McClean, J.R., Romero, J., Babbush, R., Aspuru-Guzik, A.. The theory of variational hybrid quantum-classical algorithms. New Journal of Physics 2016;18(2):023023. doi:10.1088/1367-2630/18/2/023023.

[39] Deglmann, P., Schäfer, A., Lennartz, C.. Application of quantum calculations in the chemical industry-an overview. International Journal of Quantum Chemistry 2014;115(3):107–136. doi:10.1002/qua.24811.

[40] Williams-Noonan, B.J., Yuriev, E., Chalmers, D.K.. Free energy methods in drug design: Prospects of "alchemical perturbation" in medicinal chemistry. Journal of Medicinal Chemistry 2017;61(3):638–649. doi:10.1021/acs.jmedchem.7b00681.

[41] Heifetz, A., editor. Quantum Mechanics in Drug Discovery. Springer US; 2020. ISBN 9781071602829. doi:10.1007/978-1-0716-0282-9.

[42] Continentino, M.A.. Key Methods and Concepts in Condensed Matter Physics. IOP Publishing; 2021. ISBN 9780750333955. doi:10.1088/978-0-7503-3395-5.

[43] der Ven, A.V., Deng, Z., Banerjee, S., Ong, S.P.. Rechargeable alkali-ion battery materials: Theory and computation. Chemical Reviews 2020;120(14):6977–7019. doi:10.1021/acs.chemrev.9b00601.

[44] Cao, Y., Romero, J., Aspuru-Guzik, A.. Potential of quantum computing for drug discovery. IBM Journal of Research and Development 2018;62(6):6:1–6:20. doi:10.1147/jrd.2018.2888987.

[45] Blunt, N.S., Camps, J., Crawford, O., Izsák, R., Leontica, S., Mirani, A., et al. A perspective on the current state-of-the-art of quantum computing for drug discovery applications. 2022. URL: https://arxiv.org/abs/2206.00551. doi:10.48550/ARXIV.2206.00551.

[46] Lordi, V., Nichol, J.M.. Advances and opportunities in materials science for scalable quantum computing. MRS Bulletin 2021;46(7):589–595. doi:10.1557/s43577-021-00133-0.

[47] Cao, Y., Romero, J., Olson, J.P., Degroote, M., Johnson, P.D., Kieferová, M., et al. Quantum chemistry in the age of quantum com-

puting. Chemical Reviews 2019;119(19):10856–10915. doi:10.1021/acs.chemrev.8b00803.

[48] Zhou, Y., Stoudenmire, E.M., Waintal, X.. What limits the simulation of quantum computers? Physical Review X 2020;10(4). doi:10.1103/physrevx.10.041038.

[49] Boixo, S., Isakov, S.V., Smelyanskiy, V.N., Babbush, R., Ding, N., Jiang, Z., et al. Characterizing quantum supremacy in near-term devices. Nature Physics 2018;14(6):595–600. doi:10.1038/s41567-018-0124-x.

[50] McCaskey, A.J., Parks, Z.P., Jakowski, J., Moore, S.V., Morris, T.D., Humble, T.S., et al. Quantum chemistry as a benchmark for near-term quantum computers. npj Quantum Information 2019;5(1). doi:10.1038/s41534-019-0209-0.

[51] Silverman, M.. Quantum superposition : counterintuitive consequences of coherence, entanglement, and interference. Berlin: Springer; 2008. ISBN 3540718834.

[52] Ballentine, L.E.. Quantum Superposition: Counterintuitive Consequences of Coherence, Entanglement and Interference. American Journal of Physics 2008;76(11):1078–1079. doi:10.1119/1.2973810.

[53] Helgaker, T., Jørgensen, P., Olsen, J.. Molecular Electronic-Structure Theory. John Wiley & Sons, Ltd; 2000. ISBN 9781118531471. doi:10.1002/9781119019572.

[54] Kratzer, P., Neugebauer, J.. The basics of electronic structure theory for periodic systems. Frontiers in Chemistry 2019;7. doi:10.3389/fchem.2019.00106.

[55] Li, X., Govind, N., Isborn, C., DePrince, A.E., Lopata, K.. Real-time time-dependent electronic structure theory. Chemical Reviews 2020;120(18):9951–9993. doi:10.1021/acs.chemrev.0c00223.

[56] Miceli, R., McGuigan, M.. Effective matrix model for nuclear physics on a quantum computer. In: 2019 New York Scientific Data Summit (NYSDS). 2019, p. 1–4. doi:10.1109/NYSDS.2019.8909693.

[57] Matteo, O.D., McCoy, A., Gysbers, P., Miyagi, T., Woloshyn, R.M., Navrátil, P.. Improving Hamiltonian encodings with the Gray code. Phys Rev A 2021;103(4). doi:10.1103/physreva.103.042405.

[58] Kiss, O., Grossi, M., Lougovski, P., Sanchez, F., Vallecorsa, S., Papenbrock, T.. Quantum computing of the ^{6}Li nucleus via ordered unitary coupled cluster. 2022. URL: https://arxiv.org/abs/2205.00864. doi:10.48550/ARXIV.2205.00864.

[59] Romero, A.M., Engel, J., Tang, H.L., Economou, S.E.. Solving nuclear structure problems with the adaptive variational quantum algorithm. 2022. URL: https://arxiv.org/abs/2203.01619. doi:10.48550/ARXIV.2203.01619.

[60] Bañuls, M.C., Blatt, R., Catani, J., Celi, A., Cirac, J.I., Dalmonte, M., et al. Simulating lattice gauge theories within quantum technologies. The European Physical Journal D 2020;74(8). URL: https://doi.org/10.1140/epjd/e2020-100571-8. doi:10.1140/epjd/e2020-100571-8.

[61] Bass, S.D., Zohar, E.. Quantum technologies in particle physics. Philosophical Transactions of the Royal Society A: Mathemat-

ical, Physical and Engineering Sciences 2021;380(2216). URL: https://doi.org/10.1098/rsta.2021.0072. doi:10.1098/rsta.2021.0072.

[62] Bauer, C.W., Davoudi, Z., Balantekin, A.B., Bhattacharya, T., Carena, M., de Jong, W.A., et al. Quantum simulation for high energy physics. 2022. URL: https://arxiv.org/abs/2204.03381. doi:10.48550/ARXIV.2204.03381.

[63] McArdle, S., Mayorov, A., Shan, X., Benjamin, S., Yuan, X.. Digital quantum simulation of molecular vibrations. Chemical Science 2019;10(22):5725–5735. URL: https://doi.org/10.1039/c9sc01313j. doi:10.1039/c9sc01313j.

[64] Sawaya, N.P.D., Huh, J.. Quantum algorithm for calculating molecular vibronic spectra. The Journal of Physical Chemistry Letters 2019;10(13):3586–3591. URL: https://doi.org/10.1021/acs.jpclett.9b01117. doi:10.1021/acs.jpclett.9b01117.

[65] Ollitrault, P.J., Baiardi, A., Reiher, M., Tavernelli, I.. Hardware efficient quantum algorithms for vibrational structure calculations. Chemical Science 2020;11(26):6842–6855. URL: https://doi.org/10.1039/d0sc01908a. doi:10.1039/d0sc01908a.

[66] Jahangiri, S., Arrazola, J.M., Quesada, N., Delgado, A.. Quantum algorithm for simulating molecular vibrational excitations. Physical Chemistry Chemical Physics 2020;22(44):25528–25537. URL: https://doi.org/10.1039/d0cp03593a. doi:10.1039/d0cp03593a.

[67] Lötstedt, E., Yamanouchi, K., Tsuchiya, T., Tachikawa, Y.. Calculation of vibrational eigenenergies on a quantum computer:

Application to the Fermi resonance in CO_2. Physical Review A 2021;103(6). URL: https://doi.org/10.1103/physreva.103.062609. doi:10.1103/physreva.103.062609.

[68] Sawaya, N.P.D., Paesani, F., Tabor, D.P.. Near- and long-term quantum algorithmic approaches for vibrational spectroscopy. Physical Review A 2021;104(6). URL: https://doi.org/10.1103/physreva.104.062419. doi:10.1103/physreva.104.062419.

[69] Mitarai, K., Nakagawa, Y.O., Mizukami, W.. Theory of analytical energy derivatives for the variational quantum eigensolver. Physical Review Research 2020;2(1). URL: https://doi.org/10.1103/physrevresearch.2.013129. doi:10.1103/physrevresearch.2.013129.

[70] Omiya, K., Nakagawa, Y.O., Koh, S., Mizukami, W., Gao, Q., Kobayashi, T.. Analytical energy gradient for state-averaged orbital-optimized variational quantum eigensolvers and its application to a photochemical reaction. Journal of Chemical Theory and Computation 2022;18(2):741–748. URL: https://doi.org/10.1021/acs.jctc.1c00877. doi:10.1021/acs.jctc.1c00877.

[71] Liu, J., Wan, L., Li, Z., Yang, J.. Simulating periodic systems on a quantum computer using molecular orbitals. Journal of Chemical Theory and Computation 2020;16(11):6904–6914. URL: https://doi.org/10.1021/acs.jctc.0c00881. doi:10.1021/acs.jctc.0c00881.

[72] Yoshioka, N., Sato, T., Nakagawa, Y.O., ya Ohnishi, Y., Mizukami, W.. Variational quantum simulation for periodic materials. Physical Review Research 2022;4(1). URL: https://doi.org/10.

1103/physrevresearch.4.013052. doi:10.1103/physrevresearch.4.013052.

[73] Manrique, D.Z., Khan, I.T., Yamamoto, K., Wichitwechkarn, V., Ramo, D.M.. Momentum-space unitary coupled cluster and translational quantum subspace expansion for periodic systems on quantum computers. 2020. URL: https://arxiv.org/abs/2008.08694. doi:10.48550/ARXIV.2008.08694.

[74] Lubasch, M., Joo, J., Moinier, P., Kiffner, M., Jaksch, D.. Variational quantum algorithms for nonlinear problems. Physical Review A 2020;101(1). URL: https://doi.org/10.1103/physreva.101.010301. doi:10.1103/physreva.101.010301.

[75] Joseph, A., White, T., Chandra, V., McGuigan, M.. Quantum computing of Schwarzschild-de Sitter black holes and Kantowski-Sachs cosmology. 2022. URL: https://arxiv.org/abs/2202.09906. doi:10.48550/ARXIV.2202.09906.

[76] Nielsen, M.A., Chuang, I.L.. Quantum Computation and Quantum Information. Cambridge: Cambridge University Press; 2009. ISBN 9780511976667. doi:10.1017/cbo9780511976667.

[77] Rayleigh, J.W.. In finding the correction for the open end of an organ-pipe. Phil Trans 1870;161(77).

[78] Ritz, W.. Über eine neue methode zur lösung gewisser Variationsprobleme der mathematischen Physik. J reine angew Math 1908;135:1–61.

[79] Arfken, G.B., Weber, H.J.. Rayleigh-Ritz variational technique. In:

Mathematical Methods for Physicists, 3rd ed. Orlando, FL: Academic Press. ISBN 9781483277820; 1985, p. 957–961.

[80] Lee, J., Huggins, W.J., Head-Gordon, M., Whaley, K.B.. Generalized unitary coupled cluster wave functions for quantum computation. Journal of Chemical Theory and Computation 2018;15(1):311–324. doi:10.1021/acs.jctc.8b01004.

[81] Anand, A., Schleich, P., Alperin-Lea, S., Jensen, P.W.K., Sim, S., Díaz-Tinoco, M., et al. A quantum computing view on unitary coupled cluster theory. 2021. arXiv:2109.15176v1.

[82] Elfving, V.E., Broer, B.W., Webber, M., Gavartin, J., Halls, M.D., Lorton, K.P., et al. How will quantum computers provide an industrially relevant computational advantage in quantum chemistry? 2020. arXiv:2009.12472v1.

[83] Kühn, M., Zanker, S., Deglmann, P., Marthaler, M., Weiß, H.. Accuracy and resource estimations for quantum chemistry on a near-term quantum computer. Journal of Chemical Theory and Computation 2019;15(9):4764–4780. doi:10.1021/acs.jctc.9b00236.

[84] Gonthier, J.F., Radin, M.D., Buda, C., Doskocil, E.J., Abuan, C.M., Romero, J.. Identifying challenges towards practical quantum advantage through resource estimation: the measurement roadblock in the variational quantum eigensolver. 2020. arXiv:2012.04001v1.

[85] Tilly, J., Chen, H., Cao, S., Picozzi, D., Setia, K., Li, Y., et al. The variational quantum eigensolver: A review of methods and best practices. Physics Reports 2022;986:1–128. URL: https://doi.org/

10.1016/j.physrep.2022.08.003. doi:10.1016/j.physrep.2022.08.003.

[86] Parr, R.G.. On the genesis of a theory. International Journal of Quantum Chemistry 1990;37(4):327–347. doi:10.1002/qua.560370407.

[87] Friesner, R.A.. *Ab initio* quantum chemistry: Methodology and applications. PNAS 2005;102(19):6648–6653. doi:10.1073/pnas.0408036102.

[88] Szabo, A., Ostlund, N.. Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory. Dover Books on Chemistry; Dover Publications; 1996. ISBN 9780486691862.

[89] Jensen, F.. Introduction to computational chemistry. Chichester, UK Hoboken, NJ: Wiley; 2017. ISBN 978-1-118-82595-2.

[90] Nemoshkalenko, V.. Computational Methods in Solid State Physics. Amsterdam: Gordon and Breach Science Publishers; 1998. ISBN 9789056990947.

[91] Martin, R.M.. Electronic Structure. Cambridge University Press; 2004. doi:10.1017/cbo9780511805769.

[92] Marder, M.P.. Condensed Matter Physics. John Wiley & Sons, Inc.; 2010. doi:10.1002/9780470949955.

[93] Clark, J.W., Ristig, M.L., editors. Theory of Spin Lattices and Lattice Gauge Models. Springer Berlin Heidelberg; 1997. doi:10.1007/bfb0104298.

[94] Pauli, W.. Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren. Zeitschrift für Physik 1925;31(1):765–783. doi:10.1007/bf02980631.

[95] Griffiths, D., Griffiths, P.. Introduction to Quantum Mechanics. Pearson International Edition; Pearson Prentice Hall; 2005. ISBN 978-0-13-111892-8. URL: https://books.google.co.uk/books?id=-BsvAQAAIAAJ.

[96] McArdle, S., Endo, S., Aspuru-Guzik, A., Benjamin, S.C., Yuan, X.. Quantum computational chemistry. Reviews of Modern Physics 2020;92(1):15003–15003. doi:10.1103/revmodphys.92.015003.

[97] Jordan, P., Wigner, E.. Über das Paulische äquivalenzverbot. Zeitschrift für Physik 1928;47(9):631–651. doi:10.1007/BF01331938.

[98] Wecker, D., Hastings, M.B., Troyer, M.. Progress towards practical quantum variational algorithms. Physical Review A 2015;92(4):042303. doi:10.1103/physreva.92.042303.

[99] Rubin, N.C., Babbush, R., McClean, J.. Application of fermionic marginal constraints to hybrid quantum algorithms. New Journal of Physics 2018;20(5):053020. doi:10.1088/1367-2630/aab919.

[100] Arrasmith, A., Cincio, L., Somma, R.D., Coles, P.J.. Operator sampling for shot-frugal optimization in variational algorithms. 2020. arXiv:2004.06252v1.

[101] Gokhale, P., Angiuli, O., Ding, Y., Gui, K., Tomesh, T., Suchara, M., et al. Minimizing state preparations in variational quantum eigensolver by partitioning into commuting families. 2019. arXiv:1907.13623v1.

[102] Hamamura, I., Imamichi, T.. Efficient evaluation of quantum observables using entangled measurements. npj Quantum Information 2020;6(1):56. doi:10.1038/s41534-020-0284-2.

[103] Huggins, W.J., McClean, J.R., Rubin, N.C., Jiang, Z., Wiebe, N., Whaley, K.B., et al. Efficient and noise resilient measurements for quantum chemistry on near-term quantum computers. npj Quantum Information 2021;7(1):23. doi:10.1038/s41534-020-00341-7.

[104] Torlai, G., Mazzola, G., Carleo, G., Mezzacapo, A.. Precise measurement of quantum observables with neural-network estimators. Physical Review Research 2020;2(2):022060(R). doi:10.1103/physrevresearch.2.022060.

[105] Huang, H.Y., Kueng, R., Preskill, J.. Predicting many properties of a quantum system from very few measurements. Nature Physics 2020;16(10):1050–1057. doi:10.1038/s41567-020-0932-7.

[106] Holmes, Z., Sharma, K., Cerezo, M., Coles, P.J.. Connecting ansatz expressibility to gradient magnitudes and barren plateaus. 2021. arXiv:2101.02138v1.

[107] Nakaji, K., Yamamoto, N.. Expressibility of the alternating layered ansatz for quantum computation. Quantum 2021;5:434. doi:10.22331/q-2021-04-19-434.

[108] Cerezo, M., Sone, A., Volkoff, T., Cincio, L., Coles, P.J.. Cost function dependent barren plateaus in shallow parametrized quantum circuits. Nature Communications 2021;12(1):1791. doi:10.1038/s41467-021-21728-w.

[109] McClean, J.R., Boixo, S., Smelyanskiy, V.N., Babbush, R., Neven, H.. Barren plateaus in quantum neural network training landscapes. Nature Communications 2018;9(1). doi:10.1038/s41467-018-07090-4.

[110] Crooks, G.E.. Gradients of parameterized quantum gates using the parameter-shift rule and gate decomposition. 2019. arXiv:1905.13311v1.

[111] Izmaylov, A.F., Lang, R.A., Yen, T.C.. Analytic gradients in variational quantum algorithms: Algebraic extensions of the parameter-shift rule to general unitary transformations. 2021. arXiv:2107.08131v2.

[112] Haug, T., Bharti, K., Kim, M.S.. Capacity and quantum geometry of parametrized quantum circuits. 2021. arXiv:2102.01659v2.

[113] Haug, T., Kim, M.S.. Optimal training of variational quantum algorithms without barren plateaus. 2021. arXiv:2104.14543v3.

[114] Nakanishi, K.M., Fujii, K., Todo, S.. Sequential minimal optimization for quantum-classical hybrid algorithms. Physical Review Research 2020;2(4). doi:10.1103/physrevresearch.2.043158.

[115] Ostaszewski, M., Grant, E., Benedetti, M.. Structure optimization for parameterized quantum circuits. Quantum 2021;5:391. doi:10.22331/q-2021-01-28-391. arXiv:1905.09692.

[116] Koczor, B., Benjamin, S.C.. Quantum natural gradient generalised to non-unitary circuits. 2019. arXiv:1912.08660v4.

[117] Hartree, D.R.. The wave mechanics of an atom with a non-Coulomb central field. Part II. Some results and discussion. Math Proc Camb Philos Soc 1928;24(1):111–132. doi:10.1017/s0305004100011920.

[118] Slater, J.C.. The self consistent field and the structure of atoms. Phys-Rev 1928;32(3):339–348. doi:10.1103/physrev.32.339.

[119] Gaunt, J.A.. A theory of Hartree's atomic fields. Math Proc Camb Philos Soc 1928;24(2):328–342. doi:10.1017/s0305004100015851.

[120] Hartree, D.R.. Self-consistent field, with exchange, for beryllium. Proc Royal Soc Lond A 1 1935;150(869):9–33. doi:10.1098/rspa.1935.0085.

[121] Grant, E., Wossnig, L., Ostaszewski, M., Benedetti, M.. An initialization strategy for addressing barren plateaus in parametrized quantum circuits. Quantum 2019;3:214. doi:10.22331/q-2019-12-09-214.

[122] Patti, T.L., Najafi, K., Gao, X., Yelin, S.F.. Entanglement devised barren plateau mitigation. Physical Review Research 2021;3(3). doi:10.1103/physrevresearch.3.033090.

[123] Cerezo, M., Arrasmith, A., Babbush, R., Benjamin, S.C., Endo, S., Fujii, K., et al. Variational quantum algorithms. Nature Reviews Physics 2021;3(9):625–644. doi:10.1038/s42254-021-00348-9.

[124] Farhi, E., Goldstone, J., Gutmann, S.. A quantum approximate optimization algorithm. 2014. arXiv:1411.4028v1.

[125] Bravo-Prieto, C., LaRose, R., Cerezo, M., Subasi, Y., Cincio, L., Coles, P.J.. Variational quantum linear solver. 2019. URL: https://arxiv.org/abs/1909.05820. doi:10.48550/ARXIV.1909.05820.

[126] Pellow-Jarman, A., Sinayskiy, I., Pillay, A., Petruccione, F.. A comparison of various classical optimizers for a variational quantum linear solver. Quantum Information Processing 2021;20(6):202. doi:10.1007/s11128-021-03140-x.

[127] Kwak, Y., Yun, W.J., Jung, S., Kim, J.. Quantum neural networks: Concepts, applications, and challenges. 2021. URL: https://arxiv.org/abs/2108.01468. doi:10.48550/ARXIV.2108.01468.

[128] García, D.P., Cruz-Benito, J., García-Peñalvo, F.J.. Systematic literature review: Quantum machine learning and its applications. 2022. URL: https://arxiv.org/abs/2201.04093. doi:10.48550/ARXIV.2201.04093.

[129] Abbas, A., Sutter, D., Zoufal, C., Lucchi, A., Figalli, A., Woerner, S.. The power of quantum neural networks. Nature Computational Science 2021;1(6):403–409. URL: https://doi.org/10.1038/s43588-021-00084-1. doi:10.1038/s43588-021-00084-1.

[130] Schuld, M., Killoran, N.. Quantum machine learning in feature Hilbert spaces. Physical Review Letters 2019;122(4). URL: https://doi.org/10.1103/physrevlett.122.040504. doi:10.1103/physrevlett.122.040504.

[131] Huang, H.Y., Broughton, M., Mohseni, M., Babbush, R., Boixo, S., Neven, H., et al. Power of data in quantum machine learning. Nature Communications 2021;12(1). URL: https://doi.org/10.1038/s41467-021-22539-9. doi:10.1038/s41467-021-22539-9.

[132] Preskill, J.. Quantum computing and the entanglement frontier. 2012. arXiv:1203.5813v3.

[133] Pednault, E., Gunnels, J.A., Nannicini, G., Horesh, L., Wisnieff, R.. Leveraging secondary storage to simulate deep 54-qubit Sycamore circuits. 2019. arXiv:1910.09534v2.

[134] Liu, Y.A., Liu, X.L., Li, F.N., Fu, H., Yang, Y., Song, J., et al. Closing the "quantum supremacy" gap. Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis 2021;doi:10.1145/3458817.3487399.

[135] Jaschke, D., Montangero, S.. Is quantum computing green? an estimate for an energy-efficiency quantum advantage. 2022. URL: https://arxiv.org/abs/2205.12092. doi:10.48550/ARXIV.2205.12092.

[136] Barkoutsos, P.K., Gonthier, J.F., Sokolov, I., Moll, N., Salis, G., Fuhrer, A., et al. Quantum algorithms for electronic structure calculations: Particle-hole Hamiltonian and optimized wave-function expansions. Physical Review A 2018;98(2):022322. doi:10.1103/physreva.98.022322.

[137] O'Malley, P., Babbush, R., Kivlichan, I., Romero, J., McClean, J., Barends, R., et al. Scalable quantum simulation of molecular energies. Physical Review X 2016;6(3):031007. doi:10.1103/physrevx.6.031007.

[138] Wiersema, R., Zhou, C., de Sereville, Y., Carrasquilla, J.F., Kim, Y.B., Yuen, H.. Exploring entanglement and optimization within the Hamiltonian variational ansatz. PRX Quantum 2020;1(2). doi:10.1103/prxquantum.1.020319.

[139] Malone, F.D., Parrish, R.M., Welden, A.R., Fox, T., Degroote, M., Kyoseva, E., et al. Towards the simulation of large scale protein-ligand interactions on NISQ-era quantum computers. 2021. arXiv:2110.01589v1.

[140] Kirsopp, J.J.M., Paola, C.D., Manrique, D.Z., Krompiec, M., Greene-

Diniz, G., Guba, W., et al. Quantum computational quantification of protein-ligand interactions. 2021. arXiv:2110.08163v1.

[141] Bittel, L., Kliesch, M.. Training variational quantum algorithms is NP-hard. Physical Review Letters 2021;127(12). doi:10.1103/physrevlett.127.120502. arXiv:2101.07267v1; phys. Rev. Lett. 127, 120502 (2021).

[142] Anschuetz, E.R., Kiani, B.T.. Beyond barren plateaus: Quantum variational algorithms are swamped with traps. 2022. URL: https://arxiv.org/abs/2205.05786. doi:10.48550/ARXIV.2205.05786.

[143] Krentel, M.W.. The complexity of optimization problems. Journal of Computer and System Sciences 1988;36(3):490–509. doi:10.1016/0022-0000(88)90039-6.

[144] Uvarov, A.V., Biamonte, J.D.. On barren plateaus and cost function locality in variational quantum algorithms. Journal of Physics A: Mathematical and Theoretical 2021;54(24):245301. doi:10.1088/1751-8121/abfac7. arXiv:2011.10530v1; journal of Physics A: Mathematical and Theoretical 54 (24), 245301 (2021).

[145] Barron, G.S., Wood, C.J.. Measurement error mitigation for variational quantum algorithms. 2020. arXiv:2010.08520v1.

[146] Sagastizabal, R., Bonet-Monroig, X., Singh, M., Rol, M.A., Bultink, C.C., Fu, X., et al. Experimental error mitigation via symmetry verification in a variational quantum eigensolver. Physical Review A 2019;100(1):010302(R). doi:10.1103/physreva.100.010302.

[147] Nam, Y., Chen, J.S., Pisenti, N.C., Wright, K., Delaney, C., Maslov, D., et al. Ground-state energy estimation of the water molecule on a trapped-ion quantum computer. npj Quantum Information 2020;6(1). doi:10.1038/s41534-020-0259-3.

[148] and Frank Arute, , Arya, K., Babbush, R., Bacon, D., Bardin, J.C., Barends, R., et al. Hartree-Fock on a superconducting qubit quantum computer. Science 2020;369(6507):1084–1089. doi:10.1126/science.abb9811.

[149] Tilly, J., Sriluckshmy, P.V., Patel, A., Fontana, E., Rungger, I., Grant, E., et al. Reduced density matrix sampling: Self-consistent embedding and multiscale electronic structure on current generation quantum computers. Physical Review Research 2021;3(3). doi:10.1103/physrevresearch.3.033230. arXiv:2104.05531v1; phys. Rev. Research 3, 033230 (2021).

[150] Benfenati, F., Mazzola, G., Capecci, C., Barkoutsos, P.K., Ollitrault, P.J., Tavernelli, I., et al. Improved accuracy on noisy devices by nonunitary variational quantum eigensolver for chemistry applications. Journal of Chemical Theory and Computation 2021;17(7):3946–3954. doi:10.1021/acs.jctc.1c00091.

[151] Takagi, R., Endo, S., Minagawa, S., Gu, M.. Fundamental limits of quantum error mitigation. 2021. arXiv:2109.04457v2.

[152] Fontana, E., Fitzpatrick, N., Ramo, D.M., Duncan, R., Rungger, I.. Evaluating the noise resilience of variational quantum algorithms. Physical Review A 2021;104(2). doi:10.1103/physreva.104.022403.

[153] Evangelista, F.A., Chan, G.K.L., Scuseria, G.E.. Exact parameterization of fermionic wave functions via unitary coupled cluster theory. The Journal of Chemical Physics 2019;151(24):244112. doi:10.1063/1.5133059.

[154] Hohenberg, P., Kohn, W.. Inhomogeneous electron gas. Physical Review 1964;136(3B):B864–B871. doi:10.1103/physrev.136.b864.

[155] Levy, M.. Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the v-representability problem. Proceedings of the National Academy of Sciences 1979;76(12):6062–6065. doi:10.1073/pnas.76.12.6062.

[156] Vignale, G., Rasolt, M.. Density-functional theory in strong magnetic fields. Physical Review Letters 1987;59(20):2360–2363. doi:10.1103/physrevlett.59.2360.

[157] Kohn, W., Sham, L.J.. Self-consistent equations including exchange and correlation effects. Physical Review 1965;140(4A):A1133–A1138. doi:10.1103/physrev.140.a1133.

[158] Parr, R.G., Weitao, Y.. Density-Functional Theory of Atoms and Molecules. Oxford University Press; 1995. ISBN 9780197560709. doi:10.1093/oso/9780195092769.001.0001.

[159] Bagayoko, D.. Understanding density functional theory (DFT) and completing it in practice. AIP Advances 2014;4(12):127104. doi:10.1063/1.4903408.

[160] Eriksen, J.J., Anderson, T.A., Deustua, J.E., Ghanem, K., Hait, D., Hoffmann, M.R., et al. The ground state electronic energy of ben-

zene. The Journal of Physical Chemistry Letters 2020;11(20):8922–8929. doi:10.1021/acs.jpclett.0c02621.

[161] Williams, K.T., Yao, Y., Li, J., Chen, L., Shi, H., Motta, M., et al. Direct comparison of many-body methods for realistic electronic Hamiltonians. Physical Review X 2020;10(1). doi:10.1103/physrevx.10.011041.

[162] Babbush, R., Berry, D.W., McClean, J.R., Neven, H.. Quantum simulation of chemistry with sublinear scaling in basis size. npj Quantum Information 2019;5(1). doi:10.1038/s41534-019-0199-y.

[163] Sherrill, C.D., Schaefer, H.F.. The configuration interaction method: Advances in highly correlated approaches. In: Advances in Quantum Chemistry. Elsevier; 1999, p. 143–269. doi:10.1016/s0065-3276(08)60532-8.

[164] Vogiatzis, K.D., Ma, D., Olsen, J., Gagliardi, L., de Jong, W.A.. Pushing configuration-interaction to the limit: Towards massively parallel MCSCF calculations. The Journal of Chemical Physics 2017;147(18):184111. doi:10.1063/1.4989858. arXiv:https://doi.org/10.1063/1.4989858.

[165] Ross, I.G.. Calculations of the energy levels of acetylene by the method of antisymmetric molecular orbitals, including σ-π interaction. Transactions of the Faraday Society 1952;48:973. doi:10.1039/tf9524800973.

[166] Foresman, J.. Exploring chemistry with electronic structure methods. Pittsburgh, PA: Gaussian, Inc; 1996. ISBN 978-0963676931.

[167] White, S.R.. Density matrix formulation for quantum renormalization groups. Physical Review Letters 1992;69(19):2863–2866. doi:10.1103/physrevlett.69.2863.

[168] Sharma, S., Chan, G.K.L.. Spin-adapted density matrix renormalization group algorithms for quantum chemistry. The Journal of Chemical Physics 2012;136(12):124121. doi:10.1063/1.3695642.

[169] Tubman, N.M., Lee, J., Takeshita, T.Y., Head-Gordon, M., Whaley, K.B.. A deterministic alternative to the full configuration interaction quantum Monte Carlo method. The Journal of Chemical Physics 2016;145(4):044112. doi:10.1063/1.4955109.

[170] Booth, G.H., Alavi, A.. Approaching chemical accuracy using full configuration-interaction quantum Monte Carlo: A study of ionization potentials. The Journal of Chemical Physics 2010;132(17):174104. doi:10.1063/1.3407895.

[171] Thomas, R.E., Booth, G.H., Alavi, A.. Accurate *ab initio* calculation of ionization potentials of the first-row transition metals with the configuration-interaction quantum Monte Carlo technique. Physical Review Letters 2015;114(3). doi:10.1103/physrevlett.114.033001.

[172] Li, J., Otten, M., Holmes, A.A., Sharma, S., Umrigar, C.J.. Fast semistochastic heat-bath configuration interaction. The Journal of Chemical Physics 2018;149(21):214110. doi:10.1063/1.5055390.

[173] Peterson, K.A., Feller, D., Dixon, D.A.. Chemical accuracy in *ab initio* thermochemistry and spectroscopy: current strategies and future

challenges. Theoretical Chemistry Accounts 2012;131(1). doi:10.1007/s00214-011-1079-5.

[174] Tubman, N.M., Freeman, C.D., Levine, D.S., Hait, D., Head-Gordon, M., Whaley, K.B.. Modern approaches to exact diagonalization and selected configuration interaction with the adaptive sampling CI method. Journal of Chemical Theory and Computation 2020;16(4):2139–2159. doi:10.1021/acs.jctc.8b00536.

[175] Hait, D., Tubman, N.M., Levine, D.S., Whaley, K.B., Head-Gordon, M.. What levels of coupled cluster theory are appropriate for transition metal systems? A study using near-exact quantum chemical values for 3d transition metal binary compounds. Journal of Chemical Theory and Computation 2019;15(10):5370–5385. doi:10.1021/acs.jctc.9b00674.

[176] Levine, D.S., Hait, D., Tubman, N.M., Lehtola, S., Whaley, K.B., Head-Gordon, M.. CASSCF with extremely large active spaces using the adaptive sampling configuration interaction method. Journal of Chemical Theory and Computation 2020;16(4):2340–2354. doi:10.1021/acs.jctc.9b01255.

[177] Petruzielo, F.R., Holmes, A.A., Changlani, H.J., Nightingale, M.P., Umrigar, C.J.. Semistochastic projector Monte Carlo method. Physical Review Letters 2012;109(23). doi:10.1103/physrevlett.109.230201.

[178] Holmes, A.A., Changlani, H.J., Umrigar, C.J.. Efficient heat-bath sampling in Fock space. Journal of Chemical Theory and Computation 2016;12(4):1561–1571. doi:10.1021/acs.jctc.5b01170.

[179] Holmes, A.A., Tubman, N.M., Umrigar, C.J.. Heat-bath configura-

tion interaction: An efficient selected configuration interaction algorithm inspired by heat-bath sampling. Journal of Chemical Theory and Computation 2016;12(8):3674–3680. doi:10.1021/acs.jctc.6b00407.

[180] Sharma, S., Holmes, A.A., Jeanmairet, G., Alavi, A., Umrigar, C.J.. Semistochastic heat-bath configuration interaction method: Selected configuration interaction with semistochastic perturbation theory. Journal of Chemical Theory and Computation 2017;13(4):1595–1604. doi:10.1021/acs.jctc.6b01028.

[181] Smith, J.E.T., Mussard, B., Holmes, A.A., Sharma, S.. Cheap and near exact CASSCF with large active spaces. Journal of Chemical Theory and Computation 2017;13(11):5468–5478. doi:10.1021/acs.jctc.7b00900.

[182] Holmes, A.A., Umrigar, C.J., Sharma, S.. Excited states using semistochastic heat-bath configuration interaction. The Journal of Chemical Physics 2017;147(16):164111. doi:10.1063/1.4998614.

[183] Cleland, D., Booth, G.H., Overy, C., Alavi, A.. Taming the first-row diatomics: A full configuration interaction quantum Monte Carlo study. Journal of Chemical Theory and Computation 2012;8(11):4138–4152. doi:10.1021/ct300504f.

[184] Anderson, R.J., Booth, G.H.. Four-component full configuration interaction quantum Monte Carlo for relativistic correlated electron problems. The Journal of Chemical Physics 2020;153(18):184103. doi:10.1063/5.0029863.

[185] Booth, G.H., Cleland, D., Thom, A.J.W., Alavi, A.. Breaking the

carbon dimer: The challenges of multiple bond dissociation with full configuration interaction quantum Monte Carlo methods. The Journal of Chemical Physics 2011;135(8):084104. doi:10.1063/1.3624383.

[186] Blunt, N.S., Thom, A.J.W., Scott, C.J.C.. Preconditioning and perturbative estimators in full configuration interaction quantum Monte Carlo. Journal of Chemical Theory and Computation 2019;15(6):3537–3551. doi:10.1021/acs.jctc.9b00049.

[187] White, S.R.. Density-matrix algorithms for quantum renormalization groups. Physical Review B 1993;48(14):10345–10356. doi:10.1103/physrevb.48.10345.

[188] White, S.R., Martin, R.L.. *Ab initio* quantum chemistry using the density matrix renormalization group. The Journal of Chemical Physics 1999;110(9):4127–4130. doi:10.1063/1.478295.

[189] Mitrushenkov, A.O., Fano, G., Ortolani, F., Linguerri, R., Palmieri, P.. Quantum chemistry using the density matrix renormalization group. The Journal of Chemical Physics 2001;115(15):6815–6821. doi:10.1063/1.1389475.

[190] Legeza, O., Röder, J., Hess, B.A.. Controlling the accuracy of the density-matrix renormalization-group method: The dynamical block state selection approach. Phys Rev B 2003;67:125114. doi:10.1103/PhysRevB.67.125114.

[191] Chan, G.K.L., Head-Gordon, M.. Highly correlated calculations with a polynomial cost algorithm: A study of the density matrix renormaliz-

ation group. The Journal of Chemical Physics 2002;116(11):4462–4476. doi:10.1063/1.1449459.

[192] Chan, G.K.L., Sharma, S.. The density matrix renormalization group in quantum chemistry. Annual Review of Physical Chemistry 2011;62(1):465–481. doi:10.1146/annurev-physchem-032210-103338.

[193] Olivares-Amaya, R., Hu, W., Nakatani, N., Sharma, S., Yang, J., Chan, G.K.L.. The *ab initio* density matrix renormalization group in practice. The Journal of Chemical Physics 2015;142(3):034102. doi:10.1063/1.4905329.

[194] Wouters, S., Neck, D.V.. The density matrix renormalization group for *ab initio* quantum chemistry. The European Physical Journal D 2014;68(9). doi:10.1140/epjd/e2014-50500-1.

[195] Yanai, T., Kurashige, Y., Mizukami, W., Chalupský, J., Lan, T.N., Saitow, M.. Density matrix renormalization group for *ab initio* calculations and associated dynamic correlation methods: A review of theory and applications. International Journal of Quantum Chemistry 2014;115(5):283–299. doi:10.1002/qua.24808.

[196] Knecht, S., Hedegård, E.D., Keller, S., Kovyrshin, A., Ma, Y., Muolo, A., et al. New approaches for *ab initio* calculations of molecules with strong electron correlation. CHIMIA International Journal for Chemistry 2016;70(4):244–251. doi:10.2533/chimia.2016.244.

[197] Biamonte, J., Bergholm, V.. Tensor networks in a nutshell. 2017. URL: https://arxiv.org/abs/1708.00006. doi:10.48550/ARXIV.1708.00006.

[198] Orús, R.. Tensor networks for complex quantum systems. Nature Reviews Physics 2019;1(9):538–550. URL: https://doi.org/10.1038/s42254-019-0086-7. doi:10.1038/s42254-019-0086-7.

[199] Schneider, M., Ostmeyer, J., Jansen, K., Luu, T., Urbach, C.. The Hubbard model with fermionic tensor networks. 2021. URL: https://arxiv.org/abs/2110.15340. doi:10.48550/ARXIV.2110.15340.

[200] Kaneko, R., Danshita, I.. Tensor-network study of correlation-spreading dynamics in the two-dimensional Bose-Hubbard model. Communications Physics 2022;5(1). URL: https://doi.org/10.1038/s42005-022-00848-9. doi:10.1038/s42005-022-00848-9.

[201] Toulouse, J., Assaraf, R., Umrigar, C.J.. Chapter fifteen - Introduction to the variational and diffusion Monte Carlo methods. In: Hoggan, P.E., Ozdogan, T., editors. Electron Correlation in Molecules – *ab initio* Beyond Gaussian Quantum Chemistry; vol. 73 of *Advances in Quantum Chemistry*. Academic Press; 2016, p. 285–314. doi:https://doi.org/10.1016/bs.aiq.2015.07.003.

[202] Needs, R.J., Towler, M.D., Drummond, N.D., López Ríos, P., Trail, J.R.. Variational and diffusion quantum Monte Carlo calculations with the CASINO code. The Journal of Chemical Physics 2020;152(15):154106. doi:10.1063/1.5144288. arXiv:https://doi.org/10.1063/1.5144288.

[203] Pfau, D., Spencer, J.S., Matthews, A.G.D.G., Foulkes, W.M.C.. *Ab initio* solution of the many-electron Schrödinger equation with deep neural networks. Physical Review Research 2020;2(3). doi:10.1103/physrevresearch.2.033429.

[204] Eriksen, J.J., Matthews, D.A., Jørgensen, P., Gauss, J.. Assessment of the accuracy of coupled cluster perturbation theory for open-shell systems. I. Triples expansions. The Journal of Chemical Physics 2016;144(19):194102. doi:10.1063/1.4948780.

[205] Eriksen, J.J., Matthews, D.A., Jørgensen, P., Gauss, J.. Assessment of the accuracy of coupled cluster perturbation theory for open-shell systems. II. Quadruples expansions. The Journal of Chemical Physics 2016;144(19):194103. doi:10.1063/1.4948781.

[206] Zhang, I.Y., Grüneis, A.. Coupled cluster theory in materials science. Frontiers in Materials 2019;6. doi:10.3389/fmats.2019.00123.

[207] Bartlett, R.J., Musiał, M.. Coupled-cluster theory in quantum chemistry. Reviews of Modern Physics 2007;79(1):291–352. doi:10.1103/revmodphys.79.291.

[208] Wang, Q., Duan, M., Xu, E., Zou, J., Li, S.. Describing strong correlation with block-correlated coupled cluster theory. The Journal of Physical Chemistry Letters 2020;11(18):7536–7543. doi:10.1021/acs.jpclett.0c02117.

[209] Lyakh, D.I., Musiał, M., Lotrich, V.F., Bartlett, R.J.. Multireference nature of chemistry: The coupled-cluster view. Chemical Reviews 2011;112(1):182–243. doi:10.1021/cr2001417.

[210] Bartlett, R.J., Kucharski, S.A., Noga, J.. Alternative coupled-cluster ansätze II. The unitary coupled-cluster method. Chemical Physics Letters 1989;155(1):133–140. doi:10.1016/s0009-2614(89)87372-5.

[211] Taube, A.G., Bartlett, R.J.. New perspectives on unitary coupled-cluster theory. International Journal of Quantum Chemistry 2006;106(15):3393–3401. doi:10.1002/qua.21198.

[212] Filip, M.A., Thom, A.J.W.. A stochastic approach to unitary coupled cluster. The Journal of Chemical Physics 2020;153(21):214106. doi:10.1063/5.0026141.

[213] Grimsley, H.R., Economou, S.E., Barnes, E., Mayhall, N.J.. An adaptive variational algorithm for exact molecular simulations on a quantum computer. Nature Communications 2019;10(1). doi:10.1038/s41467-019-10988-2.

[214] Liu, J., Li, Z., Yang, J.. An efficient adaptive variational quantum solver of the Schrödinger equation based on reduced density matrices. The Journal of Chemical Physics 2021;154(24):244112. URL: https://doi.org/10.1063/5.0054822. doi:10.1063/5.0054822.

[215] Grimsley, H.R., Barron, G.S., Barnes, E., Economou, S.E., Mayhall, N.J.. ADAPT-VQE is insensitive to rough parameter landscapes and barren plateaus. 2022. URL: https://arxiv.org/abs/2204.07179. doi:10.48550/ARXIV.2204.07179.

[216] Gard, B.T., Zhu, L., Barron, G.S., Mayhall, N.J., Economou, S.E., Barnes, E.. Efficient symmetry-preserving state preparation circuits for the variational quantum eigensolver algorithm. npj Quantum Information 2020;6(1):10. doi:10.1038/s41534-019-0240-1.

[217] Grimsley, H.R., Claudino, D., Economou, S.E., Barnes, E., Mayhall, N.J.. Is the Trotterized UCCSD ansatz chemically well-

defined? Journal of Chemical Theory and Computation 2019;16(1):1–6. doi:10.1021/acs.jctc.9b01083.

[218] Carleo, G., Troyer, M.. Solving the quantum many-body problem with artificial neural networks. Science 2017;355(6325):602–606. doi:10.1126/science.aag2302.

[219] Nomura, Y., Darmawan, A.S., Yamaji, Y., Imada, M.. Restricted Boltzmann machine learning for solving strongly correlated quantum systems. Physical Review B 2017;96(20). doi:10.1103/physrevb.96.205152.

[220] Choo, K., Carleo, G., Regnault, N., Neupert, T.. Symmetries and many-body excitations with neural-network quantum states. Physical Review Letters 2018;121(16). doi:10.1103/physrevlett.121.167204.

[221] Glielmo, A., Rath, Y., Csányi, G., Vita, A.D., Booth, G.H.. Gaussian process states: A data-driven representation of quantum many-body physics. Physical Review X 2020;10(4). doi:10.1103/physrevx.10.041026.

[222] Luo, D., Clark, B.K.. Backflow transformations via neural networks for quantum many-body wave functions. Physical Review Letters 2019;122(22). doi:10.1103/physrevlett.122.226401.

[223] Choo, K., Mezzacapo, A., Carleo, G.. Fermionic neural-network states for *ab initio* electronic structure. Nature Communications 2020;11(1). doi:10.1038/s41467-020-15724-9.

[224] Hermann, J., Schätzle, Z., Noé, F.. Deep-neural-network solution of

the electronic Schrödinger equation. Nature Chemistry 2020;12(10):891–897. doi:10.1038/s41557-020-0544-y.

[225] Kitaev, A.Y.. Quantum measurements and the Abelian stabilizer problem. 1995. arXiv:quant-ph/9511026v1.

[226] Abrams, D.S., Lloyd, S.. Simulation of many-body Fermi systems on a universal quantum computer. Physical Review Letters 1997;79(13):2586–2589. doi:10.1103/physrevlett.79.2586.

[227] Abrams, D.S., Lloyd, S.. Quantum algorithm providing exponential speed increase for finding eigenvalues and eigenvectors. Physical Review Letters 1999;83(24):5162–5165. doi:10.1103/physrevlett.83.5162.

[228] Cleve, R., Ekert, A., Macchiavello, C., Mosca, M.. Quantum algorithms revisited. Proceedings of the Royal Society of London Series A: Mathematical, Physical and Engineering Sciences 1998;454(1969):339–354. doi:10.1098/rspa.1998.0164.

[229] Aspuru-Guzik, A., Dutoi, A.D., Love, P.J., Head-Gordon, M.. Simulated quantum computation of molecular energies. Science 2005;309(5741):1704–1707. doi:10.1126/science.1113479.

[230] McClean, J.R., Babbush, R., Love, P.J., Aspuru-Guzik, A.. Exploiting locality in quantum computation for quantum chemistry. The Journal of Physical Chemistry Letters 2014;5(24):4368–4380. doi:10.1021/jz501649m.

[231] Reiher, M., Wiebe, N., Svore, K.M., Wecker, D., Troyer, M.. Elucidating reaction mechanisms on quantum computers. Proceedings of the

National Academy of Sciences 2017;114(29):7555–7560. doi:10.1073/pnas.1619152114.

[232] Childs, A.M., Maslov, D., Nam, Y., Ross, N.J., Su, Y.. Toward the first quantum simulation with quantum speedup. Proceedings of the National Academy of Sciences 2018;115(38):9456–9461. doi:10.1073/pnas.1801723115.

[233] Babbush, R., Wiebe, N., McClean, J., McClain, J., Neven, H., Chan, G.K.L.. Low-depth quantum simulation of materials. Physical Review X 2018;8(1):011044. doi:10.1103/physrevx.8.011044.

[234] Tubman, N.M., Mejuto-Zaera, C., Epstein, J.M., Hait, D., Levine, D.S., Huggins, W., et al. Postponing the orthogonality catastrophe: Efficient state preparation for electronic structure simulations on quantum devices. 2018. arXiv:1809.05523.

[235] Sugisaki, K., Nakazawa, S., Toyota, K., Sato, K., Shiomi, D., Takui, T.. Quantum chemistry on quantum computers: A method for preparation of multiconfigurational wave functions on quantum computers without performing post-hartree–fock calculations. ACS Cent Sci 2018;5(1):167–175. doi:10.1021/acscentsci.8b00788.

[236] Sugisaki, K., Yamamoto, S., Nakazawa, S., Toyota, K., Sato, K., Shiomi, D., et al. Open shell electronic state calculations on quantum computers: A quantum circuit for the preparation of configuration state functions based on Serber construction. Chem Phys Letts 2019;737:100002. doi:10.1016/j.cpletx.2018.100002.

[237] Motta, M., Sun, C., Tan, A.T.K., O'Rourke, M.J., Ye, E., Min-

nich, A.J., et al. Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution. Nature Phys 2019;16(2):205–210. doi:10.1038/s41567-019-0704-4.

[238] Murta, B., Fernández-Rossier, J.. Gutzwiller wave function on a digital quantum computer. Phys Rev B 2021;103(24). doi:10.1103/physrevb.103.l241113.

[239] Albash, T., Lidar, D.A.. Adiabatic quantum computation. Rev Mod Phys 2018;90(1). doi:10.1103/revmodphys.90.015002.

[240] Matsuura, S., Yamazaki, T., Senicourt, V., Huntington, L., Zaribafiyan, A.. VanQver: the variational and adiabatically navigated quantum eigensolver. New Journal of Physics 2020;22(5):053023. doi:10.1088/1367-2630/ab8080.

[241] Du, J., Xu, N., Peng, X., Wang, P., Wu, S., Lu, D.. NMR implementation of a molecular hydrogen quantum simulation with adiabatic state preparation. Physical Review Letters 2010;104(3). doi:10.1103/physrevlett.104.030502.

[242] Lanyon, B.P., Whitfield, J.D., Gillett, G.G., Goggin, M.E., Almeida, M.P., Kassal, I., et al. Towards quantum chemistry on a quantum computer. Nature Chemistry 2010;2(2):106–111. doi:10.1038/nchem.483.

[243] Li, Z., Yung, M.H., Chen, H., Lu, D., Whitfield, J.D., Peng, X., et al. Solving quantum ground-state problems with nuclear magnetic resonance. Scientific Reports 2011;1(1). doi:10.1038/srep00088.

[244] Wang, Y., Dolde, F., Biamonte, J., Babbush, R., Bergholm, V., Yang, S., et al. Quantum simulation of helium hydride cation in a solid-state spin register. ACS Nano 2015;9(8):7769–7774. doi:10.1021/acsnano.5b01651.

[245] Paesani, S., Gentile, A., Santagati, R., Wang, J., Wiebe, N., Tew, D., et al. Experimental Bayesian quantum phase estimation on a silicon photonic chip. Physical Review Letters 2017;118(10). doi:10.1103/physrevlett.118.100503.

[246] Santagati, R., Wang, J., Gentile, A.A., Paesani, S., Wiebe, N., McClean, J.R., et al. Witnessing eigenstates for quantum simulation of Hamiltonian spectra. Science Advances 2018;4(1). doi:10.1126/sciadv.aap9646.

[247] Beinert, H., Holm, R.H., Münck, E.. Iron-sulfur clusters: Nature's modular, multipurpose structures. Science 1997;277(5326):653–659. doi:10.1126/science.277.5326.653.

[248] Low, G.H., Chuang, I.L.. Hamiltonian simulation by qubitization. Quantum 2019;3:163. doi:10.22331/q-2019-07-12-163.

[249] Li, Z., Li, J., Dattani, N.S., Umrigar, C.J., Chan, G.K.L.. The electronic complexity of the ground-state of the FeMo cofactor of nitrogenase as relevant to quantum simulations. J Chem Phys 2019;150(2):024302. doi:10.1063/1.5063376.

[250] Lee, J., Berry, D.W., Gidney, C., Huggins, W.J., McClean, J.R., Wiebe, N., et al. Even more efficient quantum computations of

chemistry through tensor hypercontraction. PRX Quantum 2021;2(3). doi:10.1103/prxquantum.2.030305.

[251] Babbush, R., Gidney, C., Berry, D.W., Wiebe, N., McClean, J., Paler, A., et al. Encoding electronic spectra in quantum circuits with linear T complexity. Phys Rev X 2018;8(4). doi:10.1103/physrevx.8.041015.

[252] Kivlichan, I.D., Gidney, C., Berry, D.W., Wiebe, N., McClean, J., Sun, W., et al. Improved fault-tolerant quantum simulation of condensed-phase correlated electrons via Trotterization. Quantum 2020;4:296. doi:10.22331/q-2020-07-16-296.

[253] Wang, D., Higgott, O., Brierley, S.. Accelerated variational quantum eigensolver. Physical Review Letters 2019;122(14). doi:10.1103/physrevlett.122.140504.

[254] Somma, R.D.. Quantum eigenvalue estimation via time series analysis. New Journal of Physics 2019;21(12):123025. URL: https://doi.org/10.1088/1367-2630/ab5c60. doi:10.1088/1367-2630/ab5c60.

[255] Jiang, Z., Kalev, A., Mruczkiewicz, W., Neven, H.. Optimal fermion-to-qubit mapping via ternary trees with applications to reduced quantum states learning. Quantum 2020;4:276. doi:10.22331/q-2020-06-04-276.

[256] Setia, K., Bravyi, S., Mezzacapo, A., Whitfield, J.D.. Superfast encodings for fermionic quantum simulation. Physical Review Research 2019;1(3). doi:10.1103/physrevresearch.1.033033.

[257] Uvarov, A., Biamonte, J.D., Yudin, D.. Variational quantum ei-

gensolver for frustrated quantum systems. Phys Rev B 2020;102(7). doi:10.1103/physrevb.102.075104.

[258] Sawaya, N.P.D., Smelyanskiy, M., McClean, J.R., Aspuru-Guzik, A.. Error sensitivity to environmental noise in quantum circuits for chemical state preparation. Journal of Chemical Theory and Computation 2016;12(7):3097–3108. doi:10.1021/acs.jctc.6b00220.

[259] Derby, C., Klassen, J., Bausch, J., Cubitt, T.. Compact fermion to qubit mappings. Physical Review B 2021;104(3). doi:10.1103/physrevb.104.035118.

[260] Derby, C., Klassen, J.. A compact fermion to qubit mapping part 2: Alternative lattice geometries. 2021. arXiv:2101.10735v1.

[261] Yen, T.C., Izmaylov, A.F.. Cartan sub-algebra approach to efficient measurements of quantum observables. 2021. arXiv:2007.01234.

[262] Kandala, A., Mezzacapo, A., Temme, K., Takita, M., Brink, M., Chow, J.M., et al. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. Nature 2017;549(7671):242–246. doi:10.1038/nature23879.

[263] Hempel, C., Maier, C., Romero, J., McClean, J., Monz, T., Shen, H., et al. Quantum chemistry calculations on a trapped-ion quantum simulator. Physical Review X 2018;8(3). doi:10.1103/physrevx.8.031022.

[264] Kokail, C., Maier, C., van Bijnen, R., Brydges, T., Joshi, M.K., Jurcevic, P., et al. Self-verifying variational quantum simulation of lattice models. Nature 2019;569(7756):355–360. doi:10.1038/s41586-019-1177-4.

[265] Izmaylov, A.F., Yen, T.C., Ryabinkin, I.G.. Revising the measurement process in the variational quantum eigensolver: is it possible to reduce the number of separately measured operators? Chemical Science 2019;10(13):3746–3755. doi:10.1039/c8sc05592k.

[266] Verteletskyi, V., Yen, T.C., Izmaylov, A.F.. Measurement optimization in the variational quantum eigensolver using a minimum clique cover. The Journal of Chemical Physics 2020;152(12):124114. doi:10.1063/1.5141458.

[267] O'Brien, T.E., Streif, M., Rubin, N.C., Santagati, R., Su, Y., Huggins, W.J., et al. Efficient quantum computation of molecular forces and other energy gradients. 2021. URL: https://arxiv.org/abs/2111.12437. doi:10.48550/ARXIV.2111.12437.

[268] Yordanov, Y.S., Armaos, V., Barnes, C.H.W., Arvidsson-Shukur, D.R.M.. Iterative qubit-excitation based variational quantum eigensolver. 2020. arXiv:2011.10540v1.

[269] Tang, H.L., Shkolnikov, V., Barron, G.S., Grimsley, H.R., Mayhall, N.J., Barnes, E., et al. Qubit-ADAPT-VQE: An adaptive algorithm for constructing hardware-efficient ansätze on a quantum processor. PRX Quantum 2021;2(2):020310. doi:10.1103/prxquantum.2.020310.

[270] Vidal, J.G., Theis, D.O.. Calculus on parameterized quantum circuits. 2018. URL: https://arxiv.org/abs/1812.06323. doi:10.48550/ARXIV.1812.06323.

[271] Parrish, R.M., Iosue, J.T., Ozaeta, A., McMahon, P.L.. A Jacobi diag-

onalization and anderson acceleration algorithm for variational quantum algorithm parameter optimization. 2019. arXiv:1904.03206v1.

[272] Watanabe, H.C., Raymond, R., Ohnishi, Y.y., Kaminishi, E., Sugawara, M.. Optimizing parameterized quantum circuits with free-axis selection. 2021. URL: https://arxiv.org/abs/2104.14875. doi:10.48550/ARXIV.2104.14875.

[273] Wada, K., Raymond, R., Ohnishi, Y.y., Kaminishi, E., Sugawara, M., Yamamoto, N., et al. Simulating time evolution with fully optimized single-qubit gates on parameterized quantum circuits. 2021. URL: https://arxiv.org/abs/2111.05538. doi:10.48550/ARXIV.2111.05538.

[274] Spall, J.C.. A stochastic approximation technique for generating maximum likelihood parameter estimates. In: 1987 American Control Conference. 1987, p. 1161–1167. doi:10.23919/ACC.1987.4789489.

[275] Spall, J.. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. IEEE Transactions on Automatic Control 1992;37(3):332–341. URL: https://doi.org/10.1109/9.119632. doi:10.1109/9.119632.

[276] Cade, C., Mineh, L., Montanaro, A., Stanisic, S.. Strategies for solving the Fermi-Hubbard model on near-term quantum computers. Physical Review B 2020;102(23). doi:10.1103/physrevb.102.235122.

[277] Bonet-Monroig, X., Wang, H., Vermetten, D., Senjean, B., Moussa, C., Bäck, T., et al. Performance comparison of optimization methods

on variational quantum algorithms. 2021. URL: https://arxiv.org/abs/2111.13454. doi:10.48550/ARXIV.2111.13454.

[278] Miháliková, I., Friák, M., Pivoluska, M., Plesch, M., Saip, M., Šob, M.. Best-practice aspects of quantum-computer calculations: A case study of the hydrogen molecule. Molecules 2022;27(3):597. URL: https://doi.org/10.3390/molecules27030597. doi:10.3390/molecules27030597.

[279] Cai, Z., Babbush, R., Benjamin, S.C., Endo, S., Huggins, W.J., Li, Y., et al. Quantum error mitigation. 2022. URL: https://arxiv.org/abs/2210.00921. doi:10.48550/ARXIV.2210.00921.

[280] Berry, D.W., Kieferová, M., Scherer, A., Sanders, Y.R., Low, G.H., Wiebe, N., et al. Improved techniques for preparing eigenstates of fermionic Hamiltonians. npj Quantum Information 2018;4(1). doi:10.1038/s41534-018-0071-5.

[281] Seeley, J.T., Richard, M.J., Love, P.J.. The Bravyi-Kitaev transformation for quantum computation of electronic structure. The Journal of Chemical Physics 2012;137(22):224109. doi:10.1063/1.4768229.

[282] Bravyi, S.B., Kitaev, A.Y.. Fermionic quantum computation. Annals of Physics 2002;298(1):210–226. doi:10.1006/aphy.2002.6254.

[283] Tranter, A., Sofia, S., Seeley, J., Kaicher, M., McClean, J., Babbush, R., et al. The Bravyi-Kitaev transformation: Properties and applications. International Journal of Quantum Chemistry 2015;115(19):1431–1441. doi:10.1002/qua.24969.

[284] Bravyi, S., Gambetta, J.M., Mezzacapo, A., Temme, K.. Tapering off qubits to simulate fermionic Hamiltonians. 2017. arXiv:1701.08213v1.

[285] Setia, K., Chen, R., Rice, J.E., Mezzacapo, A., Pistoia, M., Whitfield, J.D.. Reducing qubit requirements for quantum simulations using molecular point group symmetries. Journal of Chemical Theory and Computation 2020;16(10):6091–6097. doi:10.1021/acs.jctc.0c00113.

[286] Kirby, W.M., Tranter, A., Love, P.J.. Contextual subspace variational quantum eigensolver. Quantum 2021;5:456. doi:10.22331/q-2021-05-14-456.

[287] Gokhale, P., Chong, F.T.. $O(N^3)$ measurement cost for variational quantum eigensolver on molecular Hamiltonians. 2019. arXiv:1908.11857v1.

[288] Crawford, O., van Straaten, B., Wang, D., Parks, T., Campbell, E., Brierley, S.. Efficient quantum measurement of Pauli operators in the presence of finite sampling error. Quantum 2021;5:385. doi:10.22331/q-2021-01-20-385.

[289] Ryabinkin, I.G., Lang, R.A., Genin, S.N., Izmaylov, A.F.. Iterative qubit coupled cluster approach with efficient screening of generators. Journal of Chemical Theory and Computation 2020;16(2):1055–1063. doi:10.1021/acs.jctc.9b01084.

[290] Zhang, K., Hsieh, M.H., Liu, L., Tao, D.. Toward trainability of quantum neural networks. 2020. arXiv:2011.06258v2.

[291] Claudino, D., Wright, J., McCaskey, A.J., Humble, T.S.. Benchmark-

ing adaptive variational quantum eigensolvers. Frontiers in Chemistry 2020;8. doi:10.3389/fchem.2020.606863.

[292] Stokes, J., Izaac, J., Killoran, N., Carleo, G.. Quantum natural gradient. Quantum 2020;4:269. doi:10.22331/q-2020-05-25-269; arXiv: 1909.02108.

[293] McArdle, S., Jones, T., Endo, S., Li, Y., Benjamin, S.C., Yuan, X.. Variational ansatz-based quantum simulation of imaginary time evolution. npj Quantum Information 2019;5(1):75. doi:10.1038/s41534-019-0187-2; arXiv: 1804.03023.

[294] Köppl, C., Werner, H.J.. Parallel and low-order scaling implementation of Hartree–Fock exchange using local density fitting. J Chem Theory Comput 2016;12(7):3122–3134. doi:10.1021/acs.jctc.6b00251.

[295] Havlíček, V., Troyer, M., Whitfield, J.D.. Operator locality in the quantum simulation of fermionic models. Physical Review A 2017;95(3). doi:10.1103/physreva.95.032332.

[296] Cai, Z.. Resource estimation for quantum variational simulations of the Hubbard model. Physical Review Applied 2020;14(1). URL: https://doi.org/10.1103/physrevapplied.14.014059. doi:10.1103/physrevapplied.14.014059.

[297] Sokolov, I.O., Barkoutsos, P.K., Ollitrault, P.J., Greenberg, D., Rice, J., Pistoia, M., et al. Quantum orbital-optimized unitary coupled cluster methods in the strongly correlated regime: Can quantum algorithms outperform their classical equivalents? The Journal of Chemical Physics 2020;152(12):124107. doi:10.1063/1.5141835.

[298] Setia, K., Whitfield, J.D.. Bravyi-Kitaev superfast simulation of electronic structure on a quantum computer. The Journal of Chemical Physics 2018;148(16):164104. doi:10.1063/1.5019371.

[299] Vogt, N., Zanker, S., Reiner, J.M., Eckl, T., Marusczyk, A., Marthaler, M.. Preparing symmetry broken ground states with variational quantum algorithms. 2020. arXiv:2007.01582v2.

[300] Choquette, A., Paolo, A.D., Barkoutsos, P.K., Sénéchal, D., Tavernelli, I., Blais, A.. Quantum-optimal-control-inspired ansatz for variational quantum algorithms. Physical Review Research 2021;3(2). doi:10.1103/physrevresearch.3.023092.

[301] Fan, C., Wu, F.Y.. General lattice model of phase transitions. Phys Rev B 1970;2(3):723–733. doi:10.1103/physrevb.2.723.

[302] Marro, J., Dickman, R.. Nonequilibrium Phase Transitions in Lattice Models. Cambridge University Press; 1999. doi:10.1017/cbo9780511524288.

[303] Stanislavchuk, T.N., Litvinchuk, A.P., Hu, R., Jeon, Y.H., Ji, S.D., Cheong, S.W., et al. Optical properties, lattice dynamics, and structural phase transition in hexagonal2H-BaMnO3single crystals. Phys Rev B 2015;92(13). doi:10.1103/physrevb.92.134308.

[304] Choi, J.M., Dar, F., Pappu, R.V.. LASSI: A lattice model for simulating phase transitions of multivalent proteins. PLoS Comput Biol 2019;15(10):e1007028. doi:10.1371/journal.pcbi.1007028.

[305] Li, Y.N., Wang, S., Wang, T., Gao, R., Geng, C.Y., Li, Y.W., et al. Energies and spin states of $FeS^{0/-}$, $FeS_2^{0/-}$, $Fe_2S_2^{0/-}$, $Fe_3S_4^{0/-}$, and

$Fe_4S_4^{0/-}$ clusters. ChemPhysChem 2013;14(6):1182–1189. URL: https://doi.org/10.1002/cphc.201201043. doi:10.1002/cphc.201201043.

[306] Kjaergaard, M., Schwartz, M.E., Braumüller, J., Krantz, P., Wang, J.I.J., Gustavsson, S., et al. Superconducting qubits: Current state of play. Annual Review of Condensed Matter Physics 2020;11(1):369–395. doi:10.1146/annurev-conmatphys-031119-050605.

[307] Mineh, L., Montanaro, A.. Accelerating the variational quantum eigensolver using parallelism. 2022. URL: https://arxiv.org/abs/2209.03796. doi:10.48550/ARXIV.2209.03796.

[308] Sim, S., Johnson, P.D., Aspuru-Guzik, A.. Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms. Advanced Quantum Technologies 2019;2(12):1900070. doi:10.1002/qute.201900070.

[309] Hubregtsen, T., Pichlmeier, J., Stecher, P., Bertels, K.. Evaluation of parameterized quantum circuits: on the relation between classification accuracy, expressibility, and entangling capability. Quantum Machine Intelligence 2021;3(1). doi:10.1007/s42484-021-00038-w.

[310] Rasmussen, S.E., Loft, N.J.S., Bækkegaard, T., Kues, M., Zinner, N.T.. Reducing the amount of single-qubit rotations in VQE and related algorithms. Advanced Quantum Technologies 2020;3(12):2000063. URL: https://doi.org/10.1002/qute.202000063. doi:10.1002/qute.202000063.

[311] Funcke, L., Hartung, T., Jansen, K., Kühn, S., Stornati, P.. Dimensional expressivity analysis of parametric quantum circuits. Quantum

2021;5:422. URL: https://doi.org/10.22331/q-2021-03-29-422. doi:10.22331/q-2021-03-29-422.

[312] Marrero, C.O., Kieferová, M., Wiebe, N.. Entanglement induced barren plateaus. 2020. arXiv:2010.15968v2.

[313] Sharma, K., Cerezo, M., Cincio, L., Coles, P.J.. Trainability of dissipative perceptron-based quantum neural networks. 2020. arXiv:2005.12458v1.

[314] Wang, S., Fontana, E., Cerezo, M., Sharma, K., Sone, A., Cincio, L., et al. Noise-induced barren plateaus in variational quantum algorithms. 2020. arXiv:2007.14384v3.

[315] Arrasmith, A., Holmes, Z., Cerezo, M., Coles, P.J.. Equivalence of quantum barren plateaus to cost concentration and narrow gorges. 2021. arXiv:2104.05868v1.

[316] Arrasmith, A., Cerezo, M., Czarnik, P., Cincio, L., Coles, P.J.. Effect of barren plateaus on gradient-free optimization. 2020. arXiv:2011.12245v1.

[317] Napp, J.. Quantifying the barren plateau phenomenon for a model of unstructured variational ansätze. 2022. URL: https://arxiv.org/abs/2203.06174. doi:10.48550/ARXIV.2203.06174.

[318] Cerezo, M., Coles, P.J.. Higher order derivatives of quantum neural networks with barren plateaus. Quantum Science and Technology 2021;6(3):035006. doi:10.1088/2058-9565/abf51a.

[319] Romero, J., Babbush, R., McClean, J.R., Hempel, C., Love, P.J., Aspuru-Guzik, A.. Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz. Quantum Science and Technology 2018;4(1):014008. doi:10.1088/2058-9565/aad3e4.

[320] Sauvage, F., Sim, S., Kunitsa, A.A., Simon, W.A., Mauri, M., Perdomo-Ortiz, A.. FLIP: A flexible initializer for arbitrarily-sized parametrized quantum circuits. 2021. arXiv:2103.08572v2.

[321] Kulshrestha, A., Safro, I.. Beinit: Avoiding barren plateaus in variational quantum algorithms. 2022. URL: https://arxiv.org/abs/2204.13751. doi:10.48550/ARXIV.2204.13751.

[322] Volkoff, T., Coles, P.J.. Large gradients via correlation in random parameterized quantum circuits. Quantum Science and Technology 2021;6(2):025008. doi:10.1088/2058-9565/abd891.

[323] Skolik, A., McClean, J.R., Mohseni, M., van der Smagt, P., Leib, M.. Layerwise learning for quantum neural networks. Quantum Machine Intelligence 2021;3(1). doi:10.1007/s42484-020-00036-4.

[324] Campos, E., Nasrallah, A., Biamonte, J.. Abrupt transitions in variational quantum circuit training. Physical Review A 2021;103(3). doi:10.1103/physreva.103.032607.

[325] Sack, S.H., Medina, R.A., Michailidis, A.A., Kueng, R., Serbyn, M.. Avoiding barren plateaus using classical shadows. 2022. URL: https://arxiv.org/abs/2201.08194. doi:10.48550/ARXIV.2201.08194.

[326] Wu, A., Li, G., Ding, Y., Xie, Y.. Mitigating noise-induced gradient

vanishing in variational quantum algorithm training. 2021. URL: https://arxiv.org/abs/2111.13209. doi:10.48550/ARXIV.2111.13209.

[327] Pesah, A., Cerezo, M., Wang, S., Volkoff, T., Sornborger, A.T., Coles, P.J.. Absence of barren plateaus in quantum convolutional neural networks. 2020. arXiv:2011.02966v1.

[328] Bonet-Monroig, X., Sagastizabal, R., Singh, M., O'Brien, T.E.. Low-cost error mitigation by symmetry verification. Physical Review A 2018;98(6):062339. doi:10.1103/physreva.98.062339.

[329] McArdle, S., Yuan, X., Benjamin, S.. Error-mitigated digital quantum simulation. Physical Review Letters 2019;122(18):1–17. doi:10.1103/physrevlett.122.180501.

[330] Cai, Z.. Multi-exponential error extrapolation and combining error mitigation techniques for NISQ applications. npj Quantum Information 2021;7(1). doi:10.1038/s41534-021-00404-3. arXiv:2007.01265v2; npj Quantum Inf 7, 80 (2021).

[331] Temme, K., Bravyi, S., Gambetta, J.M.. Error mitigation for short-depth quantum circuits. Physical Review Letters 2017;119(18):1–15. doi:10.1103/physrevlett.119.180509.

[332] Li, Y., Benjamin, S.C.. Efficient variational quantum simulator incorporating active error minimization. Physical Review X 2017;7(2):021050–021050. doi:10.1103/physrevx.7.021050.

[333] Dumitrescu, E., McCaskey, A., Hagen, G., Jansen, G., Morris, T., Papenbrock, T., et al. Cloud quantum computing of an atomic

nucleus. Physical Review Letters 2018;120(21):210501. doi:10.1103/physrevlett.120.210501.

[334] He, A., Nachman, B., de Jong, W.A., Bauer, C.W.. Zero-noise extrapolation for quantum-gate error mitigation with identity insertions. Physical Review A 2020;102(1):012426. doi:10.1103/physreva.102.012426. arXiv:2003.04941.

[335] Endo, S., Benjamin, S.C., Li, Y.. Practical quantum error mitigation for near-future applications. Physical Review X 2018;8(3). doi:10.1103/physrevx.8.031027.

[336] Koczor, B.. Exponential error suppression for near-term quantum devices. Physical Review X 2021;11(3):031057. doi:10.1103/physrevx.11.031057.

[337] Huggins, W.J., McArdle, S., O'Brien, T.E., Lee, J., Rubin, N.C., Boixo, S., et al. Virtual distillation for quantum error mitigation. 2020. arXiv:2011.07064v3.

[338] Vovrosh, J., Khosla, K.E., Greenaway, S., Self, C., Kim, M.S., Knolle, J.. Simple mitigation of global depolarizing errors in quantum simulations. Physical Review E 2021;104(3). URL: https://doi.org/10.1103/physreve.104.035309. doi:10.1103/physreve.104.035309.

[339] Smith, A.W.R., Khosla, K.E., Self, C.N., Kim, M.S.. Qubit readout error mitigation with bit-flip averaging. Science Advances 2021;7(47). URL: https://doi.org/10.1126/sciadv.abi8009. doi:10.1126/sciadv.abi8009.

[340] Roos, B.O., Taylor, P.R., Sigbahn, P.E.. A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach. Chemical Physics 1980;48(2):157–173. doi:https://doi.org/10.1016/0301-0104(80)80045-0.

[341] Sun, Q., Yang, J., Chan, G.K.L.. A general second order complete active space self-consistent-field solver for large-scale systems. Chemical Physics Letters 2017;683:291–299. doi:10.1016/j.cplett.2017.03.004.

[342] Mañeru, D.R., Pal, A.K., de P. R. Moreira, I., Datta, S.N., Illas, F.. The triplet–singlet gap in the m-xylylene radical: A not so simple one. Journal of Chemical Theory and Computation 2013;10(1):335–345. doi:10.1021/ct400883m.

[343] Li, Y., Wang, L., Guo, X., Zhang, J.. A CASSCF/CASPT2 insight into excited-state intramolecular proton transfer of four imidazole derivatives. Journal of Computational Chemistry 2015;36(32):2374–2380. doi:10.1002/jcc.24216.

[344] Olsen, J.. The CASSCF method: A perspective and commentary. International Journal of Quantum Chemistry 2011;111(13):3267–3272. doi:10.1002/qua.23107.

[345] Fertitta, E., Booth, G.H.. Rigorous wave function embedding with dynamical fluctuations. Physical Review B 2018;98(23). doi:10.1103/physrevb.98.235132.

[346] Fertitta, E., Booth, G.H.. Energy-weighted density matrix embedding

of open correlated chemical fragments. The Journal of Chemical Physics 2019;151(1):014115. doi:10.1063/1.5100290.

[347] Sriluckshmy, P.V., Nusspickel, M., Fertitta, E., Booth, G.H.. Fully algebraic and self-consistent effective dynamics in a static quantum embedding. Physical Review B 2021;103(8). doi:10.1103/physrevb.103.085131.

[348] Levine, B.G., Durden, A.S., Esch, M.P., Liang, F., Shu, Y.. CAS without SCF—why to use CASCI and where to get the orbitals. Journal of Chemical Physics 2021;154(9):090902. doi:10.1063/5.0042147.

[349] Takeshita, T., Rubin, N.C., Jiang, Z., Lee, E., Babbush, R., McClean, J.R.. Increasing the representation accuracy of quantum simulations of chemistry without extra quantum resources. Phys Rev X 2020;10(1). doi:10.1103/physrevx.10.011004.

[350] Yalouz, S., Senjean, B., Günther, J., Buda, F., O'Brien, T.E., Visscher, L.. A state-averaged orbital-optimized hybrid quantum–classical algorithm for a democratic description of ground and excited states. Quantum Science and Technology 2021;6(2):024004. doi:10.1088/2058-9565/abd334.

[351] Abe, M., Gopakmar, G., Nakajima, T., Hirao, K.. Relativistic multireference perturbation theory: Complete active-space second-order perturbation theory (CASPT2) with the four-component Dirac Hamiltonian. In: Radiation Induced Molecular Phenomena in Nucleic Acids. Springer Netherlands; 2008, p. 157–177. doi:10.1007/978-1-4020-8184-2_6.

[352] Rossmannek, M., Barkoutsos, P.K., Ollitrault, P.J., Tavernelli, I.. Quantum HF/DFT-embedding algorithms for electronic structure calculations: Scaling up to complex molecular systems. The Journal of Chemical Physics 2021;154(11):114105. doi:10.1063/5.0029536.

[353] Schade, R., Bauer, C., Tamoev, K., Mazur, L., Plessl, C., Kühne, T.D.. Parallel quantum chemistry on noisy intermediate-scale quantum computers. 2022. URL: https://arxiv.org/abs/2202.02417. doi:10.48550/ARXIV.2202.02417.

[354] Knizia, G., Chan, G.K.L.. Density matrix embedding: A strong-coupling quantum embedding theory. Journal of Chemical Theory and Computation 2013;9(3):1428–1432. doi:10.1021/ct301044e.

[355] Wouters, S., Jiménez-Hoyos, C.A., Sun, Q., Chan, G.K.L.. A practical guide to density matrix embedding theory in quantum chemistry. Journal of Chemical Theory and Computation 2016;12(6):2706–2719. doi:10.1021/acs.jctc.6b00316.

[356] Georges, A., Kotliar, G., Krauth, W., Rozenberg, M.J.. Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions. Rev Mod Phys 1996;68(1):13–125. doi:10.1103/revmodphys.68.13.

[357] Rubin, N.C.. A hybrid classical/quantum approach for large-scale studies of quantum systems with density matrix embedding theory. 2016. arXiv:1610.06910v2.

[358] Yamazaki, T., Matsuura, S., Narimani, A., Saidmuradov, A., Zaribafiyan, A.. Towards the practical application of near-term quantum

computers in quantum chemistry simulations: A problem decomposition approach. 2018. arXiv:1806.01305v1.

[359] Ma, H., Govoni, M., Galli, G.. Quantum simulations of materials on near-term quantum computers. npj Computational Materials 2020;6(1). doi:10.1038/s41524-020-00353-z.

[360] Mineh, L., Montanaro, A.. Solving the Hubbard model using density matrix embedding theory and the variational quantum eigensolver. Physical Review B 2022;105(12). URL: https://doi.org/10.1103/physrevb.105.125117. doi:10.1103/physrevb.105.125117.

[361] Li, W., Huang, Z., Cao, C., Huang, Y., Shuai, Z., Sun, X., et al. Toward practical quantum embedding simulation of realistic chemical systems on near-term quantum computers. 2021. arXiv:2109.08062v1.

[362] Maier, T., Jarrell, M., Pruschke, T., Hettler, M.H.. Quantum cluster theories. Rev Mod Phys 2005;77:3, 1027–1080. URL: https://link.aps.org/doi/10.1103/RevModPhys.77.1027.

[363] Bauer, B., Wecker, D., Millis, A.J., Hastings, M.B., Troyer, M.. Hybrid quantum-classical approach to correlated materials. Phys Rev X 2016;6(3). doi:10.1103/physrevx.6.031045.

[364] Endo, S., Kurata, I., Nakagawa, Y.O.. Calculation of the Green's function on near-term quantum computers. Physical Review Research 2020;2(3):033281. doi:10.1103/physrevresearch.2.033281.

[365] Rungger, I., Fitzpatrick, N., Chen, H., Alderete, C.H., Apel, H., Cowtan, A., et al. Dynamical mean field theory algorithm and experiment on quantum computers. 2019. arXiv:1910.04735v2.

[366] Kreula, J.M., García-Álvarez, L., Lamata, L., Clark, S.R., Solano, E., Jaksch, D.. Few-qubit quantum-classical simulation of strongly correlated lattice fermions. EPJ Quantum Technol 2016;3(1). doi:10.1140/epjqt/s40507-016-0049-1.

[367] Jamet, F., Agarwal, A., Lupo, C., Browne, D.E., Weber, C., Rungger, I.. Krylov variational quantum algorithm for first principles materials simulations. 2021. arXiv:2105.13298v2.

[368] Jamet, F., Agarwal, A., Rungger, I.. Quantum subspace expansion algorithm for Green's functions. 2022. arXiv:2205.00094.

[369] Boguslawski, K., Tecmer, P., Barcza, G., Legeza, A., Reiher, M.. Orbital entanglement in bond-formation processes. Journal of Chemical Theory and Computation 2013;9:7, 2959–2973. URL: https://doi.org/10.1021/ct400247p.

[370] Mitarai, K., Yan, T., Fujii, K.. Generalization of the output of a variational quantum eigensolver by parameter interpolation with a low-depth ansatz. Physical Review Applied 2019;11(4). doi:10.1103/physrevapplied.11.044087.

[371] Lengsfield, B.H.. General second order MCSCF theory: A density matrix directed algorithm. J Chem Phys 1980;73:1, 382–390.

[372] Werner, H.J., Meyer, W.. A quadratically convergent multiconfiguration–self-consistent field method with simultaneous optimization of orbitals and ci coefficients. J Chem Phys 1980;73:5, 2342–2356.

[373] Werner, H.J., Knowles, P.J.. A second order multiconfiguration SCF procedure with optimum convergence. J Chem Phys 1985;82:11, 5053–5063.

[374] Werner, H.J.. Matrix-formulatoin direct multiconfiguration self-consistent field and multi-configuration reference configuration-interaction methods. Adv Chem Phys 1987;69:1–62.

[375] Olsen, J., Yeager, D.L., Jørgensen, P.. Optimization and characterization of a multiconfigurational self-consistent field (MCSCF) state. Adv Chem Phys 1983;54:1–176.

[376] Jensen, H.J.A., Jørgensen, P.. A direct approach to second-order MCSCF calculations using a norm extended optimization scheme. J Chem Phys 1984;80:3, 1204–1214.

[377] Shepard, R.. The multiconfiguration self-consistent field method. Adv Chem Phys 1987;69:63–200.

[378] Ten-no, S., Iwata, S.. Multiconfiguration self-consistent field procedure employing linear combination of atomic-electron distributions. J Chem Phys 1996;105:9, 3604–3611.

[379] Schmidt, M.W., Gordon, M.S.. The construction and interpretation of MCSCF wavefunctions. Annu Rev Phys Chem 1998;:1, 233–266.

[380] Nakano, H., Hirao, K.. A quasi-complete active space self-consistent field method. Chem Phys Lett 2000;317:12, 90–96.

[381] Aquilante, F., Pedersen, T.B., Lindh, R., Roos, B.O., de Meras, A.S., Koch, H.. Accurate *ab initio* density fitting for multiconfigurational self-consistent field methods. J Chem Phys 2008;129:2, 024113.

[382] Szalay, P.G., Müller, T., Gidofalvi, G., Lischka, H., Shepard, R.. Multiconfiguration self-consistent field and multireference configuration interaction methods and applications. Chem Rev 2012;112:1, 108–181.

[383] Győrffy, W., Shiozaki, T., Knizia, G., Werner, H.J.. Analytical energy gradients for second-order multireference perturbation theory using density fitting. J Chem Phys 2013;138:10, 104104.

[384] Kim, I., Parker, S.M., Shiozaki, T.. Orbital optimization in the active space decomposition model. Journal of chemical theory and computation 2015;11:8, 3636–3642.

[385] Hohenstein, E.G., Luehr, N., Ufimtsev, I.S., Martinez, T.J.. An atomic orbital-based formulation of the complete active space self-consistent field method on graphical processing units. J Chem Phys 2015;142:22, 224103.

[386] Larsson, S.. Applications of CASSCF. International Journal of Quantum Chemistry 2011;111(13):3424–3430. URL: https://doi.org/10.1002/qua.23016. doi:10.1002/qua.23016.

[387] Reiher, M., Wiebe, N., Svore, K.M., Wecker, D., Troyer, M.. Elucidating reaction mechanisms on quantum computers. Proceedings of the National Academy of Sciences 2017;114:29, 7555–7560.

[388] Jeong, W., Stoneburner, S.J., King, D., Li, R., Walker, A., Lindh, R., et al. Automation of active space selection for multireference methods via machine learning on chemical bond dissociation. Journal of Chemical Theory and Computation 2020;16:4, 2389–2399. URL: https://doi.org/10.1021/acs.jctc.9b01297.

[389] Sayfutyarova, E.R., Sun, Q., Chan, G.K.L., Knizia, G.. Automated construction of molecular active spaces from atomic valence orbitals. J Chem Theory Comput 2017;13:9, 4063–4078.

[390] Thomas, R.E., Sun, Q., Alavi, A., Booth, G.H.. Stochastic multiconfigurational self-consistent field theory. Journal of Chemical Theory and Computation 2015;11:11, 5316–5325.

[391] Anderson, R.J., Shiozaki, T., Booth, G.H.. Efficient and stochastic multireference perturbation theory for large active spaces within a full configuration interaction quantum Monte Carlo framework. J Chem Phys 2020;152:5, 054101. doi:10.1063/1.5140086.

[392] Ghosh, D., Hachmann, J., Yanai, T., Chan, G.K.L.. Orbital optimization in the density matrix renormalization group, with applications to polyenes and Î²-carotene. The Journal of Chemical Physics 2008;128:14, 144117. doi:10.1063/1.2883976.

[393] Zgid, D., Nooijen, M.. The density matrix renormalization group self-consistent field method: Orbital optimization with the density matrix renormalization group method in the active space. The Journal of Chemical Physics 2008;128:14, 144116. doi:10.1063/1.2883981.

[394] Yeter-Aydeniz, K., Gard, B.T., Jakowski, J., Majumder, S., Barron, G.S., Siopsis, G., et al. Benchmarking quantum chemistry computations with variational, imaginary time evolution, and Krylov space solver algorithms. 2021. arXiv:2102.05511.

[395] Sun, Q., Berkelbach, T.C., Blunt, N.S., Booth, G.H., Guo, S., Li,

Z., et al. PySCF: the Python-based simulations of chemistry framework. WIREs Computational Molecular Science 2018;8:1, 1340.

[396] Sun, Q., Zhang, X., Banerjee, S., Bao, P., Barbry, M., Blunt, N.S., et al. Recent developments in the PySCF program package. The Journal of Chemical Physics 2020;153:2, 024109. doi:10.1063/5.0006074.

[397] Li Manni, G., Smart S, D., Alavi, A.. Combining the complete active space self-consistent field method and the full configuration interaction quantum Monte Carlo within a super-CI framework, with application to challenging metal-porphyrins. Journal of Chemical Theory and Computation 2016;12:3, 1245–1258. doi:10.1021/acs.jctc.5b01190.

[398] Halson, J.J., Anderson, R.J., Booth, G.H.. Improved stochastic multireference perturbation theory for correlated systems with large active spaces. Molecular Physics 2020;118:19–20, e1802072. doi:10.1080/00268976.2020.1802072.

[399] Scuseria, G.E., Miller, M.D., Jensen, F., Geertsen, J.. The dipole moment of carbon monoxide. The Journal of Chemical Physics 1991;94(10):6660–6663. URL: https://doi.org/10.1063/1.460293. doi:10.1063/1.460293.

[400] Angeli, C., Pastore, M., Cimiraglia, R.. New perspectives in multireference perturbation theory: the n-electron valence state approach. Theor Chem Acc 2007;117:5, 743–754. URL: https://doi.org/10.1007/s00214-006-0207-0.

[401] Vallury, H.J., Jones, M.A., Hill, C.D., Hollenberg, L.C.L.. Quantum

computed moments correction to variational estimates. Quantum 2020;4:373. doi:10.22331/q-2020-12-15-373.

[402] Cohen, J., Mori-Sánchez, P.,, , Yang, W.. Challenges for density functional theory. Chemical Reviews 2012;112:289–320.

[403] Kotliar, G., Savrasov, S.Y., Haule, K., Oudovenko, V.S., Parcollet, O.,, , et al. Electronic structure calculations with dynamical mean-field theory. Rev Mod Phys 2006;78:865–951.

[404] Sun, Q., Chan, G.K.L.. Quantum embedding theories. Accounts of Chemical Research 2016;49:12, 2705–2712. doi:10.1021/acs.accounts.6b00356.

[405] Knizia, G., Chan, G.K.L.. Density matrix embedding: A simple alternative to dynamical mean-field theory. Phys Rev Lett 2012;109:186404.

[406] Zheng, B.X., Kretchmer, J.S., Shi, H., Zhang, S., Chan, G.K.L.. Cluster size convergence of the density matrix embedding theory and its dynamical cluster formulation: A study with an auxiliary-field quantum Monte Carlo solver. Phys Rev B 2017;95:045103.

[407] Keen, T., Maier, T., Johnston, S., Lougovski, P.. Quantum-classical simulation of two-site dynamical mean-field theory on noisy quantum hardware. Quantum Science and Technology 2019;5:3. URL: http://arxiv.org/abs/1910.09512.

[408] Yao, Y., Zhang, F., Wang, C.Z., Ho, K.M., Orth, P.P.. Gutzwiller Hybrid Quantum-Classical Computing Approach for Correlated Materials. Phys Rev Research 2021;3:013184.

[409] Dhawan, D., Metcalf, M., Zgid, D.. Dynamical Self-energy Mapping (DSEM) for quantum computing. arXiv:201005441v2 2020;.

[410] Kawashima, Y., Coons, M.P., Nam, Y., Lloyd, E., Matsuura, S., Garza, A.J., et al. Efficient and accurate electronic structure simulation demonstrated on a trapped-ion quantum computer. arXiv:210207045 2021;.

[411] Sun, S.N., Motta, M., Tazhigulov, R.N., Tan, A.T., Chan, G.K.L., Minnich, A.J.. Quantum computation of finite-temperature static and dynamical properties of spin systems using quantum imaginary time evolution. PRX Quantum 2021;2(1):010317. doi:10.1103/prxquantum.2.010317.

[412] Bulik, I.W., Scuseria, G.E., Dukelsky, J.. Density matrix embedding from broken symmetry lattice mean fields. Physical Review B 2014;89(3). URL: https://doi.org/10.1103/physrevb.89.035140. doi:10.1103/physrevb.89.035140.

[413] Zheng, B.X., Chan, G.K.L.. Ground-state phase diagram of the square lattice Hubbard model from density matrix embedding theory. Physical Review B 2016;93(3). URL: https://doi.org/10.1103/physrevb.93.035126. doi:10.1103/physrevb.93.035126.

[414] LeBlanc, J., Antipov, A.E., Becca, F., Bulik, I.W., Chan, G.K.L., Chung, C.M., et al. Solutions of the two-dimensional Hubbard model: Benchmarks and results from a wide range of numerical algorithms. Physical Review X 2015;5(4). URL: https://doi.org/10.1103/physrevx.5.041041. doi:10.1103/physrevx.5.041041.

[415] Sandhoefer, B., Chan, G.K.L.. Density matrix embedding theory for interacting electron-phonon systems. Physical Review B 2016;94(8). URL: https://doi.org/10.1103/physrevb.94.085115. doi:10.1103/physrevb.94.085115.

[416] Tsuchimochi, T., Welborn, M., Voorhis, T.V.. Density matrix embedding in an antisymmetrized geminal power bath. The Journal of Chemical Physics 2015;143(2):024107. URL: https://doi.org/10.1063/1.4926650. doi:10.1063/1.4926650.

[417] Fan, Z., lin Jie, Q.. Cluster density matrix embedding theory for quantum spin systems. Physical Review B 2015;91(19). URL: https://doi.org/10.1103/physrevb.91.195118. doi:10.1103/physrevb.91.195118.

[418] Bulik, I.W., Chen, W., Scuseria, G.E.. Electron correlation in solids via density embedding theory. The Journal of Chemical Physics 2014;141(5):054113. URL: https://doi.org/10.1063/1.4891861. doi:10.1063/1.4891861.

[419] Metzner, W., Vollhardt, D.. Correlated lattice fermions in $d = \infty$ dimensions. Phys Rev Lett 1989;62:3, 324–327. URL: https://link.aps.org/doi/10.1103/PhysRevLett.62.324.

[420] Gujrati, P.D.. Bethe or Bethe-like lattice calculations are more reliable than conventional mean-field calculations. Phys Rev Lett 1995;74:5, 809–812. URL: https://link.aps.org/doi/10.1103/PhysRevLett.74.809.

[421] Eckstein, M., Kollar, M., Byczuk, K., Vollhardt, D.. Hopping on the

Bethe lattice: Exact results for densities of states and dynamical mean-field theory. Phys Rev B 2005;71:23, 235119. URL: https://link.aps.org/doi/10.1103/PhysRevB.71.235119.

[422] Nusspickel, M., Booth, G.H.. Frequency-dependent and algebraic bath states for a dynamical mean-field theory with compact support. Phys Rev B 2020;101:4, 045126. URL: https://link.aps.org/doi/10.1103/PhysRevB.101.045126.

[423] Kollar, M.. Construction of a dispersion relation from an arbitrary density of states. International Journal of Modern Physics B 2002;16:23, 3491–3501.

[424] Bulla, R.. Zero temperature metal-insulator transition in the infinite-dimensional Hubbard model. Phys Rev Lett 1999;83:136–139. URL: https://link.aps.org/doi/10.1103/PhysRevLett.83.136. doi:10.1103/PhysRevLett.83.136.

[425] Karim, K.M.R., Ong, H.R., Abdullah, H., Yousuf, A., Cheng, C.K., Khan, M.M.R.. Photoelectrochemical reduction of carbon dioxide to methanol on p-type CuFe2o4 under visible light irradiation. International Journal of Hydrogen Energy 2018;43(39):18185–18193. doi:10.1016/j.ijhydene.2018.07.174.

[426] Zeinalipour-Yazdi, C.D., Hargreaves, J.S.J., Catlow, C.R.A.. Low-t mechanisms of ammonia synthesis on Co3Mo3N. The Journal of Physical Chemistry C 2018;122(11):6078–6082. doi:10.1021/acs.jpcc.7b12364.

[427] Matsika, S., Krylov, A.I.. Introduction: Theoretical modeling of excited

state processes. Chem Rev 2018;118(15):6925–6926. doi:10.1021/acs.chemrev.8b00436.

[428] Stanton, J.F., Bartlett, R.J.. The equation of motion coupled-cluster method. a systematic biorthogonal approach to molecular excitation energies, transition probabilities, and excited state properties. The Journal of Chemical Physics 1993;98(9):7029–7039. doi:10.1063/1.464746.

[429] Monkhorst, H.J.. Calculation of properties with the coupled-cluster method. International Journal of Quantum Chemistry 2009;12(S11):421–432. doi:10.1002/qua.560120850.

[430] Jeziorski, B., Monkhorst, H.J.. Coupled-cluster method for multideterminantal reference states. Physical Review A 1981;24(4):1668–1681. doi:10.1103/physreva.24.1668.

[431] Yeter-Aydeniz, K., Pooser, R.C., Siopsis, G.. Practical quantum computation of chemical and nuclear energy levels using quantum imaginary time evolution and lanczos algorithms. npj Quantum Information 2020;6(1):63. doi:10.1038/s41534-020-00290-1.

[432] McClean, J.R., Kimchi-Schwartz, M.E., Carter, J., de Jong, W.A.. Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states. Physical Review A 2017;95(4):042308. doi:10.1103/physreva.95.042308.

[433] Ollitrault, P.J., Kandala, A., Chen, C.F., Barkoutsos, P.K., Mezzacapo, A., Pistoia, M., et al. Quantum equation of motion for computing molecular excitation energies on a noisy quantum processor. Physical Review Research 2020;2(4). doi:10.1103/physrevresearch.2.043140.

[434] Parrish, R.M., Hohenstein, E.G., McMahon, P.L., Martínez, T.J.. Quantum computation of electronic transitions using a variational quantum eigensolver. Physical Review Letters 2019;122(23):230401–230401. doi:10.1103/physrevlett.122.230401.

[435] Colless, J., Ramasesh, V., Dahlen, D., Blok, M., Kimchi-Schwartz, M., McClean, J., et al. Computation of molecular spectra on a quantum processor with an error-resilient algorithm. Physical Review X 2018;8(1):011021–011021. doi:10.1103/physrevx.8.011021.

[436] Blunt, N.S., Alavi, A., Booth, G.H.. Nonlinear biases, stochastically sampled effective Hamiltonians, and spectral functions in quantum Monte Carlo methods. Physical Review B 2018;98(8). doi:10.1103/physrevb.98.085118.

[437] Epperly, E.N., Lin, L., Nakatsukasa, Y.. A theory of quantum subspace diagonalization. 2021. arXiv:2110.07492.

[438] Ryabinkin, I.G., Genin, S.N., Izmaylov, A.F.. Constrained variational quantum eigensolver: Quantum computer search engine in the Fock space. Journal of Chemical Theory and Computation 2018;15(1):249–255. doi:10.1021/acs.jctc.8b00943.

[439] Wang, L.W., Zunger, A.. Solving Schrödinger's equation around a desired energy: Application to silicon quantum dots. The Journal of Chemical Physics 1994;100(3):2394–2397. URL: https://doi.org/10.1063/1.466486. doi:10.1063/1.466486.

[440] Liu, S., Zhang, S.X., Hsieh, C.Y., Zhang, S., Yao, H.. Probing

many-body localization by excited-state VQE. 2021. URL: https://arxiv.org/abs/2111.13719. doi:10.48550/ARXIV.2111.13719.

[441] Nakanishi, K.M., Mitarai, K., Fujii, K.. Subspace-search variational quantum eigensolver for excited states. Physical Review Research 2019;1(3):1–7. doi:10.1103/physrevresearch.1.033062.

[442] Higgott, O., Wang, D., Brierley, S.. Variational quantum computation of excited states. Quantum 2019;3:156. doi:10.22331/q-2019-07-01-156.

[443] Jones, T., Endo, S., McArdle, S., Yuan, X., Benjamin, S.C.. Variational quantum algorithms for discovering Hamiltonian spectra. Physical Review A 2019;99(6). doi:10.1103/physreva.99.062304.

[444] Chan, H.H.S., Fitzpatrick, N., Segarra-Marti, J., Bearpark, M.J., Tew, D.P.. Molecular excited state calculations with adaptive wavefunctions on a quantum eigensolver emulation: Reducing circuit depth and separating spin states. 2021. arXiv:2105.10275.

[445] Kottmann, J.S., Anand, A., Aspuru-Guzik, A.. A feasible approach for automatically differentiable unitary coupled-cluster on quantum computers. Chemical Science 2021;12(10):3497–3508. URL: https://doi.org/10.1039/d0sc06627c. doi:10.1039/d0sc06627c.

[446] Wakaura, H., Suksmono, A.B.. Tangent vector variational quantum eigensolver: A robust variational quantum eigensolver against the inaccuracy of derivative. 2021. arXiv:2105.01141.

[447] Zhang, D.B., Yuan, Z.H., Yin, T.. Variational quantum eigensolvers by variance minimization. 2020. arXiv:2006.15781.

[448] Bartlett, J.H., Gibbons, J.J., Dunn, C.G.. The normal helium atom. Physical Review 1935;47(9):679–680. doi:10.1103/physrev.47.679.

[449] Siringo, F., Marotta, L.. A variational method from the variance of energy. The European Physical Journal C 2005;44(2):293–298. doi:10.1140/epjc/s2005-02358-x.

[450] Umrigar, C.J., Filippi, C.. Energy and variance optimization of many-body wave functions. Physical Review Letters 2005;94(15). doi:10.1103/physrevlett.94.150201.

[451] Khemani, V., Pollmann, F., Sondhi, S.. Obtaining highly excited eigenstates of many-body localized Hamiltonians by the density matrix renormalization group approach. Physical Review Letters 2016;116(24). doi:10.1103/physrevlett.116.247204.

[452] Pollmann, F., Khemani, V., Cirac, J.I., Sondhi, S.L.. Efficient variational diagonalization of fully many-body localized Hamiltonians. Physical Review B 2016;94(4). doi:10.1103/physrevb.94.041116.

[453] Vicentini, F., Biella, A., Regnault, N., Ciuti, C.. Variational neural-network ansatz for steady states in open quantum systems. Physical Review Letters 2019;122(25). doi:10.1103/physrevlett.122.250503.

[454] Zhang, F., Gomes, N., Yao, Y., Orth, P.P., Iadecola, T.. Adaptive variational quantum eigensolvers for highly excited states. Physical Review B 2021;104(7). doi:10.1103/physrevb.104.075159. arXiv:2104.12636v1; phys. Rev. B 104, 075159 (2021).

[455] Tilly, J., Jones, G., Chen, H., Wossnig, L., Grant, E.. Computation of molecular excited states on IBM quantum computers using a discrim-

inative variational quantum eigensolver. Physical Review A 2020;102(6). doi:10.1103/physreva.102.062425.

[456] Hylleraas, E.A., Undheim, B.. Numerische berechnung der 2s-terme von ortho- und par-helium. Zeitschrift fur Physik 1930;65(11-12):759–772. URL: https://doi.org/10.1007/bf01397263. doi:10.1007/bf01397263.

[457] MacDonald, J.K.L.. Successive approximations by the Rayleigh-Ritz variation method. Physical Review 1933;43(10):830–833. URL: https://doi.org/10.1103/physrev.43.830. doi:10.1103/physrev.43.830.

[458] Lloyd, S., Weedbrook, C.. Quantum generative adversarial learning. Physical Review Letters 2018;121(4). doi:10.1103/physrevlett.121.040502.

[459] Benedetti, M., Grant, E., Wossnig, L., Severini, S.. Adversarial quantum circuit learning for pure state approximation. New Journal of Physics 2019;21(4):043023. doi:10.1088/1367-2630/ab14b5.

[460] Fuchs, C.A.. Distinguishability and accessible information in quantum theory. PhD Thesis, arXiv preprint quant-ph/9601020 1996;URL: https://arxiv.org/abs/quant-ph/9601020.

[461] Riedmiller, M., Braun, H.. A direct adaptive method for faster backpropagation learning: the RPROP algorithm. In: IEEE International Conference on Neural Networks. IEEE; 1993, p. 586–591. URL: https://doi.org/10.1109/icnn.1993.298623. doi:10.1109/icnn.1993.298623.

[462] Cross, A.W., Bishop, L.S., Sheldon, S., Nation, P.D., Gambetta, J.M.. Validating quantum computers using randomized model circuits. Phys Rev A 2019;100:032328. URL: https://journals.aps.org/pra/abstract/10.1103/PhysRevA.100.032328. doi:10.1103/PhysRevA.100.032328.

[463] Stenger, J.P.T., Gunlycke, D., Hellberg, C.S.. Expanding variational quantum eigensolvers to larger systems by dividing the calculations between classical and quantum hardware. Physical Review A 2022;105(2). URL: https://doi.org/10.1103/physreva.105.022438. doi:10.1103/physreva.105.022438.

[464] Okada, K.N., Osaki, K., Mitarai, K., Fujii, K.. Identification of topological phases using classically-optimized variational quantum eigensolver. 2022. URL: https://arxiv.org/abs/2202.02909. doi:10.48550/ARXIV.2202.02909.

[465] Zhang, Y., Cincio, L., Negre, C.F.A., Czarnik, P., Coles, P., Anisimov, P.M., et al. Variational quantum eigensolver with reduced circuit complexity. 2021. arXiv:2106.07619v1.

[466] Gambetta, J., McClure, D.. Hitting a quantum volume chord: IBM quantum adds six new systems with quantum volume 32. 2020. IBM Research Blog.