



The Variational Quantum Eigensolver: A review of methods and best practices

Jules Tilly^{a,b,*}, Hongxiang Chen^{b,c}, Shuxiang Cao^{b,d}, Dario Picozzi^{a,b}, Kanav Setia^e, Ying Li^f, Edward Grant^{b,c}, Leonard Wossnig^{b,c}, Ivan Rungger^g, George H. Booth^h, Jonathan Tennyson^a

^a University College London, Department of Physics and Astronomy, WC1E 6BT London, United Kingdom

^b Odyssey Therapeutics, WC1X 8BB London, United Kingdom

^c University College London, Department of Computer Science, WC1E 6BT London, United Kingdom

^d University of Oxford, Department of Physics, Clarendon Laboratory, OX1 3PU, United Kingdom

^e qBraid, 5235, Harper Court, Chicago, IL, United States of America

^f Graduate School of China Academy of Engineering Physics, Beijing 100193, China

^g National Physical Laboratory, TW11 0LW Teddington, United Kingdom

^h King's College London, Department of Physics, Strand, London, WC2R 2LS, United Kingdom

ARTICLE INFO

Article history:

Received 9 November 2021

Received in revised form 29 August 2022

Accepted 30 August 2022

Available online 23 September 2022

Editor: Giovanna Morigi

Keywords:

Quantum computing

Variational methods

Noisy intermediate scale quantum devices

Electronic structure

Quantum chemistry

Many-body physics

ABSTRACT

The variational quantum eigensolver (or VQE), first developed by Peruzzo et al. (2014), has received significant attention from the research community in recent years. It uses the variational principle to compute the ground state energy of a Hamiltonian, a problem that is central to quantum chemistry and condensed matter physics. Conventional computing methods are constrained in their accuracy due to the computational limits facing exact modeling of the exponentially growing electronic wavefunction for these many-electron systems. The VQE may be used to model these complex wavefunctions in polynomial time, making it one of the most promising near-term applications for quantum computing. One important advantage is that variational algorithms have been shown to present some degree of resilience to the noise in the quantum hardware. Finding a path to navigate the relevant literature has rapidly become an overwhelming task, with many methods promising to improve different parts of the algorithm, but without clear descriptions of how the diverse parts fit together. The potential practical advantages of the algorithm are also widely discussed in the literature, but with varying conclusions. Despite strong theoretical underpinnings suggesting excellent scaling of individual VQE components, studies have pointed out that their various pre-factors could be too large to reach a quantum computing advantage over conventional methods.

This review aims at disentangling the relevant literature to provide a comprehensive overview of the progress that has been made on the different parts of the algorithm, and to discuss future areas of research that are fundamental for the VQE to deliver on its promises. All the different components of the algorithm are reviewed in detail. These include the representation of Hamiltonians and wavefunctions on a quantum computer, the optimization process to find ground state energies, the post processing mitigation of quantum errors, and suggested best practices. We identify four main areas of future research: (1) optimal measurement schemes for reduction of circuit repetitions required; (2) large scale parallelization across many quantum computers; (3) ways to overcome the potential appearance of vanishing gradients in the optimization process for large systems, and how the number of iterations required for the optimization scales

* Correspondence to: Department of Physics and Astronomy, University College London, Gower St, London WC1E 6BT.

E-mail address: jules.tilly.18@ucl.ac.uk (J. Tilly).

with system size; (4) the extent to which VQE suffers for quantum noise, and whether this noise can be mitigated in a tractable manner. The answers to these open research questions will determine the routes for the VQE to achieve quantum advantage as the quantum computing hardware scales up and as the noise levels are reduced.

© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Contents

1.	Introduction.....	4
2.	Overview of VQE	7
2.1.	A formal definition of the VQE	7
2.2.	The VQE pipeline	8
2.3.	Advantage argument, assumptions, and limitations of the VQE	9
2.4.	VQE and conventional computational chemistry.....	11
2.4.1.	Full configuration interaction	11
2.4.2.	Efficient approximate wavefunction parameterizations for conventional computation	12
2.5.	VQE and quantum phase estimation.....	13
2.5.1.	Overview of the quantum phase estimation	13
2.5.2.	Discussion and comparison.....	15
2.6.	Our suggested best practices for VQE and their scaling assessment.....	15
2.6.1.	Best practices for <i>ab initio</i> electronic structure of molecular systems.....	16
2.6.2.	Best practices for lattice models	19
2.7.	Resource estimate for VQE	20
2.7.1.	Cost and runtime estimates for VQE	20
2.7.2.	Parallelization potential of the VQE.....	21
2.7.3.	Distribution of resources between quantum and conventional computation	22
3.	Hamiltonian representation	22
3.1.	The electronic structure hamiltonian.....	22
3.1.1.	The <i>ab initio</i> molecular hamiltonian.....	22
3.1.2.	Construction of the wavefunction.....	23
3.1.3.	Hamiltonian quantization.....	24
3.2.	Other Hamiltonian models	26
3.2.1.	Lattice Hamiltonians	26
3.2.2.	Vibrational Hamiltonian model	26
3.2.3.	Periodic systems.....	27
4.	Fermionic space to spin space transformations	27
4.1.	Generalized encodings	28
4.1.1.	The Jordan–Wigner encoding	28
4.1.2.	The parity encoding.....	29
4.1.3.	The Bravyi–Kitaev encoding.....	30
4.1.4.	Optimal general encoding based on ternary trees.....	32
4.1.5.	Discussion on generalized encodings	32
4.2.	Lattice model tailored encoding	33
4.2.1.	Auxiliary fermion schemes	34
4.2.2.	Superfast encoding/loop stabilizer encodings.....	35
4.2.3.	Discussion on lattice tailored encodings	39
4.3.	Reducing qubit requirements.....	39
5.	Efficient grouping and measuring strategies	40
5.1.	Scaling of shot numbers in VQE	40
5.1.1.	Overall scaling of measurements	40
5.1.2.	Measurement weighting.....	41
5.2.	Pauli string groupings, and other joint measurement strategies.....	42
5.2.1.	Inference methods.....	42
5.2.2.	Hamiltonian partitioning based on commutativity.....	44
5.2.3.	Unitary partitioning of the hamiltonian:	46
5.2.4.	Decomposed interactions for efficient joint measurements:.....	48
5.2.5.	Grouping heuristics.....	49
5.3.	Discussion on measurement strategies and grouping methods	50
6.	Ansatz selection and construction.....	51
6.1.	The barren plateau problem.....	52
6.1.1.	Drivers of the barren plateau problem.....	52
6.1.2.	Methods to address barren plateau problem	54
6.1.3.	Comments on barren plateau in the context of the VQE	55

6.2.	Fixed structure ansätze.....	55
6.2.1.	Hardware-efficient ansatz (HEA).....	56
6.2.2.	The Unitary Coupled Cluster (UCC) ansatz and extensions.....	57
6.2.3.	Symmetry-preserving methods.....	61
6.2.4.	The Hamiltonian Variational Ansatz and extensions.....	64
6.3.	Adaptative structure ansätze.....	65
6.3.1.	Iterative ansatz growth methods (ADAPT-VQE and extensions).....	66
6.3.2.	Iterative Hamiltonian dressing (iterative Qubit Coupled Cluster (iQCC) and extensions).....	68
6.3.3.	Ansatz structure optimization.....	71
6.4.	Discussion on ansätze used in the VQE context.....	71
7.	Optimization strategies.....	72
7.1.	Background and notation.....	72
7.2.	Gradient evaluation.....	73
7.2.1.	Stochastic approximation methods.....	73
7.2.2.	Analytical gradient calculation.....	74
7.3.	Gradient-based searching strategy.....	76
7.3.1.	First order optimizers.....	76
7.3.2.	Second order optimizers.....	77
7.4.	Gradient-free searching strategy.....	79
7.4.1.	Gradient-free optimizers.....	79
7.4.2.	Analytical optimization.....	79
7.5.	Engineering cost function.....	83
7.6.	Discussion.....	84
8.	Error mitigation for VQE.....	85
8.1.	Symmetry verification.....	85
8.2.	Extrapolation based methods.....	86
8.2.1.	Method to systematically increase the noise.....	87
8.2.2.	Modeling the noise and extrapolating from measured data.....	89
8.3.	Probabilistic error cancellation.....	91
8.4.	Exponential error suppression.....	93
8.5.	Measurement readout error mitigation.....	94
8.5.1.	Measurement calibration matrix.....	94
8.5.2.	Correcting measurement outcomes.....	94
8.5.3.	Exploiting state-dependent noise.....	95
8.6.	Other error mitigation methods.....	95
8.7.	Impact of error mitigation on the scaling of VQE.....	96
8.8.	Noise robustness of VQE algorithms.....	97
9.	Beyond the ground state of isolated molecules: Extensions of VQE.....	98
9.1.	Excited states VQE.....	98
9.2.	VQE as a solver of correlated subspaces in multiscale methods.....	99
10.	Conclusion and outlook.....	101
	Declaration of competing interest.....	102
	Acknowledgments.....	102
	Appendix A. Qubit encodings and fenwick trees.....	102
	Appendix B. Hadamard test.....	103
	Appendix C. Error mitigation appendix.....	104
C.1.	Common noise models.....	104
C.1.1.	Relaxation rates.....	104
C.1.2.	Over-rotation.....	104
C.1.3.	Depolarizing noise model.....	105
C.1.4.	Dephasing noise model.....	105
C.1.5.	Damping error.....	105
C.2.	Example for probabilistic error cancellation.....	105
C.3.	Implementation of exponential error suppression.....	106
C.3.1.	Ancilla assisted method.....	106
C.3.2.	Diagonalization method.....	107
C.3.3.	Dual state purification method.....	107
C.3.4.	Shadow tomography based method.....	108
	References.....	109

1. Introduction

Quantum computing has undergone rapid development over recent years: from first conceptualization 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.

The Variational Quantum Eigensolver (VQE) was originally developed by Peruzzo et al. [37], and its theoretical framework was extended and formalized 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], 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 theorized 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 Kantowski–Sachs cosmology [75].

The VQE starts with an initialized 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 optimizing the parameters of the ansatz in order to minimize 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 [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, which is discussed in more detail in Section 6.2.2). Since then, many alternatives have been proposed, with ansätze 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.

While the design of the ansatz is core to the VQE and to the theoretical underpinnings of its potential advantage over conventional methods, there are many other parts of the algorithm that directly affect the cost and feasibility of the approach. We provide a discussion for each of these parts in our review, and briefly outline them here. To begin with, once one has chosen a quantum system to study, one must decide on how to construct its Hamiltonian. It is a requirement of the VQE that the Hamiltonian can be written as a sum of individual terms, where the number of terms increases with system size relatively slowly. This is a feature of electronic Hamiltonians, where the Coulomb interaction, a sum of two-body terms, scales polynomially in the system size [81]. Often there is considerable flexibility and freedom in the specific choice or how a Hamiltonian is represented mathematically. This step is critical, as it impacts all aspects of the VQE: number of qubits, depth of the ansatz, and the total number of measurements required. Once a Hamiltonian is constructed, one must translate it into operators that can be directly measured on a quantum computer (spin or Pauli operators), which can again impact the depth of the ansatz and number of measurements. The next step is to decide on an ansatz, which must be expressive enough to be able to approximately model the ground state wavefunction, but not so much that it results in circuits that become too deep, or parameterizations that are too complex to efficiently train. Once an ansatz is selected, one must choose an appropriate optimizer, which significantly impacts the convergence rate of the VQE optimization and the overall cost of the algorithm. Measurements on quantum devices are effectively stochastic [76], meaning that to obtain the expectation value of an operator, one must repeat measurements enough times to achieve a given level of precision. The number of measurements required is not only defined by the required precision, but also by the number of operators in the Hamiltonian. Adopting efficient measurement strategies is critical for controlling the implementation cost of VQE. Finally, to reduce the possible impact of quantum noise on the accuracy of the result, one can introduce error mitigation strategies into the individual measurements, which again need consideration of their associated computational cost.

This review holds two main purposes: (1) provide an overview for each of the different components of the VQE algorithm outlined above, along with a suggestion for best practices found in the literature. Because of the large variety of quantum systems the VQE can handle, we propose sets of best practices for two different and fairly general types of systems: *ab initio* molecular systems and spin–lattice models, though it is likely that some of the conclusions can be generalized to other Hamiltonians, including impurity models, condensed matter, vibrational spectroscopy or nuclear structure; (2) identify a list of open research questions regarding the future applicability of the VQE, and about individual aspects of the algorithm presented above. We identify four significant (and inter-related) hurdles, which impact the future viability of this method. These are presented here at a high level and discussed throughout the review.

Firstly, several studies have considered the scaling of the number of measurements as the system size increases [82–85]. Despite scaling polynomially [38], the number of measurements required can nonetheless rapidly become too large for the method to be viable [82,83,85]. As a consequence, many studies have called for developments in efficient measurement schemes (for instance, Ref. [85]), with approaches to find compact representations of the Hamiltonian and judicious use of joint measurement of commuting observables an ongoing source of developments in this endeavor (see Section 5.2.4).

Secondly, an alternative approach to the measurement problem could lie in the enormous potential for parallelization of VQE. This potential was clearly called for in the original VQE paper of Peruzzo et al. [37], but has received relatively little consideration from the community, in particular with respect to efficient strategies for this parallelization and possible communication overheads (further discussion on these points is found in Section 2.7.2).

Thirdly, variational quantum algorithms suffer from a common pitfall in the challenge of parameter optimization: the so-called barren plateau problem [82,86,87]. The barren plateau is a result of the phenomenon that gradients of the VQE parameters vanish exponentially with the number of qubits used in the model [86], in the overall expressibility of the ansatz [88], degree of entanglement of the wavefunction modeled [89,90], or non-locality of the cost function [87,91,92]. This of course would prevent tractable optimization in larger systems. Many methods have been proposed to address this problem, but it is currently unclear whether these methods allow one to fully contain the barren plateau problem in the context of VQE (further discussions on this topic are found in Section 6.1). A related point is that the complexity of the optimization landscape for the VQE remains poorly understood. Bittel and Kliesch [93] characterized specific features of this landscape and demonstrated the risk of optimization being undermined by local minima and non-convex features. The underlying question here is the convergence rate of different optimizers, and the overall scaling of the number of iterations required to reach convergence as a function of the system size and entanglement in the wavefunction.

Finally, the extent to which VQE is resilient to quantum noise and its ability to be mitigated in a tractable manner remains unclear. Variational algorithms have the advantage of exhibiting some ability to learn away certain types of systematic noise [94,95]. While it was also shown that the cost of several error mitigation methods may largely outweigh the benefits [96], the question of the impact of quantum noise on VQE results and the predictability of these errors should be studied in much further depth than has been done so far. Overall, despite a number of impediments, the VQE has the potential to be among the first useful methods to be implemented on NISQ devices. However, as the hardware continues to progress, so must the theoretical underpinning of this approach, as well as the efficiency and robustness of the software and algorithms, to guarantee that the potential benefits of methods such as VQE can be tapped as early as possible.

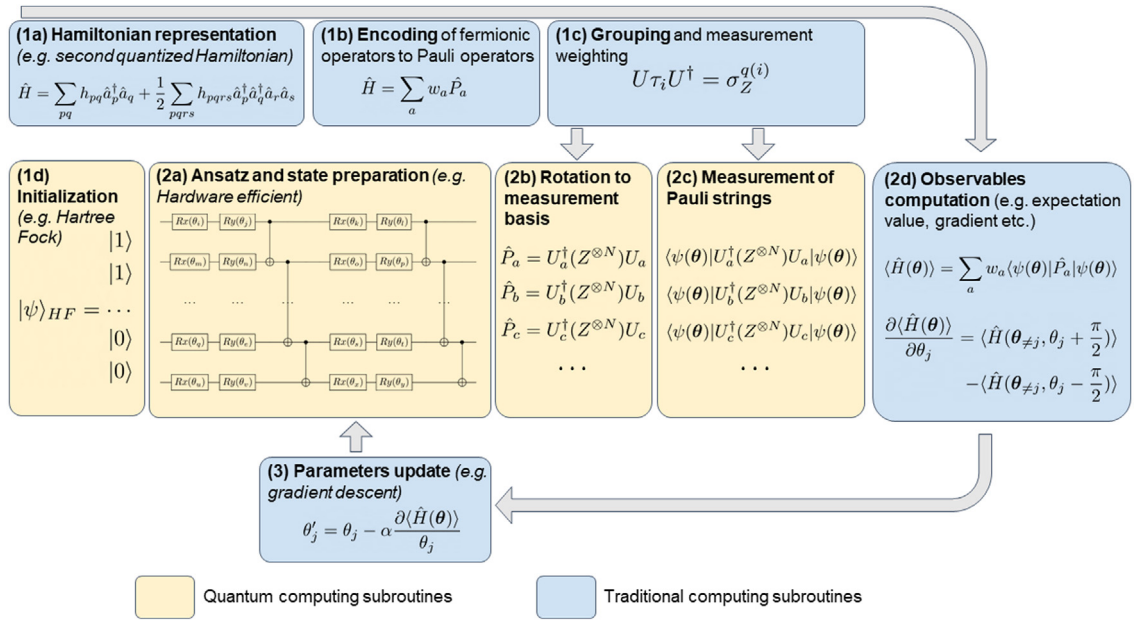


Fig. 1. The VQE Pipeline – Formulas are illustrative and do not necessarily correspond to best practices. **(1) Pre-processing:** (a) **Hamiltonian representation:** First pre-processing step of the VQE in which a set of basis functions is defined for the Hamiltonian to be expressed as a quantum observable of the electronic wave function (Section 3); (b) **Encoding:** Second pre-processing step of the VQE in which 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 (Section 4); (c) **Grouping and measurement strategy:** Third step in the pre-processing, where 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 diagonalized. It is also the step in which we decide the measurement weighting strategy (Section 5); (d) **State initialization:** In this last pre-processing step, we decide how the state to which the ansatz is applied is initialized. In general, the Hartree-Fock wavefunction is used [97–101], but other options are briefly discussed in Section 6.4 – **(2) The VQE loop:** (a) **Ansatz and trial state preparation:** The first step of the VQE loop is to apply the ansatz to the initialized qubit register, before the first iteration of the VQE all the parameters of the ansatz also need to be initialized (randomly or using a specific method, e.g. Refs. [90,102]) (Section 6); (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 (quantum computers in general measure in the Z basis), or a diagonal basis of a specific group of Pauli strings (c) **Observable computation:** The expectation value to be computed depends on the optimization strategy used, in any case however these are reconstituted by weighted summation on conventional hardware, or using a machine learning technique; (d) **Parameters update:** Based on the observables computed, and the optimization strategy, we can compute and apply updates to the ansatz parameters and begin a new iteration of the VQE (Section 7) – **(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 (Section 8).

Structure of the review: Section 2 presents a more formal definition of the VQE, its potential advantages and limitations are discussed, and our assessment of the state-of-the-art VQE is presented. The following sections provide reviews of each of the components of the VQE pipeline (as also summarized in Fig. 1). The first step in the VQE process is to choose a representation for the molecular Hamiltonian (Section 3), this is followed by a description of the methods used to map fermionic operators to spin operators, therefore allowing the Hamiltonian operators to be directly measured on a quantum device (Section 4). The next section covers efficiencies that can be used to reduce the number of measurements required to estimate the expectation value of a Hamiltonian (Section 5). We then turn to detailing the various quantum circuit structures representing wave function ansätze that have been proposed as the core component of the VQE (Section 6). This is followed by a presentation of the most important optimization methods and how they relate to the VQE (Section 7). Finally, we list and discuss the main error mitigation techniques that could improve the overall accuracy of the algorithm (Section 8). We close our review with a discussion of relevant extensions of the VQE for quantum chemistry applications (Section 9).

Additional resources and other reviews: Nielsen and Chuang [76] provide an introductory text detailing the fundamentals of quantum computing and quantum information which remains a seminal work on the topic. For an introductory overview of quantum computing and a pragmatic discussion of the current state of the technology, we recommend the review by Whitfield et al. [103]. For a more compact version of introductory topics in quantum computing, we point readers to the review by McArdle et al. [104]. This latter reference also provides a comprehensive review of quantum computing methods for quantum chemistry and material sciences. Two reviews on the same topic are provided by Bauer et al. [105] and of Motta and Rice [106]. For a review focused on general variational quantum algorithms we recommend the work of Cerezo et al. [107], and for the wider domain of NISQ algorithms, we recommend the review of Bharti et al. [108]. There

are also reviews dedicated to narrower topics relevant to the VQE, in particular, we recommend two reviews of ansätze, which complement our analysis: that by Fedorov et al. [109], and the review of unitary coupled cluster based-ansätze by Anand et al. [110]. Finally, for an overview of the fundamentals of quantum chemistry, we recommend the book of Szabo and Ostlund [81].

The VQE field is expanding rapidly, the current review only attempts to survey the literature to the end of May 2022.

2. Overview of VQE

In this section, we aim to provide sufficient information for our reader to acquire a technical understanding of the VQE, and an appreciation of where the algorithm is positioned compared to both conventional electronic structure, and other quantum methods. We also provide an outline for the suggested best practices, collected from the remainder of the review, and a perspective on the overall resources that could be required for the VQE to achieve quantum advantage.

2.1. A formal definition of the VQE

The VQE was first presented in Ref. [37] and its theoretical framework was significantly extended in Ref. [38]. It is grounded in the variational principle (and more precisely in the Rayleigh–Ritz functional [77–79]), which optimizes an upper bound for the lowest possible expectation value of an observable with respect to a trial wavefunction. Namely, providing a Hamiltonian $\hat{H}$, and a trial wavefunction $|\psi\rangle$, the ground state energy associated with this Hamiltonian, E_0 , is bounded by

$$E_0 \leq \frac{\langle \psi | \hat{H} | \psi \rangle}{\langle \psi | \psi \rangle}. \quad (1)$$

The objective of the VQE is therefore to find a parameterization of $|\psi\rangle$, such that the expectation value of the Hamiltonian is minimized. 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 minimization 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 parameterized unitary operations (see Section 6). We hence express $|\psi\rangle$ as the application of a generic parameterized unitary $U(\theta)$ to an initial state for N qubits, with θ denoting a set of parameters taking values in $(-\pi, \pi]$. The qubit register is generally initialized as $|0\rangle^{\otimes N}$, written as $|\mathbf{0}\rangle$ for simplicity, although low-depth operations can be performed for alternative initializations before the unitary is applied. Noting that $|\psi\rangle$ (as well as any $U(\theta)|\psi\rangle$) is necessarily a normalized wavefunction, we can now write the VQE optimization problem as

$$E_{\text{VQE}} = \min_{\theta} \langle \mathbf{0} | U^\dagger(\theta) \hat{H} U(\theta) | \mathbf{0} \rangle. \quad (2)$$

Eq. (2) is also referred to as the cost function of the VQE optimization problem, which is a terminology inherited from the machine learning and optimization 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 (Section 3 shows how the Hamiltonian can be defined, and Section 4 how it can be mapped to 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

$$\hat{H} = \sum_a^{\mathcal{P}} w_a \hat{P}_a, \quad (3)$$

with w_a a set of weights, and $\mathcal{P}$ the number of Pauli strings in the Hamiltonian. Eq. (2) becomes

$$E_{\text{VQE}} = \min_{\theta} \sum_a^{\mathcal{P}} w_a \langle \mathbf{0} | U^\dagger(\theta) \hat{P}_a U(\theta) | \mathbf{0} \rangle, \quad (4)$$

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

2.2. The VQE pipeline

The VQE, as presented using Eq. (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 summarize the iterative process (and the main VQE loop) in Fig. 1 to provide a schematic visual 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 [81,101,111,112], a solid-state system [42,113–115], a spin lattice model [116], 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 it 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 [81]. Additional complexity arises when specifically looking at electrons: following the Pauli exclusion principle [117,118] 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 quantization [81]. In second quantization 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. We provide an introduction to Hamiltonian construction and discuss the wider implications of particular representation choices in Section 3.
- 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 quantization, the operators can be directly translated into spin operators that can be measured on quantum computers [104], as they are not used to enforce antisymmetry of the wavefunction — we briefly outline how this is done in Section 3.1.3. In second quantization 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 [119], 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. We provide a list of the most relevant encodings for second quantized Hamiltonians in Section 4. 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 (Section 6). Cases of encoding particles others than fermions (e.g. bosons), and which do not require antisymmetry to be enforced are far simpler and are presented directly with the relevant Hamiltonian representation in Section 3.
- Measurement strategy and grouping:** The next step in the VQE pipeline is to determine how measurements are distributed and organized 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 [82,120,121]. This can be further optimized 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 [122–124] (see Section 5.3). 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 [125,126], with further details in Section 5.

- **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 parameterized quantum circuit, denoted as *ansatz*. It is used to produce the trial state, with which the Hamiltonian can be measured. Upon successful optimization 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 [127,128], defining the maximum accuracy its approximation of relevant low-energy states can achieve (assuming all parameters can be perfectly optimized). Its trainability describes the practical ability of the *ansatz* to be optimized using techniques tractable on quantum devices [87,127] (related to the total number of parameters, their linear dependence, the structure of the optimization surface, and to the related concept of barren plateaus [86], which can arise when gradients almost vanish thereby preventing optimization). 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 Section 6.
- **Parameter optimization:** 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 optimization is critical for at least three main reasons: (1) it directly impacts the number of measurements required to complete an optimization 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) [129,130]; (2) certain optimizers have been designed to alleviate specific optimization issues, such as the barren plateau problem [131–135]; (3) it directly impacts the number of iterations required to reach convergence (if it allows for convergence to be reached at all) [133–135]. We present the detailed description of the optimizers most relevant to the VQE in Section 7.
- **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 Section 8), 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) [107]. 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 Optimization Algorithms [136] or the Variational Quantum Linear Solver [137,138]). As such, the process of encoding the Hamiltonian is peculiar 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) [139,140]. 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 generalizing [141], 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 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) [142,143] 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 [144]. It was demonstrated on a tailored sampling task by several research teams [20–23], although the magnitude of the advantage has also been contested [145,146]. Quantum advantage is in general used interchangeably with quantum supremacy, however we propose to 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 initialization 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. [147]).

The VQE allows for the probabilistic measurement of observables over certain classes of parameterized 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. We provide a more detailed description of these specific classes of wavefunctions (the ansatz of the VQE) in Section 6, while a comparison to alternative wavefunction classes, which can admit polynomial complexity on conventional computing resources, are described in Section 2.4. 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 Section 2.6, that under certain assumptions this condition is theoretically achievable [37,80,82,148–150].

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 review, 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 [101], 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 [151,152]): 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. [83] 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 [82,83,85]. 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 Section 2.7). The field of research is fast-moving however, and much research has been devoted to efficient operator sampling (summarized in Section 5). Using the parallelization potential of VQE (see Section 2.7.2) 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 optimization problem. As such, to understand the true cost of implementing VQE, one needs to assess the complexity of the optimization process. The true cost and scaling are dependent on the optimizer and on the optimization landscape of the specific problem studied. While some optimizers have been shown to converge in polynomial time for convex cost functions, the VQE is far from having such a favorable landscape [93,153]. In fact, the VQE optimization is shown to be NP-hard [93], 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 optimization problems can suffer from the same issue [154]. The key open question is to know whether VQE can be optimized heuristically in a polynomial number of iterations and converge to an approximate yet accurate enough solution.
- 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 parameterizations of systems, due to the barren plateau problem [86] (discussed in detail in Section 6.1). A number of mitigating methods have been proposed, such as the identity block initialization [102] or the use of a local encoding for the Hamiltonian [87,91]. 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 [155–160]), it can significantly increase the resource requirements of the algorithm (see Section 8). 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 [96] 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 optimization 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 [94]. 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 parameterization. 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) [161]. 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. [101]) or density functional theory (DFT) calculation [162–165] (for comprehensive reviews of DFT, see Refs. [166,167]). These orbitals are used to define the representation of the Hamiltonian (see Section 3), 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 [168]. 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 [169]. 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. [170]), 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 [81,171]. This results in the approach being in most cases intractable, with practical limits for *ab initio* systems currently being 18 orbitals [172]. 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 optimized single Slater determinant, as appropriate for closed-shell systems [101], 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 diagonalize 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

$$|\psi\rangle_{\text{FCI}} = c_0|\psi\rangle_{\text{HF}} + \sum_{ia} c_{ia}\hat{a}_a^\dagger\hat{a}_i|\psi\rangle_{\text{HF}} + \sum_{ijab} c_{ij,ab}\hat{a}_a^\dagger\hat{a}_b^\dagger\hat{a}_j\hat{a}_i|\psi\rangle_{\text{HF}} + \dots, \quad (5)$$

where the first sum represents ‘singly-excited’ configurations where an occupied spin-orbital, denoted by the indices $i, j, \dots$ is depopulated, and a virtual spin-orbital, denoted by the indices $a, b, \dots$ 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 quantized operators, $\hat{a}^{(\dagger)}$, with more details of these found in Section 3.1.3. 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 parameterization, ensures that the minimization of the Ritz functional (Eq. (1)) can be written as a diagonalization of the full Hamiltonian in this basis [173,174]. 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 minimizing the loss in accuracy resulting from this truncation [81]. 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. [168]). To this end, a large number of approximate parameterizations of the FCI wavefunction have been explored, which differ in their accuracy, scaling, functional form, and method of optimization. Many of these approaches have enabled chemical accuracy and beyond compared the FCI result to be routinely reached in systems far larger than those accessible by FCI [175–180]. 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) [181] 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. [83] for a thorough discussion of chemical accuracy vs. precision in the context of computational chemistry).

The considerations described in devising an effective parameterization 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 parameterizations 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 parameterizations for conventional computation

While the complexity of the exact FCI ansatz (Eq. (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 [168]. 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. [169], with applications to transition metal systems, again showing excellent agreement between the most accurate wavefunction methodologies in these systems.

To rationalize some of these parameterizations, an obvious first approximation to Eq. (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 [81]. More recent adaptive, selective or stochastic inclusion of desired configurations exploit the sparsity in the optimized amplitudes, and can extend the ansatz further in accuracy, resulting in methods such as Adaptive Sampling CI (ASCI) [177,182–184], Semistochastic Heat-Bath CI (SHCI) [88,180,185–189], or Full Configuration Interaction Quantum Monte Carlo [178,190–193]. 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 [81]. 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 optimized within the density matrix renormalization group (DMRG), and can also yield accurate and systematically improvable approximations to FCI both in the case of *ab initio* molecular Hamiltonian and for lattice models [175,176,194–203]. More broadly, the development of tensor network theory [204,205] and the use of matrix product states has resulted in significant improvements of methods for the resolution of lattice models (for instance [206,207]). Finally, it is worth mentioning that a larger class of approximate wavefunction ansatz are able to be optimized 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 [208–210]. Within this criterion, the parameters of the ansatz can be optimized via Monte Carlo

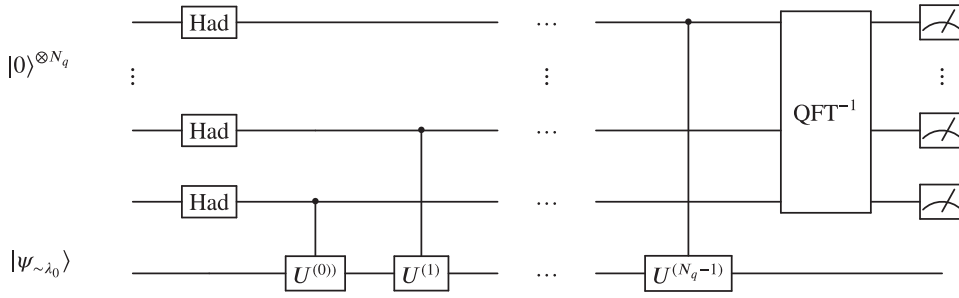


Fig. 2. Quantum circuit for Quantum Phase Estimation.

integration of the energy functional in a very general framework, albeit with the necessity of controlling for stochastic errors and care in optimization 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. (5). This results in an appropriately sized consistent method, with an excellent accuracy vs. cost balance [211–213]. 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 [214], while other approximate coupled-cluster forms suitable for stronger correlation effects have also been developed (for recent examples, see Refs. [215,216]). The coupled cluster is also the motivation of the use of the *unitary* coupled cluster ansatz of VQE [161], where a similar structure of exponentiated excitations based around a reference configuration is constructed (see Section 6.2.2), 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 [217] (Section 6.3.1). Furthermore, the Efficient Symmetry Preserving ansatz [218] looks to ensure the ability to systematically improve its span of the FCI description of Eq. (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 parameterizations 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. This analysis of the errors in different systems from 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 [161,219].

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 achieve chemical accuracy but with 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 parameterizations suitable for conventional devices, such as the rapidly emerging field of machine-learning inspired ansätze [210,220–226], 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) [227–231] 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 [232]. It does so however using quantum circuits of depths that are far beyond what is achievable in the NISQ era of quantum computing [233]. 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, and the process can be described as follows (adapting the descriptions in Refs. [76,104]):

- 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) 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:

$$e^{i\hat{H}}|\lambda_j\rangle = e^{iE_j}|\lambda_j\rangle. \quad (6)$$

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

$$|\psi_0\rangle = \sum_{j=0}^{2^{N_q}-1} \alpha_j |\lambda_j\rangle, \quad (7)$$

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 initialized 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 (for readers requiring a brief introduction to key quantum computing concepts and operations, we recommend Section 2.A of Ref. [104]). 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

$$|\psi_{\text{anc}}\rangle = \frac{1}{(\sqrt{2})^{N_a}} \sum_{x=0}^{2^{N_a}-1} |\text{bin}(x)\rangle, \quad (8)$$

where x represents integers from 0 to $2^{N_a} - 1$ and $\text{bin}(x)$ the binary representation 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

$$|\psi_{\text{tot}}\rangle = \left(\frac{1}{(\sqrt{2})^{N_a}} \sum_{x=0}^{2^{N_a}-1} |\text{bin}(x)\rangle \right) \otimes \left(\sum_{j=0}^{2^{N_q}-1} \alpha_j |\lambda_j\rangle \right) = \frac{1}{(\sqrt{2})^{N_a}} \sum_{j=0}^{2^{N_q}-1} \sum_{x=0}^{2^{N_a}-1} \alpha_j |\text{bin}(x)\rangle \otimes |\lambda_j\rangle. \quad (9)$$

- 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

$$U^{(k)} = \left(e^{i\hat{H}} \right)^{2^k}, \quad (10)$$

with k an arbitrary number for the time being. Following Eq. (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. (10) it is applied $5 = 1 \cdot 2^2 + 1 \cdot 2^0$ times). We obtain the state

$$|\psi_{\text{tot}}\rangle = \frac{1}{(\sqrt{2})^{N_a}} \sum_{j=0}^{2^{N_q}-1} \sum_{x=0}^{2^{N_a}-1} e^{2\pi i\theta_j x} \alpha_j |\text{bin}(x)\rangle |\lambda_j\rangle. \quad (11)$$

- The next step reduces the number of superposition instances 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)

$$\text{QFT}|\text{bin}(y)\rangle = \sum_{x=0}^{2^{N_a}-1} e^{2\pi i(xy/2^{N_a})} |\text{bin}(x)\rangle. \quad (12)$$

If we set $y = 2^{N_a} \theta_i$, we can observe that applying the inverse QFT to the ancilla register in Eq. (11) results in

$$(\text{QFT}^{-1} \otimes I^{\otimes N_q}) |\psi_{\text{tot}}\rangle = \sum_{j=0}^{2^{N_q}-1} \alpha_j |\text{bin}(2^{N_a} \theta_j)\rangle |\lambda_j\rangle, \quad (13)$$

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 $\text{bin}(2^{N_a} \theta_i)$ 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 $|\text{bin}(2^{N_a} \theta_i)\rangle |\lambda_j\rangle$.

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 we recommend Refs. [47,234]). Provided a non-restricted pool of qubits, Babbush et al. [235] show that by engineering the correct Hamiltonian representation (namely, a plane wave basis in first quantization, see Section 3), one can achieve sub-linear scaling in the number of basis elements for Hamiltonian simulation.

As mentioned above, QPE is only as likely to succeed as the fidelity of the input state to the unknown target eigenstate approaches one. In turn, this implies that using a randomized 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 [232]. 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 [236–240]) or in adiabatic quantum computation (for example [231,241,242]).

There have been a number of successful implementations of QPE on quantum devices [149,243–248]. 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 [249]. Reiher et al. [233] 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 qubitization (a method which aims at transforming the evolution operator into a quantum walk) [250] to reduce the gate requirements to $\mathcal{O}(10^{11})$ Toffoli gates, despite an extended active space [251]. Lee et al. [252] 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 [253,254]), 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. [83] 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 h. 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. [83]).

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. [255], 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. [255] 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.

2.6. Our 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 [42,113,115,116] and *ab initio* molecular systems [81,101,111,112]. These two categories usually require different encodings, measurement strategies, and ansätze. Table 1 summarizes 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 parameterized 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.

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 (see Section 3). 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 quantization (antisymmetry maintained by the wavefunction) or second quantization (antisymmetry maintained by the operators, see Section 3). The number of qubits in first quantization scales as $\mathcal{O}(m \log_2(n))$ [228,278], with m the number of electrons and n the number of basis functions, against $\mathcal{O}(n)$ in second quantization [119]. The former also requires additional depth to enforce the antisymmetry of the wavefunction. There has not been, to the best of our knowledge, a rigorous study of the efficiency of using first quantization in VQE. While the scaling for first quantization could be advantageous in systems with few electrons and large basis sets (e.g. if a plane wave basis is used [235]), second quantization is generally preferred.

Encoding: The next decision to take is that of the mapping used to transform the fermionic, second quantized Hamiltonian into a weighted sum of Pauli operators (see Section 4). The most relevant encodings for *ab initio* molecular system include Jordan–Wigner [119], the Parity [279], Bravyi–Kitaev [279–281], and ternary tree mappings [256] (all are explained in detail in Section 4.1). Out of these, the most promising encoding is the ternary tree mapping [256], 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 [87,91]. It is however still unclear whether this lower circuit depth does indeed result in more noise resilience, as pointed out in Ref. [259]. We also recommend that the resulting qubit Hamiltonian is further reduced using tapering off methods based on symmetries [282–284] (see Section 4.3).

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 [82,83,85]. Deciding on an efficient strategy for grouping and measuring Hamiltonian terms can go a long way in reducing this bottleneck. The decomposed interactions methods [124, 262] provide on balance the most promising means to jointly measure the Hamiltonian. They allow measuring an entire molecular Hamiltonian with $\mathcal{O}(N)$ groups (for more details about this method, see Section 5.2.4) 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. [123,285]) require fewer shots to achieve a given accuracy [262] in some numerical studies (in particular when using the Sorted Insertion heuristic [286]), this reduction will likely not be worth the additional circuit depth scaling $\mathcal{O}(N^2)$ [122] required to perform the joint measurements. It is also worth noting that among the decomposed interactions methods, the Variance-estimate Greedy Full Rank Optimization [262] appears to perform best, although it requires minimization search of decomposition parameters. While this cost could be tractable there has been, to the best of our knowledge, no thorough research on how it would scale on large systems. For this reason, the Basis Rotation Group methods [124], 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 [82,120,121]. 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 [126] 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 [268]. Further numerical studies will be required to establish the true performance of classical shadows compared to grouping methods.

Table 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		<i>ab initio</i> Molecular systems	Lattice models
Hamiltonian construction (Section 3)	Method	Second quantization	Second quantization
	Scaling	$\mathcal{O}(n^4)$ Hamiltonian terms and $N = \mathcal{O}(n)$ qubits, with n number of basis functions	<i>idem</i>
	Comments	First quantization could be advantageous on some systems, but further research is needed	<i>idem</i>
Fermion to spin encoding (Section 4)	Method	Ternary tree encoding [256]	Generalized superfast encoding [257]
	Scaling	Number of operators: $\mathcal{O}(N^4)$, Pauli weight $\mathcal{O}(\log_3(2N))$,	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
	Comments	Low weight encodings could result in more resilience to barren plateau [87,91,258], and more compact ansätze, though there is at least some suggestions that it may increase impact of quantum noise [259]	<i>idem</i> , the Compact encoding [260,261] results in a lower number of qubits for $D = 2, 3$ but has not yet been generalized to higher dimensions.
Grouping and measurement strategy (Section 5)	Method	Decomposed interactions [124,262]	Qubit-wise commutation [38,120,122,123,157,263–267]
	Scaling	Operators to measure reduced to $\mathcal{O}(N)$; additional basis rotation circuit depth $\mathcal{O}(N/2)$	Operators to measure reduced by a scalar, Additional basis rotation circuit depth $\mathcal{O}(1)$
	Comments	Full rank optimization (in particular its extensions) [262] seem to achieve better overall measurement reduction for a given precision ϵ than the basis rotation method [124], but cost scaling remains unclear. Classical shadows [126] have been shown in at least one case [268] to outperform the scaling of decomposed interactions, though further numerical studies will be required to demonstrate dominance.	QWC grouping benefits from low Pauli weight encoding and comes at virtually no cost
Ansatz (Section 6)	Method	k-UpCCGSD [80]	Hamiltonian variational ansatz (HVA) [82,150]
	Scaling	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)
	Comments	Promising scaling, and good accuracy [80,219] but uncertainty remains for applications on large highly correlated systems. Uncertainty around k , the number of repetitions required. Adaptive ansätze [217,269,270] may perform better, but their scaling requires more investigation.	HVA has shown resilience to barren plateau and efficacy on lattice models [150]. k is the number of repetition of the ansatz required to reach the desired accuracy.
Optimizer (Section 7)	Method	Rotosolve [133,134,271] or Fraxis [272,273]	Rotosolve [133,134,271] or Fraxis [272,273]
	Scaling	Requires sampling three values for each parameter at each step	<i>idem</i>
	Comments	Some indication of faster convergence [134], but does not allow for full potential for parallelization of VQE, and requires more values to sample than most optimizers	<i>idem</i>
Error mitigation strategy (Section 8)	Method	Symmetry verification [274,275] and extrapolation based methods [276,277]	<i>idem</i>
	Scaling	Exponential with respect to the circuit depth	<i>idem</i>
	Comments	The recommendation reflects the <i>de facto</i> method of choice for experiments. A fair comparison of the performance and cost between different error mitigation methods requires further research.	<i>idem</i>

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 [219], is the Unitary paired Generalized Coupled Cluster Singles and Doubles (UpCCGSD) ansatz [80] (see Section 6 for a detailed description). 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 [217], iterative Qubit Coupled Cluster [287] and Cluster VQE [288]) are also promising, as they may provide resilience against barren plateaus. Their main drawback is that these adaptive methods come at the cost of selecting an operator to grow the ansatz (or Hamiltonian) and the need to fully re-optimize the ansatz at each iteration. Numerical studies have suggested that additional measurements may be required compared to fixed structure ansätze [289], although further research is required to provide an exhaustive costs and benefits analysis.

Optimizer: It is challenging to systematically compare different optimizers since no thorough large scale studies of their convergence rate have been conducted. For the time being, the Rotosolve optimizer [133,134,271] (see Section 7.4.2) has been shown to converge significantly faster than several gradient based optimizers [134]. It offers the advantage of not relying on any meta-parameters (such as a learning rate), which makes it a very easy optimizer 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 optimization of the previous parameter, but this could result in correlated noise); secondly, parameters must be updated sequentially, thereby restricting the scope for parallelization 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 optimizers that have been shown to have superior convergence rates, and given we do not expect that there will be a sufficient number of quantum computers to fully exploit the parallelization potential of the VQE in the NISQ era (see Section 2.7.2), we propose the Rotosolve/Fraxis optimizers over other alternatives (although it is worth noting that the Quantum Natural Gradient [135] has been shown to perform well and to benefit from resilience to barren plateaus [290,291], albeit at a significant cost [135]).

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). 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 $1.6 \text{ mE}_\text{H} \approx 10^{-3} \text{ E}_\text{H}$. 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 Section 6.1). In this situation, estimating gradients may require a more precise ϵ and more measurements, but it also means that optimization may rapidly become impossible. If the k-UpCCGSD ansatz is chosen, this scales as $\mathcal{O}(kN)$, while choosing to use the decomposed interactions of Section 5.2.4 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 optimizer. As this optimizer 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 parallelization, and $\mathcal{O}(kN)$ with full parallelization (the circuit depth). This perfect parallelization 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 Section 2.7.2). 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 optimization.

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 optimizer used, and the system studied. One important point to note is that convergence tends to be rapid at the beginning of the optimization 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 optimization. As such the last few iterations of the VQE are likely orders of magnitude more expensive than the rest of the optimization, 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 $\mathcal{O}(n^4)$ [292], 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 diagonalize operators and reduce measurements requires rewriting the Hamiltonian in a different basis [124,262]. 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 [38,124,262,286] 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 (see Section 4.2 for further details). These mappings are in general designed to minimize 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) [293], 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 Generalized Superfast Encoding method [257] 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 [260,261] requires a lower number of qubits ($\sim 1.5n$), but has not yet been generalized for regular lattices of more than three dimensions. Beyond their relationship to the resilience to barren plateaus in the optimization [87,91,258], 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–Kitaev 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 in Section 4.3 [282,283].

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), as detailed in Section 4.2. Because it is in general possible to reach low Pauli-weight encodings for lattice models (see Section 4.2), qubit-wise commutativity (QWC) grouping [38,120,122,123,157,263–267] could possibly offer significant potential for reductions in the number of terms. 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 generalized encodings such as Jordan–Wigner (see for instance, Ref. [294]). However, we note 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 [161]. 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) [82,150] (see Section 6.2.4). 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 [257,260,261,295]. HVA was also shown in [150] to be particularly resilient to the appearance of barren plateaus in the optimization 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 [235] could be very efficient on certain models (in particular jellium in Ref. [235], which is a continuous model, but with the Hamiltonian defined by a single parameter). Symmetry breaking HVAs [296,297] are also a promising avenue, and though numerical tests from Ref. [297] show excellent fidelities of the state produced, results in Ref. [296] show some instabilities of the ansatz. Overall, until further research is conducted, we consider 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.

Optimizer: Similar to our proposal for *ab initio* molecular systems presented earlier, we find that Rotosolve [133,134,271] and Fraxis [272,273] currently have the best supporting evidence among the optimizers available.

Scaling: Based on the above, we 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 Refs. [298–301]). 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 we consider this an area of open research. Therefore, we maintain the previous scaling and estimate that a single energy evaluation would scale $\mathcal{O}(k(nD/\epsilon)^2)$.

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 $\mathcal{O}((knD)^3/(\epsilon)^2)$. As noted previously, if the full parallelization 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. It is worth noting that there has been so far no evidence that qubit-wise commutativity grouping can reduce the overall scaling of the number of operators to measure, and therefore we have not included it in this scaling analysis. Similarly, while the depth of HVA scales with the number of commutative groups in the Hamiltonian, we have not yet found an argument on whether this would reduce scaling below that of the number of terms $\mathcal{O}(nD)$.

2.7. Resource estimate for VQE

2.7.1. 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. [82] develop the Hamiltonian Variational Ansatz (HVA), and present a numerical study of the accuracy of certain 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 characterized in Ref. [86], 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. [302]). The literature has progressed significantly since this 2015 study, and there are now more efficient ansätze (see Section 6) and grouping methods (see Section 5) that may change this conclusion.

Kühn et al. [84] 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 run time for a VQE implementation would be impractical, even using full parallelization potential, without unfortunately providing more details about how this conclusion is reached.

Gonthier et al. [85] 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 ($\text{C}_2\text{H}_6\text{O}$), 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 (see Section 5), and different optimization methods (see Section 5.1).

It is worth noting that the studies mentioned above do not take into account the three obstacles we listed at the beginning of this section (namely, the complexity of the optimization, the barren plateau problem, and the impact of quantum noise). At the same time, Refs. [83–85] do not discuss the potential for parallelization (with the exception of [83] which touches upon it briefly). For instance, the runtime estimates of 1.9 to 71 days presented in [85] can be parallelized 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 Section 2.7.2 for a discussion).

We provide our estimated runtimes for the steps in the VQE for a representative example system in Table 2, including the general scaling estimate for such types of systems. The example system considered corresponds to the one proposed in Ref. [83], 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 review. To estimate the depth we compile a 52 qubit version

Table 2

Indicative estimates of the run time 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.

Operation	Scaling	Formula	Runtime
Single shot (using k-UpCCG SD, $k = 1$)	$\mathcal{O}(kN)$	$L \times T$	3 ms
One expectation at $\epsilon = 10^{-3}$ mE _H (using decomposed interactions methods)	$\mathcal{O}(\frac{kN^2}{\epsilon^2})$	$L \times T \times \mathcal{P} \times S$	25 days
Full iteration using Rotosolve	$\mathcal{O}(\frac{k^2 N^4}{\epsilon^2})$	$L \times T \times \mathcal{P} \times S \times 3p$	35 years
Full iteration using gradient based method	$\mathcal{O}(\frac{k^2 N^4}{\epsilon^2})$	$L \times T \times \mathcal{P} \times S \times 2p$	24 years

Table 3

Indicative estimates of the run time and number of quantum computers used for one iteration of the VQE assuming perfect parallelization of the method can be achieved and neglecting any communication overheads – using the same assumptions stated in Table 2.

Operation	Time scaling	Runtime	Scaling sets of qubits	Sets of qubits (here, $N = 52$)
Single shot (using k-UpCCGSD, $k = 1$)	$\mathcal{O}(kN)$	3 ms	$\mathcal{O}(1)$	1
One expectation at $\epsilon = 10^{-3}$ mE _H (using decomposed interactions methods)	$\mathcal{O}(kN)$	3 ms	$\mathcal{O}(\frac{N}{\epsilon^2})$	$\sim 800 \times 10^6$
Full iteration using Rotosolve	$\mathcal{O}(k^2 N^3)$	0.5 s	$\mathcal{O}(\frac{N}{\epsilon^2})$	$\sim 2500 \times 10^6$
Full iteration using gradient based method	$\mathcal{O}(kN)$	3 ms	$\mathcal{O}(\frac{kN^3}{\epsilon^2})$	$\sim 280 \times 10^9$

of k-UpCCGD assuming $k = 1$ and full connectivity of the qubit register. We find 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 find that the depth required is increased by more than tenfold). Note that choosing $k = 1$ is likely insufficient [80,219]. Gate time is assumed to be $T = 100$ ns (similar to what is presented in [85], which itself refers to Table 1 in [303]), 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 generalized 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 in 10^{-3} mE_H, which is close to chemical precision (1.6 mE_H) [181] 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 of the optimization, 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 parallelization.

2.7.2. Parallelization potential of the VQE

The potential for parallelization of the VQE was already identified in the initial paper by Peruzzo et al. [37] and subsequently mentioned in many VQE papers, although an in-depth study is lacking. Parallelism is however critical for the viability 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 2 assuming that perfect parallelization is possible, and present the results in Table 3.

It is clear that parallelization will be a critical part of any future success of the VQE method. Broad availability of quantum computers with increasing number of qubits could therefore significantly speed-up the VQE process, however there are significant caveats to that. One key observation is that full parallelization would require a number of quantum computers cores (or threads) that scales $\mathcal{O}(pN^4/\epsilon^2)$, 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 parallelization 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 synchronization cost, data aggregation and communication (possibly latency if the different sets of qubits are connected through the cloud). Second, parallelization could result in higher noise levels. We note two possible sources of additional noise: (1) if parallelization 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 [38,94] – 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.

2.7.3. Distribution of resources between quantum and conventional computation

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, and performing the parameter updates on a conventional device. This approach however ignores the excellent scaling of the simulation of certain families of quantum states on conventional computers (for example, see Ref. [48]).

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. [304] propose to partially solve a many-body Hamiltonian on a conventional computer before performing a quantum based VQE. Okada et al. [305] alternatively show how one can perform a classical optimization of the VQE for local-interaction states, leaving quantum sampling only for measurement of global quantities. Methods of Hamiltonian dressing such as ClusterVQE [306] (presented in Section 6.3.2) 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.

3. Hamiltonian representation

Before studying any methods relating directly to implementing the VQE, we must turn towards the definition of the problem itself. As written before, our aim is to find the lowest eigenvalue of a Hermitian matrix, or stated in quantum chemistry terms, we are trying to find an approximation of the ground state energy of an interacting Hamiltonian. In this section, we provide details about how Hamiltonian can be constructed, and how choices in certain freedoms in expressing this Hamiltonian may impact the remainder of the VQE pipeline. Many different forms of Hamiltonians exist in physics and chemistry; we begin with a presentation of the electronic structure Hamiltonian and construction of the eigenvalue problem. This is followed by examples of other applications, namely lattice models, vibrational spectroscopy, and periodic adaption of electronic structure problems. Unlike the other models, the electronic structure Hamiltonian is in general known and the choices related to its construction are aiming at reducing the complexity of the eigenvalue problem. In the case of lattice models, and vibrational spectroscopy Hamiltonian assumptions are made to define the Hamiltonian.

3.1. The electronic structure hamiltonian

3.1.1. The *ab initio* molecular hamiltonian

This represents the operator for the total energy of an arbitrary molecular system defined in terms of its atomic composition, and the relative positions of the nuclei. From this geometrical definition, also referred to as a conformation, one must determine the correlated probability amplitudes of the electrons in the space surrounding the nuclei, i.e. the electronic wavefunction, which has the lowest energy: the ground state energy.

In a non-relativistic settings and following the Born–Oppenheimer approximation (which assumes that the motion of the nuclei can be neglected, as they are much heavier than the electrons) the electronic Hamiltonian depends parametrically on the nuclear positions $\mathbf{R}_k$. Up to a constant (given by the nuclear–nuclear repulsion energy) the electronic Hamiltonian can be written as

$$\hat{H} = \hat{T}_e + \hat{V}_{ne} + \hat{V}_{ee}, \quad (14)$$

where

$$\hat{T}_e = - \sum_i \frac{\hbar^2}{2M_i} \nabla_i^2, \quad (15)$$

$$\hat{V}_{ne} = - \sum_{i,k} \frac{e^2}{4\pi\epsilon_0} \frac{Z_k}{|\mathbf{r}_i - \mathbf{R}_k|}, \quad (16)$$

$$\hat{V}_{ee} = \frac{1}{2} \sum_{i \neq j} \frac{e^2}{4\pi\epsilon_0} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}, \quad (17)$$

with, $\mathbf{r}_i$ is the position of electron i , M_i its mass, Z_k is the atomic number of nucleus k , e is the elementary charge, $\hbar$ is the reduced Planck constant, and ∇_i^2 is the Laplace operator for electron i .

3.1.2. Construction of the wavefunction

One needs to define a basis in which to represent the electronic wavefunction. A number of possible types of basis functions exist. Given the number of qubits required scales as a function of the number of basis functions (see Section 4), choosing a basis that is compact, but yet provides an accurate description of the system studied, is critical for an efficient implementation of VQE. Basis elements, functions or orbitals describe the probability distribution of a single electron. For *ab initio* systems, these have primarily been built from parameterized atom-centered Gaussian ('atomic') orbitals, with the majority of research to date in VQE using minimal sized Slater Type Orbitals (STO) (for example [37,80,263]). These basis functions are defined as a weighted sum of Gaussian functions to provide approximately the right radial distribution, long-range behavior, and nuclear cusp conditions for each atom [307,308]. As an illustration, the radial component of the minimal STO-3G basis for each atom is constructed from three Gaussians such that an atomic orbital is given by

$$\chi(r) = c_1\gamma_1(r) + c_2\gamma_2(r) + c_3\gamma_3(r) \quad (18)$$

where γ_i are Gaussian functions, r the distance of the electron from the nucleus, and c_i are fitted weight parameters. These basis sets are often called minimal basis sets as they only include orbitals necessary to represent the valence shell of an atom. For an appropriate treatment of the correlation, it is essential in real systems to enlarge the basis set to include higher energy atomic orbitals, allowing for additional polarization and diffuse functions, and higher angular momentum functions which are required to build in flexibility to describe the correlated positions of the electrons [53]. An example of these larger basis sets commonly used for correlated calculations are the correlation-consistent polarized Valence n-Zeta (cc-pVnZ) basis sets [309], which allow for a systematic (and extrapolatable) expansion in terms of the cardinal number of the basis set [310]. These basis sets have been used in some VQE research (for example [84,159,311]), but their additional size can limit the size of the systems treated. Alternative basis sets have been considered in the context of VQE. For example, plane wave basis has been used as a mean to construct a compact ansatz for certain models which naturally exploit the translational symmetry of certain models [235]. Alternatively, a grid of points in real space make a natural representation to enforce locality and enable a sparse representation of the Hamiltonian. However grid representations of these basis functions generally require a significantly larger qubit count, and therefore their use is limited in NISQ applications and for the VQE [104,312–318].

These non-orthogonal atomic orbitals are generally linearly combined into 'molecular' orbitals before use, which constitutes an orthonormal set of delocalized basis functions which can no longer be assigned to a particular atomic site. The Hamiltonian is then expressed within this molecular basis by way of a transformation of the matrix elements. Overwhelmingly, this transformation is obtained via a mean-field (generally Hartree–Fock) calculation, which produces this rotation to the molecular orbital basis, and additionally provides an energy measure for each single-particle molecular orbital (for a description of this, see Ref. [101]). However, given the constraints on qubit numbers and therefore the size of the single particle basis, there is also research in the use of further contractions of molecular orbitals to a more suitable and compact basis for subsequent correlated calculations. These are often based on approximate correlated treatments in order to truncate to frozen natural orbitals, which has been favorably suggested by Verma et al. [319], Mochizuki et al. [320] and showed to potentially resulting in computation cost reductions for VQE by Gonthier et al. [85]. Furthermore, self-consistent active space approaches also optimize the set of molecular orbitals within the correlated treatment, to optimally span this correlated physics, and are considered further in Section 9.2.

Once the single-particle basis functions have been selected, the many-body basis for the electronic wavefunction is constructed from products of these functions. For a non-interacting Hamiltonian, the solution is given as a single many-body basis function with optimized orbitals, which is the principle behind the Hartree–Fock and other mean-field methods. In addition, following the Pauli exclusion principle, the electronic wavefunction must be antisymmetric, meaning that the exchange of any two electrons changes the sign of the wavefunction. To account for this, these many-body basis functions can be formally written as Slater determinants, which for a wavefunction of n occupied orbitals can be formally written as

$$\psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = \frac{1}{\sqrt{n!}} \begin{vmatrix} \phi_1(\mathbf{x}_1) & \phi_2(\mathbf{x}_1) & \cdots & \phi_n(\mathbf{x}_1) \\ \phi_1(\mathbf{x}_2) & \phi_2(\mathbf{x}_2) & \cdots & \phi_n(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(\mathbf{x}_n) & \phi_2(\mathbf{x}_n) & \cdots & \phi_n(\mathbf{x}_n) \end{vmatrix}, \quad (19)$$

where $\phi_j(\mathbf{x}_j)$ denotes a spin-orbital of the chosen basis, with the variable $\mathbf{x}_j = (\mathbf{r}_j, \sigma_j)$ subsuming both the associated spatial and spin indices. As a shorthand for this, we can write an n -electron many-body basis function as

$$|\psi\rangle = |\phi_1\phi_2 \dots \phi_n\rangle, \quad (20)$$

where $\phi_j = 1$ means that the j th basis function $\phi_j(\mathbf{x})$ is occupied, and $\phi_j = 0$ is unoccupied. This representation can be simply encoded in the qubit register, with an occupied spin-orbital denoted by an up-spin, and an unoccupied orbital by a down spin. This defines a many-body Hilbert space in which the correlated wavefunction can be expanded, since the correlations will ensure that the state can no longer be written as a single Slater determinant, but rather a linear combination over the space. This Hilbert space of electron configurations has a well-defined inner product between these many-body basis states, as $\langle\psi_a|\psi_b\rangle = \delta_{a,b}$.

There are commonly two ways that antisymmetry condition can be met in practice: one can do so by enforcing it in the construction of the wavefunction, where the Schrodinger equation is typically written in real space and is generally referred to as first quantization; alternatively, one can enforce antisymmetry through enforcing antisymmetry within the commutation relations of the operators for expectation values, and is traditionally referred to as second quantization. The choice of representation has significant implications for the resources required to implement VQE, and is the topic we turn to in Section 4.

3.1.3. Hamiltonian quantization

First quantization: First quantization is most commonly used in quantum mechanics and is a direct adaption of classical mechanics to quantum theory by quantizing variables such as energy and momentum. It was the method of used in early quantum computing research of Hamiltonian simulation [228,312–315,321,322], and has seen a resurgence in recent years due to promising scaling [170,323]. Following the definition of the wavefunction in terms of Slater determinant Eq. (20), one must ensure that the antisymmetry of the wavefunction is verified:

$$|\sigma(\phi_1\phi_2\ldots\phi_n)\rangle = (-1)^{\pi(\sigma)}|\phi_1\phi_2\ldots\phi_n\rangle, \quad (21)$$

where $\pi(\sigma)$ is the parity of a given permutation σ on a set of basis functions.

As the antisymmetry is addressed in the wavefunction, the Hamiltonian can be constructed by simple projection onto the single particle basis function. If we project the Hamiltonian over the space spanned by $\{\phi_i(\mathbf{x}_i)\}$ which we assume to be orthonormal, we obtain (using the Slater–Condon rules [324,325]) the one and two body integrals. For the one-electron terms matrix elements h_{pq} :

$$\begin{aligned} h_{pq} &= \langle\phi_p|\hat{T}_e + \hat{V}_{ne}|\phi_q\rangle \\ &= \int d\mathbf{x} \phi_p^*(\mathbf{x}) \left(-\frac{\hbar^2}{2m_e} \nabla^2 - \sum_k \frac{e^2}{4\pi\epsilon_0} \frac{Z_k}{|\mathbf{r} - \mathbf{R}_k|} \right) \phi_q(\mathbf{x}) \end{aligned} \quad (22)$$

And for the two-electron interaction terms we obtain the matrix elements h_{pqrs} :

$$\begin{aligned} h_{pqrs} &= \langle\phi_p\phi_q|\hat{V}_{ee}|\phi_r\phi_s\rangle \\ &= \frac{e^2}{4\pi\epsilon_0} \int d\mathbf{x}_1 d\mathbf{x}_2 \frac{\phi_p^*(\mathbf{x}_1) \phi_q^*(\mathbf{x}_2) \phi_r(\mathbf{x}_2) \phi_s(\mathbf{x}_1)}{|\mathbf{r}_1 - \mathbf{r}_2|}. \end{aligned} \quad (23)$$

The first quantized form of the Hamiltonian can therefore be written directly in the single particle basis:

$$\hat{H} = \sum_{i=1}^m \sum_{p,q=1}^n h_{pq} |\phi_p^{(i)}\rangle \langle\phi_q^{(i)}| + \frac{1}{2} \sum_{i \neq j}^m \sum_{p,q,r,s=1}^n h_{pqrs} |\phi_p^{(i)}\phi_q^{(j)}\rangle \langle\phi_r^{(i)}\phi_s^{(j)}|. \quad (24)$$

Following the method described in [228], one can map these n single-particle basis functions binary numbers and to $\log_2(n)$ qubits. For instance, qubit state $|\phi_1\rangle = |00\dots 00\rangle$ would represent function $\phi_1(\mathbf{x}_1)$, $|\phi_2\rangle = |10\dots 00\rangle$ would represent function $\phi_2(\mathbf{x}_2)$, and so on. Given m electrons in the wavefunction we are trying to model, and given each electron can occupy at most one basis function (represented by $\log_2(n)$ qubits), one can model any product state with $N = m \log_2(n)$ qubits. As discussed before however, in first quantization the antisymmetry is enforced directly through the wavefunction. Procedures have been developed to maintain this requirement using an ancilla qubit register of $\mathcal{O}(m \log_2(n))$ qubits, and a circuit depth of $\mathcal{O}(\log_2^c(m) \log_2(\log_2(n)))$, where $c \geq 1$ depends on a choice of sorting network [278]. This additional depth would be considered acceptable in contrast to the scaling of most VQE ansätze (see Section 6).

Translation of the Hamiltonian operators (e.g. $|\phi_p^{(i)}\rangle \langle\phi_q^{(i)}|$) into operators that can be measured directly on quantum computers is fairly straightforward (also clearly explained in Ref. [104]). Each operator can be re-written into a tensor product of four types of single qubit operators: $|0\rangle\langle 0|$, $|0\rangle\langle 1|$, $|1\rangle\langle 0|$, and $|1\rangle\langle 1|$ which can be mapped to Pauli operators as follows:

$$\begin{aligned} |0\rangle\langle 0| &= \frac{1}{2}(I + Z) \\ |0\rangle\langle 1| &= \frac{1}{2}(X + iY) \\ |1\rangle\langle 0| &= \frac{1}{2}(X - iY) \\ |1\rangle\langle 1| &= \frac{1}{2}(I - Z). \end{aligned} \quad (25)$$

From Eq. (24) one can see that the number of operators in the Hamiltonian will scale with the number of two body terms $\mathcal{O}(n^4 m^2)$, as there is one term for each combination of 4 spin-orbitals and 2 electrons. Each spin-orbital is represented by $\log_2(n)$ qubits, hence for the two body terms we have $2 \log_2(n)$ tensored qubit outer products (e.g. $|0\rangle\langle 1|$), each of them composed of two Pauli operators. This results in a sum of $2^{2 \log_2(n)} = n^2$ Pauli strings each acting on up to $2 \log_2(n)$ qubits

(hence the Pauli weight in first quantization), for every operator in the Hamiltonian. This implies that the scaling of Pauli strings of the Hamiltonian in first quantization is $\mathcal{O}(n^6 m^2)$.

A few points are worth noting with respect to the use of first quantization for quantum computing. Firstly, to the best of our knowledge, there are no publications studying the usage of this method within the context of VQE or any other NISQ algorithm. Secondly, it is very clear that while the ancilla qubits do not change the overall scaling of the method, it bears a significant cost for NISQ devices, in addition to having to compute all spatial integrals on the quantum computer [326]. Finally it is very important to note that in general first quantization can be advantageous on systems which require a very large number of basis functions compared to the number of electrons (due to the logarithmic scaling of the number of qubits in n). This is the case in particular when a plane wave basis is selected, as it usually requires several orders of magnitude more functions than the molecular basis to achieve equivalent accuracy [235] — plane wave basis in first quantization has been shown to bring significant scaling advantages for in fault tolerant quantum simulation of chemical systems [170]. Overall, despite offering clear promise for fault tolerant quantum computing, it appears for the moment that first quantization is on balance too costly for the NISQ era.

Second quantization: Second quantization distinguishes itself from first quantization in that it enforces antisymmetry through the construction of its operators, rather than through the wavefunction. As such, the operators used to construct the Hamiltonian must abide by certain properties. The action of the operators must also allow moving a particle from one basis function to another (e.g. from an occupied orbital to a virtual orbital). In particular, while in first quantization this action (e.g. $|\phi_p^{(i)}\rangle\langle\phi_q^{(i)}|$) straightforwardly acts as moving an electron, operators in second quantization additionally need to verify antisymmetric properties. These operators are often referred to as fermionic creation ($\hat{a}_p^\dagger$) and annihilation ($\hat{a}_p$) operators. In their 1928 paper [119], Jordan and Wigner introduced the canonical fermionic anti-commutation relation:

$$\begin{aligned}\{\hat{a}_p, \hat{a}_q^\dagger\} &= \delta_{pq}, \\ \{\hat{a}_p^\dagger, \hat{a}_q^\dagger\} &= \{\hat{a}_p, \hat{a}_q\} = 0,\end{aligned}\tag{26}$$

from where one can also derive the commutation relations:

$$\begin{aligned}[\hat{a}_p, \hat{a}_q] &= -2\hat{a}_q\hat{a}_p, \\ [\hat{a}_p, \hat{a}_q^\dagger] &= \delta_{pq} - 2\hat{a}_q\hat{a}_p.\end{aligned}\tag{27}$$

These allow to re-write the Slater determinants as

$$|\psi\rangle = \prod_i (\hat{a}_i^\dagger)^{\phi_i} |\text{vac}\rangle = (\hat{a}_1^\dagger)^{\phi_1} (\hat{a}_2^\dagger)^{\phi_2} \dots (\hat{a}_n^\dagger)^{\phi_n} |\text{vac}\rangle,\tag{28}$$

where $|\text{vac}\rangle$ is the special vacuum state which disappears after any operation by the annihilation operator,

$$\hat{a}_j |\text{vac}\rangle = 0.\tag{29}$$

With such definition, $\hat{a}_p^\dagger$ acts on the unoccupied p th orbital to make it occupied, and $\hat{a}_p$ acts on the occupied p th orbital to make it unoccupied. Specifically, it is straightforward to show that

$$\begin{aligned}\hat{a}_j^\dagger |\phi_1 \phi_2 \dots\rangle &= \begin{cases} 0, & \phi_j = 1 \\ s_p |\phi_1 \phi_2 \dots 1_j \dots\rangle, & \phi_j = 0 \end{cases} \\ \hat{a}_j |\phi_1 \phi_2 \dots\rangle &= \begin{cases} 0, & \phi_j = 0 \\ s_p |\phi_1 \phi_2 \dots 0_j \dots\rangle, & \phi_j = 1 \end{cases}\end{aligned}\tag{30}$$

where s_p is the parity of p th orbital, i.e. s_p is 1 or -1 when the number of occupied orbitals up to and not including the p th orbital is even or odd:

$$p = (-1)^{\sum_{i=1,2,\dots,j-1} \phi_i}.\tag{31}$$

Fermionic operators allow re-writing the Hamiltonian presented in Eq. (24), using the one and two body integrals given in Eqs. (22) and (23):

$$\hat{H} = \sum_{pq} h_{pq} \hat{a}_p^\dagger \hat{a}_q + \frac{1}{2} \sum_{pqrs} h_{pqrs} \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_r \hat{a}_s,\tag{32}$$

providing a second quantized form of the molecular Hamiltonian. Lattice Hamiltonians can also be written using these operators. We can project the electronic coordinates into a basis set in order to define the two-body reduced density matrix (RDM):

$$\Gamma_{pqrs} \equiv \langle \psi | \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_r \hat{a}_s | \psi \rangle,\tag{33}$$

with other rank RDMs defined equivalently, and where the indices $p, q, \dots$ label spin-orbital degrees of freedom. In this example, the partial trace down to the one-body RDM can then be written as

$$\gamma_{pr} = \frac{1}{n-1} \sum_q \Gamma_{pq, rq}.\tag{34}$$

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.

Because, unlike first quantization operators, the fermionic operators are not defined explicitly, there could be a number of different ways to define them in terms of explicit Pauli operators (the reference to Jordan and Wigner's 1928 research is one of them). We have dedicated Section 4 to detailing numerous methods used to explicitly defined these operators.

3.2. Other Hamiltonian models

3.2.1. Lattice Hamiltonians

Rather than defining a Hamiltonian based on atomic configurations, lattice models assumes a number of sites organized along a lattice, which could be a one-dimensional chain, a two-dimensional lattice of various geometries, or any higher dimensional graph. Here we consider electrons as the “particles moving in this discretized space. Note that if one considers bosonic instead of fermionic particles, this would result in much simpler representation and encoding, since the fermionic antisymmetry relationships described in the following section are not required. Lattice models are widely used in condensed matter physics to model phenomenological properties of certain materials, such as electronic band structures [113,115,327] or phase transitions [328,329]. There exist a number of lattice models, here we only describe a few examples briefly:

- Hubbard/Fermi–Hubbard [330]:

$$\hat{H} = -t \sum_{\sigma} \sum_{\langle p, q \rangle} (\hat{a}_{p, \sigma}^{\dagger} \hat{a}_{q, \sigma} + \hat{a}_{q, \sigma}^{\dagger} \hat{a}_{p, \sigma}) + U \sum_p \hat{a}_{p, \uparrow}^{\dagger} \hat{a}_{p, \uparrow} \hat{a}_{p, \downarrow}^{\dagger} \hat{a}_{p, \downarrow}, \quad (35)$$

where the sum $\langle p, q \rangle$ is only taken for neighboring lattice sites, with $\hat{a}_{p, \sigma}$ denoting second quantized electron annihilation operators on site p with $\sigma \in \{\alpha, \beta\}$ an index for the spin of the electron, as discussed in more detail in Section 3.1.3.

- The Anderson impurity model used within the dynamical mean field theory [331] (see Section 9):

$$\begin{aligned} \hat{H} &= \hat{H}_{\text{imp}} + \hat{H}_{\text{bath}} + \hat{H}_{\text{mix}} \\ \hat{H}_{\text{imp}} &= \sum_{\alpha \beta} t_{\alpha \beta} a_{\alpha}^{\dagger} a_{\beta} + \sum_{\alpha \beta \gamma \delta} U_{\alpha \beta \gamma \delta} a_{\alpha}^{\dagger} a_{\beta}^{\dagger} a_{\gamma} a_{\delta} \\ \hat{H}_{\text{mix}} &= \sum_{\alpha i} \left(V_{\alpha i} a_{\alpha}^{\dagger} a_i + \bar{V}_{\alpha i} a_i^{\dagger} a_{\alpha} \right) \\ \hat{H}_{\text{bath}} &= \sum_i t_i' a_i^{\dagger} a_i \end{aligned} \quad (36)$$

where indices $\{\alpha, \beta, \gamma, \delta\}$ refer to the impurity sites, index i refers to bath sites, and $\{t_{\alpha \beta}, U_{\alpha \beta \gamma \delta}, V_{\alpha i}, t_i'\}$ parameterize the impurity model. $\hat{H}_{\text{imp}}$ is a general Hamiltonian describing the local correlation of the impurity, though this is also often approximated to have Hubbard-like interactions. $\hat{H}_{\text{mix}}$ describes the hopping between the impurity and the bath, which is taken in this example in a particular ‘star’ geometry, and $\hat{H}_{\text{bath}}$ is the non-interacting Hamiltonian of the bath sites.

- Spin Hamiltonians, such as the Heisenberg model [332,333]:

$$\hat{H} = J \sum_{\langle p, q \rangle} \hat{\mathbf{S}}_p \cdot \hat{\mathbf{S}}_q \quad (37)$$

where again $\langle p, q \rangle$ denotes a sum over neighboring pairs of sites on the lattice, J is the positive exchange constant, $\hat{\mathbf{S}}_p = (\hat{S}_p^x, \hat{S}_p^y, \hat{S}_p^z)$ is the three spin-1/2 angular momentum operators on site p . Note that the spin-1/2 matrices are related to Pauli matrices by $(\hat{S}^x, \hat{S}^y, \hat{S}^z) = \frac{\hbar}{2}(X, Y, Z)$.

3.2.2. Vibrational Hamiltonian model

The vibrational Hamiltonian is the nuclear-motion counterpart to the electronic Hamiltonian within the Born–Oppenheimer approximation. The Hamiltonian describes the nuclear motions resulting from bond vibration or rotation of the whole molecule. Example of vibrations will include bond stretching, or bending. The vibrational modes are the number of possibility for a molecule to undergo vibrational tensions. For a molecule with N atoms, each atom can move along three dimensions resulting in a total of $3N$ degrees of freedom. However, three degrees of freedom correspond to translations of the molecule in 3D space, and three degrees of freedom correspond to rotations of the molecule (of course, if the molecule is linear, there are only two rotational modes). This results in a $3N - 6$ for molecules ($3N - 5$ for linear molecule).

Vibrational Hamiltonian have far more options in their constructions than molecular Hamiltonians [334] resulting in a large variety of Hamiltonian representations. The simplest example, presented here is that of the harmonic oscillator,

suppose a molecule with $V = 3N - 6$ vibrational modes. We can provide an example of an effective nuclear Hamiltonian following McArdle et al. [63], once the electronic Hamiltonian has been solved (or estimated), as:

$$\hat{H} = \frac{\mathbf{p}^2}{2} + V_s(\mathbf{q}), \quad (38)$$

where $\mathbf{q}$ represent the nuclear positions, $\mathbf{p}$ the nuclear momenta, and $V_s(\cdot)$ the nuclear potential which is dependent on the electronic potential energy surface. Noting ω_i the harmonic frequency of vibrational mode i , the Hamiltonian in Eq. (38) can be approximated as a sum of independent harmonic oscillators:

$$\hat{H} = \sum_i \omega_i \hat{a}_i^\dagger \hat{a}_i. \quad (39)$$

The accuracy of results can be improved by adding anharmonic terms, however raising the complexity of the computation. In this case, the annihilation and creation act differently than in the case of fermionic operators, instead they represent transitions between different eigenstate of a single mode harmonic oscillator Hamiltonian. Suppose we consider d eigenstates of $\hat{h}_i = \omega_i \hat{a}_i^\dagger \hat{a}_i$. In Ref. [63], two means of encoding such Hamiltonians in qubits are proposed: in the first one (called the direct mapping) each eigenstate $|s\rangle$, $s \in \{0, d-1\}$ requires d qubits, and only qubit s is equal to 1 for state $|s\rangle$. The resulting creation operator is:

$$\hat{a} = \sum_{s=0}^{d-2} \sqrt{s+1} |0\rangle \langle 1|_s \otimes |1\rangle \langle 0|_{s+1}. \quad (40)$$

In the second case (called the compact mapping), each eigenstate is encoded in binary form, with each requiring $\log_2(d)$ qubits which results in the following creation operator:

$$\hat{a} = \sum_{s=0}^{d-2} \sqrt{s+1} |s+1\rangle \langle s|. \quad (41)$$

From this representation, the Hamiltonian of Eq. (39) (or its extension to anharmonic terms) can be mapped to Pauli operators in the same way as the electronic Hamiltonian in first quantization (see Section 3.1.3). Given V vibrational modes, the direct mapping requires Vd qubits, and if the anharmonic terms up to order k are included the Hamiltonian will have $\mathcal{O}(V^k d^k)$ Pauli terms. The compact mapping requires $V \log_2(d)$ qubits, and $\mathcal{O}(V^k d^{2k})$ terms in the Hamiltonian [63].

3.2.3. Periodic systems

Periodic system Hamiltonians aim at representing the energetic behavior of solid state systems, and periodic materials. As such, they are very similar to the definition of the molecular electronic structure Hamiltonian, but with the addition of boundary conditions defining the periodicity of the system. In second-quantized form (as described in Refs. [72,73]), the crystal Hamiltonian is given by:

$$\hat{H} = \sum_{pq} \sum_{\mathbf{k}} h_{pq}^{\mathbf{k}} \hat{a}_{p\mathbf{k}}^\dagger \hat{a}_{q\mathbf{k}} + \frac{1}{2} \sum_{pqrs} \sum_{\mathbf{k}_p \mathbf{k}_q \mathbf{k}_r \mathbf{k}_s} h_{pqrs}^{\mathbf{k}_p \mathbf{k}_q \mathbf{k}_r \mathbf{k}_s} \hat{a}_{p\mathbf{k}_p}^\dagger \hat{a}_{q\mathbf{k}_q}^\dagger \hat{a}_{r\mathbf{k}_r} \hat{a}_{s\mathbf{k}_s}, \quad (42)$$

where $h_{pq}^{\mathbf{k}}$, and $h_{pqrs}^{\mathbf{k}_p \mathbf{k}_q \mathbf{k}_r \mathbf{k}_s}$ are the one and two body integrals of the periodic system. In this representation, the one-electron integrals tensor is diagonal in $\mathbf{k}$ and the two-electron integrals follow $\mathbf{k}_p + \mathbf{k}_q - \mathbf{k}_r - \mathbf{k}_s = \mathbf{G}$, where $\mathbf{G}$ is a reciprocal lattice vector [73,335].

4. Fermionic space to spin space transformations

The core part of the VQE algorithm is the measurement of the expectation value of the Hamiltonian operator (Eq. (32)) with respect to a parameterized ansatz wavefunction. As mentioned in the previous section, in the NISQ era, with a limited number of qubits available, second quantization formalism has been favored over first quantization for quantum simulation. The fermionic creation and annihilation operators in second quantization formalism obey the anti-commutation algebra (Eq. (26)).

Qubits in comparison are spin-1/2 objects, and as such, the only operators that can, in general, be directly measured on QPUs are spin-operators (or the Pauli operators, X , Y , and Z) which obey a different algebra specified by their Lie bracket. This means that before starting the VQE loop, one must transform the Hamiltonian in second quantization into a linear combination of Pauli strings (tensor product of Pauli operators on multiple qubits).

As discussed in Section 3.1.3 however, operators in second quantization must enforce the antisymmetry of the wavefunction and as such must obey anti-commutation rules as detailed in Eq. (26). Pauli operators do not naturally obey these relationships and therefore specific cares must be brought to the mapping of fermionic operators to spin operators in the case of second quantization.

All transformations can be formalized as

$$\mathcal{T} : \mathcal{F}_n \rightarrow (\mathbb{C}^2)^{\otimes N}, \quad (43)$$

where a transformation $\mathcal{T}$ maps the space of operators acting on Fock states of n spin orbitals, $\mathcal{F}_n$, to the Hilbert space $(\mathbb{C}^2)^{\otimes N}$ of operators acting on spin states of N qubits. The most important feature of the transformation is that it must maintain the anti-commutation of fermionic operators. Conveniently, Jordan and Wigner [119] showed long before Quantum Computers were even conceptualized that an isomorphism exists between fermionic space and qubit space, maintaining the algebraic structure.

It is important to note that the mappings described in this section not only affect the operators measured when computing the expectation value of the Hamiltonian, but also the construction of an ansatz that is initially defined in fermionic terms (for instance the Unitary Coupled Cluster ansätze, see Section 6.2.2). These ansätze must be transformed into a series of Pauli operators exponentials which can be implemented as quantum gates.

In this section, we present a number of such transformations. There are three main characteristics relevant to deciding on a specific encoding, although it is worth noting that these are not necessarily independent from each other:

- **Number of qubits:** The number of qubits required to represent the electronic wavefunction. In general, the number of qubits is directly proportional to the number of spin-orbitals or sites considered. However, several techniques have been developed to concentrate the information held in the wavefunction to as few qubits as possible [282–284,336,337]. These methods generally rely on symmetries of the Hamiltonian and are presented in the last part of this section.
- **Pauli weight:** The maximum number of qubits on which each Pauli string produced act, i.e. the maximum number of non-identity operators in any Pauli string produced by the mapping. This is, in general, referred to as Pauli weight. Many features throughout the VQE pipeline are affected by this number. Firstly, low-weight encodings result in lower depth ansätze and lower circuit construction costs [293,338,339]. This is in part because a few operators that act locally on different qubit subsets can be implemented in parallel, in other parts because operators acting on local qubit subsets require fewer entangling gates than non-local qubit subsets. Secondly, it has been shown that using low-weight operators as observable in the VQE cost function provides some resilience against the barren plateau problem [87] (see Section 6.1). Finally, the lower the Pauli weight, the lower is the overall probability of readout error. This is obvious, as identity operators do not need to be measured, and as such the probability of measurement error increases with the Pauli weight [124].
- **Number of Paul-strings:** The number of different Pauli strings resulting from the mapping. The number of Pauli strings directly impacts the cost of implementing a VQE. As such, one should always prefer to have the lowest number of Pauli strings to measure. In general, we see that this number of strings scales $\mathcal{O}(n^4)$ for molecular Hamiltonian, and with the number of edges for lattice models.

We can distinguish two main families of fermion-to-spin encoding. The first one concerns itself with being as general as possible and directly encodes the entire Fock space. The mappings it includes tend to be more relevant for *ab initio* molecular systems, with their main drawback being the non-locality of the operators produced (see Section 4.1). The second one uses the specific geometry of the system studied to try to minimize the Pauli weight of the operators produced, at the cost of using additional qubits. These mappings tend to be more relevant for low degree lattice models (for instance the 2-dimensional Hubbard model) as they allow capping the Pauli weight (see Section 4.2).

4.1. Generalized encodings

In this section, we are concerned with encodings that map the entire Fock space in which the Hamiltonian is expressed. As such, the number of qubits N required from these encodings is in principle equal to the number of spin orbitals n in the Hamiltonian considered (without taking into consideration possible use of symmetries to reduce the number of qubits required, as presented in Section 4.3). These encodings are the most general, in that they are not tailored for specific Hamiltonian structures. They are agnostic to the degree of connectivity of the Hamiltonian graph (the maximum number of fermionic operators linking one spin orbital to another) and are in general better suited for *ab initio* molecular systems than to lattice models. One recurrent issue related to these encodings is the fact that they transform one and two-body fermionic operators which are local (act only on up to four spin orbitals at the time) into Pauli strings which are in general non-local and therefore have either high Pauli weight or Pauli weight that scales with the number of spin orbitals in the system. It also means that high connectivity on the qubit lattice may be required to efficiently implement fermionic operator based ansätze without relying on a large number of entangling gates. A summarized comparison of these mappings is presented in Table 4.

4.1.1. The Jordan–Wigner encoding

The Jordan–Wigner mapping encodes the electronic wavefunction in an array of qubits by mapping the occupation number of spin orbitals in qubits. The occupation number of n spin orbitals is stored in N qubits. $|0\rangle_j$ corresponds to the j th spin orbital being unoccupied and $|1\rangle_j$ corresponds to the j th spin orbital being occupied (where once again, j merges the indices of spatial and spin orbitals). The theoretical foundations of this mapping lie in the 1928 Jordan–Wigner transformation [119] using spin-1/2 operators to explicitly describe fermionic ladder operators. Its inverse can therefore be used as a means to simulate fermions on a Quantum Computer [231,341–343].

Table 4

Overview and comparison of the generalized encodings. n represents the number of fermionic mode in the Hamiltonian considered. All mappings in the table use by default n qubits, and produce a number of operators scaling $\mathcal{O}(n^4)$. These operators are counted assuming an *ab initio* molecular Hamiltonian in second quantized form.

Method	Pauli weight	Comments
Jordan–Wigner [119]	$\mathcal{O}(n)$	Most commonly used mapping. Encodes orbital occupation directly and locally onto qubits.
Parity [279]	$\mathcal{O}(n)$	Encodes orbital parity directly and locally onto qubits.
Bravyi–Kitaev [279,280]	$\mathcal{O}(\log_2(n))$	Focuses on minimizing the Pauli weight by mixing occupation and parity encoding. Usually results in lower gate depth [281,295,340], but not necessarily higher noise resilience [259].
Optimal general encoding on ternary trees [256]	$\mathcal{O}(\log_3(2n))$	Achieves optimal Pauli weight asymptotically. Little to no benchmarking of actual applications compared to other mapping invites to further investigation.

As mentioned in Section 3.1.3, Jordan and Wigner [119] introduced the canonical fermionic anti-commutation relation given in Eq. (26) and showed how one can go about making the identification between fermionic operators and spin operators.

Considering the case of a single orbital, we can write the fermionic operators actions presented in Eq. (30) on the j th qubit as spin operators:

$$\begin{aligned}\hat{a}_j^\dagger &\xrightarrow{?} |1\rangle\langle 0|_j = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = \frac{X_j - iY_j}{2} \\ \hat{a}_j &\xrightarrow{?} |0\rangle\langle 1|_j = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \frac{X_j + iY_j}{2},\end{aligned}\quad (44)$$

where X_j and Y_j are Pauli gates acting on the j th qubit. These Pauli operators do not enforce the fermionic sign prescription in Eq. (30) (or equivalently the anticommutation relation of Eq. (26)) which is critical to preserve the structure of the algebra through the transformation.

We can fix this by upgrading the definition in Eq. (44) to include a string of Z operators $Z_0 \otimes \cdots \otimes Z_{j-1}$ acting on each of the other qubits up to the j th. Indeed we have that $Z_0 \otimes \cdots \otimes Z_{j-1}$ has eigenvalue $+1$ for states with an even number of occupied orbitals up to the j th, and eigenvalue -1 for states with an odd number of occupied orbitals up to the j th, restoring the fermionic sign prescription and providing the Jordan–Wigner transformation:

$$\begin{aligned}\hat{a}_j^\dagger &\rightarrow \frac{X_j - iY_j}{2} \otimes Z_0 \otimes \cdots \otimes Z_{j-1} \\ \hat{a}_j &\rightarrow \frac{X_j + iY_j}{2} \otimes Z_0 \otimes \cdots \otimes Z_{j-1}.\end{aligned}\quad (45)$$

One can easily verify that this also enforces the anticommutation relation of Eq. (26). Indeed for $i \neq j$ we have that $\hat{a}_i$ and $\hat{a}_j^\dagger$ anticommute:

$$\hat{a}_i \hat{a}_j^\dagger = -\hat{a}_j^\dagger \hat{a}_i, \quad (46)$$

and the definition in Eq. (45) gives us the correct anticommutation relation as the Z_j operator anticommutes with both X_j and Y_j and hence anticommutes with $|0\rangle\langle 1|_j$ and $|1\rangle\langle 0|_j$.

The direct consequence of adding the Z strings is that the Pauli weight of Jordan–Wigner mapping scales $\mathcal{O}(N)$. This makes the mapping relatively costly with regard to the number of entangling gates required to simulate fermionic based ansätze. However, it is worth noting that some ansätze include spin operator level parameters [344]. This allows the VQE to variationally learn the anti-commuting relationship and therefore allows forgoing the Z strings at the cost of additional circuit parameters. This approach caps the Pauli weights of the Jordan–Wigner mapping to four (one for each fermionic operator in two-body terms) and significantly reduces the number of entangling gates required.

Another feature of the Jordan–Wigner mapping that is worth discussing is the number of Pauli strings produced. From Eq. (45), it is clear that each fermionic operator results in two Pauli strings. As such, the number of Pauli strings scales at the same rate as the number of fermionic operators in the system considered. For a second quantized Hamiltonian, the number of fermionic operators scales with the number of two-body fermionic terms, or $\mathcal{O}(n^4)$, and therefore, the number of strings in the Jordan–Wigner mapping would scale $\mathcal{O}(N^4)$. The Jordan–Wigner mapping is by design applicable only to qubits, encoded from $SU(2)$ fermions. However a generalized form of Jordan–Wigner, applicable to $SU(N)$ fermions was proposed by Consiglio et al. [345].

4.1.2. The parity encoding

If instead of using j th qubit to encode directly the information of whether the j th orbital is occupied, we use it to encode information about the parity of the orbitals up to the j th, we obtain the parity encoding. The parity mapping was

explicitly defined in [279]. We have $|0\rangle_j$ if the number of orbitals up to and including the j th that are occupied is even and $|1\rangle_j$ if it is odd. Given this definition, if we are given the fermionic state $|v_0 v_1 \dots v_n\rangle$ we can translate it to the qubit state $|p_0 p_1 \dots p_n\rangle$ in the parity encoding by

$$\begin{aligned} p_i &= \sum_{j \leq i} v_j \pmod{2} \\ &= \sum_j [\pi_n]_{ij} v_j \pmod{2}, \end{aligned} \quad (47)$$

or:

$$|\mathbf{p}\rangle = |\pi_n(\mathbf{v})\rangle \pmod{2}, \quad (48)$$

where addition is taken modulo 2, and π_n is the $n \times n$ matrix defined as

$$\pi_n = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix}. \quad (49)$$

If the parity changes at the j th position (that is if we have either $|0\rangle_{j-1}$ and $|1\rangle_j$ or $|1\rangle_{j-1}$ and $|0\rangle_j$) then we know that the j th orbital is occupied. Conversely if the parity does not change (that is if we have either $|0\rangle_{j-1}$ and $|0\rangle_j$ or $|1\rangle_{j-1}$ and $|1\rangle_j$) the j is unoccupied.

Moreover, in the parity representation, we already have the information on the parity of the qubits up to the j th encoded in the $(j-1)$ th qubit. Hence we know that following Eq. (30) the operators $\hat{a}_j^\dagger$ and $\hat{a}_j$ give a minus sign when we have $|1\rangle_{j-1}$, and a plus sign when we have $|0\rangle_{j-1}$. We can thereby define mappings for isolated fermionic operators as

$$\begin{aligned} \hat{a}_j^\dagger &\xrightarrow{?} |01\rangle\langle 00|_{j-1,j} - |10\rangle\langle 11|_{j-1,j} \\ &= \frac{Z_{j-1} \otimes X_j - iY_j}{2} \\ \hat{a}_j &\xrightarrow{?} |00\rangle\langle 01|_{j-1,j} - |11\rangle\langle 10|_{j-1,j} \\ &= \frac{Z_{j-1} \otimes X_j + iY_j}{2}. \end{aligned} \quad (50)$$

However because we are adding or removing an electron in the j th orbital and hence changing the parity encoded in the qubits that follow the j th we also need to flip those qubits. We can update the parity by means of a string of X operators acting on the qubits that follow the j th qubit, $X_{j+1} \otimes \dots \otimes X_{n-1}$, so that the parity transformation is given by

$$\begin{aligned} \hat{a}_j^\dagger &\rightarrow \frac{Z_{j-1} \otimes X_j - iY_j}{2} \otimes X_{j+1} \otimes \dots \otimes X_{n-1} \\ \hat{a}_j &\rightarrow \frac{Z_{j-1} \otimes X_j + iY_j}{2} \otimes X_{j+1} \otimes \dots \otimes X_{n-1} \end{aligned} \quad (51)$$

The Pauli weight of the parity mapping scales similarly to that of the Jordan–Wigner mapping: $\mathcal{O}(N)$. It also results in the same operator scaling: $\mathcal{O}(N^4)$.

4.1.3. The Bravyi–Kitaev encoding

The theory of the Bravyi–Kitaev encoding was first given in Ref. [280] and then the encoding was explicitly constructed in Ref. [279] (and revisited later in Ref. [281]). The motivation for this encoding comes from the desire to lower the Pauli weight of the qubit operators. The Bravyi–Kitaev encoding achieves lower Pauli weight by storing a combination of parity and occupation number in qubits. This produces $\mathcal{O}(N^4)$ spin operator terms, with a maximum Pauli weight of $\mathcal{O}(\log_2(N))$.

We can define an encoding matrix for the Bravyi–Kitaev mapping, similar to the one proposed for the parity mapping in Eq. (49). It is produced recursively by defining blocks of increasing sizes and combining them. If we set x as the recursive index, the block β of index x has size 2^x . Therefore, as the final matrix must be $N \times N$, the recursion stops whenever $x \geq \log_2(N)$ – the matrix rows and columns greater than N can be discarded.

Starting from $\beta_1 = [1]$, the recursion rule can be defined as

$$\beta_{2^x} = \left[\begin{array}{c|c} \beta_{2^{x-1}} & 0 \\ \hline 0 & \beta_{2^{x-1}} \\ \hline \leftarrow 1 \rightarrow & \end{array} \right], \quad (52)$$

where the bottom-left block is a $2^{x-1} \times 2^{x-1}$ matrix of all zeros except for the bottom row.

For example, the Bravyi–Kitaev block for $x = 3$ matrix acting on 8 qubits β_8 is given by

$$\beta_8 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}. \quad (53)$$

From here, it is easy to notice that within each block, the j th qubit encodes the occupancy of the j th orbital when j is even, and it encodes information on the parity of a set of orbitals and occupancy of the j th orbital combined when j is odd. Regarding the parity information, we can distinguish two cases: if j is odd but different from a power of 2, minus 1 (since we index from 0) it encodes the parity of orbitals up to the j th orbital within a block; if it is equal to a power of 2 minus 1, it encodes the parity of the entire set of orbitals up to and including j .

To be able to write down expressions for the representations of the $\hat{a}_j^\dagger$ and $\hat{a}_j$ operators we first need to consider four qubit sets (for a thorough and formal definition of these sets and how to obtain them recursively, we recommend [279, 281]). These sets are defined based on index j and determine the behavior of qubits of other indices whenever an operator acts on qubit j :

- **The update set** $U(j)$ includes qubits that are dependent on the occupation of orbital j . Because even-index qubits encode occupancy, they are never part of this set. Odd qubits are included if they encode the parity of orbital j . Intuitively, it corresponds to the nonzero elements in the j th column of the Bravyi–Kitaev transformation matrix, excluding the j th qubit itself (for example, Eq. (52)).
- **The parity set** $P(j)$ are the qubits that determine the parity of the set of orbitals up to, but excluding, j (and therefore the fermionic sign, Eq. (30)), when an operator acts on qubit j .
- **The flip set** $F(j)$ is a subset of the parity set which determines whether qubit j is equal or opposite to orbital j of the fermionic basis. Of course, this set is always empty for even qubits which encode occupancy.
- **The remainder set** $R(j)$ are the qubits that are part of the parity set, but not the flip set.

$$R(j) = P(j) \setminus F(j) \quad (54)$$

For a visual interpretation of how these sets can be constructed, one can refer to the Fenwick tree construction of the Bravyi–Kitaev mapping presented in Ref. [293] (we also provided a brief description in Appendix A).

With this in mind, we now look at how to express fermionic operators in the Bravyi–Kitaev mapping. These differ depending on whether they act on qubits that encode occupancy, or qubits that encode some parity information. When considering the definition of the qubit sets above, one can observe that due to the recursive definition of the Bravyi–Kitaev transformation matrix, the maximum number of qubits included in any set scales $\mathcal{O}(\log_2(N))$. The Pauli weight of this encoding follows from this observation and the construction of the operators presented below.

Operators for occupancy qubits: as explained above, these operators apply to even-index qubits. For individual qubits, the creation and annihilation operators are identical to those initially formulated for the Jordan–Wigner mapping (Eq. (44)). We now need to incorporate the relevant actions on the remainder of the wavefunction, namely: we need to flip all the qubits above index j that are affected by an action on qubit j (i.e. the qubits from the update set $U(j)$), and we need to apply the correct fermionic sign, depending on the qubits that define the parity of the wavefunction up to orbital j (the qubits from the parity set j). The former requires X gates, the latter requires Z gates. As such, we can write the operators acting on even qubits as

$$\begin{aligned} \hat{a}_j^\dagger &\rightarrow Z_{P(j)} \otimes |1\rangle\langle 0|_j \otimes X_{U(j)} \\ &= \frac{1}{2} Z_{P(j)} \otimes (X_j - iY_j) \otimes X_{U(j)} \\ \hat{a}_j &\rightarrow Z_{P(j)} \otimes |0\rangle\langle 1|_j \otimes X_{U(j)} \\ &= \frac{1}{2} Z_{P(j)} \otimes (X_j + iY_j) \otimes X_{U(j)}. \end{aligned} \quad (55)$$

Operators for parity qubits: When j is odd, a creation operator, which in the case of occupancy only excites qubits from $|0\rangle$ to $|1\rangle$, could flip a qubit from $|0\rangle$ to $|1\rangle$ instead depending on the parity of the wavefunction up until j (if the parity is 1 until j , and if qubit j is in state $|0\rangle$ it means orbital j is occupied as the parity is flipped). Recall that the flip set determines whether qubit j is equal, or opposite to orbital j . Two cases are possible, either the number of qubits equal to $|1\rangle$ in the flip set is even, in which case qubit j is equal to the orbital occupancy (the parity flips if qubit $j = |1\rangle$, and vice-versa); or the number of $|1\rangle$ in the flip set is odd, and qubit j is opposite to the orbital occupancy (the parity flips if qubit $j = |0\rangle$,

and vice-versa). Following this, one can define two projectors onto the even and odd states of the flip set:

$$\begin{aligned} E_{F(j)} &= \frac{1}{2} (I^{\otimes N} + Z_{F(j)}) \\ O_{F(j)} &= \frac{1}{2} (I^{\otimes N} - Z_{F(j)}). \end{aligned} \quad (56)$$

Therefore, action on an isolated qubit can be described as follows:

$$\begin{aligned} \hat{a}_j^\dagger &\xrightarrow{?} E_{F(j)} \otimes |1\rangle\langle 0|_j + O_{F(j)} \otimes |0\rangle\langle 1|_j \\ &= \frac{X_j - iZ_{F(j)} \otimes Y_j}{2} \\ \hat{a}_j &\xrightarrow{?} E_{F(j)} \otimes |0\rangle\langle 1|_j + O_{F(j)} \otimes |1\rangle\langle 0|_j \\ &= \frac{X_j + iZ_{F(j)} \otimes Y_j}{2}. \end{aligned} \quad (57)$$

Just like in the case of even j , our creation and annihilation operators also need to flip all the qubits in the update set $U(j)$ by applying X Pauli gates on them and to enforce the correct fermionic sign by applying Z operators to the qubits given by the parity set $P(j)$. And because $Z^2 = I$ we have $Z_{F(j)} \otimes Z_{P(j)} = Z_{P(j) \setminus F(j)} = Z_{R(j)}$, so that

$$\begin{aligned} \hat{a}_j^\dagger &\rightarrow \frac{Z_{P(j)} \otimes X_j \otimes X_{U(j)} - iZ_{R(j)} \otimes Y_j \otimes X_{U(j)}}{2} \\ \hat{a}_j &\rightarrow \frac{Z_{P(j)} \otimes X_j \otimes X_{U(j)} + iZ_{R(j)} \otimes Y_j \otimes X_{U(j)}}{2} \end{aligned} \quad (58)$$

4.1.4. Optimal general encoding based on ternary trees

The mapping proposed by Jiang et al. [256] offers an optimal scaling of Pauli weights of $\mathcal{O}(\log_3(2n + 1))$, with n the number of fermionic modes, outperforming Bravyi–Kitaev in this respect. It is optimal in the sense that for a Hamiltonian for which fermionic modes are fully connected, it achieves the minimum average Pauli weight possible. It organizes qubits along with ternary trees [346] and relies on the definition of the second quantized Hamiltonian in terms of Majorana fermions.

Majorana fermions are theorized particles, which act as their own antiparticle [347,348]. Most of the Standard Model fermions are known not to behave like Majorana particles, at least at low energy, except the neutrino for which the question is still open [349–352]. Bound Majorana fermions have been shown to appear composed of several particles in condensed matter physics [353]. Formally, this means that creation and annihilation operators for Majorana fermions are identical $\hat{\gamma}_i^\dagger = \hat{\gamma}_i$. These can also be expressed in terms of ordinary fermionic operators as

$$\begin{aligned} \hat{\gamma}_{2j} &= \hat{a}_j + \hat{a}_j^\dagger \\ \hat{\gamma}_{2j+1} &= -i(\hat{a}_j - \hat{a}_j^\dagger). \end{aligned} \quad (59)$$

There are two Majorana operators for each fermionic operator. These must anticommute if they are of different indices and commute otherwise.

To build this encoding, one must first map the qubits to the vertices of a ternary tree (a tree that splits into three edges after each vertex) as presented in Fig. 3. For any path $\mathbf{p}$ in the tree, one can define the following operators:

$$A_{\mathbf{p}} = \bigotimes_{l=0}^{h-1} \sigma_{\alpha}^{(v)}, \quad (60)$$

where h is the height of the tree, v is the qubit index on path $\mathbf{p}$ and α corresponding to X , Y or Z depending on whether the path follows the left, central or right edge respectively after qubit v .

These operators clearly obey the same anti-commutator relationships as Majorana operators: $\{A_{\mathbf{p}}, A_{\mathbf{q}}\} = 0$ if $\mathbf{p} \neq \mathbf{q}$, and $A_{\mathbf{p}}^2 = \mathbb{1}$. Given there are $2n + 1$ distinct path in the ternary tree, and we need $2n$ Majorana operators, Jiang et al. [256] propose mapping each $A_{\mathbf{p}}$ to a single Majorana operator, and these operators have a maximum Pauli weight equal to the height of the tree (h), hence $\log_3(2n + 1)$.

To construct the Hamiltonian, one needs to first transform the fermionic operators into Majorana operators and re-write the Hamiltonian, and then decide on an allocation of the Pauli operators defined above to the Majorana operators [256]. Similar to the encodings previously defined, the number of qubits N is equal to the number of fermionic modes n , and the scaling of the number of Pauli operators is the same as that of the two body-terms in the second quantized Hamiltonian $\mathcal{O}(N^4)$.

4.1.5. Discussion on generalized encodings

Before drawing a comparison of the different generalized encodings mentioned above, we would like to raise two relevant points regarding this type of encoding:

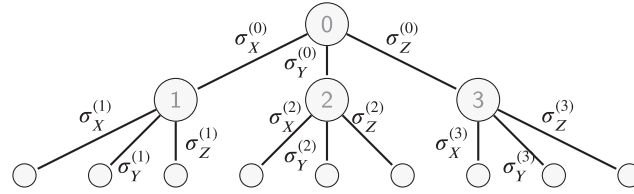


Fig. 3. Ternary tree structure in the case of a 4 fermionic modes system. Labeled nodes represent qubits, while unlabeled nodes are empty and added only to complete the edges from the previous level. We note h the height of the tree (here we have $h = 2$). The operators are placed on edges but apply to the qubit at the origin of each edge (i.e. the qubit with index corresponding to the node from which the edge start). By convention, X applies to left edges, Y to central edges, and Z to right edges.

- All the encoding mentioned above can be constructed in terms of Fenwick trees [293], providing an opportunity to define them differently and possibly optimize the encoding to specific fermionic models. We included a description of Fenwick trees construction (similar to the presentation in Ref. [293]) and their relation to encodings in Appendix A. Fenwick trees are also relevant in the context of optimizing the Bravyi–Kitaev mapping for application to 2-dimensional lattice models [293].
- Steudtner et al. [337] propose three additional generalized encodings which allow the number of qubits to be reduced using symmetries at the cost of additional entangling gates. Because the method presented in Ref. [282] and Ref. [283] addresses these symmetries without significant additional gate cost (see Section 4.3), it is in general preferred. We nonetheless encourage readers interested in going deeper in the knowledge of these general encodings to go through Ref. [337] as it discusses the relevant theory in depth.

The key metric for comparing the four mappings presented above is their Pauli weight: while Jordan–Wigner and Parity scale $\mathcal{O}(N)$, Bravyi–Kitaev scales $\mathcal{O}(\log_2(N))$ and the optimal ternary tree encoding scales $\mathcal{O}(\log_3(2N))$ (which is asymptotically the better one). The number of qubits is the same across all these mappings and equates to the number of fermionic modes ($N = n$). These mappings can also benefit from the same amount of qubit reduction using symmetries (see Section 4.3). While the number of operators may vary, it has also been shown numerically that in general applying grouping strategies (presented in Section 5) to these operators results in a very similar number of total operators to measure [123].

Because the Bravyi–Kitaev and optimal ternary tree mappings have lower Pauli weight, they are in theory less subject to read-out errors. However, this becomes invalid once the grouping of the Pauli operators for joint measurements is introduced. This is because, for most groups, the entire register of qubits needs to be measured in the Z basis. The main advantage of the $\log_2(N)$ or $\log_3(Nn)$ Pauli weight is in using the mapping to construct an ansatz based on fermionic operators (such as the Unitary Coupled Cluster ansatz, UCC, see Section 6.2.2). Numerical studies looking at the use of Bravyi–Kitaev mapping in the context of Quantum Phase Estimation have consistently shown that it results in a significant reduction in the number of entangling gates required [281,295,340], without impacting the overall accuracy of the result in noiseless simulations. It is clear these results also translate to applications of the UCC ansatz in the context of the VQE and are likely to extend to the optimal ternary tree mapping. It is important to note that the relative impact of low Pauli weight is very much dependent on the degree of connectivity of the qubit register, and limited connectivity (e.g. qubits placed in a line), could result in a significant increase in the number of entangling gates required in a low Pauli weight encoding.

A final point to note is that a recent study [259] numerically tested that UCC ansatz constructed for VQE with the Bravyi–Kitaev mapping could result in slightly higher sensitivity to quantum noise channels (see Appendix C.1) than the Jordan–Wigner mapping, despite lower gate depth. Sawaya et al. [259] conjecture that this could be because occupation numbers are stored locally in the Jordan–Wigner mapping: a single qubit error only impacts the result of one orbital, while for the Bravyi–Kitaev mapping, it could affect several. This also implies that occupation number errors create greater errors in the energy than parity errors [259]. One could expect the parity and ternary tree mappings to be affected by the same phenomenon, although we are aware of any further research on this topic.

4.2. Lattice model tailored encoding

The mappings defined above are oftentimes inefficient for lattice models. This is because they assume full connectivity between the different fermionic modes, and translating this into spin-operators results in non-local operations. Instead, the mappings presented in this section are concerned about being as efficient as possible for a given lattice model, where fermionic modes are attached to a specific geometry. The earliest explicit definition of fermion to qubit mapping was indeed an application to the Hubbard model [228]. The literature is divided into two main approaches for doing so: auxiliary fermion schemes [354,355] and Loop-Stabilized Bravyi–Kitaev (LSBK), also known as Superfast Encoding (SFE) [280]. Maintaining low Pauli weights comes at a cost of additional qubits for all the methods mentioned in this section. However, it has been shown that all these methods also allow for at least some degree of error correction. A summary of the key metrics of each encoding is presented in Table 5.

Table 5

Comparative summary of lattice tailored encodings. d represents the degree of the Hamiltonian graph, v and h respectively the vertical and horizontal dimensions of a 2-dimensional lattice (we set $v \leq h$), n is the number of fermionic modes/sites on the lattice. The number of operators scales linearly with E , the number of edges. We have $E = h(v - 1) + v(h - 1)$ for a 2-dimensional regular lattice, and $E = \sum_i^D [(n_i - 1) \prod_{j \neq i}^D n_j]$ for an regular lattice of dimension D and n_i the number of sites along the i th dimension.

Method	Pauli weight	Qubits	Comments
Jordan–Wigner (snake pattern) [354]	$\mathcal{O}(2v)$	n	Optimal direct application of the Jordan–Wigner mapping to a 2-dimensional lattice.
Bravyi–Kitaev (Fenwick tree lattice mapping) [293]	$\mathcal{O}(\log(v))$	n	Optimal direct application of the Bravyi–Kitaev mapping to a 2-dimensional lattice.
Auxiliary fermion scheme [354]	4 (2-dim.)	$2n$, ($n = vh$)	Uses auxiliary fermion/qubit registers to create operators that cancel-out Z strings in the Jordan–Wigner mapping
Superfast Bravyi–Kitaev [280,293,295]	$\mathcal{O}(2d)$	$\mathcal{O}(nd/2)$	Relies on stabilizers formalism to define an efficient encoding with cost dependent on the degree of the Hamiltonian graph
Generalized superfast encoding [257]	$\mathcal{O}(\log_2(d))$	$\mathcal{O}(nd/2)$	Extension of Superfast BK, optimizing Pauli weight and offering better opportunities for error corrections
Compact encoding [260,261]	3 (2-dim.), 4 (3-dim.)	$\mathcal{O}(1.5n)$	Modifies the stabilizer formalism used in Superfast BK to optimize the number of qubits required. So far limited to 2 and 3-dimensional lattices

All the mappings presented in this section will exhibit a similar scaling in the number of operators produced, which will be linearly proportional to the number of edges E in the lattice. For example, consider a regular lattice of dimension D , with n_i the number of sites along the i th dimension we easily obtain

$$E = \sum_i^D \left[(n_i - 1) \prod_{j \neq i}^D n_j \right]. \quad (61)$$

We also have the total number of sites $n = \prod_j^D n_j$ and therefore a number of edges is capped below $\mathcal{O}(nD)$. An additional pre-factor should be included to account for the fact that some encoding require auxiliary sites (and hence auxiliary operators), however this does not change the overall scaling.

4.2.1. Auxiliary fermion schemes

Suppose a 1-dimensional spin-chain, with only nearest neighbor interactions. All operators of interest are of the form $\hat{a}_i^\dagger \hat{a}_j$, with $i = j \pm 1$. Therefore, under the Jordan–Wigner mapping all operators have the form

$$\hat{a}_{i \pm 1}^\dagger \hat{a}_i = \frac{X_{i \pm 1} - iY_{i \pm 1}}{2} \otimes \frac{X_i + iY_i}{2}. \quad (62)$$

One can note that the Z strings required to maintain the anticommuting relationship have canceled out, and are no longer necessary: the Jordan–Wigner mapping preserves operator locality in one-dimensional systems. Auxiliary Fermion Schemes answer the question of how to maintain operator locality in mappings such as Jordan–Wigner when the lattice model considered has higher dimensions.

A first example of encoding tailored for lattice model was independently proposed in [354,355]. As described above (both grounded in [356,357]), it can be seen as a Jordan–Wigner encoding optimized for rectangular lattice problems. It avoids the need for strings of Z operators in interaction terms built out of creation and annihilation operators thereby optimizing the Pauli weight, but at the cost of increasing the number of qubits. This is referred to in the literature alternatively as the Ball–Verstraete–Cirac (BVC) encoding (for example [358]).

Suppose a rectangular spin lattice, such as the 2-dimensional Hubbard model. One could nearly map it to a one-dimensional system by ordering the operators (for example, as the ‘snake pattern’ presented in Fig. 4). Along with this ordering (i.e. considering only the operator connections it covers), one can easily implement a local version of the Jordan–Wigner mapping. The issue however is that some connections are not covered and cannot be directly expressed with local operators using the Jordan–Wigner mapping (consider for example the dotted line between site 1 and 6, representing operators $\hat{a}_1^\dagger \hat{a}_6$ and transpose in the figure, which would require Z strings in this ordering).

The BVC mapping provides a means to encode operators along the edges not covered in the ordering locally. In order to do so, it defines an auxiliary Hamiltonian composed of a weighted sum of interaction operators which are constructed using an alternative set of fermionic operators $\hat{b}_i$ such that

$$\hat{\mathcal{H}}_{\text{aux}} = \sum_{\{i,j\}} \hat{h}_{ij}^{(\text{aux})} = \sum_{\{i,j\}} -(\hat{b}_i + \hat{b}_i^\dagger)(\hat{b}_j - \hat{b}_j^\dagger), \quad (63)$$

which we note can be re-written in terms of Majorana operators (see Eq. (59)):

$$\hat{\mathcal{H}}_{\text{aux}} = \sum_{\{i,j\}} i\hat{\gamma}_i \hat{\gamma}_{j+1}. \quad (64)$$

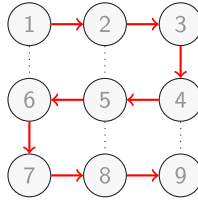


Fig. 4. Example of a 2-dimensional spin lattice with nine sites. Edges represent connections between the different fermionic modes. Red arrows are connections that provide an example of a ‘snake ordering’ of the fermionic modes in a 1-dimensional pattern, while dotted lines show the connections missed by the ordering.

The fermions are configured to be in the ground state of this Hamiltonian, denoted $|\chi\rangle$. The auxiliary Hamiltonian sites match those of the physical Hamiltonian. The overall system is now composed of $2n$ fermions which we can order as $1, 1', 2, 2', \dots, n, n'$.

For notational simplicity, indices of the auxiliary fermionic system are primed. From there, the edges of the physical lattice Hamiltonian which are not covered by the initial ordering of fermions into a 1-dimensional system (indexed here by $\{p, q\}$), can be modified as follows:

$$\hat{a}_p^\dagger \hat{a}_q \rightarrow \hat{a}_p^\dagger \hat{a}_q \hat{h}_{pq}^{(\text{aux})} = \hat{a}_p^\dagger \hat{a}_q (i\gamma_{p'} \gamma_{q'+1}). \quad (65)$$

It is important to note, that because any of the vertical hopping terms in the physical Hamiltonian commute with the auxiliary Hamiltonian, the operation described above does not modify the ground state of the system. The consequence is that when now applying the Jordan–Wigner transformation to the new, joint Hamiltonian, the Z strings of the original, non-covered edges are canceled out by those of the auxiliary terms by which they have been modified. Note that using Jordan–Wigner, the auxiliary Hamiltonian presented in Eq. (64) can be non-local, though this can be addressed easily by substitution of the non-local operators with local ones (see Ref. [354] for details). The main advantage of this method is that, as long as only nearest-neighbor interactions are allowed, it caps the Pauli weights to 4 for hopping terms (basically following Eq. (62)), and to 2 for Coulomb terms. Suppose a rectangular lattice of $h \times v$ sites, (we set $v \leq h$), a naive implementation of a Jordan–Wigner mapping would result in a maximum Pauli weight of $\mathcal{O}(2v)$ (consider for example the mapping of the hopping term between site 1 and 6 in Fig. 4). This benefit comes at the cost of doubling the number of qubits required due to the introduction of the auxiliary Hamiltonian, hence a total of $2vhL$ qubits as initially one is required for each fermionic mode.

Whitfield et al. [358] extend the theory of the BVC mapping presented in [354,355], by showing that there is a range of possible choices of the auxiliary coupling operators, which are not required to be Majorana operators. They also show that the number of auxiliary modes required for each fermionic mode in a lattice grows as $D - 1$, with D the dimension of the lattice. Further improvements were also proposed in [359].

An adaptation of the auxiliary fermion scheme to the Bravyi–Kitaev encoding is also proposed by Havlíček et al. [293], thereby optimizing it for rectangular lattice models. Havlíček et al. propose to first build the Bravyi–Kitaev mapping into a data structure using a Fenwick tree and then map the connections of this Fenwick tree to a fermion lattice. Because sites that are connected through the tree (either by means of the parity set or the update set) only require local operations, non-local operations are restricted to the mapping of connections on the lattice which are not included on the tree. By doing so, they show that with optimal mapping of the Fenwick tree to the lattice, the Pauli weight of their adapted Bravyi–Kitaev encoding scales as $\mathcal{O}(\log(v))$, with v the smallest side of the lattice.

4.2.2. Superfast encoding/loop stabilizer encodings

The encoding concept developed in this section was initially presented in [280] and is based on a graph representation of the Hamiltonian. Of course, any lattice models (or *ab initio* molecular system) can be represented as a graph, where each fermionic mode is a vertex, and each interaction with another fermionic mode is a weighted edge in the graph. Unlike the mappings presented so far, in this mapping, qubits are used to encode interactions between fermionic modes rather than the state of the fermionic modes themselves. It is referred to in the literature as the Superfast Bravyi–Kitaev (SFBK) or alternatively, Loop-Stabilized Bravyi–Kitaev (LSBK) [293] (we use SFBK as it is more generally used in the literature).

The overall advantage of this method is that, similar to the auxiliary qubit scheme, the Pauli weight of the qubit operators corresponding to the fermionic operators does not depend on the number of fermionic modes in the Hamiltonian but instead depends on the degree of the interaction graph. The number of qubits required depends on the total number of edges which is a function of the total number of fermionic modes and the degree of the interaction graph. This makes Superfast Encoding more suitable for lattice based models. We first present the SFBK mapping, as theorized in [280], and further developed in [293,295,358], and we then turn to extensions developed based on this encoding [257,260,261,360].

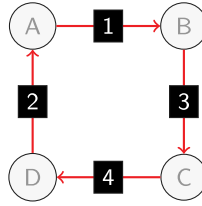


Fig. 5. A graph corresponding to the SFBK encoding defined from the H_2 Hamiltonian in a minimal basis set: fermions in the physical system of interest correspond to the vertices (gray circles) of the graph; qubits (the black rectangles) in the quantum computer are associated to the edges (red arrows) on the graph; the vertices and edges the graph are used to define the edge and vertex fermionic operators $\hat{A}_{jk}$ and $\hat{B}_j$, which in turn correspond to edge and vertex qubit operators $\hat{A}_{jk}$ and $\hat{B}_j$.

Background and definition of SFBK: Given the fermionic Hamiltonian an interaction graph can be constructed. The set of edges, E for the interaction graph correspond to the excitation, number excitation and double excitation operators (e.g. Fig. 5). In SFBK, the qubits are placed on the edges of the interaction graph, and the Fock space of the fermionic modes is associated with a codespace of a stabilizer code defined on the qubits on the edges. The qubit indices are defined based on two fermionic indices where a connection exists (e.g. (jk)). To define the codespace and the fermionic operators in terms of the qubit operators acting on qubits on the edges of the interaction graph, it is helpful to work with the Majorana modes defined in terms of Fermionic modes as in Eq. (59). These operators are Hermitian and satisfy

$$\hat{\gamma}_j \hat{\gamma}_k + \hat{\gamma}_k \hat{\gamma}_j = 2\delta_{jk} \quad (66)$$

Using Majorana modes, the following vertex operator and edge operator can be defined as

$$\hat{B}_j = -i\hat{\gamma}_j \hat{\gamma}_{j+1}, \quad \hat{A}_{jk} = -i\hat{\gamma}_j \hat{\gamma}_k. \quad (67)$$

The vertex operator $\hat{B}_j$ only acts on fermionic mode j , while the edge operator $\hat{A}_{jk}$ connects fermionic mode j to mode k . These operators satisfy the following algebraic relations:

$$[\hat{B}_k, \hat{B}_l] = 0, \quad (68)$$

$$\hat{A}_{(jk)} \hat{B}_l = (-1)^{\delta_{(jl)} + \delta_{(kl)}} \hat{B}_l \hat{A}_{(jk)}, \quad (69)$$

$$\hat{A}_{(jk)} \hat{A}_{(ls)} = (-1)^{\delta_{(jl)} + \delta_{(js)} + \delta_{(kl)} + \delta_{(ks)}} \hat{A}_{(ls)} \hat{A}_{(jk)}, \quad (70)$$

and the loop condition:

$$(i)^p \hat{A}_{(j_0 j_1)} \hat{A}_{(j_1 j_2)} \dots \hat{A}_{(j_{p-1} j_p)} \hat{A}_{(j_p j_0)} = \mathbb{1}, \quad (71)$$

with $j_0 \dots j_p$ any closed loop over the vertex indices.

It can be shown that the vertex and the edge operators satisfying the above relations generate the algebra of the physical operators. This means that any physical Hamiltonian with even fermionic modes can be represented in terms of these edge and vertex operators. The fermionic creation and annihilation operators can be converted into Majorana modes which can then be converted to edge and vertex operators. Table 6 gives the edge and vertex operators expressions for various Hamiltonian terms (with computation initially presented in [295]). Using the algebraic relations for the edge and vertex operators we can define the qubit operators for the edge and vertex operators. As we will see through the construction, it is also possible to define the codespace for the simulation.

Starting from the vertex operator, we note that $\hat{B}_k$ obeys an additional constraint on the excitation parity when considering an even number of fermions in the system, for a total of V vertices, the fermion parity operator is

$$\prod_{k \in V} \hat{B}_k = \mathbb{1} \quad (72)$$

If the number of particles is instead odd, one can always split the Hamiltonian in even and odd sectors, the odd sector can be simulated by changing the sign of the mapping as described below in Eq. (73) [360]. From Eq. (72) and the rule defined in Eq. (68), one can easily create a corresponding definition for the qubit version vertex operator $\tilde{B}_k$:

$$\tilde{B}_k = \bigotimes_{j \in n(k)} Z_{(kj)}, \quad (73)$$

where $n(j)$ is the set of vertices connected to mode j by an edge, and the pair (kj) indices a qubit. This definition meets the constraint of Eq. (72) as each edge is connected to exactly two vertices. We note that the Pauli weight for these operators is capped at the degree of the Hamiltonian graph.

For the edge operator $\hat{A}_{(jk)}$, we start by defining an ordering of the vertices and related direction of edges, such that $\epsilon_{jk} = 1$ when $j > k$, and $\epsilon_{kj} = -1$. Looking first at the rule of Eq. (69), in the case where $l \neq j$ and $l \neq k$ operators $\hat{A}_{(jk)}$ and

Table 6

Molecular Hamiltonian operators in second quantized form and in the corresponding vertex and edge operator form used in the Superfast Bravyi–Kitaev encoding.

Operator	Second quantized form	Vertex and edge operator form
Number operator	$h_{pp} \hat{a}_p^\dagger \hat{a}_p$	$\frac{1-\hat{B}_p}{2}$
Coulomb/exchange operators	$h_{pqqp} \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_q \hat{a}_p$	$h_{pqqp} \frac{(1-\hat{B}_p)(1-\hat{B}_q)}{4}$
Excitation operator	$h_{pq}(\hat{a}_p^\dagger \hat{a}_q + \hat{a}_q^\dagger \hat{a}_p)$	$-h_{pq} \frac{i}{2}(\hat{A}_{pq} \hat{B}_q + \hat{B}_p \hat{A}_{pq})$
Number-excitation operator	$h_{pqqr}(\hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_q \hat{a}_r + \hat{a}_r^\dagger \hat{a}_q^\dagger \hat{a}_q \hat{a}_p)$	$-h_{pqqr} \frac{i}{4}(\hat{A}_{pr} \hat{B}_r + \hat{B}_p \hat{A}_{pr})(1 - B_q)$
Double excitation operator	$h_{pqrs}(\hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_r \hat{a}_s + \hat{a}_s^\dagger \hat{a}_q^\dagger \hat{a}_r \hat{a}_p)$	$\frac{h_{pqrs}}{8} \hat{A}_{pq} \hat{A}_{rs} (-1 - \hat{B}_p \hat{B}_q + \hat{B}_p \hat{B}_r + \hat{B}_p \hat{B}_s + \hat{B}_q \hat{B}_s - \hat{B}_r \hat{B}_s - \hat{B}_q - \hat{B}_p \hat{B}_r \hat{B}_s)$

$\hat{B}_l$ must commute, and hence given the definition of $\tilde{B}_l$ from Eq. (73) all qubit operators representing edges adjacent to j or k (except for (jk) itself) must be either $\mathbb{1}$ or Z , as l could be one of the connected vertices. If $l = j$ or k , then $\hat{A}_{(jk)}$ and $\hat{B}_l$ must anticommute. Following the same reasoning as above, the qubit operator on the edge connecting l to the other vertex (j or k) must anticommute with Z : it must be either X or Y .

The last commutation rule that must be met is Eq. (70), which states that two edge operators anticommute if they share a single vertex (they commute otherwise). All these conditions, together can be met by the following qubit operator as defined in [280,293,295]:

$$\tilde{A}_{(jk)} = \epsilon_{jk} X_{(jk)} \bigotimes_{l < k}^{n(j)} Z_{(lj)} \bigotimes_{s < j}^{n(k)} Z_{(sk)}, \quad (74)$$

where the Z operators on vertices either connected and directed into j or from k , and the X operator on vertex (jk) enforce the rule of Eq. (69), and the phase factor ϵ_{jk} enforce the rule of Eq. (70).

The qubit operator representation of vertex and edge operators $\tilde{B}_j$ and $\tilde{A}_{(jk)}$ do not satisfy the loop condition defined in Eq. (71). And this condition lets us define the codespace. Using the qubit operator representation of the edge operator, $\tilde{A}_{jk}$, we can define a set of stabilizers corresponding to the loops in the graph and define the codespace to be the subspace where the loop condition defined in Eq. (71) holds. The loop operator is defined as follows:

$$\tilde{A}(\eta) \equiv (i)^p \hat{A}_{(\eta_0 \eta_1)} \hat{A}_{(\eta_1 \eta_2)} \cdots \hat{A}_{(\eta_{p-1} \eta_p)} \hat{A}_{(\eta_p \eta_0)}, \quad (75)$$

with η defining a loop of length p on the Hamiltonian graph. This definition of the loop operator generates an Abelian group of all the stabilizers (all necessary properties of this group are explored and demonstrated in [257]). The number of independent loops s in the graph is equal to the number of edges, minus the number of vertices plus one, or recalling N the number of qubits, and n the number of fermionic mode: $s = N - n + 1$. Therefore the code space defined by the stabilizer group encodes $N - s = n - 1$ logical qubits, into N physical qubits. This code space effectively restricts operators to the even parity subspace of the fermionic Fock space [257,280]. Once this restriction is applied to $\tilde{B}_j$ and $\hat{A}_{(jk)}$, one can defined an encoded qubit Hamiltonian.

The total number of qubits required for this encoding is proportional to the number of edges in the Hamiltonian graph. For regular lattice models, the number of edges itself is a direct multiple of the graph degree and of the number of sites. Since the Pauli weight of vertex and edge operators scales as $O(d)$, where d is the degree of the interaction graph, the Pauli weight of each term in the transformed Hamiltonian also scales as $O(d)$.

SFBK was applied to the 2-dimensional Hubbard model in Ref. [293] and was shown to be applicable to *ab initio* molecular systems in [295]. In both cases, it is clear that for a Hamiltonian graph of degree d , and a number of fermionic modes n , the Pauli weight of SFBK scales $O(2d)$ and the number of qubits required scales $O(nd/2)$.

Generalized superfast encoding: The Abelian stabilizer group defined using the loop operators given in Eq. (75) defines the codespace for the simulation. Since the edge and vertex operators commute with the stabilizer group, a state initialized within the codespace remains in codespace through the action of edge or vertex operators. Any operation that moves the state out of the codespace is not valid and can be considered as an error. The motivation for the Generalized Superfast Encoding was to come up with a modified version of SFBK that can detect all the errors. The number of qubits required for GSE is the same as SFBK, but the Pauli weight of qubit operator representation of edge and vertex operator is lower and scales as $O(\log(d))$.

In contrast to SFBK, where qubits are placed on the edges, GSE places them on vertices which consequently modifies the operators used for the encoding. For each fermionic mode j (which are assumed to be even in GSE), one must use $d^{(j)}/2$ qubits, where $d^{(j)}$ is the degree of the vertex corresponding to the mode j . Hence, it has the same scaling of the number of qubits as for SFBK: $O(nd/2)$.

In the GSE, a vertex j with $d/2$ qubits encodes d Majorana modes $\hat{\gamma}j, 1, \hat{\gamma}j, 2, \dots, \hat{\gamma}j, d$. A procedure to construct these Majorana modes can be found in [295]. The vertex and edge operator for qubits can be reformulated for the GSE as follows:

$$\tilde{B}_j = (-i)^{d^{(j)}/2} \hat{\gamma}j, 1 \hat{\gamma}j, 2 \dots \hat{\gamma}j, d^{(j)}, \quad (76)$$

$$\tilde{A}_{(jk)} = \epsilon_{jk} \hat{\gamma}j, p \hat{\gamma}k, q, \quad (77)$$

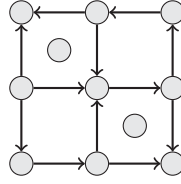


Fig. 6. Example of qubit placement and edge orientation in the compact mapping. Each gray dot represents a qubit, only those connected by edges correspond to sites, the others correspond to odd lattice faces.

where p is the p th local Majorana mode on the site j , and q is the q th local Majorana mode on the site k . These operators can also be restricted to the even-parity subspace, using the same stabilizers definition based on the loop operator of Eq. (75) [257]. It can be shown using Fenwick tree encoding [293] (see Appendix A) that a Majorana mode $\hat{\gamma}_j, p$ can be encoded into Pauli operators of weight $\log_2 d^{(j)}$. This means that operators $A_{(jk)}$ of Eq. (77) have a maximum Pauli weight of $2 \log_2 d$, while operators B_j (Eq. (76)) have a Pauli weight of 1 as it requires a single Z string [293]. This shows a key advantage of GSE over SFBK.

As mentioned above, it is shown in [257] that GSE additionally allows use of the code space defined as part of the encoding as a means of performing error correction. This property is applicable from $d^{(j)} \geq 6$ and catches single-qubit errors. So far, the error threshold tolerance of this error mitigation method has not been estimated.

Setia et al. [257] also showed that SFBK cannot provide single-qubit error correction for Hamiltonian graph of degree $d^{(j)} \leq 6$. The Majorana Loop Stabilizer Code (MLSC) was introduced in [360] as a means to allow addressing single-qubit errors on 2D square lattices (hence with vertex degree equal to 4) while preserving a local encoding. This encoding is still dependent on further research to be extended to higher-dimensional systems. These ideas have been further generalized in [361].

Compact mappings: The so-called compact mapping [260] is another extension of SFBK which does not concern itself error correcting properties, but focuses on minimizing the Pauli weight while reducing the number of qubits required in the methods previously mentioned. Derby and Klassen [260] present application of the method to square and hexagonal lattices, and extend it in Ref. [261] to uniform lattices of degree less than 4 and cubic lattices.

In the compact encoding, one must first assign a qubit to each vertex j . In a 2D lattice, each square of four vertices (referred to as a face), is defined as even or odd in a checkerboard pattern. Edges also need to be given an orientation. The orientation recommended in Ref. [261] is to set the orientation anticlockwise for even faces of the lattice, orientation on odd faces follows from completion. A schematic is presented in Fig. 6.

From this allocation of qubits, we can already see that the number of qubits for this encoding is capped to 1.5 times the number of sites (or number of fermionic mode, n). We first define the edge operators: for a given edge (i, j) , oriented i to j , at most one of the adjacent faces is odd, in which case we index the related qubit with (i, j) . We have for a downward edge

$$\tilde{A}_{(ij)} = X_i Y_j X_{(i,j)}, \quad (78)$$

for an upward edge:

$$\tilde{A}_{(ij)} = -X_i Y_j X_{(i,j)}, \quad (79)$$

and for a horizontal edge:

$$\tilde{A}_{(ij)} = X_i Y_j Y_{(i,j)}. \quad (80)$$

When the edge is on a boundary of the lattice with no odd face adjacent, there is no qubit (i, j) , and therefore the relevant Pauli operator can be omitted.

Vertex operators are defined as

$$\tilde{B}_j = Z_j. \quad (81)$$

It is easy to verify that these operators meet the conditions set for SFBK in Eqs. (68), (69), and (70). As for SFBK, the loop condition in Eq. (71) must also be met. The key feature of the compact mapping is that it does so in a way that allow to avoid the parity condition set by Eq. (72), and with it the need to have a Pauli weight equal to the degree of the graph.

Considering first the stabilizers for odd faces, defining a, b, c, d as the four vertices (for instance consider the first face in Fig. 6), we have

$$\begin{aligned} \tilde{A}_{(ab)} &= (X_b Y_a Y_{(a,b)}) \\ \tilde{A}_{(bc)} &= (X_b Y_c X_{(b,c)}) \\ \tilde{A}_{(cd)} &= (X_d Y_c Y_{(c,d)}) \end{aligned} \quad (82)$$

$$\begin{aligned}
\tilde{A}_{(dc)} &= (-X_d Y_a X_{(d,a)}) \\
\tilde{A}_{(ab)} \tilde{A}_{(bc)} \tilde{A}_{(cd)} \tilde{A}_{(dc)} &= -(Y_{(a,b)} X_{(a,b)}) (Y_{(a,b)} X_{(a,b)}) \\
&= -(-iZ_{(a,b)}) (-iZ_{(a,b)}) \\
&= 1
\end{aligned}$$

as of course, the qubit with index (a, b) , is the same as with index (b, c) , (c, d) , and (d, a) .

For even faces however, the stabilizers are non-trivial and therefore impose a constraint on the operators. Construction of these constraints are detailed in [260,261]. Extensions of this method to other uniform lattices of degree up to 4 and cubic lattices can be found in [261]. In addition, it was shown in [362] that the compact encoding can also be used to detect a large proportion of single-qubit errors.

4.2.3. Discussion on lattice tailored encodings

There are three main metrics that are worth discussing when comparing the different lattice tailored mapping: the number of qubits they require, their Pauli weights, and their capacity to mitigate or even correct errors resulting from quantum noise.

The number of qubits required appears to be a trade-off for the two other features mentioned above. All mappings from this section considered, the compact encoding from Ref. [260] offer the most advantageous combination of low number of qubits and low Pauli weights for systems of dimensions up to 3. Extension of the compact encoding to higher dimensions requires further research. In the meantime, the GSE [257] provides the best performance in terms of Pauli weight but does require that the number of qubits increases linearly with the degree d of the graph (recalling $d = 2D$) and the number of sites n .

While we found no comprehensive studies on the matter, these would very likely result in the most compact ansatz. However, as noted in [259] when comparing the Jordan–Wigner and Bravyi–Kitaev mappings, lower gate depth does not necessarily translate to an overall higher resilience to quantum noise.

An overall question that remains pending regarding all the lattice mappings considered is their efficacy at realizing error mitigation or error correction. For example, while the GSE demonstrates some error correcting properties, it is still unclear what quantum noise threshold is tolerable for these properties to be useful. In particular, such assessment is required to fully compare these encoding like-for-like. For example, the compact encoding could require additional quantum resources from error mitigation techniques (see Section 8) to achieve an accuracy comparable to what can be achieved when relying on the error correcting properties of other mappings such as SFBK or GSE.

This point also extends to applications of methods to *ab initio* molecular systems. Ref. [295] shows that SFBK results in lower gate depth than Jordan–Wigner, but higher than Bravyi–Kitaev on a small molecular system. Chien et al. [363] also show that Jordan–Wigner can outperform SFBK in terms of gate count under certain conditions on the Hamiltonian studied. One can expect GSE to perform better due to its lower Pauli weight, however for a relevant comparison, one would need to incorporate equivalent error mitigation techniques into Jordan–Wigner or Bravyi–Kitaev to balance quantum resources required to achieve a given accuracy. Overall, the question of the real impact of the error mitigating and correcting properties of lattice tailored encodings could be critical in determining their future applicability.

4.3. Reducing qubit requirements

Several methods have focused on reducing the overall number of qubits required to encode a specific Hamiltonian [282–284,336,337]. These methods usually come at limited additional costs and can therefore be used to significantly improve the efficiency of VQE. Bravyi et al. [282] formulate several proposals designed to reduce (‘taper off’) the number of qubits required to simulate fermionic systems for variational quantum algorithms. In particular, they find that it is possible to reduce qubit requirements in encodings where $\mathbb{Z}_2$ symmetries are present. This results in halving the dimension of the Hilbert space of the system considered for each qubit removed. The starting point of the proposal in Ref. [282] is the observation that in some encodings, some qubits are representing a conserved quantity of the molecular system, and should therefore never change. One example is the last qubit in the parity mapping which encodes the parity of the entire wavefunction and is directly dependent on the electron number.

By definition, a symmetry is an operator which leaves the Hamiltonian invariant when acting on it (i.e. it must commute with the Hamiltonian) [364]. Hence one can always find a common eigenbasis for the operator corresponding to the symmetry and the Hamiltonian. As the qubit Hamiltonian is written as a weighted sum of Pauli strings, if all Pauli strings commute with an operator, it is a symmetry operator. As such, given any initial encoding, one can find a certain number of operators, the symmetry operators τ_i , that commute with each of the Pauli strings in the Hamiltonian and between themselves. This can be done by finding the kernel of the check matrix that corresponds to the Pauli terms [282]. Although only symmetries that can be expressed as tensor products of Pauli operators can be found by this procedure. We can then define a change of basis, represented by a unitary operator U such that for each symmetry τ_i we have

$$U \tau_i U^\dagger = Z_{q(i)}. \quad (83)$$

In this basis the symmetry operator acts as a Z operator on the $q(i)$ th qubit. The matrix corresponding to U can be found as the product $U = \prod_i U_i$, where each unitary U_i is defined as

$$U_i = \frac{Z_{q(i)} + \tau_i}{\sqrt{2}} \quad (84)$$

In a similar fashion to what we have done before, we can remove each $q(i)$ th qubit and replace it in each Pauli term in the Hamiltonian by its eigenvalue. Setia et al. [283] show that symmetry operators can be identified using molecular point group symmetries [365] for the case of quantum chemistry simulations. One can remove the qubits corresponding to each member of an Abelian (commutative) subgroup of the molecular point group describing the symmetries of the molecule under consideration. It is shown in Ref. [283] that the number of qubits that can be removed is either higher or the same as the method proposed in Ref. [282]. For instance, CO_2 on 30 qubits, C_2H_2 on 24 qubits, BeH_2 on 14 qubits, can all be reduced by 5 qubits using the method presented in Ref. [283]. It is worth noting that a similar approach is briefly outlined in [366], along with a proposed method to identify point-group symmetries.

The Contextual Subspace VQE (CSVQE) [284], proposes to separate the expectation value of the Hamiltonian into two contributions: a contextual part, computed using VQE, and a non-contextual part, computed using a conventional computer. A set of Pauli strings observables is considered non-contextual if it is possible to measure and assign value to them simultaneously without contradiction [284,367–373]. While this technique can introduce an approximation in the energy computed, it does so with a significant reduction in the number of qubits, and can also be applied after using the qubit reduction technique presented above for further efficiency gains [284].

5. Efficient grouping and measuring strategies

One of the key challenges possibly holding back the VQE is the very large amount of samples that are required to accurately compute the relevant values of the algorithm. There are two main aspects to manage for efficiently sampling these expectation values: the number of terms in the Hamiltonian cost functions (computed using the mappings presented in Section 4), and the number of shots required to sample an expectation value at a certain level of accuracy. It is worth noting that the level of accuracy required changes throughout the optimization process. This could be because of the accuracy needed for computation of the final output of the VQE, but also most importantly because optimizer gradients (or optimizer steps in gradient-free methods) must be estimated precisely enough to be distinguished from one another when the optimization landscape flattens. An extreme case of the latter is the barren plateau problem [86], see Section 6.1. We first discuss strategies for optimal sampling error reduction (Section 5.1), and then we discuss methods used to reduce the number of Pauli strings that are necessary to separately measure (Section 5.2).

5.1. Scaling of shot numbers in VQE

In this section, we discuss methods that have been proposed to reduce the impact of the denominator in Eq. (86) on the total number of measurements required for each energy evaluation in the VQE, and in particular for each estimation within a gradient estimation. Namely how to minimize the number of measurements required to achieve a given level of precision, established by a target standard error on the measurement: ϵ .

5.1.1. Overall scaling of measurements

To get an idea of the scaling in sampling requirements, let us first consider the scaling of the output from the mapping methods presented in the previous section (Section 4). As seen previously, generalized mappings for molecular Hamiltonians result in $\mathcal{P} \sim \mathcal{O}(n^4)$ distinct Pauli strings to estimate. For mappings tailored to lattice models, this scales with the number of edges in the lattice (as an illustration, for a regular square lattice Hubbard model of dimension D , the number of edges scales $\mathcal{O}(nD)$).

With this in mind, let us consider the number of shots required to achieve a given precision. In any sampling experiment, the standard error is equal to $\epsilon = \sigma/\sqrt{S}$, where σ is the population standard deviation, and S is the experimental sample size, in our case, the number of shots also noted S (for a general introduction to statistical theory, we recommend [374]). This means that the number of times an experiment needs to be repeated to achieve a given expected error ϵ goes as $\mathcal{O}(1/\epsilon^2)$. More specifically, when measurements are distributed optimally among the different Pauli strings, such that the variance is minimized with respect to a given precision ϵ , the number of measurements required is upper-bounded by

$$S \leq \left(\frac{\sum_a^{\mathcal{P}} w_a}{\epsilon} \right)^2, \quad (85)$$

where w_a are the weights of the Pauli strings in the Hamiltonian [82,120].

As a result, for a given level of accuracy for each Pauli string measured independently, the overall scaling of the number of shots required for an energy estimation is:

$$\mathcal{O}\left(\frac{N^4}{\epsilon^2}\right). \quad (86)$$

In the context of quantum chemistry, successful computing methods are expected to produce results within a precision of $\epsilon = 1.6 \text{ mE}_\text{H}$ [181] to the target. When results obtained numerically are within this level of precision to experimental results, the simulations is deemed to reach chemical accuracy. This metric can be used as a bound for target precision in the VQE context (for an excellent discussion of precision vs. accuracy in this context, we direct readers to Eflving et al. [83]). One should be cautious however not to assume too much of a relationship between this number and the number of shots required to perform VQE. That is because the key bottleneck of VQE optimization is not the estimation of the wavefunction itself but the estimation of gradients and in particular the difference between these gradients (which allow for the optimization step to be performed reliably). This difference may be orders of magnitude smaller than the chemical precision threshold in a barren plateau, requiring that many more measurements [86], see Section 6.1. While polynomial in scaling, it has been pointed out on several occasions that the number of shots required to accurately compute a VQE optimization process rapidly becomes unmanageable [82,83,85], suggesting the method might be unable to compete with its conventional computing counterparts [83]. Ref. [82] estimates (though conservatively) that simulating the energy of Fe_2S_2 ferredoxin (in STO-3G basis with $N = 112$ spin-orbitals) using a VQE would require a total of $O(10^{19})$ shots based on the upper bound defined in Eq. (85). As such, a significant amount of effort has been devoted to finding solutions that reduce the pre-factor for the number of shots required.

5.1.2. Measurement weighting

Uniform distribution of measurements: The variance of Pauli strings can be computed easily as they are self-inverse. In particular:

$$\begin{aligned}\text{Var}[\hat{P}_a] &= \langle \psi | \hat{P}_a^2 | \psi \rangle - \langle \psi | \hat{P}_a | \psi \rangle^2 \\ &= 1 - \langle \psi | \hat{P}_a | \psi \rangle^2 \leq 1.\end{aligned}\quad (87)$$

Measurements of different Pauli operators (when not grouped) are independent and therefore uncorrelated, resulting in mean squared error for the energy estimate given a total of $\mathcal{P}$ Pauli strings of [120]:

$$\begin{aligned}\epsilon &= \sqrt{\sum_a^{\mathcal{P}} \frac{w_a^2 \text{Var}[\hat{P}_a]}{S_a}} \\ &= \sqrt{\sum_a^{\mathcal{P}} w_a^2 \left(\frac{1 - \langle \psi | \hat{P}_a | \psi \rangle^2}{S_a} \right)},\end{aligned}\quad (88)$$

where S_a is the number of shots used to measure the expectation value of each Pauli string $\hat{P}_a$, which has weight w_a , and such that $S = \sum_a S_a$. Assuming uniformly distributed shots across all Pauli strings, we can rearrange this result to showcase the number of measurements required to achieve a target standard error [121]:

$$S = \mathcal{P} \sum_a \frac{w_a^2 \text{Var}[\hat{P}_a]}{\epsilon^2}.\quad (89)$$

This particular distribution of shots could be considered optimal in the special case where $\sqrt{\text{Var}[\hat{P}_a]} \propto 1/|w_a|$ [120,121]. This is in general not the case and therefore further methods have been developed to distribute measurements to optimally reduce estimation variance [121]. The methods outlined below aim to reduce S , the total number of shots, while maintaining a given precision ϵ .

Weighted distribution of measurements: An alternative to uniform distribution of shots is to focus on measuring more precisely the operators which contribute most to the total variance of the expectation value estimated. With a given shot budget, one can improve the overall precision of measurement by distributing these shots towards specific operators.

A straightforward manner to distribute these shots is to simply weight them with respect to the Pauli strings weights ($|w_a|$) in the Hamiltonian [82]. When looking at Eq. (89), we can easily see that reducing the number of shots on strings contributing less to the total energy estimate (with lower $|w_a|$ value) and adding these to strings that contribute the most reduces total variance, as long as $\text{Var}[\hat{P}_a]$ are similar for all a [120]. Rubin et al. [120] indeed show that $S_a \propto |w_a| \sqrt{\text{Var}[\hat{P}_a]}$, is optimal (although $\text{Var}[\hat{P}_a]$ might not be easily accessible), while Arrasmith et al. [121] show numerically that when considering random states, variations in $|w_a|$ tend to be higher than variation in $\text{Var}[\hat{P}_a]$, resulting in the weight pro-rata distribution (where $S_a \propto |w_a|$) of measurements outperforming the uniform distribution in most cases. To address cases in which the number of shots is limited (i.e. there are so few shots, that each one could create a bias in the energy estimate), it is proposed in Ref. [121] to perform measurements on Pauli strings randomly, with probabilities proportional to $|w_a| \sqrt{\text{Var}[\hat{P}_a]}$, thereby allowing unbiased estimates even with a low total shot number.

Rubin et al. [120] further use fermionic marginals, and N -representability constraints to determine optimized measurements distributions for Hamiltonian estimates. The idea of optimizing shots distributions was also merged to optimization strategy by designing specific optimizers aiming to balance their optimization-per-shot cost [121,375].

Term truncation: Another approach to consider is to remove from measurement scope terms that have contributions significantly below the error tolerance threshold ϵ [38]. This method has been shown to significantly reduce the cost of quantum chemistry calculations with negligible impact on accuracy [232].

To implement this, one must observe that the contribution of any Pauli observable to the final energy estimate is bounded by the absolute value of its associated weight: $|\langle \psi | w_a \hat{P}_a | \psi \rangle| \leq |w_a|$. By ordering these contributions in ascending order, one can construct a partial sum of the $k \leq \mathcal{P}$ smallest contributor:

$$e_k = \sum_a^k |w_a|. \quad (90)$$

From there, one can choose a constant $C \in [0, 1]$, and include in the partial sum the terms up to an index k that verify: $e_k < C\epsilon$. This method [38] introduces a bias in total energy estimation, and as such the key to implement successfully is to pick a constant C such that the truncation bias is lower than the mean square error reduction from measurements added to the remaining terms. McClean et al. [38] present an adjusted estimate for the number of shots required to achieve ϵ , to be contrasted with Eq. (89):

$$S = (\mathcal{P} - k) \sum_{a=k+1}^{\mathcal{P}} \frac{w_a^2 \text{Var}[\hat{P}_a]}{(1 - C^2)\epsilon^2}. \quad (91)$$

If the expected number of shots is lower than without term truncation (Eq. (89)), then the method provides an improvement regarding the precision to measurement cost ratio.

5.2. Pauli string groupings, and other joint measurement strategies

The methods described below work with the idea that measuring a given Pauli string a :

$$\begin{aligned} \hat{P}_a &= \sigma_{p_1^{(a)}} \otimes \sigma_{p_2^{(a)}} \otimes \cdots \otimes \sigma_{p_{N-1}^{(a)}} \otimes \sigma_{p_N^{(a)}} \\ p_i^{(k)} &\in \{I, X, Y, Z\}, \end{aligned} \quad (92)$$

provides information about other Pauli strings which have overlapping Pauli elements (i.e. $p_j^{(a)} = p_j^{(b)}$) (for instance [38]). In essence, all of these methods target the same information gathering optimization. As such, there are a number of incompatibilities between them. In particular, General Commutativity [123,285,376], Unitary [288,377] and decomposed interactions [124] based grouping are aimed at diagonalizing a set of Pauli strings (i.e. rotate them so that $\forall i, p_j^{(a)} \in \{I, Z\}$), they exhaust all of the information that can be gained from inference methods (e.g. [125,126]). It is not the case however for Qubit-Wise Commutation based grouping [38,120,122,123,157,263–267] which can also be used with inference methods.

5.2.1. Inference methods

Inference methods all seek to recover the expectation values of an observable from a restricted set of operators, rather than the complete basis in which it can be measured. They are extensions built upon the theory of Quantum State Tomography (for some literature on the topic, we recommend: [378–384]) aiming at achieving a target precision ϵ with a minimum number of shots S using the structure of the Lie algebra in which Pauli strings are defined [385].

Methods for low Pauli weight Hamiltonians: Restricting the problem of tomography to the estimation of a Hamiltonian can however significantly reduce the cost of measurements. For instance, taking into account the Pauli weight of the Hamiltonian can help reduce the sampling requirements. It is worth noting that both methods described in this paragraph are initially designed for Hamiltonian characterization rather than ground state estimation, and as such may not be directly optimized for use within the context of the VQE.

One such method is Quantum Overlapping Tomography (QOT) [386] which aims at efficiently estimating all k -body operators for a given quantum state (a k -body qubit operator is computed in a manner that is similar to the one- and two-body RDM presented in Section 3.1.3, but replacing the fermionic operators with computational basis elements), which in turn allows computing the expectation of an observable having up to k Pauli weight. They observe that complete tomography on a k -qubit state grows exponentially in k [382,384] (namely 3^k), and that there are $\binom{N}{k}$ k -body reduced density operators to measure for an N qubit state, bringing the cost of a naive measurement of all w -body operators to $\sim e^{\mathcal{O}(w)} \binom{N}{w}$. However, one can also use the fact that many of these operators can be overlapped and measured in parallel on the same qubit register. Cotler et al. [386] show that this can result in a significant reduction of the number of shots required to estimate all k -qubit reduced density matrices within a precision of ϵ , to:

$$S \sim k e^{\mathcal{O}(k)} \left(\frac{\log(N)}{\epsilon} \right)^2, \quad (93)$$

with the detail process on how to optimally allocate this measurements being described in Ref. [386]. An interesting research question would be to study whether this method lends itself well to *ab initio* molecular systems, and lattice

models with high connectivity – this is because QOT does not necessarily take into account pre-existing structures of the Hamiltonian, for which in principle $\mathcal{O}(n^4)$ and $\mathcal{O}(nD)$ operators are normally needed to be estimated respectively, instead of the 3^k for each k -body reduced density matrix.

Another such method that is worth mentioning briefly is Bayesian Hamiltonian Learning [387] which can infer the value of $\mathcal{P}$ Pauli strings with high success probability using $\mathcal{O}(3^k \log(\mathcal{P})/\epsilon^2)$ parallel shots.

Shadow tomography, classical shadow, and locally-biased classical shadow: The concept of shadow tomography was initially presented in [388,389]. It describes the task of predicting certain properties of a quantum state without conducting full tomography on the state. In particular, it was shown that an exponential number of target functions (for instance, computing the expectation value of an operator) can be predicted from a polynomial number of shots. It was highlighted however in Ref. [126] that this method requires exponential depth in the quantum circuit and access to quantum memory, thereby rendering it on balance too costly for NISQ algorithms such as VQE.

The concept of classical shadow [126] is an attempt to extend the idea of shadow tomography and address some of these caveats. It aims at efficiently learning a classical sketch $\tilde{\rho}$ of an unknown quantum state ρ , that is then used to predict arbitrary linear functions of that state (for instance, the expectation value of an operator: $\langle \hat{O} \rangle = \text{Tr}[\hat{O}\rho]$) using a median-of-means protocol [390,391]. The classical shadow is beneficial if it can be constructed with a tractable number of measurements and such that $\mathbb{E}[f(\tilde{\rho})] = f(\rho)$, with f any of the aforementioned functions. Applied to the case of the expectation value of an observable, measurement of the classical shadow produces a random variable whose expectation value must match the expectation value of the observable with respect to the quantum state:

$$\mathbb{E}[\langle \tilde{O} \rangle] = \mathbb{E}[\text{Tr}(O\tilde{\rho})] = \text{Tr}[O\rho]. \quad (94)$$

To construct a classical shadow, one must first produce an instance of the state ρ and apply a random unitary U taken from an ensemble $\mathcal{U}$ (defined by its elements and a probability rule for picking each element) to rotate it before performing a measurement. The measurement returns a N -bit measurement vector $|\tilde{b}\rangle \in \{0, 1\}^N$. The conjugate of the unitary can then be applied back to measurement vector, producing $U^\dagger |\tilde{b}\rangle \langle \tilde{b}| U$.

From this point, we can define the linear map $\mathcal{M}$ (or measurement channel), which transforms ρ to the expectation value of $U^\dagger |\tilde{b}\rangle \langle \tilde{b}| U$ over both any unitary in $\mathcal{U}$ and all measurement outcome possible $|\tilde{b}\rangle$. We have:

$$\begin{aligned} \mathcal{M}(\rho) &= \mathbb{E}_{U \in \mathcal{U}} \sum_{b \in \{0,1\}^N} \langle b|U\rho U^\dagger |b\rangle U^\dagger |b\rangle \langle b|U \\ &= \mathbb{E} \left[U^\dagger |\tilde{b}\rangle \langle \tilde{b}| U \right], \end{aligned} \quad (95)$$

where Born's rule was used: $\text{Pr}[\tilde{b} = b] = \langle b|U\rho U^\dagger |b\rangle$. The linear map admits a unique inverse (this requires Tomographic completeness, please refer to the Supplementary materials of [126] for details) which can be applied through conventional post-processing, producing a classical shadow, such that:

$$\tilde{\rho}_i = \mathcal{M}^{-1} \left(U_i^\dagger |\tilde{b}_i\rangle \langle \tilde{b}_i| U_i \right). \quad (96)$$

This procedure is repeated T times, to produce an T -sized classical shadow array: $S(\rho, T) = \{\tilde{\rho}_1, \dots, \tilde{\rho}_T\}$. Huang et al. [126] report that the use of classical shadow allows scaling measurements required for the desired precision logarithmically in the number of operators to measure. This can be combined with measurements weighting methods [392] to avoid exponential scaling of required classical shadow in case of non-local observables. Low weight derandomization strategy, presented in Ref. [393] (and independently Ref. [394]) expand on the ideas of shadow tomography and classical shadows. It removes the need to draw unitaries (rotating the measurement basis) randomly from a pool, and the need for median-of-means predictions, by progressively and deterministically selecting Pauli strings that have the highest impact on narrowing a given confidence bound. An adaptive process to reduce computational cost of the low-weight derandomization strategy was also proposed by Hadfield [395].

Two questions central to implementing classical shadows are the choice of the unitary ensemble $\mathcal{U}$ to draw from, and the construction of the inverse linear map $\mathcal{M}^{-1}$. Clifford ensembles were initially proposed [126], but alternative ensembles have also been discussed, such as unitaries corresponding to time evolution of a random Hamiltonian [396], unitary ensembles defined through locally scrambled quantum dynamics [397] achieving a lower tomography complexity compared to Clifford based methods, and in the case of fermionic states, a discrete group of fermionic Gaussian unitaries [268,398]. Bu et al. [399] propose to use Pauli-invariant unitary ensembles (unitary ensembles that are invariant under multiplication by a Pauli operator), a class which includes both Clifford ensembles and locally scrambled unitary ensembles. They also provide an explicit formula for the inverse linear map corresponding to these ensembles.

Classical shadows can be applied to study several aspects of quantum states, in particular for energy estimation as part of a VQE [268,392,395,400], and though it remains unclear how it can perform over other efficient measurement schemes numerical studies have already suggested superior empirical scaling over methods such as Basis Rotation Grouping [124] (described below). These methods have also been studied in the presence of noise and modifications to the scheme have been shown to render the method resilient to quantum noises [401,402]. Finally, classical shadows have been shown to be useful as a mean to mitigate barren plateaus [403].

Neural network tomography: Another type of method proposed to reduce the number of shots required to achieve a given precision level on the measurement of a Hamiltonian is to use Machine Learning on a series of shot outputs to decrease the variance of the expectation value. A first example is presented in Ref. [125], where Torlai et al. present a method to learn a mock of the state produced by an ansatz using an unsupervised restricted Boltzmann machine (RBM) [404] then used to compute expectation values of quantum observables. RBMs have been shown as successful models to represent quantum states in the field of condensed matter physics [405,406].

RBMs are in general composed of two layers of binary-valued neurons, a visible layer composed of i units equal to the input size, noted $\mathbf{v}$ (in this case, it would be N) and a hidden layer composed of j units, noted $\mathbf{h}$. The two layers are connected by a set of weights (a matrix noted $\mathbf{W}$, where entry W_{ij} connects unit v_i to unit h_j), and units within each layer are not connected (unlike in general Boltzmann machines). Each unit also has a bias weight $b_i^{(v)}$ and $b_j^{(h)}$ ($\mathbf{b}^{(v)}$ and $\mathbf{b}^{(h)}$ in vector notation), such that the energy of the RBM can be expressed as:

$$E(\mathbf{v}, \mathbf{h}) = -\mathbf{b}^{(v)T} \mathbf{v} - \mathbf{b}^{(h)T} \mathbf{h} - \mathbf{v}^T \mathbf{W} \mathbf{h}. \quad (97)$$

In the case of representing a many-body wavefunction, the network parameters $\lambda = \{\mathbf{b}^{(v)}, \mathbf{b}^{(h)}, \mathbf{W}\}$ are complex valued [220]. In addition, following [125], the energy of the RBM in Eq. (97) is modified to represent a wavefunction dependent on the network parameters and a binary vector $\mathbf{v}$ of size N :

$$\psi_{\lambda}(\mathbf{v}) = e^{\mathbf{b}^{(v)T} \mathbf{v}} e^{\sum_j \log \cosh \sum_i W_{ij} v_i + b_j^{(h)}} \quad (98)$$

The aim of the RBM is to train λ such that given an element of the computational basis $|\mathbf{v}\rangle$:

$$\psi_{\lambda}(\mathbf{v}) = \langle \mathbf{v} | \psi(\theta) \rangle, \quad (99)$$

where $|\psi(\theta)\rangle$ is the output state of the ansatz so that RBM approximates the probability distribution of measuring the output state. Otherwise said, the RBM is trained so that the output of the RBM energy function (Eq. (98)) is equal to the amplitude of the quantum circuit output state with respect to the basis element $|\mathbf{v}\rangle$.

To train this RBM, suppose you have a pool of single-qubit measurement basis operators $\mathbf{p} = \bigotimes_i p_i$, with $p_i = \{X, Y, Z\}$. For each $\mathbf{p}$, there exist a set of binary vectors $\{\mathbf{v}^{\mathbf{p}}\}$ corresponding to the possible measurement outcomes of $\mathbf{p}$ by state $|\psi(\theta)\rangle$. As such, following Born's rule we can define the probability of measuring each of $\mathbf{v}^{\mathbf{p}}$ as $P(\mathbf{v}^{\mathbf{p}}) = |\langle \mathbf{v}^{\mathbf{p}} | \psi(\theta) \rangle|^2$. We note $P_{\lambda}(\mathbf{v}^{\mathbf{p}}) = |\psi_{\lambda}(\mathbf{v}^{\mathbf{p}})|^2$, the amplitude square of the output of the RBM given $\mathbf{v}^{\mathbf{p}}$ as input and λ as parameters. Torlai et al. [125] suggest to use the extended Kullback–Leibler (KL) divergence as cost function for the RBM:

$$\begin{aligned} C_{\lambda} &= \sum_{\mathbf{p}} \sum_{\mathbf{v}^{\mathbf{p}}} P(\mathbf{v}^{\mathbf{p}}) \log \frac{P(\mathbf{v}^{\mathbf{p}})}{P_{\lambda}(\mathbf{v}^{\mathbf{p}})} \\ &\approx - \sum_{\mathbf{p}} \sum_{\mathbf{v}^{\mathbf{p}}} P(\mathbf{v}^{\mathbf{p}}) \log P_{\lambda}(\mathbf{v}^{\mathbf{p}}), \end{aligned} \quad (100)$$

where the numerator of the log has been discarded in the second equality as it does not depend on λ and therefore does not impact the optimization process. This formulation is of course intractable as the set $\{\mathbf{p}\}$ scales 3^N and each set $\{\mathbf{v}^{\mathbf{p}}\}$ scales 2^N . Ref. [125] bypasses these exponential sums by restricting $\{\mathbf{p}\}$ to the set of Pauli strings already included in the Hamiltonian decomposition, and restricting the measurement outputs $\{\mathbf{v}^{\mathbf{p}}\}$ to a set of finite size measurements $\mathcal{D}$ such that:

$$C_{\lambda} = - \sum_{\mathbf{v}^{\mathbf{p}} \in \mathcal{D}} P(\mathbf{v}^{\mathbf{p}}) \log P_{\lambda}(\mathbf{v}^{\mathbf{p}}) \quad (101)$$

5.2.2. Hamiltonian partitioning based on commutativity

The premise of this section, and the methods that follow, is that an Abelian group of Pauli strings can be simultaneously diagonalized through a unitary rotation of the measurement basis [118], thereby reducing the number of terms that need to be measured to accurately compute the expectation value of a Hamiltonian. This implies that single measurement values for all the terms in a given Abelian group can be inferred from a single *joint measurement* (a measurement that simultaneously assesses multiple Pauli operators) of the complete qubit register. This principle derives from the stabilizer theory (for a review of stabilizer theory, we recommend: Refs. [76,407]). First, we need to identify a set of generators of the Abelian group, $\{\tau_i\}$. From there, we know that there exists a unitary U , such that we can write:

$$U \tau_i U^{\dagger} = \sigma_z^{q(i)}, \quad (102)$$

where $q(i)$ maps generator index i to a unique address in the qubit register. This means that after applying U to a given quantum circuit, obtaining the expectation value of generator τ_i can be done by measuring the expectation value of $\sigma_z^{q(i)}$. Finding the unitary U for an Abelian group therefore allows to measure all of its generators by simply measuring σ_z on all qubits (the expectation values of all the Pauli terms within the group, which are not generators, can be also be recovered, from the data gathered from the generators measurements).

Qubit-wise commutativity (QWC) or Tensor Product Basis (TPB) groups: Two of Pauli strings are said to be QWC if each Pauli operator in the first commutes with the Pauli operator of the second one that has the same index. Generally speaking, that would be any group where, any given Pauli operator in any Pauli string has an index such that, all the operators of the same index across all the other Pauli strings in the group are either the same Pauli operator, or the identity (for example, XI , IZ , and XZ are altogether QWC). The term TPB refers to the same idea, all the Pauli strings in the group can be diagonalized simultaneously in a joint tensor product basis with no entanglement.

This basis for grouping terms has been widely used and studied [38,120,122,123,157,263–267]. It allows performing joint measurements more efficiently. In particular, Gokhale et al. [122] found that this method reduces the pre-factor for the number of Pauli terms to be measured by about three, without however changing its asymptotic scaling (see also Ref. [376]). It is worth noting that the process for partitioning for QWC is significantly cheaper computationally [122] than for the General Commutativity (GC) rule which we present later on.

Another key advantage is that the basis rotation used to conduct the joint measurements only requires a circuit of depth 1. To achieve this, we need to find the unitary U_{QWC} which rotates all the Pauli strings in a given QWC group into a basis in which they are all diagonalized. This is a straightforward process as any individual Pauli operator can be rotated in the Z basis with one single-qubit operation as follows:

$$\begin{aligned} Z &= \text{Ry}\left(-\frac{\pi}{2}\right) X \text{Ry}\left(-\frac{\pi}{2}\right)^\dagger \\ Z &= \text{Rx}\left(\frac{\pi}{2}\right) Y \text{Rx}\left(\frac{\pi}{2}\right)^\dagger \end{aligned} \quad (103)$$

This method of grouping and joint measurement is therefore relatively cheap to implement and allows for significant savings in the number of shots required to complete a VQE, although without changing the overall scaling.

General Commutativity (GC), or entangled measurements: This method uses the same logic as the QWC groupings, however rather than allowing for qubit-wise commutation groupings, it allows for grouping of any Pauli strings that generally commute. There has been a number of independent research producing similar results [123,285,376]. From here it is obvious that the set of QWC relationships is included in the set of GC relationships (if two Pauli strings Qubit-Wise commute, they necessarily generally commute, the reverse is not true).

Consider two Pauli strings ν and η , each composed of N Pauli/identity tensored operators (noted ν_i and η_i for $i \in [1, N]$). If two Pauli operators do not commute we have: $\nu_i \eta_i = -\eta_i \nu_i$, and $\nu_i \eta_i = \eta_i \nu_i$ if they commute. This implies that:

$$\nu \eta = (-1)^k \eta \nu, \quad (104)$$

where k is the number of index-wise Pauli operators that do not commute. As such, two Pauli strings generally commute if they comprise an even number k of index-wise operators that do not commute (for example, XYZ , XZY , and ZIX altogether are GC). The interesting aspect of this approach is that it reduces the scaling of the number of separate terms to be measured from $O(N^4)$ to $O(N^3)$ [285,376,408] in the case of *ab initio* molecular Hamiltonian, providing a significant advantage to VQE optimization.

The problem of building appropriate unitaries for joint measurements of the groups identified is more complicated than in the QWC case. The objective is the same: finding a unitary rotation that simultaneously diagonalizes all the elements of a given group. The measurement basis however is more complex than a TPB, and requires entangled measurements [123]. As for QWC, the qubit register is entirely measured in the Z basis and therefore the unitary performing the basis rotation itself must include non-local, entangling operations. This requires careful circuit design, which has been thoroughly explained in Ref. [122], and we strongly recommend readers that are interested in building their implementation to refer back to this article, or to the “CZ” construction (based on Ref. [409]) and “CNOT” construction (based on Refs. [410,411]) proposed in Ref. [286]. These two latter methods have the advantage of explicitly treating the case where the number of independent operators in a group, k , is strictly less than N (as we have in all cases $k \leq N$). It is worth noting that while not explicitly covered in Ref. [122], this case can also be addressed using the former method.

Gokhale et al. [122] also point out that the number of gates in the circuits scales $O(N^2)$ (with some gate parallelization possible, making this a worst-case for depth), while Ref. [376] shows a gate scaling of $O(N^2/\log(N))$. The methods presented in [286] have a number of two-qubit gates that in the worst-case scale:

$$\begin{aligned} u_{CZ}(k, N) &= kN - k(k+1)/2 \\ u_{CNOT}(k, N) &= O(kN/\log(k)). \end{aligned} \quad (105)$$

In all cases, it can be considered negligible compared to the scaling of most ansätze, but not all, as linearly scaling ansätze such as UpCCGSD [80] could see their cost become negligible compared to the required basis rotation (see Section 6). As such, the decision on whether this method should be used depends on the type of ansatz used and the quantum cost (circuit depth) one is willing to reduce the computing time (number of repeated measurements).

Simple example: As an example, consider the arbitrary set of Pauli strings covering 4 qubits presented in Table 7

From this set, one can apply one of the heuristics presented in Section 5.2.5 to decompose the set into commutative groups. As an example, one can identify the commutative group presented in Table 8.

Table 7

Example set of Pauli strings.

$IIXX, YXII, XYII, YXYY, YXXX, XYYY, XYXX, IIYY, XXYX, IIYX, YYYX, IIXY, XXII, YYII, XXXY, YYXY$
--

Table 8

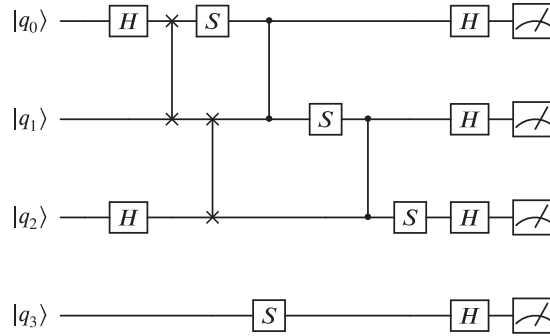
Example set of a commutative group extracted from Table 7.

$XXYX, IIYX, YYYX, IIXY, XXII, YYII, XXXY, YYXY$
--

Table 9

Example set of generators for the commutative from Table 8.

$XXYX, IIYX, YYYX, IIXY$

**Fig. 7.** Example quantum circuit for the realization of the relationships defined in Eq. (106), following the method presented in [122].

The next step is to identify a minimum set of generators of the multiplicative group defined – for instance, we can observe that $(XXYX) \times (IIYX) = XXII$. Ref. [122] suggests using Gaussian elimination to achieve this. There are of course more than one possibility. The set presented in Table 9 is an example of generators identified.

The final step is to construct the unitary U that maps each of the Pauli strings above to single-qubit measurements, such that:

$$\begin{aligned} U^\dagger (XXYX) U &= ZIII \\ U^\dagger (IIYX) U &= IZII \\ U^\dagger (YYYX) U &= IIZI \\ U^\dagger (IIXY) U &= IIIZ \end{aligned} \quad (106)$$

The quantum circuit realizing this unitary can be produced for example following the method presented in Ref. [122]. For the example above, an example of realization for this unitary is given in Fig. 7.

From this, we can see that half of the Pauli strings set presented in Table 7 can be jointly measured through the operator $ZZZZ$.

5.2.3. Unitary partitioning of the hamiltonian:

An alternative to the GC rule as a connection system for grouping Hamiltonian terms is to connect anti-commuting terms. In essence, this is the complement to the GC connections (if the GC relationships for Hamiltonian terms are defined in terms of a graph, the anti-commuting relationships are the complementary graph). It offers however a convenient property: it allows partitioning the Hamiltonian as a weighted sum of a minimum (or close to minimum) number of unitary operators [377].

To see this, consider that while Pauli strings are hermitian unitary, a weighted sum of Pauli strings generally is not. The conditions for a weighted sum to be unitary however can be recovered through anti-commuting grouping. In particular, for any operator expressed as a Pauli sum:

$$\hat{O} = \sum_a^{\mathcal{P}} w_a \hat{P}_a, \quad (107)$$

to be unitary, we must have:

$$\text{Im}(w_a^\dagger w_b) = 0, \quad (108)$$

$$\sum_a |w_a|^2 = 1, \quad (109)$$

and:

$$\{\hat{P}_a, \hat{P}_b\} = 2\delta_{ab}. \quad (110)$$

The first two conditions are easily met. Given w_a weights are always real, Eq. (108) is always verified in the VQE context. The normalization condition in Eq. (109) can easily be engineered by extracting a normalization factor w such that

$$w = \left(\sum_a |w_a|^2 \right)^{\frac{1}{2}}, \quad (111)$$

and

$$\hat{O} = w \sum_a \frac{w_a}{w} \hat{P}_a. \quad (112)$$

By selecting the Pauli strings in $\hat{O}$ such that they anti-commute, we can re-write the above equation as

$$\hat{O} = w\hat{U}, \quad (113)$$

with $\hat{U}$ a unitary operator. Therefore, by partitioning the Hamiltonian following anti-commuting operator it can be re-written as a weighted sum of $\mathcal{U}$ unitary operators as

$$\hat{H} = \sum_a^{\mathcal{U}} w_a \hat{U}_a. \quad (114)$$

This grouping premise has been shown to allow a linear reduction in the number of terms in the Hamiltonian, similar to what can be achieved with GC grouping [377,412,413].

The next step is to implement the joint measurement of all the elements of each unitary group identified. Unlike the groupings based on QWC and GC, the groups produced based on unitary partitioning can be directly implemented as quantum circuits. To see how this is done, one can follow the method in Ref. [377] starting by incorporating the partitioned Hamiltonian from Eq. (114) into the VQE optimization problem (Eq. (4)):

$$E_{\text{VQE}} = \min_{\theta} \sum_a^{\mathcal{U}} w_a \langle \psi(\theta) | \hat{U}_a | \psi(\theta) \rangle. \quad (115)$$

The unitary group cannot be directly measured jointly. However, as the observable is a unitary, one can transform Eq. (115) into an overlap measurement problem (rather than an observable measurement problem). Using the fact that a unitary observable has the same expectation value as its complex conjugate, we have

$$\begin{aligned} E_{\text{VQE}} &= \frac{1}{2} \sum_a^{\mathcal{U}} w_a \langle \psi(\theta) | \hat{U}_a | \psi(\theta) \rangle + \langle \psi(\theta) | \hat{U}_a^\dagger | \psi(\theta) \rangle \\ &= \frac{1}{2} \sum_a^{\mathcal{U}} w_a (\langle \mathbf{0} | \Psi_a \rangle + \langle \Psi_a | \mathbf{0} \rangle) \\ &= \sum_a^{\mathcal{U}} w_a \text{Re}(\langle \mathbf{0} | \Psi_a \rangle), \end{aligned} \quad (116)$$

where $|\mathbf{0}\rangle$ is the initial qubit register (which can be replaced with the Hartree–Fock wavefunction) and we have set $|\Psi_a\rangle = U_a^\dagger U_a U_\theta |\mathbf{0}\rangle$ (and of course $|\psi(\theta)\rangle = U_\theta |\mathbf{0}\rangle$). The real part of the overlap in Eq. (116) can be computed using a Hadamard test (as presented in Appendix B).

In terms of the cost of the circuit add-on, one should first note that the ansatz used to prepare the trial state is used twice therefore at least doubling circuit depth. In addition, one must take into account the depth required to implement U_a . As a sum of Pauli operators, it can be implemented using $2(\tilde{q} - 1)$ CNOT gates per operator, where q is the Pauli weight of the operator. Hence, U_a takes a total of $2L(\tilde{q} - 1)$, where L is the number of terms in U_a (and as such, on average $L \approx N$ given there are $\mathcal{O}(N^3)$ groups). This gives a total CNOT scaling of $\mathcal{O}(\tilde{q}N^2)$ [377].

Ref. [412] shows that unitary groups can also be jointly measured in a manner similar to commutative groups by appending the unitary at the end of the ansatz which rotates the unitary group into a diagonal basis. Their method requires entangling gate scaling of $\mathcal{O}(qL)$, with q the maximum Pauli weight in the Hamiltonian, and L the number of terms in each unitary group. As discussed before, this implies a scaling of $\mathcal{O}(N^2)$ using Jordan–Wigner, and $\mathcal{O}(N \log(N))$ using Bravyi–Kitaev. Zhang et al. [412] also show that the circuit depth can be reduced further using ancilla qubits. These methods of partitioning the Hamiltonian into unitary groups have been tested in an implementation on a quantum computer by Ralli et al. [413].

5.2.4. Decomposed interactions for efficient joint measurements:

Basis rotation grouping: The method proposed in Ref. [124], referred to as "Basis Rotation Grouping" is based on a tensor decomposition of the two-body operator. It details how to significantly reduce the overall number of (joint) terms to measure in the Hamiltonian, down to a linear number with system size. This same decomposition has also been used to reduce the total gate depth of the full UCCSD Ansatz as well as Trotter steps in Ref. [414] and is based on a two-stage decomposition of the interaction tensor originally proposed in Ref. [415]. It also provides a large improvement in the noise resilience of mappings with high Pauli weight such as Jordan–Wigner (see Section 4). The price to pay for this is that the measurement has to take place in a different basis for each term, necessitating an additional $\mathcal{O}(N)$ gate depth before measurement to implement this orbital rotation for each grouped term of this decomposed Hamiltonian (which remains much less than $\mathcal{O}(N^2)$ required for general commutative grouping).

The first step in this method is to re-write the second-quantized Hamiltonian into a factorized form through decomposition of the two-electron integral tensor [124,414,416]. Starting from Eq. (32) in the basis of spin-orbitals, we can rewrite the two-electron part as

$$V = \frac{1}{2} \sum_{pqrs=1}^n h_{pqrs} \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_r \hat{a}_s \quad (117)$$

$$= \frac{1}{2} \sum_{pqrs=1}^n h_{ps,qr} (\hat{a}_p^\dagger \hat{a}_s \hat{a}_q^\dagger \hat{a}_r - \hat{a}_p^\dagger \hat{a}_r \delta_{qs}) = V' + S, \quad (118)$$

where S is an additional one-body operator, and $h_{ps,qr} = (ps|qr)$ is a representation of the two-body integrals in the conventional 'chemists' notation. The positive-definite super-operator $h_{ps,qr}$ can be decomposed in a low-rank spectral decomposition:

$$V' = \sum_{pqrs} \sum_{l=1}^L v_{ps}^{(l)} v_{qr}^{(l)} \hat{a}_p^\dagger \hat{a}_s \hat{a}_q^\dagger \hat{a}_r. \quad (119)$$

This decomposition has a long history in quantum chemistry, and this low-rank factorized form can be directly constructed, avoiding an explicit diagonalization (which would scale as $\mathcal{O}(N^6)$), through techniques such as density fitting or Cholesky decomposition [415,417–421] (which is generally accepted to scale $\mathcal{O}(N^3)$). It has long been known that the decomposed form is not of full rank, with a number of terms $L = \mathcal{O}(N)$, which is a sufficient description of the system in the case of arbitrary basis quantum chemistry [417]. It is worth noting that in special representations where the Coulomb operator is diagonal, as demonstrated for the plane wave basis and dual basis in Ref. [235], this can be rigorously $L = 1$.

Further eigendecomposition of the resulting matrix for each value of l is possible, with

$$v_{ps}^{(l)} \hat{a}_p^\dagger \hat{a}_s = \sum_{i=1}^{\rho_l} U_{pi}^{(l)} \lambda_i^{(l)} U_{si}^{(l)} \hat{a}_p^\dagger \hat{a}_s, \quad (120)$$

where U denote single-particle unitary operators. These can be combined with a decomposition of the one-body part into a final doubly factorized Hamiltonian form of

$$\hat{H} = U^{(0)} \left(\sum_i g_i \hat{n}_i \right) (U^{(0)})^\dagger + \sum_{l=1}^L U^{(l)} \left(\sum_{ij} g_{ij}^{(l)} \hat{n}_i \hat{n}_j \right) (U^{(l)})^\dagger, \quad (121)$$

where $\hat{n}_i = \hat{a}_i^\dagger \hat{a}_i$ and $g_{ij}^{(l)} = \frac{\lambda_i^{(l)} \lambda_j^{(l)}}{2}$ are scalars constructed from absorbing all relevant weights. The single-particle unitaries $U^{(0)}$ and $U^{(l)}$ implement the orbital basis change and can be applied to the prepared state before measurement. The key is that all $\hat{n}_i$ and $\hat{n}_i \hat{n}_j$ for a given l commute, and thus can be measured simultaneously, resulting in $L \sim \mathcal{O}(N)$ separate terms to estimate, but additionally requiring the change in measurement basis for each term increasing the gate depth by $\mathcal{O}(N)$. This method was tested in Ref. [85] and showed clear superiority of basis rotation method compared to QWC, achieving a significant scaling reduction in the number of measurements required.

Full rank optimization (FRO): Yen and Izmaylov [262] propose a more general method for decomposition of the two body operator. Starting from V , as defined in Eq. (117), one can write

$$\begin{aligned} V &= \frac{1}{2} \sum_{\alpha=1}^L U_\alpha^\dagger \left[\sum_{ij} \lambda_{ij}^{(2,\alpha)} \hat{a}_i^\dagger \hat{a}_j^\dagger \hat{a}_i \hat{a}_j \right] U_\alpha \\ &= \frac{1}{2} \sum_{\alpha=1}^L \sum_{ij} \lambda_{ij}^{(2,\alpha)} [U_\alpha^\dagger \hat{a}_i^\dagger \hat{a}_i U_\alpha] [U_\alpha^\dagger \hat{a}_j^\dagger \hat{a}_j U_\alpha], \end{aligned} \quad (122)$$

with λ a tensor which must be discovered. The transformation unitary can be written as:

$$U_\alpha = \exp \left[\sum_{i>j}^n -i\zeta_{ij}^{(\alpha)}(\hat{a}_i^\dagger \hat{a}_j + \hat{a}_j^\dagger \hat{a}_i) + \eta_{ij}^{(\alpha)}(\hat{a}_i^\dagger \hat{a}_j - \hat{a}_j^\dagger \hat{a}_i) \right], \quad (123)$$

where ζ and η are tensors which must be discovered. The process for estimating tensors λ , ζ , and η involves re-writing Eq. (122) to express the two-body coefficients h_{pqrs} as function of these tensors. The system of equation can then be resolved using a difference minimization (for which the Broyden–Fletcher–Goldfarb–Shanno, BFGS, optimizer is suggested [422–425]), and is described in further details in Ref. [262].

Two variants of this method are also proposed in Ref. [262]. The Greedy FRO (GFRO) discovers the tensors mentioned above iteratively by starting from $\alpha = 1$. The variance-estimate GFRO (VGfro) also takes into account the variance of the operators for the discovery of the tensors. Overall, Yen and Izmaylov find that FRO achieves a much lower number of partitions than the Basis rotation group method [124] (less than half, on six different systems ranging from 4 to 20 qubits), though FRO and VGfro result in significantly more partitions (between 5 and 7 times more). The comparative analysis is however pushed further to take into account the risk for co-variances resulting from joint measurement of operators [38] (more details below), and it is found [262] that General Commutativity (GC) with Sorted Insertion [286] (presented below) performs the best in most systems. VGfro, GFRO, and the basis rotation methods perform similarly, and far better than FRO. Yen and Izmaylov [262] conjecture that the relative performance of VGfro, GFRO and basis rotation methods will improve compared to GC with Sorted Insertion as the size of the system studied increases.

5.2.5. Grouping heuristics

We have seen that defining whether two Pauli strings are QWC, GC, or anti-commuting is straightforward. However, given neither QWC, GC or anti-commutation properties are transitive (if two Pauli strings, A and B commute, and A commutes with a third one, C , B does not necessarily commute C), finding optimal groups of Pauli strings that together all meet one of these criteria out of the $\mathcal{O}(N^4)$ terms in the Hamiltonian can be a challenging process.

Before detailing grouping methods directly dedicated to Pauli strings, we briefly outline conventional heuristics that have been used in the context of grouping Pauli strings and the studies that have been conducted to compare these.

Conventional grouping heuristics used in VQE context: The problem of grouping Pauli terms that have been connected following a specific rule (QWC, GC, or AC), can be straightforwardly mapped to a graph problem. Namely, it can be mapped to the Minimum Clique Cover (MCC) problem [123,267,286,376,377,408,412] (or equivalently the graph coloring problem [426]), which aims at finding the minimum number of fully connected subgraphs in an initial input graph. In this case, we can define a graph $G(V, E)$, where V , the vertices are representing the Pauli strings, and E , the edges, are representing the connections established using one of the rules defined above. It is in general NP-hard [427] and therefore should be solved using heuristics. The problem of grouping Pauli strings can therefore be addressed using the same heuristics as for MCC like problems:

- **Largest Degree First Coloring (LDfC) Algorithm:** In this algorithm, edges represent anti-commuting relationships. It works by first assigning a color to the vertex V with the highest degree (colors are represented by integers, starting with 1). Following this step, the vertex among those remaining with the highest degree is assigned the lowest color that is not already attributed to one of its neighbors. The process is repeated iteratively until all colors have been assigned [428] (Used for Pauli grouping in [123]).
- **Smallest first:** Identical to the LDfC except for the ordering of colors allocation, which starts from the vertex with the smallest degree [429].
- **DSatur:** The degree of saturation of a vertex is defined as the number of different colors it is adjacent to. With that in mind, the DSatur algorithm functions broadly like the LDfC algorithm, albeit by attributing colors along an ordering of the degree of saturation of the remaining uncolored vertices. The first color is attributed based on the largest degree [430].
- **Independent-operator sorting algorithm:** Vallury et al. [431] propose to group Pauli strings, using a QWC relationship, by ranking them according to the number of identity operators in each string. The method starts with the Pauli string having the lowest number of identity operators. Following strings are iteratively sorted: if a string QWC with all elements of an existing TBP group, it is added to this group, otherwise a new group is created.
- **Others of note include:** Dutton and Brigham, [432], COSINE [433], Ramsey [434,435], and Connected Sequential d.f.s [436], Recursive largest first [437].

For a thorough review of graph coloring methods, we recommend [436].

All the heuristics mentioned above, except the BKT algorithm, are polynomial in scaling with respect to the number of graph vertices [377]. The number of vertices being equal to the number of Pauli strings in the Hamiltonian these heuristics, in general, scale $\mathcal{O}(N^{k^4})$, with $k \geq 2$ an integer corresponding to the respective scaling of each method. For instance, LDfC scales quadratically in the number of vertices in the graph [436], therefore the method's time complexity is $\mathcal{O}(N^8)$ for a graph built from the second quantized Hamiltonian.

An alternative has been proposed in [285], where the problem is treated as finding a Maximal Flow in a Network Flow Graph. Gokhale et al. propose to identify the commuting relationships at the fermionic operator level, basing themselves on spin orbital indices of the fermionic terms. This is only treated under the Jordan–Wigner mapping. They show that they can create partitions of size N , thereby reducing the total number of terms to measure to N^3 , using the Baranyai construction approach which has a computational cost of $\mathcal{O}(N^5 \log N) \leq \mathcal{O}(N^{k4})$. They also note that the groupings can be re-used across multiple Hamiltonian of the same sizes, which does not necessarily occur when grouping at the spin-operator level.

Several papers have drawn comparisons in the ability of these different heuristics to approach the MCC in the context of Pauli strings grouping. In particular, [377] shows that RLF and DB tend to result in the lowest number of cliques for Unitary based grouping, on systems up to 14 qubits. [267] shows similar results for GC based grouping, with Largest First and SL also performing equivalently (on systems up to 14 qubits for all methods, and up to 36 for Largest first). [286] compares a lower number of methods (and also performs analysis on the co-variance implications, which is discussed in the next paragraph) and also finds that LDFC finds a lower number of cliques against Connected d.f.s, DSATUR, and Sorted insertion.

Sorted insertion, heuristic dedicated to Pauli strings and VQE applications: The method presented in Ref. [286] stems from the observation that grouping Pauli strings can result in measurement co-variances, thereby increasing the number of measurements required to achieve the desired precision [38]. It targets grouping based on optimization of the number of measurements required rather than on the number of terms to measure. Sorted Insertion works by allocating Pauli strings to commuting groups in descending order of their absolute weights in the Hamiltonian. The complexity of implementing this grouping heuristic is capped to $\mathcal{O}(N\mathcal{P}^2)$, with $\mathcal{P}$ the number of terms in the Hamiltonian [286] – hence in the general second quantized Hamiltonian case $\mathcal{O}(N^9)$.

In addition, Crawford et al. [286] show that breaking commuting groups into smaller groups cannot reduce the variance under an optimal measurement strategy. They also define a useful figure of merit for a grouping strategy: the ratio of the minimum number of measurements required to achieve a desired precision (see Eq. (89)) in the cases where Hamiltonian terms are not grouped, over the cases in which they are. They propose an approximate version of this metric which can be computed analytically from the Hamiltonian and the grouping. Given a Hamiltonian $\hat{H}$, which can be decomposed into k operators $\hat{h}_i = \sum_j w_{ij} P_{ij}$ where all P_{ij} for a given i commute, the figure of merit is given by

$$\tilde{R} := \left(\frac{\sum_i^k \sum_j |w_{ij}|}{\sum_i^k \sqrt{\sum_j |w_{ij}|^2}} \right)^2. \quad (124)$$

It is shown that Sorted Insertion achieves a significantly higher $\tilde{R}$ score than LDFC, DSatur, Connected Sequential d.f.s. and Independent Set on a number of molecular systems. Yen and Izmaylov [262] also showed that GC grouping using Sorted Insertion achieves the highest reduction in the number of shots required for a given precision for several systems up to 16 qubits (one system of 20 qubits is tested in Ref. [262], and for which decomposed interactions methods [124,262] perform better, see 5.2.4).

5.3. Discussion on measurement strategies and grouping methods

The definition of the ‘best possible grouping method’ is not straightforward. While it is clear that aiming for the lowest number of groups possible is advantageous, it is not the only metric to take into consideration. In particular, it was shown in [38] that grouping terms, both under GC and QWC based grouping, suffers from co-variances arising from the joint measurement (thereby changing the formula presented in Eq. (88)). This covariance effects increase the sampling noise and as such the total number of measurements required at a given level of precision. Joint measurements under the Unitary grouping or the decomposed interaction methods suffer from the same issue [262]. Therefore, the total number of measurements required to achieve a given precision should be taken into consideration as figure of merit for a grouping strategy [286]. Another cost to consider is the additional quantum noise resulting from the circuit used to rotate the measurement basis which could be significant in both the case of Unitary grouping and GC grouping. Further resources may be required to mitigate these additional errors (Section 8). Finally, an important point to note is that the scaling of most grouping heuristics could end up being somewhat prohibitive (for instance, we recall that LDFC scales $\mathcal{O}(N^8)$) for large systems. A possible way to dampen this issue is to use grouping methods which can be re-used across different Hamiltonians of the same active space sizes, by applying heuristics at the fermionic operator level [122], or by relying on two-body reduced density matrices [159].

Inference methods could perform better for each of the costs listed above but could also face their own pitfalls, in particular when an additional machine learning model requires training. A thorough numerical analysis of the multiple methods that have been proposed would be an interesting avenue for future research. In the meantime, we consider that the decomposed interactions [124,262] methods have the most supporting arguments for the treatment of molecular Hamiltonians. While it is shown numerically in Ref. [262] that GC with Sorted Insertion [286] tends to be most efficient with respect to the number of measurements required to achieve a given level of precision, decomposed interactions

methods perform almost at the same level (and it appears that the gap in performance narrows as the system increases in size [262]), and requires significantly less depth to perform the required unitary transformations ($\mathcal{O}(N)$ against $\mathcal{O}(N^2)$ for GC). It is worth noting that the basis rotation group method [124] currently has a much more predictable implementation cost than FRO and its extensions. The latter requires solving a minimization problem to perform the decomposition, which could come at a significant computational cost or loss in accuracy, though further research will be required to investigate this point.

6. Ansatz selection and construction

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 maximize 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 minimizes the expectation value of the Hamiltonian). The span of possible states an ansatz can reach is referred to as its expressibility. However, performing a variational optimization on an ansatz with high expressibility could easily become intractable due to the number of parameters required to allow reaching many these different states, the number of iterations required for convergence, or the number of shots required to achieve sufficient gradient accuracy to continue the optimization. Whether an ansatz can be optimized 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 [127,128,438]. 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 [438], 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. [439] 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 $\mathbb{U} \subseteq \mathcal{U}(N)$ [87,127,128,438]. The following super operator, representing the second order difference between the Haar measure on $\mathcal{U}(N)$ and the uniformity distribution of $\mathbb{U}$ can be constructed (we follow the formalism in [127]):

$$\mathcal{A}_{\mathbb{U}}(\cdot) := \int_{\mathcal{U}(n)} d_{\mu}(V) V^{\otimes 2}(\cdot)(V^{\dagger})^{\otimes 2} - \int_{\mathbb{U}} dU U^{\otimes 2}(\cdot)(U^{\dagger})^{\otimes 2}, \quad (125)$$

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 $\mathcal{A}_{\mathbb{U}}(\hat{O}) \rightarrow 0$, 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

$$\varepsilon_{\mathbb{U}}^{\rho} := \|\mathcal{A}_{\mathbb{U}}(\rho^{\otimes 2})\|_2 \quad (126)$$

$$\varepsilon_{\mathbb{U}}^{\hat{P}} := \|\mathcal{A}_{\mathbb{U}}(\hat{P}^{\otimes 2})\|_2. \quad (127)$$

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 parameterized quantum circuits more generally [128]. In addition, several methods have been proposed to remove redundant parameters from quantum circuits without decreasing expressibility [440] or reducing the set of states that can be generated through the circuit [441].

Trainability: The trainability of an ansatz refers to the ability to find the best set of parameters of the ansatz by (iteratively) optimizing the ansatz with respect to expectation values of the Hamiltonian in a tractable time [87,127]. 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. In this section, we first provide a review of the barren plateau problem [86], which is the main known obstacle to ansatz trainability, and discuss its implication for the VQE in Section 6.1. We then describe of the most relevant fixed-structure ansatz for the VQE in Section 6.2, followed by a description of the most relevant adaptive structure ansatz in Section 6.3. Finally, we provide a discussion regarding ansatz selection for the VQE in Section 6.4. For alternative recent reviews of ansatz selection for the VQE, we recommend Refs. [109,110].

6.1. 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 initialization [86]. This refers to the risk of the cost function gradients vanishing exponentially as a function of specific properties of the optimization for a problem. McClean *et al.* [86] provide the first formal characterization of this barren plateau problem (some early numerical evidence of this problem are outlined in Ref. [82], without a characterization being provided), and show that cost function gradients are vanishing exponentially in the number of qubits in the quantum register when provided with random initialization 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 [86]:

- 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 (see Eq. (86)). 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 optimization into a random walk, overall requiring an exponential number of shots to continue optimization.
- 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 [127], degree of entanglement of the wavefunction [89,90], non-locality of the wavefunction [87,91,92], or quantum noise [442].

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 characterized in [87]). 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 these two problems are equivalent [443]. 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 [86], rendering intractable optimization away from the mean.

In the context of the VQE, the barren plateau problem can be formally characterized as follows. Consider a VQE optimization problem with cost function:

$$E(\theta) = \langle \psi(\theta) | \hat{H} | \psi(\theta) \rangle, \quad (128)$$

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

$$\Pr(|\partial_{\theta_i} E(\theta)| \geq \epsilon) \leq \mathcal{O}\left(\frac{1}{b^N}\right), \quad (129)$$

which is an immediate consequence of Chebyshev’s inequality and the result from above (for the expectation value and variance) [87]. 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 optimizers [89,444] (e.g. COBYLA, Powel, Nelder–Mead, RotoSolver, see Section 7). It is easy to understand, as gradient-free optimizers 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 optimization step.

6.1.1. Drivers of the barren plateau problem

System size and random initialization [86]: 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 initialized randomly. A layered ansatz for a random parameterized quantum circuit can be described as [86]

$$U(\theta) = \prod_{l=1}^L U_l(\theta_l) \mathcal{W}_l, \quad (130)$$

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-parameterized unitary. The cost function is as described in Eq. (128), taking $|\psi(\theta)\rangle = U(\theta)|0\rangle$. The gradient of this cost function with respect to any given parameter θ_i can be conveniently computed as

$$\partial_{\theta_i} E = i \langle 0 | U_{1 \rightarrow (i-1)}^\dagger \left[\hat{V}_i, U_{i \rightarrow L}^\dagger \hat{H} U_{i \rightarrow L} \right] U_{1 \rightarrow (i-1)} | 0 \rangle. \quad (131)$$

where $1 \rightarrow (i-1)$ represent the ansatz layers from layer index 1 to layer index $(i-1)$, and $i \rightarrow L$ represent the ansatz layers from layer index i to layer index L . From the computation of the gradient, McClean et al. [86] show that if both $U_{1 \rightarrow (i-1)}$ and $U_{i \rightarrow L}$ are 2-designs, the variance of the gradient is clearly vanishing exponentially in the system size:

$$\text{Var}[\partial_{\theta_i} E] \approx \frac{1}{2^{(3N-1)}} \text{Tr}[\hat{H}^2] \text{Tr}[\rho^2] \text{Tr}[\hat{V}^2] \quad (132)$$

Cases in which either of $U_{1 \rightarrow (i-1)}$ or $U_{i \rightarrow L}$ is not a 2-design are also addressed in Ref. [86], with similar outcomes (we 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. [445] shows additional analytical bounds for unstructured variational ansätze, moving away from the layered ansatz described above.

Expressibility [127]: 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 initialization, as above, or a very non-local cost function [87,91,92]). This observation implies that one cannot lower-bound gradients as a function of expressibility, but it can be upper-bounded.

This is shown in Ref. [127] by extending the expression of the barren plateau problem as explained in Ref. [86], 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. (130)) and resulting gradients (Eq. (131)). As an illustration, Ref. [127] finds a generalized bound for gradient variance as a function of expressibility (we encourage the reader to read the finer details directly from the source material), as

$$\text{Var}[\partial_{\theta_i}, E] \leq \frac{g(\rho, \hat{P}, U)}{2^{2N} - 1} + f(\varepsilon_L^{\hat{P}}, \varepsilon_R^{\rho}), \quad (133)$$

where the first half of the bound corresponds to the maximally expressive ansatz (and $g(\cdot)$ is a function defining a pre-factor on this expressibility, defined in detail in Appendix E of Ref. [127]). The function $f(\cdot)$ is the extra bound resulting from the expressibility (or lack thereof) of the ansatz, defined as

$$f(\varepsilon_x, \varepsilon_y) := 4\varepsilon_x \varepsilon_y + \frac{2^{N+2}(\varepsilon_x \|\hat{P}\|_2^2 + \varepsilon_y \|\rho\|_2^2)}{2^{2N} - 1}, \quad (134)$$

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 $N \rightarrow \infty$. As a result of the definitions in Eqs. (125)–(127), 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. [128] provide a study of the expressibility of the shallow alternating layer ansatz.

Cost function non-locality: [87,91,92]: In Ref. [87], Cerezo et al. show that an ansatz trained on local cost functions is 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 $\hat{O}_L = \mathbb{1} - \frac{1}{N} \sum_i |0\rangle\langle 0|_i \otimes \mathbb{1}_{\neq i}$, 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 never trainable (this is not necessarily true on other types of ansätze), ansätze trained on the local cost function are trainable if their depth scales logarithmically with the circuit width (i.e. $\mathcal{O}(\log(N))$ or below). Cerezo et al. [87] also show that ansätze with a scaling $\mathcal{O}(\text{poly}(\log(N)))$ could also be either trainable or not. Ref. [91] extends these findings to a wider range of cost functions, and Ref. [92] 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 [87] is that local encoding such as Bravyi–Kitaev [279–281] (Section 4.1.3), ternary tree encoding [256] (Section 4.1.4) or Generalized Superfast Encoding [257] (Section 4.2.2) with lower Pauli weight, would offer more resilience than encodings such as Jordan–Wigner [119] (Section 4.1.1) 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], Section 6.2.2, Fourier Transform-HVA [235] on some systems, Section 6.2.4). Uvarov et al. [258] compare 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. [258] 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. [148]) ultimately dominates and reduces gradient variance to a negligible number in either case.

Noise induced barren plateau (NIBP): Wang et al. [442] show that incorporating quantum noise in a variational optimization 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. [442]), as

$$\left| \frac{\partial \langle \hat{H}(\theta) \rangle}{\partial \theta_i} \right| \leq G(N) q^{L+1}, \quad (135)$$

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 [442]. First, the noise-induced barren plateau is independent of parameter initialization, or locality of the cost function, meaning that some of the strategies listed in Section 6.1.2 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 optimization 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 [89,90]: 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 [446].

6.1.2. Methods to address barren plateau problem

It follows from Refs. [86,127] 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 (Section 6.3) exhibit some resilience to the barren plateau problems. In addition, there are optimization methods which aim at tempering this problem (for example Ref. [132]) that are considered further in Section 7. Another means available to contain the barren plateau problem is to select a local encoding (see Section 4) with low Pauli weights, as discussed above and explained in Ref. [87].

Some methods have also been developed specifically to address barren plateaus. A first example consists of initializing 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 optimization process in a barren plateau [102]. Starting from the layered ansatz in Eq. (130), 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 $U_k(\theta_k)$, parameterized by a vector θ_k can then be split into two parts of equal depth, such that

$$U_k(\theta_k) = \prod_{d=1}^{D/2-1} U_d(\theta_{d,1}^k) W_d \prod_{d=D/2}^D U_d(\theta_{d,2}^k) W_d, \quad (136)$$

where $\theta_{d,1}^k$ can be initialized at random, but where $\theta_{d,2}^k$ are initialized 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 optimization, we have

$$U_k(\theta_k) = I_k, \quad \text{and} \quad U(\theta_{\text{init}}) = I. \quad (137)$$

Grant et al. [102] show that this parameter initialization strategy could slow down the optimization process of the VQE as the initial state produced by the circuit would have no entanglement. They propose to initialize the qubit register with a random entangled state, using a shallow random unitary which remains constant throughout the optimization process. While showcasing promising results on small systems, the method is however quite challenging to implement in practice. Identifying a block initialization is not directly possible for all ansatz structures: for example it is not straightforward for unitary coupled cluster (see Section 6.2.2) based ansatz without repeating some operators, and it is in general not possible exactly with hardware efficient ansätze.

Nonetheless, the idea of adjusting parameter initialization to improve resilience to the barren plateau problem has been extended to alternative mitigation techniques. In Ref. [447], Sauvage et al. propose to select optimal initial parameters with the help of a machine learning model (FLEXible Initializer for arbitrarily-sized Parameterized quantum circuits, or FLIP). The model is trained to identify structures of parameters that best suit specific families of quantum circuit optimization problems, and is numerically shown to provide significant improvements. Similarly, Kulshrestha and Safo show that initializing 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 [448]. They also show that adding perturbations to the parameters between each optimization step also helps in mitigating vanishing gradients.

Several additional methods have been developed for the barren plateau in the general case of parameterized quantum circuits. In Ref. [449], Volkoff et al. show that one can reduce the dimensionality of the parameter space by using spatially and temporally correlated parameterized quantum gates, resulting in higher resilience to barren plateaus. The ansatz can also be trained layer by layer to the same effect [450], though limitations of this method were shown in Ref. [451]. This is somewhat akin to adaptive ansätze (see Section 6.3), but generalized to any quantum neural network optimization. Patti et al. [90] also propose several additional mitigating methods including an alternative initialization strategy in which two qubit registers are initially not entangled, regularization on the entanglement, the addition of Langevin noise, or rotation into preferential cost function eigenbases. Sack et al. [403] showed that barren plateaus can be partially mitigated as part of a classical shadow measurement scheme. Wu et al. [452] 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. [453] show numerically that Quantum Convolutional Neural Networks exhibit natural resilience to the barren plateau problem. Similarly, Zhang et al. [288] show that Quantum Neural Networks with tree tensor structure and step-controlled architectures have gradients that vanish at most polynomially in the system size.

6.1.3. Comments on barren plateau in the context of the VQE

In Ref. [127], 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 initialization strategy such as the ones presented in Refs. [90,102,447] would allow an ansatz to begin optimization away from a barren plateau, and as such, away the target operator mean. One could argue that given the optimization problem aims at finding a minimum, a reliable optimizer should always move a state away from the mean expectation value and therefore away from the barren plateau regions. This point however could be invalidated by several aspects of the optimization process. These include the existence of local minima, the increase in entanglement of the trial wavefunction as the optimization progresses, or the presence of noise which results in vanishing of the value function amplitude.

Focusing on local encodings (see Section 4) has been shown to provide some resilience to barren plateaus [87,91,258], 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. [87], ansätze 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.

In the remainder of this section, we discuss options for ansätze in the context of VQE. We look at two different types of ansatz: fixed structure ansatz (Section 6.2) which are set at the beginning of the optimization process and remain unchanged afterward, and adaptive structure ansatz (Section 6.3) which are constructed iteratively as part of the optimization process. We can present three key metrics to compare ansätze:

- The depth of the ansatz (number of sequential operations required for the implementation), which impacts the overall runtime of the method, its resilience to noise and to barren plateaus.
- The number of parameters, which significantly influence the overall runtime of the implementation (although this cost can in theory, and in most cases, be entirely parallelized, see Section 2.7.2), and the complexity of the optimization process.
- The number of entangling gates, which is, in general, the main source of noise resulting from execution of a quantum circuit.

Of course, while these metrics can be objectively defined and provide a useful tool to compare different ansätze, they are only relevant if the ansätze studied can indeed provide an accurate description of the ground state wavefunction. This latter point is far more difficult to assess, beyond the formal considerations at the beginning of this chapter.

6.2. Fixed structure ansätze

Fixed structure ansätze are initialized at the beginning of the VQE process and remain unchanged (aside from the value of their parameters) throughout the optimization. Table 10 provides a comparative summary of these ansätze. Some fixed structure ansätze (such as the Qubit Coupled Cluster [454], and Qubit Coupled Cluster Singles and Doubles [455]) are only discussed in the adaptive ansätze section as their adaptive version tend to be more relevant for discussion.

Table 10

Summary of circuit depth, parameters and entangling gates scaling across most fixed structure ansätze reviewed. The scaling in the number of entangling gates assumes full connectivity of the qubit lattice.

Method	Depth	Parameters	Entangling gates	Comments
Hardware Efficient Ansatz (HEA) [263]	$\mathcal{O}(L)$	$\mathcal{O}(NL)$	$\mathcal{O}((N-1)L)$	L is an arbitrary number of layers, its scaling for exact ground state is unknown, exponential in the worst case (i.e. if the entire Hilbert space needs to be spanned to find the ground state)
UCCSD [37,456]	$\mathcal{O}((N-m)^2 m \tau)$	$\mathcal{O}((N-m)^2 m^2 \tau)$	$\mathcal{O}(2(\tilde{q}-1)N^4 \tau)$	$\tilde{q}$ is the average Pauli weight across the operators used to build the ansatz. As an indication, maximum Pauli weight under Jordan–Wigner is N , and $\log(N)$ under Bravyi–Kitaev. τ is the number of Trotter steps used
UCCGSD [37,80,82]	$\mathcal{O}(N^3 \tau)$	$\mathcal{O}(N^4 \tau)$	$\mathcal{O}(2(\tilde{q}-1)N^4 \tau)$	As above
k-UpCCGSD [80]	$\mathcal{O}(kN \tau)$	$\mathcal{O}(k\tau N^2/4)$	$\mathcal{O}(k\tau(\tilde{q}-1)N^2/2)$	k is an arbitrary constant which determines the accuracy of the result. Scaling is unknown. Rest is as above.
OO-UCCD [457]	$\mathcal{O}((N-m)^2 m \tau)$	$\mathcal{O}((N-m)^2 m^2 \tau)$	$\mathcal{O}(2(\tilde{q}-1)N^4 \tau)$	Same as UCCSD. It is worth noting that OO-UCCD is a nested loop between orbital (one-body terms) optimization, done on a conventional machine, and two-body terms optimization done on the quantum computer.
Symmetry preserving [148]	$\mathcal{O}((N-1)L)$	$\mathcal{O}(2(N-1)L)$	$\mathcal{O}(3(N-1)L)$	This ansatz spans a wider range of the Hilbert space than the EPS. It is therefore likely it requires more circuit resources and as such we suspect that L grows exponentially in N for exact resolution of the ground state. This logic also applies to HEA.
Efficient Symmetry Preserving (EPS) ansatz [218]	$\mathcal{O}(2^{\binom{N-1}{m}})$	$\mathcal{O}(2^{\binom{N}{m}} - 2)$	$\mathcal{O}(3^{\binom{N}{m}})$	Scaling range from linear when $m=1$, or $m=N-1$, to exponential if $m \sim N/2$
Hamiltonian Variational Ansatz [82,150]	$\mathcal{O}(\tilde{C}L)$	$\mathcal{O}(\tilde{C}L)$	$\mathcal{O}(2(\tilde{q}-1)CL)$	L represents the number of repetition of the ansatz required to achieve the desired accuracy. C is the number of terms in the Hamiltonian, and $\tilde{C}$ the number of commutative groups among these terms.

6.2.1. Hardware-efficient ansatz (HEA)

The initial motivation for the Hardware-Efficient ansatz (HEA) was for the trial state of VQE to be parameterized by quantum gates directly tailored to the quantum device on which the experiment is run [263]. Several flavors of the HEA have been proposed, however, they all follow the same logic: the ansatz is constructed by repeating blocks of interweaved single qubit, parameterized, rotation gates, and ladders of entangling gates (see Fig. 8 for an example). A generic representation of the HEA can be presented as follows:

$$|\psi(\theta)\rangle = \left[\prod_{i=1}^d U_{\text{rota}}(\theta_i) \times U_{\text{ent}} \right] \times U_{\text{rota}}(\theta_{d+1}) |\psi_{\text{init}}\rangle, \quad (138)$$

where $U_{\text{rota}}(\theta_i) = \prod_{q,p}^{N,p} R_p^q(\theta_i^{pq})$, with q the qubit addresses and $p \in \{X, Y, Z\}$, the selected set of parameterized Pauli rotations.

Both the set of rotation gates and entangling gates can vary depending on the native gate set of the device, and the complexity of the target state sought [458]. The ansatz' entangling section was originally presented as a quantum analogue task [263]. The main promise of HEA is that it can be flexibly tailored to the specific native gate set of the device used, while at the same time being highly expressive. Its versatility and ease of construction resulted in it being widely used for numerous small-scale quantum experiments [159,263,458–461].

There are however many known limitations of HEA. A first obvious issue is the fact that it must span a very large portion of the Hilbert space to guarantee that an accurate enough representation of the ground state wavefunction can be produced. HEA can therefore be quite inefficient with respect to its required Hilbert space coverage, requiring in worst cases an exponential depth (though further research would be required to verify how the accuracy of ground state representations would respond to arbitrary depth reduction and shrinkage of the Hilbert space coverage). An immediate consequence is that HEA is significantly limited by barren plateaus [127] (see Section 6.1 [86]). It also implies that a large

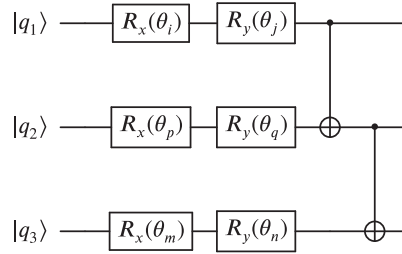


Fig. 8. Example of one block of a HEA, using R_x and R_y as rotation gates and $CNOT$ as entangling gates, on three qubits.

number of parameters need to be optimized and managed (for a numerical example of the number of parameters that HEA requires compared to other ansätze, one can refer to Table II in Ref. [297], where 198 parameters are required for a 12 qubit version of H_2O , against 8 for UCCSD detailed below). Regarding this last point, Wu et al. have shown that certain architectures of HEA provide higher expressibility and can be overall more efficient in terms of quantum resources [462]. With some caveats, Bravo-Prieto et al. [463] demonstrate that HEA type structures can provide reasonable accuracies on spin chain models such as Ising or XXZ chains using a number of layers at least linear in the system size. Overall, the HEA is a convenient tool for a proof of principle for a given algorithm or optimization strategy. It is, however, unlikely to be suitable for general larger-scale chemical problems. One therefore would rather turn toward more problem-tailored ansätze.

6.2.2. The Unitary Coupled Cluster (UCC) ansatz and extensions

The Unitary Coupled Cluster (UCC) ansatz is arguably the most studied ansatz for VQE. It featured in the initial VQE work of Peruzzo et al. [37] and has occupied an important position in the literature since then. As such, we have split this section into three parts: a presentation of the general UCC framework, a description of the extensions that have been proposed, and a discussion on the benefits and issues of the ansatz. An extensive review dedicated to this family of ansätze was recently published by [110], we direct reader in search for further information towards it.

It is worth noting that while initially designed in the context of electronic structure computation, variants of UCC have also been applied in other contexts. For example, it is been applied to variational quantum computation of ground and excited states for periodic systems [71–73], and for vibrational spectroscopy (with the Unitary Vibrational Coupled Cluster ansatz proposed in Ref. [63], and applied in Ref. [65]).

General framework

The UCC theory [464–466] stems from adapting the Coupled Cluster (CC) theory. CC is a post Hartree–Fock method that aims at recovering a portion of electron correlation energy by evolving an initial wave function (usually the Hartree–Fock wave function) under the action of parameterized excitation operators [214,467–471].

In general, these are single-electron excitations and double electron excitations, resulting in CC Single and Double (CCSD), however, these can be extended systematically to higher-order (e.g. CCSDT). Only excitation operators allowing transitions from an occupied orbital to an unoccupied orbital are traditionally included in CC (note that we use i, j, k , and l for occupied orbital indices, a, b, c , and d for virtual orbital indices and p, q, r , and s for either). The action of these operators on the initial state is performed through exponentiation of part of the cluster operator T . For ν denoting the maximum allowed excitation, we have:

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \cdots \hat{T}_\nu, \quad (139)$$

with for example the single and double excitation operators:

$$\hat{T}_1 = \sum_{ia} t_i^a \hat{a}_a^\dagger \hat{a}_i \quad (140)$$

$$\hat{T}_2 = \sum_{ijab} t_{ij}^{ab} \hat{a}_a^\dagger \hat{a}_b^\dagger \hat{a}_j \hat{a}_i. \quad (141)$$

Using the Hartree–Fock state as reference state, the CC ansatz wave function is given by

$$|\psi\rangle = e^{\hat{T}} |\psi_{HF}\rangle. \quad (142)$$

The conventional resolution method for CCSD scales $\mathcal{O}(m^2(N-m)^4)$, with m the number of electrons, and N the number of spin orbitals. CCSD is however in general not variational and has been reported to fail in numerous cases, in particular in systems with strong correlation, with possible solutions to avoid these failures usually scaling exponentially in the system size (for further reading on CC theory, failures and how they can be addressed, we direct the reader toward Refs. [472–477]). Another issue in the context of the VQE is that the operator $e^{\hat{T}}$ is not unitary, and therefore the CC ansatz cannot be implemented as a series of quantum gates.

The UCC method was developed as a way to address these limitations [464–466]. It is based on the fact that for any linear operator $\hat{T}$, the expression $(\hat{T} - \hat{T}^\dagger)$ is an anti-Hermitian operator. The exponential of an anti-Hermitian operator is a unitary operator, and the difference between the cluster operator and its complex conjugate can be used as an evolution operator to form a unitary version of CC. Elements of the truncated cluster operator for UCC are identical to that of CC, as presented in Eqs. (140) and (141).

$$|\psi\rangle = e^{\hat{T} - \hat{T}^\dagger} |\psi_{HF}\rangle. \quad (143)$$

The energy can then be evaluated using a variational approach based on the Ritz functional (see Section 2). Exact evaluation of this functional for UCC on classical resources is known to scale exponentially in the system size [465]. It is therefore natural to bring this ansatz to quantum computation.

By grouping the individual excitation operators in $\hat{T}$ with their corresponding conjugate in $\hat{T}^\dagger$, and denoting them as τ , we can obtain the parameterized version of the UCC ansatz:

$$U(\vec{t}) = e^{\sum_j t_j (\tau_j - \tau_j^\dagger)}, \quad (144)$$

where t_j correspond to the amplitude weights presented in Eqs. (140) and (141), and with j spanning all the excitation operators included. Akin to our description of CC, UCC can accept several level of excitations (UCCSD, UCCSDT).

Using any of the mappings described in Section 4 we can re-write each of the fermionic operators as

$$\tau_j - \tau_j^\dagger = i \sum_k \hat{P}_{k,j}, \quad (145)$$

where the subterm $\hat{P}_{k,j}$ is a tensor product of Pauli operators (technically, this term can actually be a sum of mutually commuting Pauli operators, as commuting groups do not need to be Trotterized). Resulting in the following expression of the ansatz:

$$U(\vec{t}) = e^{\sum_j \sum_k i t_j \hat{P}_{k,j}}, \quad (146)$$

The next step is to convert the UCC ansatz into a series of parameterized quantum gates which can directly be implemented on a quantum computer. A usual step in the process of building the UCC ansatz is to use a Trotter–Suzuki decomposition (a process also referred to as Trotterization) to separate the sum in the exponent of Eq. (146) into a product of exponentials, such that

$$U(\vec{t}) \sim U_{\text{Trotter}}(\vec{t}) = \left(\prod_{j,k} e^{i t_j \frac{\rho}{\rho} \hat{P}_{k,j}} \right)^\rho. \quad (147)$$

The Trotter number ρ defines the precision of the approximation, but also enters as a pre-factor impacting the overall depth of the quantum circuit required for its implementation (technically, $\hat{P}_{k,j}$ can actually be a sum of mutually commuting Pauli operators, as the exponential of a sum of commuting groups does not need to be Trotterized and can be directly converted to a product of exponentials). Assuming $\rho = 1$, this reduces Eq. (144) to

$$U_1(\vec{t}) = \prod_{j,k} e^{i t_j \hat{P}_{k,j}}. \quad (148)$$

This illustrates the basic framework surrounding UCC and its application to VQE [37,82,456]. The required circuit depth for this version of UCC has been shown to scale polynomially in the system size. More specifically, with η the number of electrons and N the number of spin-orbitals, the circuit depth is shown to scale $\mathcal{O}((N - m)^2 m)$ for each Trotter step [80]. It was also numerically shown that a single Trotter step is sufficient for an accurate description of the ground state in simple molecular systems [148], because variational optimization can absorb some of the Trotterization error [149,456] (though it does imply that operator ordering can impact accuracy [219], see discussion below for more details). The UCC ansatz can also be used in the context of restricted active space methods [478], considered in Ref. [456]. Finally, it has been shown that operators in the UCC ansatz contribute to the expressibility of the ansatz to different extents, allowing for approaches where some are discarded to increase efficiency [366] (a method similar to what is proposed in ADAPT-VQE [217], see Section 6.3). The remainder of the section covers several extensions which aim at improving both the precision and efficiency of UCC based ansatz.

Extensions of UCC

Generalized UCC: The first extension of the UCC theory applied in the context of the VQE is the Generalized UCC (UGCC) [80]. Unlike conventional CC and UCC which only include excitation operators that correspond to transitions from occupied to virtual orbitals, UGCC is agnostic to the orbital character in the (generally Hartree–Fock) reference state it is applied to. Therefore the single and double excitations in Eqs. (140) and (141) are re-written in the context of the UGCCSD [80] as

$$\hat{T}_1 = \sum_{pq} t_p^q \hat{a}_q^\dagger \hat{a}_p \quad (149)$$

$$\hat{T}_2 = \sum_{pqrs} t_{pq}^{rs} \hat{a}_r^\dagger \hat{a}_s^\dagger \hat{a}_q \hat{a}_p, \quad (150)$$

where we recall that p, q, r , and s are used as indices spanning both occupied and virtual orbitals. Generalized CC, was actively discussed in the conventional computing literature, though a unitary version was only mentioned in [479] and never extensively studied. Lee et al. [80] provide an in-depth study of the method and showed that the associated UCCGSD ansatz depth scales $\mathcal{O}(N^3)$ in the number of qubits N while at the time providing example showing that this method is more robust to noise and accurate than UCCSD.

Paired UCC (k -UpCCGSD): A second extension of UCC is the Unitary pair CC with Generalized Singles and Doubles (UpCCGSD) [80]. This version is based on the paired coupled-cluster method [480,481]. The key difference with CCSD is that it only includes two body terms that move pairs of opposite-spin electrons together from doubly-occupied to fully unoccupied spatial orbitals, ensuring no single occupancy in the described states. Recalling that we use i for an occupied orbital and a for virtual orbitals, the $\hat{T}_2$ terms included in UpCCSD can be described as

$$\hat{T}_2 = \sum_{ia} t_{ia}^{a\alpha a\beta} \hat{a}_\alpha^\dagger \hat{a}_\beta^\dagger \hat{a}_{i\beta} \hat{a}_{i\alpha}, \quad (151)$$

with α and β the spin up and spin down indices. The one body terms remain identical to Eq. (140). Lee et al. [80] however showed that using a unitary version of the method failed to achieve chemical accuracy. They therefore adapted it by using generalized one- and two-body excitation (see Eq. (149)) (where the indices are not constrained by whether they are occupied or not) while maintaining the spatial orbital pairing), hence the ‘G’ in UpCCGSD. They further improved the method by allowing repetitions of the ansatz (the number of repetitions referred to by the k parameter) allowing more flexibility for the wave function. This forms the k -UpCCGSD ansatz, which can be described as

$$|\psi\rangle_{k\text{-UpCCGSD}} = \prod_{l=1}^k (e^{\hat{T}^{(l)}} - e^{\hat{T}^{(l)\dagger}}) |\psi_{\text{HF}}\rangle. \quad (152)$$

The key advantage of this method is to allow a linear scaling ansatz [80], namely that the ansatz depth scales as $\mathcal{O}(kN)$ with N the number of spin-orbitals, and a reasonable quadratic scaling in the number of parameters $\mathcal{O}(kN^2)$ [482]. In this method, the parameters for each l block of the ansatz are treated independently. A few drawbacks of this method have been raised in Ref. [219]. In particular, it is shown that while the ansatz error can be arbitrarily reduced by increasing k , it also involves a nonlinear increase in the optimization burden due to the energy landscape becoming more complex from adding new parameters. The result found through this ansatz also seems to be dependent on the parameter initialization [219]. However, Lee et al. [80] show numerically that k -UpCCGSD finds better energies than UCCSD (also for excited-state calculations, see Section 9.1). These results were corroborated in Ref. [483], which studies the accuracy of UCC-based ansätze.

Pair-natural orbital-UCC (PNO-UpCCGSD): The idea behind k -UpCCGSD was further extended in Refs. [311,484] by restricting excitations to those between pair-natural orbitals (PNO) and the occupied orbitals of the reference state. PNOs are an efficient basis for modeling of the most relevant virtual states for a given occupied orbital pair [485–488] initially developed for Coupled Electron-Pair Approximations, but subsequently used in CC [489,490], and found to significantly reduce the size of the virtual orbital space in CC calculations (for example [491]). PNOs are found as the eigenvectors of a pair density matrix, often found from a lower level of theory.

One can define a set of orthonormalized PNOs S_{ij} from each possible pair $\{i, j\}$ of canonical HF orbitals. The PNO-UpCC(G)SD ansatz is constructed similarly to any UCC ansatz, but only with operators allowing excitations from occupied indices (i.e. S_{ij}) to unoccupied orbitals of the PNO set. It is also possible to extend beyond excitations from a reference orbital to create a generalized version of the ansatz. The single excitation term then allows excitations from any PNO to any PNO. Ref. [311] shows numerically that the PNO method significantly reduces the number of parameters and of CNOT gates compared to k -UpCCGSD (by a factor of 4 to 8 in both cases). This method was used to develop an ansatz tractable on conventional computers [484].

OO-UCC: A third example of an extension of UCC is the Orbital Optimized UCC (OO-UCC) (independently proposed in both Ref. [457] and Ref. [294] with minor differences). The logic behind OO-UCC is to optimize the one-body part of the ansatz using conventional computation, as it is a tractable optimization of orbital rotations. In particular, the one-body terms act as orbital rotation on the initial input state (usually the Hartree–Fock wave function), therefore if we start from the UCCSD ansatz and apply and use a single Trotter step approximation ($\rho = 1$) we can write:

$$\begin{aligned} |\psi\rangle_{\text{UCCSD}} &= e^{\hat{T}_2 + \hat{T}_1} |\psi_{\text{HF}}\rangle \\ &\approx e^{\hat{T}_2} e^{\hat{T}_1} |\psi_{\text{HF}}\rangle. \end{aligned} \quad (153)$$

With the one body operator only changing the Hartree Fock determinant to another initial wave function determinant, such that $|\psi_{\text{init}}\rangle = e^{\hat{T}_1} |\psi_{\text{HF}}\rangle$, we have the OO-UCCD ansatz:

$$|\psi\rangle_{\text{OO-UCCD}} = e^{\hat{T}_2} |\psi_{\text{init}}\rangle \quad (154)$$

The one body terms $\hat{T}_1$ are used to optimize single-orbital rotations on a conventional computer, effectively resulting in a change to the orbital definitions on which the VQE is subsequently applied at each step [294,457]. In this way, OO-UCC is bears similarities to multi-configurational self-consistent field (MCSCF) methods, as it requires an iterative optimization of the orbitals on a conventional computer coupled to the VQE on the quantum computer (for MCSCF proposals on a quantum computer see, for instance, Refs. [159,492]).

Sokolov et al. [294] propose two variants of OO-UCC. The first one (OOpUCC) only allows paired excitations in the two-body terms (similar to UpCCGSD [80]), grounded in the success of the analogous CC method [480,493,494] (and labeled OO-pUCC). The second one adapts the singlet CC method (CCD0) [495], where the two-body excitation operators are split into singlet and triplet terms, and singlet operators are kept in the ansatz. For this variant, Sokolov et al. [294] propose to either reduce operators by leveraging index symmetries in singlet terms (OO-UCCD0) or to use the full set of singlet operators (OO-UCCD0-full), akin to the methods proposed in Ref. [495]. Mizukami et al. [457] point out that, while OO-UCC requires more computation than UCC, it trades it off for a simpler quantum circuit, allowing all wave function parameters to be treated fully variationally, and allowing incorporation of correlations outside of the active space with a reduced the number of qubits (as discussed for MCSCF methods in Refs. [159,492]).

Downfolded Hamiltonian and Double Unitary Coupled Cluster (DUCC): Downfolded Hamiltonian methods were initially developed in the context of CC theory and shortly thereafter extended to UCC [496–500], to reduce the space in which the ansatz is constructed without impacting the quality of the representation (and usually referred to as DUCC formalism). The idea was adapted for implementation as part of VQE using a downfolded Hamiltonian and a Generalized UCC ansatz [501] or UCCSD ansatz [502]. Here we report on the construction of the effective Hamiltonian under the DUCC formalism, as any other UCC ansatz one might want to use). To begin with, the clusters usually used in CC theory are split into $\hat{T}_{int}$ and $\hat{T}_{ext}$ which represents excitations within, and outside the target active space respectively [497,498]. Unitary operators are then constructed using the UCC formalism:

$$\begin{aligned}\hat{\sigma}_{int} &= \hat{T}_{int} - \hat{T}_{int}^\dagger \\ \hat{\sigma}_{ext} &= \hat{T}_{ext} - \hat{T}_{ext}^\dagger.\end{aligned}\quad (155)$$

The DUCC wavefunction can be constructed by exponentiation these operators and applying them to a reference state:

$$|\psi_{DUCC}\rangle = e^{\hat{\sigma}_{ext}} e^{\hat{\sigma}_{int}} |\psi_0\rangle. \quad (156)$$

The part of the ansatz external to the target active space (or the part that one aims to fold into the effective active space Hamiltonian) is then absorbed into the Hamiltonian giving the following operator $\hat{H}_{ext}^{DUCC} = (e^{\hat{\sigma}_{ext}})^\dagger \hat{H} e^{\hat{\sigma}_{ext}}$. Because the external operators represent higher energy excitations, they can be approximated without significant loss in accuracy [497, 502]. The downfolded Hamiltonian can be constructed in the active space by defining a projector on the reference state $\hat{P}$ and on the orthogonal determinants of the target active space $\hat{Q}_{int}$ such that: $\hat{H}_{ext}^{eff-DUCC} = (\hat{P} + \hat{Q}_{int}) \hat{H}_{ext}^{DUCC} (\hat{P} + \hat{Q}_{int})$. The DUCC wavefunction can then be written as

$$|\psi_{DUCC}\rangle = E e^{\hat{\sigma}_{int}} |\psi_0\rangle, \quad (157)$$

with the remaining internal excitations being implemented using a UCC ansatz and trained with VQE [501,502].

The method was shown to recover a large proportion of correlation energy on several small molecules while saving significant quantum resources [502]. In Ref. [501], Bauman et al. also show that using UCCGSD provides a material improvement over using UCCSD in this context, bringing results close to FCI on H₂O, of course at the cost of additional parameters from the Generalized UCC.

Other UCC-based methods: Several other extensions of UCC are also worth mentioning briefly. The Unitary Cluster-Jastrow [503] ansatz uses Jastrow factors, traditionally used in Quantum Monte Carlo [504] to include the impact of cusps from the Coulomb potential in the wavefunction and adapted for CC theory in [505]. The method shows similar scaling to k-UpCCGSD and excellent results in numerical analysis [503]. It also does not require trotterization and thereby is less prone to errors due to operator ordering [219]. The multicomponent UCC [506] extends the UCC ansatz to systems that include coupling to bosonic statistic modes. It was also applied to photon number states to cover polaritonic chemistry [507]. Finally, the Low Depth Circuit Ansatz (LDCA) [508] is grounded in the Bogoliubov (or quasi-particle) CC [509–511], for which a unitary version was developed in Ref. [508]. While the Bogoliubov UCC can be implemented as an ansatz for VQE, Dallaire et al. [508] propose some additional transformations to exactly parameterize the ansatz in a linear depth circuit (but with quadratic parameter scaling). The parameter-efficient circuit training (PECT) optimizer proposer in Ref. [512], appears to provide an effective method to optimize LDCA on larger systems.

Unitary selective coupled-cluster method: The USCC method, proposed by Fedorov et al. [513] could be categorized both as a UCC extension and as an adaptive ansatz. To implement this method, one starts with the reference Hartree–Fock state. From there, one and two body excitations (from occupied to unoccupied orbitals) which have corresponding Hamiltonian matrix elements above a given threshold τ_1 are included in the ansatz. A VQE is then performed on this ansatz until convergence – the first iteration of the method is complete. For the following iterations, the threshold is update such that: $\tau_i = \tau_{i-1}/2$, and new, higher order excitations are generated by applying one and two body excitations to those already

included in the ansatz. The metric to compare to the threshold for inclusion of these excitations is obtained by multiplying the parameters obtained after optimization at the previous step with the Hamiltonian matrix element corresponding to the single or double excitation added. As per previous iterations, coefficients that are above the threshold are included in the ansatz and then optimized through VQE. The optimization continues until a convergence criterion is met or sufficient excitations have been included. Fedorov et al. [513] note that the method does not provide a wavefunction form that is as compact as can be achieved through ADAPT-VQE, but is likely to require significantly less measurements to perform and converge. It is worth noting that as an adaptive method USCC would likely also present some resilience to the barren plateau problem.

UCC discussion

Overall, it is clear that there are several advantages of using UCC-type ansatz compared to HEA. First and foremost, the number of parameters in the former, ranging from $\mathcal{O}(N^3)$ to $\mathcal{O}(kN^2)$ (k an arbitrary constant) is significantly more advantageous than in the HEA which spans a much wider part of the Hilbert space. It was also shown numerically that the UCC ansatz can provide exact parameterization for an arbitrary electronic wavefunction [161], subject to certain conditions on optimization. The accuracy of the family of ansatz and its ability to approach FCI level results on a variety of small systems was numerically established in Ref. [483] (namely H_2 , H_3 , H_4 , NH , OH^+ , CH_2 , NH_2^+). Similarly, Ref. [294] establishes that UCCSD outperforms conventional CCSD methods in accuracy on H_4 , H_2O , N_4 and on the one-dimensional Fermi–Hubbard model. It is further shown in the same study that orbital optimized methods (namely p-OO-UCCSD, OO-UCCD0 and OO-UCCD0-full) achieve results within 10 to 20 mH, showing a minor loss in accuracy compared to UCCSD, but a significant reduction in quantum resources required.

A clear drawback of UCC-type ansätze is specifically that these are not hardware efficient. All UCC ansätze are designed in a manner that is agnostic to the connectivity of the device. The number of CNOT gates for each of the exponentiated Pauli strings required in the ansatz scales with its Pauli weight q (number of non-identity operators in the string) when full connectivity is allowed. Namely, each Pauli strings exponential requires $2(q - 1)$ CNOT gates [76]. Assuming double excitations are used, this implies a scaling of $\mathcal{O}(2N^4(\bar{q} - 1))$ CNOT gates (with $\bar{q}$ the average Pauli weight across all strings) if no further action is taken to make the ansatz more efficient (some of these gates can be canceled out at compilation). However, a significant overhead in the number of CNOT gates is required if connectivity is limited to bridge entanglement between qubits that are not directly connected on the device. Connectivity limitation of the qubit register can be addressed by introducing SWAP networks to effectively relocate qubits together when q -qubit operators need to be implemented on a non-local set of qubits [514]. This allows maintaining the depth scaling of UCC ansätze by parallelizing on a single qubit register several operators that would otherwise not commute, irrelevant of the degree of connectivity of the device. It does however come at the cost of additional SWAP (or CNOT) gates required to build the SWAP network compared to a fully connected device. This is also related to a similar drawback of UCC: while the scaling is advantageous, the ansatz generally requires a very large pre-factor in the number of gates [219,297,456]. A similar method to construct a more efficient gate fabric, but with the added benefit of preserving quantum number symmetry is found in Ref. [515].

Overall, the UCC ansätze have a large pre-factor for the depth and number of entangling gates. As an example, Ref. [297] shows that to achieve similar accuracy on a 12-qubit version of H_2O , UCCSD requires 528 entangling gates, hence at the very best a depth of about 100 as one can fit at most 6 entangling gates in one time step, against 88 entangling gates for HEA, or a circuit depth of about 40 (in Ref. [297], Choquette et al. also present an ansatz which extends the Hamiltonian Variational Ansatz presented in the following section, and which requires 108 CNOT gates for the same molecular model). It is fair to assume that k-UpCCGSD would perform significantly better than UCCSD in terms of depth. The Qubit-excitations [344,455,516], and qubit-operators [270,287,454] Coupled Cluster based ansatz (see Section 6.3), are a potential answer to address the scaling of CNOT gates while preserving some of the benefits of UCC ansätze.

Another point worth mentioning about UCC-based ansatz is their dependence on operator ordering [219,517]. Ordering of fermionic operators within the ansatz significantly impacts its accuracy, in particular for highly correlated electron wave functions (which is, as Grimsley et al. [219] and Wecker et al. [82] point out, the primary target of interest for quantum computers, as low correlation wave functions can easily be treated with conventional methods – see Section 2.4). This is particularly noticeable when dissociation limits are approached [219,503], but can be in great part addressed by repeating the ansatz (i.e. setting k to a higher number for k-UpCCGSD). As pointed out in Ref. [80] and corroborated by results from Ref. [219], k-UpCCGSD does appear to be largely resilient to problems of operator order as long as one allows for k to be sufficiently high to provide the variational flexibility needed.

Finally, another issue that can happen in both HEA and UCC ansatz is the appearance of ‘kinks’ in the Potential Energy Surface (PES). While already apparent in Ref. [263], these are very well characterized in Ref. [518]. These kinks are due to the energy ordering of the electronic states and lack of preservation of certain of symmetries (e.g. electron number, total spin) when building the ansatz, and can be addressed using symmetry preserving methods, including as an additional constraint placed in the VQE cost function [518] (see Section 7.5).

6.2.3. Symmetry-preserving methods

A proposed solution to a common pitfall of HEA and UCC, is to build the ansatz in a manner that preserves certain symmetries of the system studied (though it is worth noting that UCC ansätze in general preserve particle number symmetry). There are two main benefits related to this:

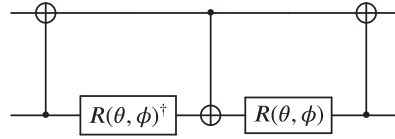


Fig. 9. Quantum circuit for realization of gate $A(\theta, \phi)$, Eq. (158). With $R(\theta, \phi) = R_z(\phi + \pi)R_y(\theta + \pi/2)$.

- It restricts the size of the Hilbert space in which the ground state can be searched while remaining largely agnostic to the system studied, potentially speeding up the optimization and reducing the risk of barren plateaus.
- It avoids risks of localized inaccuracies that have been observed as ‘kinks’ in the dissociation curves of molecules studied [263], and that are related to local minima states with a different number of electrons [518]

It is worth pointing out that symmetry preservation can also be achieved using an adjustment to the cost function [38, 518] (further details are presented in Section 7.5), and can be used as error mitigation method (see Section 8.1).

The core objective of symmetry-preserving ansätze in general, is the preservation of the electron number in the modeled wave function. A common theme in these methods is the use of exchange-type gates (or iSWAP gates) (for a description and hardware implementation of these gates, we refer readers to Refs. [519,520]). These two-qubit gates aim at allowing relative phase changes between superpositions of $|01\rangle$ and $|10\rangle$ states, while leaving $|00\rangle$ and $|11\rangle$ unchanged. In the context of a VQE ansatz, their use can therefore never alter the number of occupied orbitals in the modeled wave-function.

There exist two published versions of this type of gate. The most commonly used one is a two-parameter version, originating from Ref. [520], which allows further flexibility in the output wavefunction:

$$A(\theta, \phi) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta & e^{i\phi} \sin \theta & 0 \\ 0 & e^{-i\phi} \sin \theta & -\cos \theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (158)$$

It can be obtained with a quantum circuit composed of three CNOT gates and two rotation gates, as presented in Fig. 9. A special, single parameter version of this gate has also been presented (see Refs. [156,521])

$$A(\theta) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\theta & -i \sin 2\theta & 0 \\ 0 & -i \sin 2\theta & \cos 2\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (159)$$

It is worth noting that this approach to particle conservation in the wave function is only valid when considering certain types of fermion-to-spin mappings (see Section 4). In particular, while these gates necessarily preserve the number of excited qubits under Jordan–Wigner [218] and the Parity mappings, it is easy to see that it does not under Bravyi–Kitaev, as swapping an excited qubit with a non-excited qubit across two blocks could require changing the state of all the subsequent ‘parity’ encoding qubits within a block (see Section 4.1.3).

A first relevant use case is presented in Ref. [148], where the HEA (Section 6.2.1) is adapted to preserve the particle number. The method transforms the HEA into the particle/hole representation of the Hamiltonian (see Section 3). The ansatz in Ref. [148] is structured like the HEA, with a series of single-qubit rotations and entangling blocks, similar to that presented in Eq. (138). Each constituent block however can be composed of interlaced exchange-conserving gates, and as these are parameterized, the rotation unitaries can be absorbed in the entangling unitaries.

Two versions are proposed in Ref. [148]. The first one makes use of the two-parameter exchange-conserving gate (Eq. (158)), as presented in Fig. 10, the second one relies on the one-parameter exchange-conserving gate (Eq. (159)), as presented in Fig. 11 (note these require additional rotation gates in each block). Barkoutsos et al. [148] also provide a comparison of these two approaches against a more typical HEA method (where the entangling block is composed of a ladder of CNOT gates), and show numerically that particle conserving blocks achieve better accuracy when computing the dissociation curve of H_2O (with a slight advantage for the two-parameters gates). This approach is further studied and tested successfully on a superconducting quantum device in Ref. [522].

In Ref. [218], Gard et al. propose to address the wider set of symmetries (particle number, time reversal, total spin, spin magnetization) by creating an Efficient Symmetry-Preserving (ESP) ansatz for VQE. It uses the same two-parameters exchange-conserving building blocks as in Ref. [148] and a very similar structure to Fig. 10 with some optimization. The idea behind this ansatz is to build a structure that spans the largest part of the Hilbert space while constraining symmetries as much as possible. As such, it is more efficient than HEA, but is less restrictive than UCC, guaranteeing that the ground state is indeed in the spanned Hilbert space [218]. In that respect, we consider that this is the VQE ansatz that is closest to providing the span of FCI (see Section 2.4.1).

This ansatz can be systematically built by first applying X gates to a number of qubits equal to the number of electrons in the problem considered. For efficient circuit building, and unlike in Ref. [148], Gard et al. [218] advise that one must

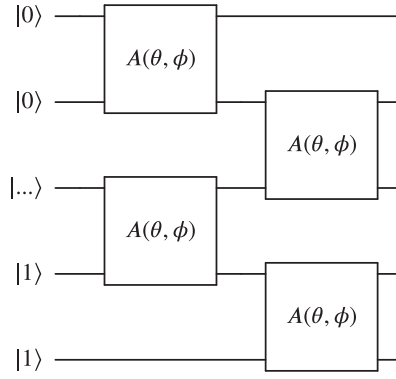


Fig. 10. Single layer of symmetry preserving ansatz presented in Ref. [148], based in two-parameters exchange conserving gates $A(\theta, \phi)$, Eq. (158).

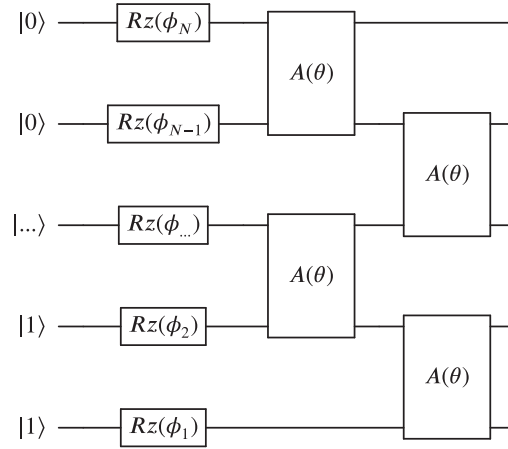


Fig. 11. Single layer of symmetry preserving ansatz presented in Ref. [148], based in one-parameters exchange conserving gates $A(\theta)$, Eq. (159).

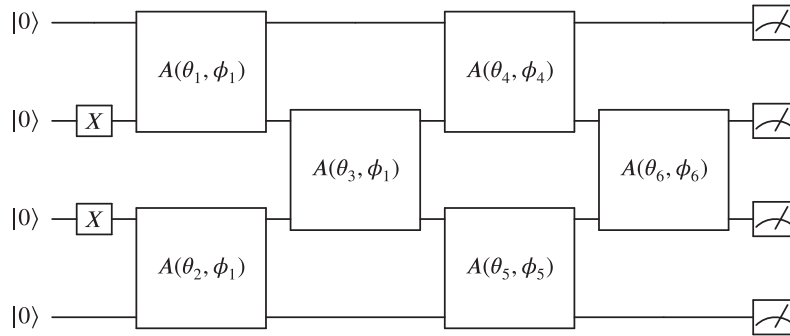


Fig. 12. Example of the ESP ansatz construction for $N = 4$ and $m = 2$, as presented in [218], based on two-parameters exchange-conserving $A(\theta, \phi)$ gates, Eq. (158). The first three ϕ_i parameters are set to be equal as the number of optimal parameters has already been reached in the ansatz and further flexibility is not required.

avoid placing these as neighboring qubits. Following this, one can apply A gates to each pair of $|1\rangle$ and $|0\rangle$ zero qubits, and subsequently, apply additional A gates to span the complete register and until $\binom{N}{m}$ A gates have been placed (N the number of orbitals or number of qubits, m the number of electrons). We have reproduced an example initially presented in [218] for $N = 4$ and $m = 2$ in Fig. 12.

The optimal number of parameters to allow full parameterization of the Hilbert space, given particle-conserving symmetry can be derived from the number of complex coefficients required to span the subspace drawn by a given Hamiltonian, Gard et al. [218] show that this number is $2^{\binom{N}{m}} - 2$. Given there are $2^{\binom{N}{m}}$ parameters in the ESP ansatz as described above, the optimal number can be obtained by setting two of the ϕ_i equal to a third one, as shown in Fig. 12.

They also point out that the method for constructing the ansatz is neither unique nor optimal in terms of the number of entangling gates and as such, further research may help extend the symmetry preserving ansatz (for instance by including ansatz options for fully-connected machines). Other symmetries can be imposed as follows:

- Time reversal symmetry can be imposed by setting all ϕ_i to zero, resulting in the number of parameters scaling $\binom{N}{m}$
- Spin symmetries require to first order the qubit such that the first half of the register represents spin up orbitals, and the second half spin down orbitals (which requires some fermion to spin mapping modifications). The ladder structure described for the ESP ansatz is maintained, with a difference that any A gate that links the first and second half of the register has parameters set to 0 (or is removed), preventing swaps of electrons from spin-up to spin-down orbitals (and vice versa), and enforcing spin symmetry.

Based on this construction, the number of CNOTs scales at most as $3\binom{N}{m}$ (without further optimization and compilation of the ansatz). This implies that it goes from linear if $m = 1$, or $m = N - 1$ to exponential if $m \sim N/2$. It can be slightly reduced using some of the symmetries mentioned above, however it remains significantly less efficient than UCC based ansatz in particular in terms of parameters, due to a larger span of the Hilbert space.

As such, we find that the cost of ESP (and other symmetry preserving ansatz) is too prohibitive to be implemented in cases where $m \sim N$. There could be scope for these ansätze to outperform the likes of UCC type ansätze when the problem can be treated accurately with m being negligible or close to N . An interesting research question would be to study this ansatz in a plane wave basis where the number of basis functions (and therefore qubits) required for each electron to treat the system accurately can be extremely large [235].

6.2.4. The Hamiltonian Variational Ansatz and extensions

The Hamiltonian Variational Ansatz (HVA): this ansatz was initially presented in Refs. [82,150], and is inspired from the quantum approximate optimization algorithm (QAOA) ansatz [136] which draws from ideas developed for adiabatic quantum computation [523]. The first step in building this ansatz is to decompose the Hamiltonian into a sum of non-commuting operators:

$$\hat{H} = \sum_a \hat{h}_a, \quad (160)$$

where $[\hat{h}_i, \hat{h}_j] \neq 0$. These terms are used alternatively over L layers to evolve an initial state $|\psi_0\rangle$.

$$|\psi_L\rangle = \prod_{l=1}^L \left(\prod_a e^{-i\theta_{a,l}\hat{h}_a} \right) |\psi_0\rangle. \quad (161)$$

The initial state $|\psi_0\rangle$ is picked as the ground state of any of the $\hat{h}_a$, except for the first one acting on it. There are a few points to note about this ansatz. The first one is that for a single repetition of the ansatz it has a number of variational parameters lower than the UCCSD ansatz, because there is only one parameter per commutative group in the Hamiltonian (we have seen in Section 5, that N^3 commutative group is a reasonable estimate for *ab initio* molecular systems, though it could be much lower for lattice models). This means that layer repetitions are likely to be required to achieve the same level of expressibility, although exact scaling is not known. In Ref. [82], Wecker et al. suggest that the ansatz is trained layer-by-layer to facilitate optimization. The scaling in the number of entangling gates is similar to UCC based ansatz, as approximately the same number of operators are incorporated for each repeated layer. This also implies a similar depth scaling.

The HVA was shown however to be particularly well suited for models of many-body physics [150]. Key points revealed in this study include the fact that the ansatz appears to exhibit strong resilience to barren plateaus, and that layer repetitions allow for over-parameterization of the ansatz and therefore smoothing out of the energy landscape to avoid local minima. It is clear that low Pauli-weight mapping (or lattice models) significantly improves the algorithm. The ansatz was also adapted to solve the Heisenberg antiferromagnetic model on the Kagome lattice in Refs. [332,333], both study showing promising results. In particular, Ref. [333] shows that VQE approaches the true ground of the lattice exponentially as a function of circuit depth on a noiseless simulation of up to 20 sites. Bosse and Montanaro [332] find a similar result on a simulation of 24 sites, though point out to the large number of variational parameters needed. Another adaptation of HVA to the Heisenberg model, accompanied by a scaling study is presented in Ref. [524]. Finally, a version of HVA for Fermi–Hubbard is studied by Stanisic et al. [525], in which the ansatz is created by first applying a number of Givens rotations (rotation in a two dimensional subspace of the Hilbert space, which can be used as partial excitation exchanges between two qubits), allowing to efficiently prepare the ground state of the non-interacting Fermi–Hubbard model [526], then followed by time evolution following the same logic as HVA. The paper shows that through this ansatz construction, few parameters are needed to reproduce key qualitative features of strongly-correlated electronic systems.

Fourier-transform HVA [235]: This ansatz was initially developed as a linear scaling method to variationally model the ground state of jellium (uniform electron gas). In this context, it was developed using a plane wave basis, allowing the one-body hamiltonian of jellium to be diagonal [235]. A slightly more general formulation is presented in Ref. [297]. The method relies on the fermionic fast Fourier transform (FFFT) [526–528] which was initially constructed to diagonalize

dynamics in many-body systems, and which has been shown to have a depth scaling of $\mathcal{O}(N)$ if implemented on a planar lattice of qubits [235]. To construct this ansatz one must first separate the Hamiltonian into kinetic and potential energy terms, $\hat{H} = \hat{T} + \hat{V}$ (as presented in Section 3). The ansatz is then constructed in a manner similar to HVA, but applying FFFT to the kinetic energy term:

$$U(\theta) = \prod_{l=1}^L \text{FFFT}^\dagger \left(\prod_a e^{-i\theta_{a,l}^{(T)} \hat{T}_a} \right) \text{FFFT} \left(\prod_b e^{-i\theta_{b,l}^{(V)} \hat{V}_b} \right) \quad (162)$$

with $\hat{T} = \sum_a \hat{T}_a$ and $\hat{V} = \sum_b \hat{V}_b$. In Ref. [235], Babbush et al. recommend to use $|\psi_0\rangle = \text{FFFT}|0\rangle$ as initial state, as this makes $|\psi_0\rangle$ an eigenstate of $\hat{T}$ in the plane wave basis. They also suggest absorbing the component of the ansatz corresponding to $\hat{T}$ in the Hamiltonian as it can be done efficiently. It is worth noting that given use of a plane wave basis, for jellium, the operators of $\hat{V}$ are mapped to diagonal Z operators using Jordan–Wigner, which provides a very convenient and compact ansatz scaling $\mathcal{O}(LN)$ on planar qubit lattices, and which can also be optimized layerwise [82,235]. To the best of our knowledge, this ansatz has not been benchmarked or tested on alternative systems.

Symmetry breaking HVA [296,297]: The salient feature of this type of ansätze is that it incorporates unitaries into the HVA to purposefully break symmetries, to allow for a better optimization path. Both the quantum-optimal-control-inspired ansätze (QOCA) [297], and the Variational Extended Hamiltonian Ansatz (VEVA) [296] are nearly identical and were initially published at the same time. The version presented in Ref. [297] is motivated by optimal control theory [529,530]. Quantum optimal control (QOC) is concerned about the efficient design of control pulses to manipulate a quantum system (for a recent review of QOC, we refer reader to Müller et al. [531]). QOC theory relies on a control Hamiltonian which is composed of a set of drive terms $\{\hat{H}_b\}$, and a drift term $\hat{H}_0$, such that $\hat{H}^{(ctrl)}(t) = \hat{H}_0 + \sum_b c_b(t) \hat{H}_b$ (where $\{c_b(t)\}$ are time-dependent parameters). To build QOCA or VEHA, one must first create a Hamiltonian incorporating the structure of the control Hamiltonian:

$$\hat{H}_{tot}(t) = \hat{H} + \sum_b c_b(t) \hat{H}_b^{(ctrl)} \quad (163)$$

where $\hat{H}$ is the problem Hamiltonian. The ansatz is constructed from this Hamiltonian in a manner similar to HVA, thereby mimicking the time evolution of QOC. It can therefore be constructed as:

$$U(\theta, \phi) = \prod_{l=1}^L \left(\prod_a e^{-i\theta_{a,l} \hat{H}_a} \right) \left(\prod_b e^{-i\phi_{b,l} \hat{H}_b^{(ctrl)}} \right) \quad (164)$$

where $\hat{H} = \sum_a \hat{H}_a$ and $\hat{H}_b^{(ctrl)}$ are a set of drive terms selected for the purpose of the problem studied (ϕ corresponds to the parameters $\{c_b(t)\}$). Following the description in Ref. [297], the problem Hamiltonian is used to explore the symmetry subspace of the Hilbert space, and the drive terms are used to provide shortcuts through the subspaces to increase efficiency. It is recommended in Ref. [297] that the drive terms do not commute with $\hat{H}$. Choquette et al. suggest that this terms could be selected adaptively (akin to adaptive ansätze described in Section 6.3), but also provide a method to identify the most relevant drive operators.

This method is applied to the Fermi–Hubbard model and compared to FT-HVA and HVA in Ref. [297]. Results show excellent fidelities achieved with a reasonable number of entangling gates up to 12 qubits (which could also suggest some degree of resilience to barren plateaus). The method is also shown to perform well on a 12-qubit version of H_2O , with fewer numbers of entangling gates compared to UCCSD, albeit a higher number of parameters [297]. The method is also tested on the Hubbard model in Ref. [296], with the stability of the ansatz being very much dependent on the problem parameters. Vogt et al. [296] extend the idea of the symmetry breaking HVA by adding classical mean-field parameterization to the ansatz (a method named Variational Mean Field Hamiltonian Ansatz, or VMFHA), which performs better where the VEHA fails.

6.3. Adaptive structure ansätze

Adaptive structure ansätze aim at creating a circuit structure tailored to the problem studied. As such, they all function similarly, in which the ansatz is grown iteratively throughout the optimization process by adding new operators at each step based on their contribution to the overall energy (or other metric). We distinguish three types of adaptive ansatz, those that grow the quantum circuit iteratively (Section 6.3.1), those that incorporate (at least part of) the ansatz into the Hamiltonian by ‘dressing’ it (Section 6.3.2), and those that variationally update the gate structure of the entire ansatz (Section 6.3.3). At this point, it is challenging to build comparative metrics for adaptive ansätze reliably, and as such we do not provide a comparative table as we have done in other sections. Instead, we direct readers to the qualitative discussion at the end of this section.

6.3.1. Iterative ansatz growth methods (ADAPT-VQE and extensions)

Fermionic-ADAPT-VQE: The idea behind ADAPT-VQE [217] is to progressively build the ansatz by sequentially incorporating into it the operators that contribute most to lowering the VQE energy towards the ground energy. Starting from an initial state (usually the Hartree–Fock wave function), and given a pool of operators $\{\hat{A}_i\}$ the ADAPT-VQE ansatz evolves as follows, with the iteration number as superscripts:

$$\begin{aligned}
 |\psi^{(0)}\rangle &= |\psi_{\text{HF}}\rangle \\
 |\psi^{(1)}\rangle &= e^{\theta_1 \hat{A}_1} |\psi_{\text{HF}}\rangle \\
 |\psi^{(2)}\rangle &= e^{\theta_2 \hat{A}_2} e^{\theta_1 \hat{A}_1} |\psi_{\text{HF}}\rangle \\
 &\dots \\
 |\psi^{(k)}\rangle &= \prod_{i=k}^1 e^{\theta_i \hat{A}_i} |\psi_{\text{HF}}\rangle.
 \end{aligned} \tag{165}$$

The operators in the pool are similar to those produced as part of the UCC ansatz (see Eq. (145)), however rather than specifying a number of possible excitation, one can include any one-, two body- or even higher body operator that is believed to be particularly relevant for the system considered: $\hat{A}_m \in \{(\hat{\tau}_p^q - (\hat{\tau}_p^q)^\dagger), (\hat{\tau}_{pq}^{rs} - (\hat{\tau}_{pq}^{rs})^\dagger), (\hat{\tau}_{pqr}^{stu} - (\hat{\tau}_{pqr}^{stu})^\dagger), \dots\}$.

To decide on which operator to incorporate, Grimsley et al. [217] recommend choosing the one which results in the highest gradient of the energy functional with respect to the operator parameter. After initializing all operator parameters to 0, to guarantee that they initially act as the identity operator, gradients can be simply computed as:

$$\frac{\partial E^{(k)}}{\partial \theta_i} = \langle \psi^{(k)} | [\hat{H}, \hat{A}_i] | \psi^{(k)} \rangle. \tag{166}$$

Once an operator has been selected, and incorporated into the ansatz, a full VQE optimization is completed until parameters all parameters are at their optimum. This implies that at the next operator selection step, all existing θ_j are such that $\partial E^{(k-1)} / \partial \theta_j = 0$, enforcing the results of Eq. (166) for all θ_i corresponding to new operators in following iterations.

The simplicity of the gradient computation at all iterations of ADAPT-VQE is key to its efficiency. However, assuming the case of a second-quantized electronic structure Hamiltonian (with $\mathcal{O}(N^4)$ terms), and given there are $\mathcal{O}(N^4)$ terms in a complete pool for up to double excitation operators, measurement scaling at each outer-loop iteration could naively reach $\mathcal{O}(N^8)$. This can be reduced to $\mathcal{O}(N^6)$ if one reformulates the energy gradients in terms of the three-body reduced density matrix [532]. This method also allows predicting gradients from the two-body reduced density matrix of the previous iteration of ADAPT-VQE, reducing overall scaling of measurements back to $\mathcal{O}(N^4)$ albeit at a cost to accuracy.

Therefore, an important research question surrounding ADAPT-VQE is the optimal size of the operator pool, as it determines the number of iterations required, and the complexity of each step. It is also important to note that the scaling of the ADAPT-VQE pool can also be adjusted the same way it is adjusted in certain variations of the UCC ansatz. For instance, one could restrict operator availability to those available in k-UpCCGSD [80], or proceed to only integrate double excitation operator on the quantum computer subroutine akin to the OO-UCC ansatz [311]. Comparative performance of these respective approaches is yet to be assessed.

Also contributing to the cost of ADAPT-VQE, a complete VQE convergence must be conducted between each incremental step. While this is largely tractable for early steps, as the ansatz remains shallow, it could rapidly become prohibitively expensive to implement. Grimsley et al. [533] nonetheless showed that the build-in “recycling” of parameters (initialization using the optimal values from the previous step) allow for rapid convergence and avoidance of local optima. Grimsley et al. [217] show that using the ADAPT-VQE method outperforms UCCSD in terms of accuracy for a given number of ansatz parameters (which is a proxy for the number of operators as we need one parameter for each operator). Using BeH_2 as an example, they show that ADAPT-VQE achieves an error of 10^{-3} kcal/mol for about 50–60 parameters (and still significantly less than 100 for 10^{-5} kcal/mol), while UCCSD requires nearly 300. They also tested against random ordering of the operator pool and showed that the gradient-based ordering described above performs significantly better.

It is difficult however to assess output at a comparable computational cost. While UCCSD only needs to be optimized once (over 300 parameters in BeH_2 example), ADAPT-VQE needs to be optimized many times as the number of operators (or parameters) included (in addition to computation of $\mathcal{O}(N^4)$ gradients at each iteration). Therefore, it is important to assess the comparative cost of ADAPT-VQE in terms of total number of measurements required, in addition to the number of parameters required to achieve a given accuracy. Large-scale numerical simulations may be required to assert the superiority of adaptive ansatz methods. Ref. [289] presents a numerical comparison of ADAPT-VQE and a UCCSD based VQE, corroborating the results presented in Ref. [217], showing that the former can achieve excellent accuracy (on H_2 , NaH , and KH), however with on average significantly more total measurements than the latter. Due to the targeted expressibility of the ansatz produced, ADAPT-VQE has also been shown to be resilient to the barren plateau problem [533] (other numerical evidence can be found in Refs. [450,534]). ADAPT-VQE could however suffer from exponential vanishing of pool operator gradients [533], though further research is needed to characterize and address this potential issue.

ADAPT-VQE also inherits some advantageous properties of the UCC ansätze such as the ease of management of particle number and time-reversal symmetries. Similar to UCC, however, it comes at the cost of a very large number of entangling

gates required to implement the fermionic excitation operators into an ansatz, in particular when the qubit lattice has low connectivity (as discussed in the UCC section, see Section 6.2.2). Qubit-ADAPT-VQE [270] and Qubit Coupled Cluster (QCC) [454] methods detailed below are methods to address this shortcoming. An alternative method called Energy Sorting VQE has also been proposed to not only compute the gradient of operator in the pool, but perform a full VQE optimization on each of them to rank them and include them in groups in the ansatz [535]. The principle of ADAPT-VQE has also been replicated in applications beyond electronic structure computation, in particular for periodic systems [71], and nuclear structure problems [59].

ADAPT-VQE for nuclear structure problems: While initially designed for electronic structure problems, ADAPT-VQE has been implemented for nuclear structures problems by Romero et al. [59]. The pool of fermionic operators can be replaced by one and two body operators for an N -nuclei wavefunction. The one body operators can be defined as X_i , Y_i , and Z_i , scaling as $3N$. For the two body operators, with $j < k$ the pool can be defined as:

$$\begin{aligned} T_+^{jk} &= \frac{1}{2}(X_j X_k - Y_j Y_k), & T_-^{jk} &= \frac{1}{2}(X_j Y_k + Y_j X_k) \\ U_+^{jk} &= \frac{1}{2}(X_j X_k + Y_j Y_k), & U_-^{jk} &= \frac{1}{2}(Y_j X_k - X_j Y_k) \\ V_+^{jk} &= X_j Z_k, & V_-^{jk} &= Y_j Z_k, & V_0^{jk} &= Z_j Z_k. \end{aligned} \quad (167)$$

This results in $7N(N-1)/2$ which results in a pool scaling $\mathcal{O}(N^2)$. This method is tested for both the Lipkin–Meshkov–Glick [536] and the nuclear shell model [537,538], up to $N = 12$ and demonstrates reasonable scaling in the number of outer-loop iteration required to reach good accuracy.

qubit-ADAPT-VQE: The Qubit-ADAPT-VQE [270] proposes an incremental selection of operators directly from the Pauli string exponentials produced from a given encoding rather than at the fermionic-operator level. Namely, this implies building an operator pool, with operators structured as

$$\hat{A}_a = i \bigotimes_i^N \hat{P}_i, \quad \hat{P}_i \in \{I, X, Y, Z\}^{\otimes N}. \quad (168)$$

From this point, the ansatz is constructed as presented in Eq. (165).

Further actions can be taken to restrict this operator pool. First restricting operations to those verifying time-reversal symmetry, removing from the pool in Eq. (168) imaginary operators which have no impact on energy if $\hat{H}$ is time-reversal symmetric. The Pauli strings included in the pool are also selected only from those produced from the fermionic excitation mapping (a mapping that must happen anyway to produce the measurable Hamiltonian). Tang et al. [270] also show analytically that a linear number of operator $(2N-2)$ is required for the ansatz to be analytically exact and describe a method to select this pool of operators. They also suggest reducing the string Pauli weights by removing the Pauli Z operators used for anti-symmetry of the wave function in the Jordan–Wigner mapping (Section 4) and show that this does not impact energy estimate. This implies a maximum Pauli weight of 4 for any Pauli string used to construct the ansatz, allowing for significant depth contraction.

While it is clear that the Pauli weight reduction from removing Pauli- Z chains is not exclusive to this method we conjecture that it works in this instance because qubit-ADAPT-VQE uses one parameter for each Pauli string (instead of one per fermionic operator) and is, therefore, able to variationally learn the anti-commuting relationships required to reach ground-state energy (a view also supported in Ref. [344]). Shkolnikov et al. [539] also showed that $(2N-2)$ is the minimum size of a complete pool comprised of Pauli strings. They also show that while minimal complete pools do not perform well on electronic structure Hamiltonian, these can be symmetry-adapted to successfully model ground state energy, achieving outer-loop measurement scaling of $\mathcal{O}(N^5)$. Overall, this approach allows a significant reduction in the number of CNOT gates (4 per operator) required for the implementation of adaptive ansatz albeit at the cost of introducing a larger pool to sample from using the gradient-based selection rules, and as shown in Ref. [270] more parameters to achieve a given accuracy. In particular, Mukherjee et al. [540] show that qubit-ADAPT-VQE where operators are taken from a pool of Hamiltonian commutators produces a significantly more compact circuit than the HVA and the UCCSD ansätze.

QEB-ADAPT-VQE: The qubit-excitation-based adaptative VQE [344] (QEB-ADAPT-VQE) offers an interesting compromise between ADAPT-VQE and qubit-ADAPT-VQE, by having fewer CNOT gates than both, yet a similar number of parameters than ADAPT-VQE. It is based on a previous ansatz labeled Qubit Coupled Cluster Singles and Doubles (QCCSD) [455] which uses the same qubit-excitations, up to second order, but without the adaptive aspect. It introduces two main changes:

- Screening process includes a new step, whereby the top k candidate operators in the pool (by gradient magnitude) are used for full VQE optimizations, with the one triggering the largest energy drop being selected for incorporation.
- It replaces the pool of operators with qubit-excitation operators [516]. These have been developed as a CNOT efficient alternative to the conventional implementation of fermionic excitation operators.

The premise for qubit-excitation operators is simply based on qubit creation and annihilation operators ($Q_i^\dagger, Q_i$), similar to the ones used in Jordan–Wigner, but without using the Pauli-Z chains maintaining anti-commutation relations (Ref. [516] shows how to re-incorporate anti-commutation properties, but this is not used in QEB-ADAPT-VQE). Therefore one can define

$$\begin{aligned} Q_i^\dagger &= \sigma_i^- = \frac{1}{2}(X_i - iY_i) \\ Q_i &= \sigma_i^+ = \frac{1}{2}(X_i + iY_i). \end{aligned} \quad (169)$$

To construct operator pools from these, Ref. [516] proceeds by producing skew-Hermitian operators (to later be exponentiated) of the type

$$\begin{aligned} T_i^k &= Q_k^\dagger Q_i - Q_i^\dagger Q_k \\ T_{ij}^{kl} &= Q_k^\dagger Q_i^\dagger Q_j Q_l - Q_i^\dagger Q_j^\dagger Q_k Q_l. \end{aligned} \quad (170)$$

From there, we can derive the parameterized single and double qubit excitations:

$$\begin{aligned} A_{ik}(\theta_a) &= e^{T_i^k(\theta_a)} \\ A_{ijkl}(\theta_a) &= e^{T_{ij}^{kl}(\theta_a)}. \end{aligned} \quad (171)$$

Operators from the pool are selected as in ADAPT-VQE, with the gradient computation rule identical to Eq. (166), and with the additional step mentioned above (although it is unclear from Ref. [344] whether this additional step has a significant impact on performance). Overall, this method was shown to achieve a lower number of CNOT gates and parameter count than qubit-ADAPT-VQE. It also shows again that the inclusion of Pauli-Z chains into ansatz operators does not matter as long as there is enough variational flexibility for the correct phase to be recovered (which is likely not applicable using fermionic operators in the pool).

Entangler pool compression: Research has also focused on addressing the wider question of operator pool optimization for adaptive ansätze. An interesting method that achieves up to nearly two orders of magnitude reduction in the operator pool size is the Mutual Information (MI) assisted Adaptive VQE [541] (it is worth noting that the method described below can be applied to any adaptive ansatz described in this review). The idea behind this method is to screen operators in the pool based on their mutual information [542–544]. Operators which have a higher correlation of information across the qubits it acts on for an approximate wave function are likely to contribute significantly more to the overall energy estimate. To implement this MI screening, one must first compute an approximation of the ground state wave function using the Density Matrix Renormalization Group (DMRG) method (see Section 2.4). From the DMRG wave function, the MI of each qubit pair must then be computed as

$$I_{ij} = \frac{1}{2} (S(\rho_i) + S(\rho_j) - S(\rho_{ij})) (1 - \delta_{ij}), \quad (172)$$

where ρ_x represents the reduced density matrix of system x , and $S(\rho)$ represents the von Neumann entropy of ρ , δ_{ij} is the Kronecker delta. The correlation strength for each operator τ_a in the pool can then be computed as

$$C(\tau_a) = \frac{1}{q_{\tau_a}(q_{\tau_a} - 1)} \sum_{ij \in Q(\tau_a), i \neq j} I_{ij}, \quad (173)$$

where q_{τ_a} is the Pauli weight of the operator (or the number of qubit it acts on), and $Q(\tau_a)$ the set of qubit indices the operator acts on. Operators can then be ranked based on both their gradient (as in other adaptive methods) and their correlation strength. In Ref. [541], Zhang et al. suggest to cut off a certain percentage of operators with low correlation strength which they optimize numerically.

6.3.2. Iterative Hamiltonian dressing (iterative Qubit Coupled Cluster (iQCC) and extensions)

The QCC ansatz [454] is not strictly speaking an adaptive ansatz in that its structure is not decided throughout the optimization process. However, because pre-processing of the relative importance of each of the operators incorporated in the ansatz is required to make it tractable, we thought it is more appropriate to include it in this section. It also makes it easier to present the iterative QCC, which is indeed an adaptive ansatz that makes use of the concept of Hamiltonian dressing to grow the Hamiltonian instead of the ansatz at each iteration.

The Qubit Coupled Cluster ansatz: QCC was introduced as a means to bypass the detrimental non-local action arising in the UCC ansatz, and the large number of two-qubit gates that ensue. It is in essence very similar to the qubit-ADAPT-VQE, although with the addition of a Qubit Mean-Field wave function, and initially presented in Ref. [454] without adaptive looping and without prior selection of a restricted pool of Pauli strings.

The central premise of QCC is that, unlike UCC, it uses spin operators directly to construct the ansatz instead of fermionic excitation operators which are then transformed. The first step to construct this ansatz is to separate the wave function into mean field and correlated components, such that

$$|\psi(\boldsymbol{\tau}, \boldsymbol{\Omega})\rangle = U(\boldsymbol{\tau})|\boldsymbol{\Omega}\rangle. \quad (174)$$

The mean field component $|\Omega\rangle$ can be simply prepared through single qubit rotations on each individual qubit in the wave function.

$$\begin{aligned} |\Omega\rangle &= \bigotimes_j^N |\Omega\rangle_j \\ |\Omega\rangle_j &= R(\theta_j, \phi_j) |0\rangle_j \\ |\Omega\rangle_j &= \cos \frac{\theta_j}{2} |0\rangle_j + e^{i\phi_j} \sin \frac{\theta_j}{2} |1\rangle_j, \end{aligned} \quad (175)$$

with θ, ϕ parameters for the rotation of each qubit. The correlation component consists in an entangling block unitary, built from amplitude parameters $\tau = \{\tau_a\}$, and a set of corresponding Pauli strings $\{P_a\}$, as

$$U(\tau) = \prod_a^P e^{-i\tau_a \hat{P}_a/2}, \quad (176)$$

with P the total number of Pauli strings used. With this we can define the VQE problem using the QCC ansatz:

$$E_{\text{QCC}} = \min_{\tau, \Omega} \langle \Omega | U^\dagger(\tau) \hat{H} U(\tau) | \Omega \rangle. \quad (177)$$

Because the number of possible Pauli strings candidates for this ansatz scales $\mathcal{O}(4^N)$, it is necessary to carefully select the appropriate operators to incorporate into the entangling block. A first proposal in Ref. [454] is to look at the (negative) contribution of using each operator individually once its amplitude parameter (τ_a) is optimized. This rapidly becomes computationally prohibitive, and as such Ryabinkin et al. [454] also propose to rank operator contributions based on their first and second order gradients. In a manner that is similar to that presented in ADAPT-VQE, the first order derivative with respect to the amplitude parameter τ_a at $\tau_a = 0$ is given by

$$\frac{\partial E(\tau_a, \hat{P}_a)}{\partial \tau_a} = \langle \Omega^* | \left(-\frac{i}{2} [\hat{H}, \hat{P}_a] \right) | \Omega^* \rangle \quad (178)$$

where $|\Omega^*\rangle$ is the optimized mean-field qubit wave function with respect to θ and ϕ . This remains exponential in scaling, and further restrictions on the Pauli strings to be included in the pool must be implemented. In particular, Ref. [287] introduces the concept of direct interaction set (DIS), which aims at excluding from the operator pool all those with zero gradients, and grouping the remaining $\mathcal{O}(4^N)$ operators into groups having identical gradients up to a sign. This results in a polynomial scaling number of groups ($\mathcal{O}(N^4)$) each composed of 2^{N-1} operators.

iQCC: The iterative Qubit Coupled Cluster method [287] is an improvement on the QCC method which aims to reduce the number of operators that can be included in the ansatz by incorporating operators into the Hamiltonian iteratively instead of increasing the ansatz as it is done for ADAPT-VQE and extensions. It works based on a slightly modified cost function, compared to Eq. (177):

$$E_{\text{iQCC}} = \min_{\tau} \{ \min_{\Omega} \langle \Omega | \hat{H}_d(\tau) | \Omega \rangle \}, \quad (179)$$

where $\hat{H}_d = U(\tau)^\dagger \hat{H} U(\tau)$ is called the *dressed Hamiltonian*. The dressed Hamiltonian can be computed recursively, starting from $\hat{H}_d^{(0)} = \hat{H}$, and each $\hat{H}_d^{(k)}$ being dependent on the first k amplitude in the vector τ , such that $\hat{H}_d^{(k)} = \hat{H}_d^{(k)}(\tau_{1 \rightarrow k})$:

$$\begin{aligned} \hat{H}_d^{(k)} &= e^{i\tau_k \hat{P}_k/2} \hat{H}_d^{(k-1)} e^{-i\tau_k \hat{P}_k/2} \\ &= \hat{H}_d^{(k-1)} - \frac{i \sin \tau_k}{2} [\hat{H}_d^{(k-1)}, \hat{P}_k] \\ &\quad + \frac{(1 - \cos \tau_k)}{2} \hat{P}_k [\hat{H}_d^{(k-1)}, \hat{P}_k], \end{aligned} \quad (180)$$

where the exponential terms $e^{i\tau_k \hat{P}_k/2}$ are called the generators of the iQCC process, and akin to the elements of the QCC ansatz presented in Eq. (176). The process for completing the iQCC optimization process works as follows:

- Inputs are the mean-field parameters of $|\Omega\rangle$ optimized at the previous iteration, and the dressed Hamiltonian of the previous iteration $\hat{H}_d^{(k-1)}$ (or $\hat{H}_d^{(0)}$ at the start).
- Using the optimized qubit mean-field wave function, and $\hat{H}_d^{(k-1)}$, use a screening procedure to select a number of generators g with the largest gradient from an operator pool. The value for g is at the discretion of the user (including $g = 1$, although Ryabinkin et al. [287] point out that increasing g accelerates convergence).
- Using the selected generators, construct the appropriate ansatz and proceed to a VQE optimization on QPU of both the mean-field parameters and g generator amplitudes τ . Ryabinkin et al. [287] suggest using random initializations for the amplitudes or set $\tau = 0$, if random initialization does not start from a lower energy than at the previous iteration.

- Once parameters are optimized, incorporate the g generators into the Hamiltonian by computing $\hat{H}_d^{(k)}$ using Eq. (180), and begin a new iteration unless convergence is reached.

Using this method, one can enforce a fixed size for the ansatz, at each iteration, determined uniquely by the constant g . This provides many advantages from a NISQ perspective as it allows to directly adjust the size of the ansatz in each iteration to the noise level acceptable (and which can be mitigated) for the device being used. This comes however with two significant caveats. First, there is no guarantee that the method can converge to the exact solution, despite successfully demonstrating the method works for small systems [287]. Secondly, the number of terms in the Hamiltonian grows in a somewhat prohibitive manner. It is estimated that the Hamiltonian grows by a factor of $\sim (3/2)^g$ at each iteration, which means that for a number of iterations k , and with $\mathcal{P}$ the number of terms in $\hat{H}_d^{(0)}$, the Hamiltonian scales $\mathcal{O}(\mathcal{P}(3/2)^{gk})$, or exponentially in the number of iterations (with an obvious upper bound of $\mathcal{O}(4^k)$). This was improved upon in Ref. [545], where Lang et al. show that using Truncation of the Baker–Campbell–Hausdorff expansion, one can restrict the increase in the number of terms by a factor $\mathcal{O}(g^2)$ at each iteration, and hence an overall scaling of $\mathcal{O}(g^{2k})$. They show, using an active space of 4 electrons on 4 molecular orbitals for H_2O , with $k = 1$, one can recover most of the electron correlation energy. Reaching chemical accuracy requires between $k = 5$ and $k = 10$ depending on the bond distance. That considered, numerical analysis of the growth in the number of terms in the Hamiltonian when using iQCC shows a plateau after a number of iterations, owing to the fact that many terms will have negligible weights, resulting in a low polynomial scaling [546]. Finally, in Ref. [547], Ryabinkin et al. propose to use ex-post corrections on the energy obtained from iQCC based on perturbation theory, to reduce the number of iteration k required for the implementation of the method.

Overall, despite the promising fixed-length ansatz, the viability of the method depends on further research regarding scalability and extension of the method to very large systems. The analysis presented by Genin et al. [546] suggests that iQCC could compete with methods such as DFT in terms of scaling and accuracy using a minimum of 72 fully-connected and corrected qubits.

ClusterVQE: In Ref. [306], Zhang et al. present a method that borrows ideas from ADAPT-VQE, iQCC and Mutual Information (MI) assisted Adaptive VQE to optimize the use of quantum resources. The first step in this method is to create a cluster of qubits (across the qubit register of the QPU used) that have strong MI within, and low MI across, each cluster. MI was shown to be a strong metric to measure the correlation between these quantum clusters [544], computed as presented in Eq. (172). As described in Section 6.3.1, the MI can be initially approximated using a DMRG approximation of the ground state wave function. Once the MI for each qubit pair has been computed, a conventional clustering algorithm can be used to create a set of clusters that minimizes the MI between clusters (sum over MIs between inter-cluster qubits).

The idea behind ClusterVQE is to separate the pool of operators (akin to those found in other adaptive methods) into entanglers that act within a cluster of qubits (noted U_{c_i} , i labeling a specific cluster) and those acting on two different clusters, coupling them (noted $U_{c_{ij}}$). The former are used to grow the ansatz, similarly to qubit-ADAPT-VQE (see Section 6.3.1), the latter are used to dress the Hamiltonian similarly to iQCC (see Section 6.3.2). The ClusterVQE optimization problem can then be written as

$$E = \min_{\theta, \phi} \langle \psi_{\text{HF}} | \prod_i U_{c_i}^\dagger(\theta_i) \hat{H}^d(\phi) \prod_i U_{c_i}(\theta_i) | \psi_{\text{HF}} \rangle, \quad (181)$$

with, the dressed Hamiltonian

$$\hat{H}^d(\phi) = \prod_{i \neq j} U_{c_{ij}}^\dagger(\phi_{ij}) \hat{H} \prod_{i \neq j} U_{c_{ij}}(\phi_{ij}). \quad (182)$$

The optimization process works similar to the methods mentioned above, where a pool of entanglers is first created (Zhang et al. [306] recommend selecting qubit-operators from the pool generated by single and double fermionic excitations) and ranked based on their gradients, computed as in Eq. (166). Entanglers that couple qubit clusters are used to dress the Hamiltonian, others are used to grow the ansatz to be implemented on the quantum computer. Once top entanglers have been selected and integrated, a VQE optimization is conducted. The wave function obtained is then used to update the MI (which can be computed from the one and two-body reduced density matrices) and a new loop begins.

A first observation to make on the ClusterVQE is that it allows separating the ansatz into smaller, shallower depth quantum circuits corresponding to each qubit cluster [306]. The quantum circuit for each cluster can be optimized separately as their gradients are independent of the rest of the ansatz, allowing for implementation of ClusterVQE on smaller quantum computers. It is worth noting however that doing so means that the number of measurements required must be multiplied by the number of qubit-clusters, though these can be parallelized. It also means that the number of terms in the Hamiltonian is lower than the iQCC method as only certain entanglers are integrated into the dressed Hamiltonian.

Finally, Zhang et al. [306] show numerically that ClusterVQE can achieve similar accuracy as iQCC but with significantly lower numbers of Hamiltonian terms. The shallower circuits also offer some noise resilience compared to VQE and qubit-ADAPT-VQE, though iQCC remains the most resilient due to its fixed ansatz depth. A few caveats are worth mentioning, however. Despite promising features, the dressed Hamiltonian still grows exponentially in the number of cross-qubit-cluster entanglers; while these have been shown to be few for small systems (e.g. LiH in Ref. [306]) there is at this stage no way to tell how this method scales on large, highly entangled systems. Also noted by Zhang et al. a potential bottleneck for the method is the inter-MI clustering process which could become too costly to compute accurately enough.

6.3.3. Ansatz structure optimization

Several methods have been developed to iteratively update the structure of the ansatz throughout the optimization process. Unlike the adaptive ansätze presented in the previous section, these methods do not propose to build the circuit layer by layer but start from an initial guess of the ansatz structure to progressively learn an optimal one. We outline a few of these methods at a high level below.

RotoSelect [134]: this method is built as an extension to the RotoSolve optimizer [128,134], which we present in more details in Section 7.4.2. Rotosolve uses the fact that the expectation value of the Hamiltonian is a 2π periodic sine-type function of a given parameter to find the optimal value. This can be done by sampling three different values of the parameter and finding the minimum using trigonometric relationships. RotoSelect extends that idea, by optimizing not only the parameter but also the Hilbert space dimension of the rotation gate. In other words, the algorithm finds which orientation, X, Y, or Z, allows for the largest reduction in energy for a given rotation gate and replaces the orientation of the gate accordingly. This is associated with an additional sampling cost on the initial iterations of the VQE, but has been shown in Ref. [134] to significantly improve convergence rate over RotoSolve and other gradient-based optimizers. These results were tested on HEA, and could be different on the likes of UCC-based ansätze.

Variable Ansatz (VAns) [534]: The VAns is another algorithm that proposes adapting the structure of the ansatz as the optimization proceeds. Unlike RotoSelect, mentioned above, it not only changes the axes of rotation gates but also incorporates and removes gates throughout the process. The first step in this method is to define a dictionary of parameterized gates. From an initial circuit configuration (which can be flexibly designed), pre-defined insertion rules stochastically select gates from the dictionary to add them to the circuit. Additional simplification rules are used to identify gates with lower impact on the resulting energy and remove them from the ansatz. The ansatz is primarily designed for use on Quantum Machine Learning tasks which cannot benefit from a problem tailored ansatz (as it is the case for VQE), but is also shown to outperform HEA on the Ising and Heisenberg models, and on H_2 and H_4 [534]. The key advantage of this approach is the large flexibility it provides in terms of designing rules, allowing for efficient design and tailoring to any learning task, while offering some degree of resilience to the barren plateau problem. The cost of implementing the method is unclear however, and while it reduces the burden on quantum resources, it could result in a large pre-factor cost for its implementation, requiring further research.

Machine learning of ansatz structure and Quantum Architecture Search (QAS): These comprise a series of methods aiming at learning an optimal ansatz structure using machine learning tools. While it would probably deserve a review in its own right, we outline here some key publications regarding this family of approaches. The Evolutionary VQE (EVQE) [548] uses evolutionary programming to dynamically generate and optimize the ansatz and is shown to produce ansätze which are nearly twenty times shallower (with twelve times fewer CNOT gates) than UCCSD. Rattew et al. [548] also show numerically a significant ability of the EVQE to learn and thereby reduce the impact of inherent quantum noise, all while providing resilience to the barren plateau problem. Similarly, the Multi-objective Genetic VQE (MoG-VQE) [549] uses a genetic algorithm to learn the circuit topology of a series of blocks constituting the ansatz. This results in blocks of gates resembling the HEA structure, but only spanning a limited number of qubits. The method achieves a significant reduction in two-qubit gates compared to HEA (showing that one can reach chemical accuracy on a 12-qubit version of LiH with only 12 CNOTs [549]).

QAS was initially developed in Ref. [550] as a means to improve both trainability and noise resilience of variational quantum algorithms. Similar to the methods described above, QAS modifies the optimization problem by incorporating the structure of the quantum circuit in the cost function. In Ref. [551], Zhang et al. extend this idea by creating the Differentiable QAS (DQAS), which is based on the differentiable neural architecture search (DARTS) [552]. The idea is to relax the discrete search space of network architectures onto a continuous and differentiable domain, accelerating QAS [551]. Additional work on QAS includes a version based on Neural Predictors [553], on Deep Reinforcement Learning [554], and on meta-learning [555]. QAS can also be used to improve noise resilience of an ansätze [556]. Overall, while very promising, machine learning based ansätze all suffer from a common pitfall: the cost of training machine learning models to optimize ansätze for large-scale systems remains largely untested. Studying the training cost and cost scaling for these methods would be an interesting research avenue to warrant their applicability on larger systems than has been analyzed so far.

6.4. Discussion on ansätze used in the VQE context

There has been a very large amount of research published on VQE ansätze over recent years, we summarize in this subsection the key findings from our research. The k-UpCCGSD ansatz [80] (and extensions) have been numerically shown to achieve excellent accuracy while offering linear scaling, thereby arguably offering the best trade-off in cost to accuracy among ansätze proposed at the time of writing. Adaptive ansätze could place themselves as a reliable alternative subject to further studies on their expected computational cost. Layered HEA are likely not suitable for large systems. As they are not physically motivated, they require a large number of parameters and need to span a significant proportion of the Hilbert space to guarantee a good ground state approximation can be reached. This also makes HEA particularly prone to barren plateaus [127]. There are good arguments however to suggest that VQE may be resilient to barren plateaus if the

right design choices are made in the ansatz (except for Noise-Induced Barren Plateau [442]). In particular, these are; local encodings for the Hamiltonian [87,91], an ansatz that is problem tailored (e.g. UCC), or is constructed adaptively during the optimization (e.g. ADAPT-VQE [217] and others), using specific initialization techniques (e.g. Refs. [90,102]), and use of a local mapping for the Hamiltonian (e.g. Bravyi–Kitaev).

Qubit-based ansätze [287,454,545] (or qubit-excitations based ansätze) could be more appropriate for the NISQ era than fermion based (or hardware efficient) ansätze. That is because they offer the option to parameterize the ansatz with respect to each Pauli string used in the construction. Given enough variational flexibility, this allows forgoing Pauli-Z chains used for maintaining the antisymmetric relationship of fermionic operators under the Jordan–Wigner mapping. As such, each operator is acting on very few qubits (up to four in the Jordan–Wigner mapping, instead of $\mathcal{O}(N)$ when using a fermionic ansatz), reducing the reliance on costly two-qubit gates and allowing depth reduction. These however tend to require a larger number of parameters and operators, and as such may be more cumbersome to optimize and have slightly worse computational scaling.

Adaptive ansätze have been a significant focus of recent research. While they have brought several critical breakthroughs, the scaling of the number of measurements required remains unclear. We find that the main advantages of adaptive ansätze are the reduction in number of parameters (though as far as we have seen, less than an order of magnitude), and numerically demonstrated resilience to barren plateau. This however comes at a significant computational cost of gradient estimations (scaling with the size of the Hamiltonian) and repeated VQE optimizations (shallow at the beginning of the process, but not at the end). The relative benefit of adaptive ansatz should be studied in more detail.

Some research has been conducted on the impact of initialization on the performance of different ansätze, in particular, Choquette et al. [297] numerically show that a computational basis initialization for HEA has a minor impact on target state fidelity compared to Hartree–Fock initialization. Of course, UCC-based ansätze are motivated as an evolution of a reference state, and, in general, apply to a qubit register initialized as the Hartree Fock wavefunction [37]. As a side note, Murta and Fernández-Rossier [240] propose a quantum routine to model the Gutzwiller wavefunction [557] on a quantum device. This wavefunction is a physically-motivated model which can be applied to a variety of lattice systems [558]. While not providing a full parameterized ansatz, it was shown to provide a better approximation than a mean-field method for the ground state of the Fermi–Hubbard model, and could therefore be used as an initialization strategy for lattice model VQE implementations [240]. Further research on the preparation of the Gutzwiller wavefunction is presented in Ref. [559].

7. Optimization strategies

The VQE is in essence an optimization problem, it aims at heuristically constructing an approximation of an electronic wavefunction through iterative learning of ansatz parameters. For the algorithm to be viable, it must be that it can learn a good enough approximation of the solution within a tractable number of learning steps. It was already demonstrated that optimization of the variational quantum ansätze is NP-hard [93], meaning that there exist at least some problems in which finding an exact solution for the VQE problem is intractable. As such, efficient optimization strategies that provide a well-approximated solution within an acceptable number of iterations are essential for any variational algorithms to be put into practice. Compared to the conventional numerical optimization problem, however, optimizing the expectation value of a variational quantum ansatz faces additional challenges:

- Sampling noise and gate noise on NISQ devices disturb the landscape of the objective function. Such noise can be detrimental to the convergence of optimization [442,560], and could limit the scope for quantum advantage [561].
- While the precision of conventional numerical optimization is generally not considered a problem, the precision of the measured expectation value is limited by the sample shot number. The cost of optimization is heavily dependent on the precision required for optimization.
- Related to the point above, the landscape of the expectation value of variational ansatz may cause the vanishing of gradient very easily as a result of the barren plateau problem [86] (see Section 6.1 for further details).

On the flip side, studies from recent years show the landscape of expectation value has some analytical properties that are useful to extract information, such as evaluating gradients directly on quantum devices [133,134,456,562]. In addition, the ansatz' landscape can be efficiently approximated to accelerate the convergence [135,563]. Utilizing such prior knowledge helps to develop efficient optimization strategies for variational quantum algorithms. This section presents the most relevant optimizers used in the context of the VQE and recent studies of optimization strategies, focusing on the strategies that target fast convergence rates. Some optimizers are developed to reduce the shot numbers [121,375], for which specific applications are discussed in Section 5.1. We also compare these proposed algorithms and discuss the current challenge facing the optimization of variational quantum algorithms.

7.1. Background and notation

The objective function of a variational algorithm is constructed conventionally based on the measurement outcome. Denote $\mathbf{O}(\boldsymbol{\theta}) = (\hat{O}_1(\boldsymbol{\theta}^{(1)}), \hat{O}_2(\boldsymbol{\theta}^{(2)}), \dots, \hat{O}_a(\boldsymbol{\theta}^{(a)}))$ are the observables used to compose the objective function and a is the number of observables. The objective function is given by

$$\mathcal{L}(\boldsymbol{\theta}) = C(\mathbf{O}(\boldsymbol{\theta})) \quad (183)$$

where C is a function that maps the observed expectation value to the objective function, and usually have the simple linear form

$$C(\mathbf{X}) = \sum_i c_i X_i \quad (184)$$

where c_i is a constant defined by the problem as the coefficient of each measurement expectation value, X_i is the i th component of $\mathbf{X}$. Note that such linear form preserves the analytical properties and it is essential for using the analytical methods to directly calculate the gradient or implement analytical gradient free optimization strategy [133–135,456,562,564].

A measurement expectation value is given by

$$\langle \hat{O}_k(\boldsymbol{\theta}^{(k)}) \rangle = \langle \psi_0 | U^{(k)\dagger}(\boldsymbol{\theta}^{(k)}) \hat{M}_k U^{(k)}(\boldsymbol{\theta}^{(k)}) | \psi_0 \rangle, \quad (185)$$

where $|\psi_0\rangle$ is the initial state on the quantum computer. $\hat{M}^{(k)}$ is a Hermitian measurement operator, usually chosen to be the tensor product of Pauli operators to match the physical measurement implementation on quantum hardware. $U_k(\boldsymbol{\theta}^{(k)})$ is the variational ansatz defined as

$$U^{(k)}(\boldsymbol{\theta}^{(k)}) = \prod_j U_j^{(k)}(\theta_j^{(k)}) \quad (186)$$

and each U_j is a quantum gate, which is generalized as

$$U_j^{(k)}(\theta_j^{(k)}) = \exp(i\theta_j^{(k)} P_j^{(k)}) \quad (187)$$

where $P_j^{(k)}$ is a Hermitian matrix, usually is a tensor product of Pauli operators.

It is sometimes convenient to utilize the superoperator formalism and consider the noise into the optimization process, the expectation value can be written as

$$\langle \hat{O}_k(\boldsymbol{\theta}^{(k)}) \rangle = \text{Tr}[\hat{M}_k \Phi^{(k)}(\boldsymbol{\theta}^{(k)}) \rho_0], \quad (188)$$

where ρ_0 is the initial state density operator and $\Phi^{(k)}(\boldsymbol{\theta}^{(k)})$ denote the transformation matrix, which is given by

$$\Phi^{(k)}(\boldsymbol{\theta}^{(k)}) = \prod_j \Phi_j^{(k)}(\theta_j^{(k)}). \quad (189)$$

For devices that support single shot readout, each sample the quantum device would yield a bit string $\mathbf{s}$. For each string $\mathbf{s}$ a measurement value could be calculated with $M_i(\mathbf{s})$, and the expectation value is the average of each $M_k(\mathbf{s})$.

$$\langle \hat{O}_k(\boldsymbol{\theta}^{(k)}) \rangle = \sum_j \text{Prob}(\mathbf{s}(\boldsymbol{\theta}^{(k)})) = b_j M_k(b_j) \quad (190)$$

where $b_j \in B$, B covers all possible single shot bit string (all binary numbers from 0 to 2^{n-1}) of the measurement outcome.

Due to the physical implementation from the quantum hardware, not all quantum computing systems support single-shot readout. Some systems can only yield expectation value by averaging the signal from the readout. For example, some NMR systems use an ensemble of molecules to implement quantum computing and cannot read the state of each single molecule [565,566]. In practice, the quantum hardware system may directly yield an expectation value, and the measurement approaches vary from different physical systems.

In the following discussion, the upper label “(k)” is simplified when there is no ambiguity.

7.2. Gradient evaluation

A significant amount of numerical optimization methods requires the knowledge of the gradient of the objective function at a given parameter value. Therefore the evaluation of gradient is rather important and we would cover this topic in a separate subsection.

7.2.1. Stochastic approximation methods

Stochastic approximation (SA) is a family of methods used to reconstruct the properties of the expectation value of a function that depends on some random variables. Instead of measuring the expectation values directly, SA recovers the properties of the expectation value with random sampling. SA has been widely used for big data and machine learning applications when the objective function is too costly to evaluate directly. In the context of variational quantum algorithms, the objective function is some function of the expectation value of the quantum state; evaluating the expectation value of a variational ansatz is expensive. Therefore SA naturally fits into the context of variational quantum algorithms and has been used to approximate the gradient of the expectation values.

Finite difference stochastic approximation (FDSA) One of the simplest methods to approximate the gradient is evaluating two function points and using its difference as the approximated gradient. Finite difference stochastic approximation (FDSA) [567] is given by

$$(g(\theta_t))_j = \frac{\mathcal{L}(\theta + c_t \mathbf{e}_j) - \mathcal{L}(\theta - c_t \mathbf{e}_j)}{2c_t}, \quad (191)$$

where $(g(\theta_t))_j$ is j th component of the gradient at point θ_t . $\mathbf{e}_j$ is the unit vector of the j th dimension, c_t is the displacement value that generally decrease with iteration step number t [567].

Simultaneous perturbation stochastic approximation (SPSA) SPSA algorithm approximates the gradient with only two measurements of the objective function. Instead of choosing the measurement points symmetrically, SPSA uses a small random vector to perturb the objective function. While FDSA calculates the gradient at one direction of the variable, SPSA calculates a direction composed of multiple variables.

$$(g(\theta_t))_j = \frac{\mathcal{L}(\theta + c_t \Delta_j) - \mathcal{L}(\theta - c_t \Delta_j)}{2c_t(\Delta_j)}, \quad (192)$$

where Δ_t is a random perturbation vector. The SPSA algorithm needs to measure two points per iteration, which is friendly to variational quantum algorithms. Also it can be applied to noisy optimization [568–570].

The SPSA methods can approximate the preconditioned gradient for second order optimization methods [571]. Explanation of second order optimization methods are covered in Section 7.3.2. Consider the preconditioned gradient $g'(\theta) = F^{-1}g(\theta)$ where F is some metric with information from second order derivatives of function $f(\theta)$. Denote $\tilde{F}$ as the approximated value for F , we have

$$\tilde{F}^{(k)} = \frac{\delta f}{2\epsilon^2} \frac{\Delta_1^{(k)} \Delta_2^{(k)T} + \Delta_2^{(k)} \Delta_1^{(k)T}}{2} \quad (193)$$

where

$$\delta f^{(k)} = f(\theta^{(k)} + \epsilon \Delta_1^{(k)} + \epsilon \Delta_2^{(k)}) \quad (194)$$

$$- f(\theta^{(k)} + \epsilon \Delta_1^{(k)}) \quad (195)$$

$$- f(\theta^{(k)} - \epsilon \Delta_1^{(k)} + \epsilon \Delta_2^{(k)}) \quad (196)$$

$$+ f(\theta^{(k)} - \epsilon \Delta_1^{(k)}) \quad (197)$$

The approximated value is estimated from all previously evaluated values with an exponentially smooth estimator

$$\bar{F}^{(k)} = \frac{k}{k+1} \bar{F}^{(k-1)} + \frac{1}{k+1} \tilde{F}^{(k)} \quad (198)$$

7.2.2. Analytical gradient calculation

The gradient of measurement observables can be evaluated on quantum computers directly by utilizing the analytical property of the ansatz. For ansätze made from a sequence of parameterized quantum gates, the partial derivative is

$$\frac{\partial \langle \hat{O}_k(\theta) \rangle}{\partial \theta_j} = 2 \operatorname{Im}(\langle \phi_0 | V_k^{j\dagger}(\theta) \hat{M}_k U(\theta) | \phi_0 \rangle), \quad (199)$$

where the operator $V_k^j(\mathbf{t})$ is defined as inserting P_k^j between the $(j-1)$ th term and j th term, which comes from the derivative of the j th term:

$$V_k^j(\theta) = e^{i\theta_{N_k} P_k^{N_k}} \dots e^{i\theta_j P_k^j} P_k^j e^{i\theta_{j-1} P_k^{j-1}} \dots e^{i\theta^1 P_k^1}. \quad (200)$$

The difficulty of evaluating the gradient directly is that $V_k^{j\dagger}(\theta) \hat{M}_k U(\theta)$ is usually not Hermitian, therefore there is no known method that can directly convert it to a quantum circuit. Two methods to analytically measure the gradient have been developed, and it is worth noticing that these two methods are measuring the gradient with the exact same mechanism, but with direct and indirect measurements [572]. Here, direct measurement means the observed quantity is encoded into a single quantum observable, while indirect measurement calculates the quantity with multiple quantum observables.

Direct analytical gradient measurement The imaginary part of $\langle \phi_0 | V_k^{j\dagger}(\theta) \hat{M}_k U(\theta) | \phi_0 \rangle$ can be directly evaluate by introducing an ancillary qubit [456]. This is done by first preparing the ancillary qubit into the $1/(\sqrt{2})(|0\rangle + |1\rangle)$ state, and preparing the rest of quantum register into $|\phi(\theta)\rangle$ state. Now the system has the state

$$|\psi\rangle = \frac{\sqrt{2}}{2}(|0\rangle + |1\rangle) \otimes |\psi(\theta)\rangle. \quad (201)$$

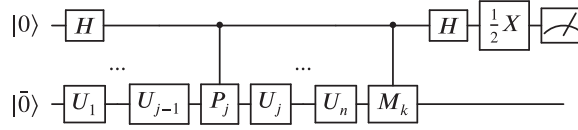


Fig. 13. Quantum circuit that evaluates $\text{Im}(\langle \phi_0 | V_k^{\dagger}(\theta) \hat{M}_k U(\theta) | \phi_0 \rangle)$.

We can then apply the P_j gate controlled by the ancillary qubit and apply the remaining unitary. Finally we apply the measurement observable controlled by the ancillary qubit. This gives the new state

$$|\psi\rangle = \frac{(|0\rangle \otimes U(\theta)|\psi_0\rangle + |1\rangle \otimes \hat{M}_k V_k^j(\theta)|\phi_0\rangle)}{\sqrt{2}}. \quad (202)$$

Now we apply the Hadamard gate to the ancillary qubit, which gives

$$|\psi\rangle = \frac{|0\rangle \otimes (U|\phi_0\rangle + \hat{M}_k V_k^j(\theta)|\phi_0\rangle) + |1\rangle \otimes (U|\phi_0\rangle - \hat{M}_k V_k^j(\theta)|\phi_0\rangle)}{2}. \quad (203)$$

The imaginary part of $\langle \phi_0 | V_k^{\dagger}(\theta) \hat{O}_i U(\theta) | \phi_0 \rangle$ is now encoded in the ancillary qubit in the Y-basis (see Fig. 13).

Indirect analytical gradient measurement To illustrate this method first we make some modifications to our notations. We now absorb the common term of U and V into the state and measurement operator, and define

$$\begin{aligned} W_{k,1}(\theta) &= \exp(i\theta_{j-1} P_k^{j-1}) \cdots \exp(i\theta^1 P_k^1) \\ W_{k,2}(\theta) &= \exp(i\theta_n P_k^n) \cdots \exp(i\theta_j P_k^j) \\ |\phi(\theta)\rangle &= W_{k,1}(\theta) |\phi_0\rangle \\ Q_k(\theta) &= W_{k,2}^{\dagger} \hat{M}_i W_{k,2}. \end{aligned} \quad (204)$$

Then we have

$$\frac{\partial \langle \hat{O}_k(\theta) \rangle}{\partial \theta_j} = 2 \text{Im}(\langle \phi(\theta) | Q_k(\theta) P_k | \phi(\theta) \rangle). \quad (205)$$

We use the following construct which is related to $\langle \phi(\theta) | P_k Q_k(\theta) | \phi(\theta) \rangle$:

$$\langle \phi(\theta) | (\mathbb{1} \mp iP_k) Q_k(\theta) (\mathbb{1} \pm iP_k) | \phi(\theta) \rangle = \mp \text{Im}(\langle \phi(\theta) | P_k Q_k(\theta) | \phi(\theta) \rangle) + C, \quad (206)$$

where $C = \langle \phi(\theta) | Q_k(\theta) | \phi(\theta) \rangle + \langle \phi(\theta) | P_k Q_k(\theta) P_k | \phi(\theta) \rangle$ is a constant value. For single qubit gate given by $U_j(\theta_j) = \exp(i\theta_j P_k)$, $P_k \in X, Y, Z$ where X, Y, Z are Pauli matrices and with exactly two eigenvalues, we have [562]

$$\mathbb{1} \pm iP_k = \exp(\mp \frac{\pi}{4} iP_k), \quad (207)$$

therefore

$$\langle \phi(\theta) | (\mathbb{1} \mp iP_k) Q_k(\theta) (\mathbb{1} \pm iP_k) | \phi(\theta) \rangle \quad (208)$$

$$= \langle \phi(\theta \pm \frac{\pi}{2} \mathbf{e}_j) | Q_k(\theta \pm \frac{\pi}{2} \mathbf{e}_j) | \phi(\theta \pm \frac{\pi}{2} \mathbf{e}_j) \rangle, \quad (209)$$

and

$$\frac{\partial \langle \hat{O}_k(\theta) \rangle}{\partial \theta_j} = \langle \hat{O}_k(\theta + \frac{\pi}{2} \mathbf{e}_j) \rangle - \langle \hat{O}_k(\theta - \frac{\pi}{2} \mathbf{e}_j) \rangle. \quad (210)$$

The parameter shift rule can be generalized to a general parameter shift rule and give gradient g_{gPSR} to obtain higher-order derivatives [562,573].

$$g_{\text{gPSR}}(\theta, \gamma_1, \gamma_2) := r[L(\theta + \gamma_1) + L(\theta + \gamma_2)] \quad (211)$$

where $2r$ is the difference of the eigenvalue of the gate generator G . For single-qubit gates with Pauli operators as the generator, $r = 1$.

$$g_{\text{gPSR}}(\theta, \gamma_1, \gamma_2) = \frac{\sin(2r\gamma_1) + \sin(2r\gamma_2)}{2} g^{(1)}(\theta) - \frac{\cos(2r\gamma_1) - \cos(2r\gamma_2)}{4r} g^{(2)}(\theta), \quad (212)$$

and because of the sinusoidal property of the expectation value, the k th order derivative has a constant multiplier different to the $k - 2$ th order derivative.

$$g^{(k+2)} = -\frac{1}{4r}g^{(k)}. \quad (213)$$

So far the analytical gradient methods only apply to parameterized single-qubit gates. The relation presented in Eq. (207) holds only when P_k has exactly two different eigenvalues. To use the parameter shift rule for arbitrary two-qubit gates, one can decompose the multi-qubit gate into a sequence of single-qubit gate and product of the same Pauli matrices, which always has two different eigenvalues [129]. When the operators have 3 different eigenvalues, the gradient can be evaluated with a modified 4-value shift rule [130]. For other operators, the value P can be polynomially expanded into a linear combination of low-rank (2 or 3 eigenvalues) operators [130]. Alternatively, the objective function can be decomposed into trigonometric polynomials, and the gradient can be evaluated with trigonometric interpolation methods [574,575].

It is worth mentioning that for VQE, the parameter shift rule can be implemented before encoding the fermionic excitation to the qubits, which reduces the required measurement amount [576].

7.3. Gradient-based searching strategy

Gradient-based optimization strategies utilize information from cost function derivatives. First-order optimizers utilize only the first-order derivatives of the cost function, and second-order optimizers utilize both first-order and second-order derivatives, at the cost of computing the second-order derivatives. In this section, we discuss some of the popular gradient-based optimizers.

7.3.1. First order optimizers

First-order optimizers for variational ansatz are mostly borrowed from the deep learning communities. These methods have been widely used for the early studies of variational quantum algorithms.

Simple gradient descent Simple gradient descent is the simplest first-order method searching for local minimum with gradient [577,578]. The algorithm takes a step towards the opposite direction of the gradient, and the step size is calculated based on the absolute value of the gradient and a meta-parameter η usually referred to as the learning rate. Almost all the gradient-based methods are developed based on the idea of simple gradient descent.

Algorithm 1: Simple gradient descent

```

 $\eta$ : Learning rate.
while Not converged do
    Calculate gradient  $g_t = g(\theta_t)$ ;
     $\theta_{t+1} = \theta_t - \eta g_t$ 
end

```

Adaptive optimizers well developed from the deep learning community are adapted to the ansatz parameter optimization.

RMSProp RMSProp is an adaptive learning rate optimization method. The RMSProp divides the gradient by the weighted moving average of the root mean square of gradient value, which enhances the direction of the gradient and reduces the significance of the gradient magnitude.

Algorithm 2: RMSProp optimizer

```

 $\gamma$ : Moving average parameter.
 $\eta$ : Learning rate.
while Not converged do
    Calculate gradient  $g_t = g(\theta_t)$ ;
     $E[g^2]_t = \gamma E[g^2]_{t-1} + (1 - \gamma)g_t^2$ ;
     $\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{E[g^2]_t + \epsilon}}g_t$ 
end

```

Adam optimizer Adaptive moment (Adam) optimizer [579] is a widely used optimizer developed from the deep learning community to solve the stochastic gradient-based optimization problem. The Adam optimizer performs efficient stochastic optimization with only first-order gradient. Adam optimizers utilize adaptive moment estimation as an extra on RMSProp.

Instead of descent to the gradient direction, it jumps towards the momentum direction, the weighted moving average of the gradient.

Algorithm 3: Adam optimizer

β_1 : Moving average parameter for past gradients.
 β_2 : Moving average parameter for past squared gradients.
while Not converged **do**
 Calculate gradient $g_t = g(\theta_t)$
 $m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t$;
 $v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$;
 $m'_t = \frac{m_t}{1 - \beta_1^t}$;
 $v'_t = \frac{v_t}{1 - \beta_2^t}$;
 $\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{v'_t} + \epsilon} m'_t$;
end

Since the Adam optimizer utilizes moving average for both momentum and magnitude estimations, it can be applied to scenarios with noisy gradients.

7.3.2. Second order optimizers

Second-order optimization techniques make use of the second-order derivative of the objective function to determine the descent direction.

Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm The BFGS algorithm [422–425] is a gradient-based iterative approach to solve standard nonlinear optimization problems. The direction of each step is obtained by preconditioning the gradient with curvature information. Preconditioning the gradient in the context of optimization means it modifies the gradient before taking the descending step. The BFGS algorithm modifies the gradient based on the curvature information obtained by the Hessian. In practice, the Hessian is difficult to calculate precisely, even for conventional problems with large parameter space. A different variant of the BFGS method has been developed to approximate the Hessian, such as Limited-memory BFGS (L-BFGS) [580].

Algorithm 4: BFGS algorithm

B_k : Approximated Hessian at step k .
 $\mathcal{L}(\theta)$: Objective function
 α_k : stepsize at step k
 $\mathbf{p}_k$: Descent direction
 Calculate B_0 from initial guess $\mathbf{x}_0$;
while Not converge **do**
 Obtain $\mathbf{p}_k$ by solving $B_k \mathbf{p}_k = -\nabla \mathcal{L}(\theta_k)$;
 Perform line search for $\alpha_k = \text{argmin} \mathcal{L}(\theta_k + \alpha \mathbf{p}_k)$;
 Set $\mathbf{s}_k = \alpha_k \mathbf{p}_k$ and update $\theta_{k+1} = \theta_k + \mathbf{s}_k$;
 Set $\mathbf{y}_k = \nabla \mathcal{L}(\theta_{k+1}) - \nabla \mathcal{L}(\theta_k)$;
 Set $B_{k+1} = B_k + \frac{\mathbf{y}_k \mathbf{y}_k^T}{\mathbf{y}_k^T \mathbf{s}_k} - \frac{B_k \mathbf{s}_k \mathbf{s}_k^T B_k^T}{\mathbf{s}_k^T B_k \mathbf{s}_k}$;
end

Although the Hessian of a variational ansatz landscape can be calculated with analytical method [581], it is still difficult to calculate the full Hessian. Therefore this method is only used in the very early age of simulating conventional algorithms or as a baseline of comparison for lately developed optimizers. The SPSA algorithm from Section 7.2.1 can be used to approximate a Hessian.

Quantum natural gradient Natural gradient descent is a well-studied second-order optimization technique for numerical optimization. Natural gradient descent uses the steepest descent direction for the information geometry instead of taking each step with the gradient from parameter space [582–584]. A similar principle can be applied to optimizing the parameter of quantum variational ansatz, which is then referred to as quantum natural gradient descent. The idea of taking steps in another manifold can also be extended beyond information geometry [585].

Before diving into the details of quantum natural gradients, it is worth noting that the pure state quantum natural gradient descent is, in fact, equivalent to imaginary time evolution [290]. The imaginary time evolution is defined by

$$|\psi(\tau)\rangle = A(\tau)e^{-H\tau}|\psi(0)\rangle. \quad (214)$$

The state at $\tau \rightarrow \infty$ is the ground state of H . Instead of evolving the state directly, variational imaginary time evolution calculates the evolution direction with McLachlan's variational principle and simulates the evolution by varying the ansatz parameters [291].

The idea of natural gradient descent can be explained as follows. Consider the objective function $\mathcal{L}(\theta)$ is a likelihood function $\mathcal{L}(\theta|x)$ where x denote the measurement outcome of the distribution. For a general optimization, the likelihood function is given by the definition of the problem. While the traditional gradient descent algorithm takes one step towards the gradient direction in the parameter space, such direction may not necessarily be the direction that most significantly changes the difference of the distribution of $\mathcal{L}(\theta|x)$. From the statistical point of view, we would like each step to be taken in the direction that maximizes the distribution difference of the cost function. The natural gradient is then invented to present the gradient of the difference of the distribution function, and the corresponding geometry space is called information geometry.

The likelihood function's distribution difference from different θ can be characterized by the Kullback–Leibler (KL) divergence [586]. The KL divergence defines a distance measure between $\mathcal{L}(\theta|x)$, and when parameter θ is mapped to the likelihood function, it creates a Riemannian manifold with a metric F , where for small vector $d\theta$ in the parameter space, the distance ds on the manifold is

$$ds^2 = d\theta^T F d\theta. \quad (215)$$

Here F is the Fisher information metric, which is the Hessian of the KL divergence. The Riemannian manifold describes the information geometry of the problem. The natural gradient descent instead takes the gradient from the information geometry and maximizes the KL divergence for each step. The natural gradient is defined as

$$\tilde{g} = F^{-1}g = F^{-1}\nabla\mathcal{L}(\theta). \quad (216)$$

So far our discussion about natural gradient descent are with real space. To implement natural gradient descent for variational quantum algorithms, we can use the Fubini–Study metric as the Fisher Information metric in Hilbert space [587,588], which is given by [290,584,589,590]

$$(F)_{ij} := \text{Re}(\langle\partial_i\psi|\partial_j\psi\rangle) - \langle\partial_i\psi|\psi\rangle\langle\psi|\partial_j\psi\rangle, \quad (217)$$

where $|\partial_i\psi\rangle := \frac{\partial U(\theta)}{\partial\theta_i}|\psi_0\rangle$. These quantities can be evaluated with similar techniques as the Hadamard test. More details on the Hadamard test have been included in [Appendix B](#).

There are a few challenges for the quantum natural gradient. The first is the quantum natural gradient is defined on pure states, while in practice we need to consider the noise and non-unitary evolution. To solve this problem, the Fubini–Study metric can be generalized to a non-unitary circuit [135].

$$(F)_{ij} = \frac{1}{2}\text{Tr}[\rho(\theta)(L_iL_j + L_jL_i)], \quad (218)$$

where L is the symmetric difference defined as

$$\partial_k\rho(\theta) =: \frac{\partial\rho(\theta)}{\partial\theta_k} = \frac{1}{2}(L_k\rho(\theta) + \rho(\theta)L_k). \quad (219)$$

Suppose the density matrix of mixed state has fidelity $f_{id} = 1 - \epsilon$, which means the eigenvalue $\lambda_1 = f_{id}$. Such density matrix is given by

$$\rho_\epsilon = (1 - \epsilon)|\psi_1\rangle\langle\psi_1| + \epsilon \sum_{k=2}^d \lambda_k |\psi_k\rangle\langle\psi_k|. \quad (220)$$

The Fisher information matrix on NISQ device can be approximated efficiently as

$$(F)_{ij} = 2\text{Tr}\left[\frac{(\partial_i\rho_\epsilon)(\partial_j\rho_\epsilon)}{f_{id}}\right] + \mathcal{O}\left(\frac{1-f_{id}}{d}\right). \quad (221)$$

Term $\frac{1}{f_{id}}$ is ignored since it is only a scale factor. $\text{Tr}[(\partial_i\rho_\epsilon)(\partial_j\rho_\epsilon)]$ is a Riemannian metric tensor and can be evaluated with SWAP tests.

Secondly, the quantum natural gradient can be ill-conditioned and might lead to unreasonably large updates due to very small eigenvalues of F . In practice, the learning rate for the initial steps needs to be chosen very small or one needs to use *Tikhonov* regularization to add a small constant to the diagonal of F before the inversion. Another method to get a stable gradient is by doing a “half-inversion”, given by [132]

$$\tilde{g} = F^{-\alpha}\nabla\mathcal{L}(\theta) \quad (222)$$

where α is considered a regularization parameter. When $\alpha = 0$, no precondition is applied to the original gradient, and when $\alpha = 1$ the formula gets back to the original natural gradient. In practice, α is often chosen to be 0.5.

The third challenge comes from the cost of constructing the Fisher information metric. To fully construct the Fisher information metric with p parameters, p^2 different expectation values need to be estimated from the quantum computer. For a large p the cost of construct the Fisher information metric can be significant. Gacon et al. [591] proposed that quantum natural gradient can also be approximated with the SPSA methods previously discussed in Section 7.2.1. Given the Fisher information metric $F(\theta_1, \theta_2) = |\langle \psi_0 | U^T(\theta_1) U(\theta_2) | \psi_0 \rangle|$, the estimated gradient $\tilde{g}$ is given by

$$\tilde{g}^{(k)} = \frac{\delta F}{2\epsilon^2} \frac{\Delta_1^{(k)} \Delta_2^{(k)T} + \Delta_2^{(k)} \Delta_1^{(k)T}}{2} \quad (223)$$

where

$$\delta F^{(k)} = F(\theta^{(k)}, \theta^{(k)} + \epsilon \Delta_1^{(k)} + \epsilon \Delta_2^{(k)}) \quad (224)$$

$$- F(\theta^{(k)}, \theta^{(k)} + \epsilon \Delta_1^{(k)}) \quad (225)$$

$$- F(\theta^{(k)}, \theta^{(k)} - \epsilon \Delta_1^{(k)} + \epsilon \Delta_2^{(k)}) \quad (226)$$

$$+ F(\theta^{(k)}, \theta^{(k)} - \epsilon \Delta_1^{(k)}), \quad (227)$$

and the exponentially smooth estimator

$$\bar{F}^{(k)} = \frac{k}{k+1} \bar{F}^{(k-1)} + \frac{1}{k+1} \tilde{F}^{(k)}. \quad (228)$$

The simulation results from [591] show that the stochastic approximated natural gradient does not perform as well as the original natural gradient, however, it still improves the convergence speed compared to the standard gradient descent.

7.4. Gradient-free searching strategy

The gradient-free search strategy is the other large category of optimization strategy. Because of the analytical property of the quantum ansatz, a few gradient-free searching strategies are developed for optimizing the variational ansatz.

7.4.1. Gradient-free optimizers

Nelder–Mead algorithm The Nelder–Mead algorithm [592] is a gradient-free heuristic search method based on a simplex. A simplex S in $\mathbb{R}^k$ is defined as the convex hull of $k+1$ vertices in $\mathbb{R}^k$. It can be considered a multidimensional version of triangles. For example, in $\mathbb{R}^2$ a simplex is a triangle, and in $\mathbb{R}^3$ a simplex is a tetrahedron.

The algorithm first generates a random simplex and then transforms the simplex S and decreases the function values at each vertex iteratively. In each iteration, the objective function value at one or more test points is measured. Then a new simplex is updated by replacing the vertices with one of the test points.

The Nelder–Mead algorithm is gradient-free, and for each iteration, only very limited points need to be measured. Therefore it is very friendly to variational quantum algorithms. A result from [138,593] shows the Nelder–Mead algorithm performs well in a low-noise environment, but not as well when a noise model is included.

Powell's conjugate direction algorithm Powell's conjugate direction algorithm is a gradient-free optimization method. Powell's algorithm takes a starting point and two non-parallel vectors as inputs. The algorithm first search bidirectionally through the direction parallel to the first vector, and uses the minimal point as the direction, and search through the next direction parallel to the second vector. A conjugate direction can be defined as the displacement between the initial starting point and the optimal point of the second iteration. The algorithm repeats this process until it converges. The linear directional search can be implemented with Golden-section search or Brent's method [594].

7.4.2. Analytical optimization

In this section, we review optimization methods that take advantage of the analytical property of the objective function landscape. Consider the ansatz circuit as a CPTP mapping as a product of individual quantum super-operators

$$\Phi(\theta) = \Phi_k(\theta_k) \dots \Phi_2(\theta_2) \Phi_1(\theta_1). \quad (229)$$

Here $\Phi_k(\theta_k)$ are k th parameterized quantum gates where $\Phi_k(\theta_k)\rho := U_k \rho U_k^\dagger$ and $U_k = \exp(-i\frac{\theta_k}{2} P_k)$. Here P_k are tensor products of single-qubit Pauli operators as $P_k \in \{I, X, Y, Z\}^{\otimes N}$. Each superoperator can be expanded around θ_0 as

$$\Phi_k(\theta_0 + \theta) = a(\theta) \Phi_{ak} + b(\theta) \Phi_{bk} + c(\theta) \Phi_{ck}, \quad (230)$$

where $a(\theta), b(\theta) = 1 \pm \cos(\theta)$ and $c(\theta) = \frac{1}{2} \sin(\theta)$.

Now we only allow n parameters to be variables and fix all the others, the whole ansatz can be expanded into

$$\Phi(\theta_0 + \theta) = \prod_{k=1}^v [a(\theta_k) \Phi_{ak} + b(\theta_k) \Phi_{bk} + c(\theta_k) \Phi_{ck}], \quad (231)$$

where θ is the displacement around the reference point θ_0 .

Algorithm 5: Nelder–Mead algorithm

α_r : Reflection coefficient.
 α_e : Expansion coefficient.
 α_c : Contraction coefficient.
 α_s : Shrink coefficient.

while Not converge **do**

Ordering the existing test points, so that $\mathcal{L}(\theta_0) \leq \mathcal{L}(\theta_1) \dots \leq \mathcal{L}(\theta_{n+1})$

Calculate θ_o , the centroid of all points except θ_{n+1} ;

Calculate reflection point $\theta_r = \theta_o + \alpha_r(\theta_o - \theta_{n+1})$ with $\alpha_r > 0$.

if The reflection point the not the best estimation, but better than the second worst point θ_n **then**

replace the worst point θ_{n+1} with the reflected point θ_r ;

continue;

end

if Reflected point θ_r is the best point **then**

Calculate expansion point $\theta_e = \theta_o + \alpha_e(\theta_r - \theta_o)$ with $\alpha_e > 1$

replace the worst point θ_{n+1} with the best point from expansion point θ_e or reflected point θ_r ;

end

Calculate contraction point $\theta_c = \theta_o + \alpha_c(\theta_{n+1} - \theta_o)$ with $0 < \alpha_c \leq 0.5$.

if Contraction point θ_c is better than the worst point θ_{n+1} **then**

Replacing the worst point θ_{n+1} with the contracted point θ_c ;

Continue ;

end

// Shrink

Replace all points except the best point θ_1 with $\theta_i = \theta_1 + \alpha_s(\theta_i - \theta_1)$

end

Algorithm 6: Powell's algorithm

x_k : Searched minimal point. x_0 is the starting point.
 $\mathcal{L}(\theta)$: Objective function
 α_k : Searching displacement
 $\mathbf{d}_k$: Searching direction. $\mathbf{d}_1$ and $\mathbf{d}_2$ are the starting direction.

while Not converged **do**

Search for α_k that minimize $\mathcal{L}(\mathbf{x}_{k-1} + \alpha_k \mathbf{d}_k)$

Set $x_k = x_{k-1} + \alpha_k d_k$.

Remove the initial vectors by assigning $\mathbf{d}_j = \mathbf{d}_{j+1}$ Set $\mathbf{d}_N = \mathbf{x}_N - \mathbf{x}_0$ Search for α_N that minimize $\mathcal{L}(\mathbf{x}_N + \alpha_N \mathbf{d}_N)$

Set $\mathbf{x}_0 = \mathbf{x}_0 + \alpha_N \mathbf{d}_N$

end

Sequential optimization with sinusoidal fitting (Rotosolve) A variety of quantum variational algorithms are dealing with linear objective functions. For example, the energy value for VQE is a linear objective function. For these problems, the objective function landscape of every single variable is a sinusoidal function. The full information of such function can be obtained by simply measuring three different points of the objective function. The sequential optimization method [133,134,271] utilizes this feature and calculates the minima of the sinusoidal function directly. By iteratively finding the minima of every single parameter while fixing the other parameters, a greedy method can be derived to efficiently optimize the parameters of the ansatz.

From Eq. (231) we can simplify $a(\theta)$, $b(\theta)$ and $b(\theta)$ into a simpler form if we only allow one single parameter to change. When all the parameters are independent, the objective function can be transformed as

$$\mathcal{L}(\theta_0 + \theta \cdot e_i) = \mathcal{L}_i(\theta)_{\theta_0} = \text{Tr}[M \Phi_i(\theta_0 + \theta \cdot e_i)] \quad (232)$$

$$= \text{Tr}[M(a(\theta)\Phi_a + b(\theta)\Phi_b + c(\theta)\Phi_c)] \quad (233)$$

$$= A \sin(\theta + \phi) + C. \quad (234)$$

The parameter can be evaluated analytically with following equation when the hermitian generator has exactly two eigenvalues.

$$\theta^* = \phi - \frac{\pi}{2} - \arctan 2(\phi_1, \phi_2) + 2\pi k, \quad (235)$$

where

$$\phi_1 = 2\mathcal{L}_i(\theta)_{\theta_0} - \mathcal{L}_i(\theta + \frac{\pi}{2})_{\theta_0} - \mathcal{L}_i(\theta - \frac{\pi}{2})_{\theta_0} \quad (236)$$

$$\phi_2 = \mathcal{L}_i(\theta + \frac{\pi}{2})_{\theta_0} - \mathcal{L}_i(\theta - \frac{\pi}{2})_{\theta_0}. \quad (237)$$

All the single qubit gates have Pauli operators as their generators, which satisfy the above equations. The cost function can also sum sinusoidal functions with different period [574,595] when the generator has more than two distinct eigenvalues.

When several gates share the same parameter, the objective function can be transformed into similar form but with different oscillation period:

$$\mathcal{L}(\theta_0 + \theta \cdot e_i) = A \sin(k\theta + \phi) + C, \quad (238)$$

where k is the number of appearances of θ_i in the ansatz. The same approach can be applied to find the analytical solution.

It is important to note that the conditions for validity of this optimization method implies that it cannot be applied to some ansätze (for instance in the case where a parameterized controlled unitary is used). In general terms, Rotosolve and its subsequent extension require that all parameterized gate subject to optimization have 2π -periodicity and full-rank for the Hermitian matrix generating the rotation. The work presented by Wierisch et al. [596], relying on alternative general parameter-shift rules, demonstrates a generalization of Rotosolve which allows for it to be used on all quantum gates with arbitrary frequencies.

Analytical Free-Axis Selection with fixed rotation angles (Fraxis) The analytical method we mentioned so far has fixed rotation axis in the ansatz, but with a flexible rotation angle. The Analytical Free-Axis Selection (Fraxis) method implemented an alternative version where angles are fixed but the axis can be flexible [272]. The method is shown to provide faster optimization than Rotosolve in some numerical tests. Wada et al. [273] further improved on this idea by optimizing the angle and the axis at the same time for time evolution simulations.

The Fraxis method considers each single qubit gate in the ansatz as

$$U_k(\theta_k) = e^{-i\frac{\theta_k}{2}\hat{n}_k \cdot \vec{P}} \quad (239)$$

where $U_k(\theta_k)$ is the k th single qubit rotation, θ_k is the rotation angle, and $\hat{n}_k$ is the direction vector characterize the rotation direction of the hermitian generator, given by

$$\hat{n}_k \cdot \vec{P} = n_{k,x}X + n_{k,y}Y + n_{k,z}Z \text{ for } \hat{n}_k \in \mathcal{R}^3, |\hat{n}_k| = 1. \quad (240)$$

Here X, Y, Z are Pauli matrices.

With Lagrange multiplier method, the optimal $\hat{n}_k$ is found to satisfy following linear equations:

$$\sin^2(\frac{\theta_k}{2})\mathbf{R} - 2\lambda^* \hat{n}_k = -\alpha_{\theta_k} \mathbf{b} \quad (241)$$

where

$$\mathbf{b} = (\text{tr}(M[\rho, X]), \text{tr}(M[\rho, Y]), \text{tr}(M[\rho, Z]))^T \quad (242)$$

and

$$r_x \equiv \text{tr}(MX\rho X), \quad (243)$$

$$r_y \equiv \text{tr}(MY\rho Y), \quad (244)$$

$$r_z \equiv \text{tr}(MZ\rho Z), \quad (245)$$

$$r_{(x+y)} \equiv \text{tr}\left(M\left(\frac{X+Y}{\sqrt{2}}\right)\rho\left(\frac{X+Y}{\sqrt{2}}\right)\right), \quad (246)$$

$$r_{(x+z)} \equiv \text{tr}\left(M\left(\frac{X+Z}{\sqrt{2}}\right)\rho\left(\frac{X+Z}{\sqrt{2}}\right)\right), \quad (247)$$

$$r_{(y+z)} \equiv \text{tr}\left(M\left(\frac{Y+Z}{\sqrt{2}}\right)\rho\left(\frac{Y+Z}{\sqrt{2}}\right)\right). \quad (248)$$

$$\mathbf{R} \equiv \begin{pmatrix} 2r_x & 2r_{(x+y)} - r_x - r_y & 2r_{(x+z)} - r_x - r_z \\ 2r_{(x+y)} - r_x - r_y & 2r_y & 2r_{(y+z)} - r_y - r_z \\ 2r_{(x+z)} - r_x - r_z & 2r_{(y+z)} - r_y - r_z & 2r_z \end{pmatrix}. \quad (249)$$

The optimum value $\hat{n}_k^*$ can be solved by measuring all elements in $\mathbf{R}$ and solving the linear Eq. (242). When $\theta_k = \pi$, the method can be further simplified and $\hat{n}_k^*$ becomes the eigenvector of $\mathbf{R}$.

Quantum analytical descent The full landscape of the expectation value could be too expensive to construct. However, it can be approximated. Previous works from Sung et al. [597] approximate the local region with a predefined model, such as a quadratic polynomial. The quantum analytical descent [135] approximates the landscape properly and utilizes a similar idea of sequential optimization with sinusoidal fitting. Instead of fitting the sinusoidal function for each parameter iteratively, the entire landscape of the linear objective function can be approximated efficiently with a quadratic number of parameters and quadratic number of measurements.

The expansion of the entire ansatz has 3^p terms, which cannot be efficiently evaluated. The full expansion of the ansatz can be approximated into

$$\begin{aligned} \Phi(\theta) = A(\theta)\Phi^{(A)} + \sum_{k=1}^p [B_k(\theta)\Phi_k^{(B)} + C_k(\theta)\Phi_k^{(C)}] \\ + \sum_{l>k}^p [D_{kl}(\theta)\Phi_{kl}^{(D)}] + O(\sin^3 \epsilon). \end{aligned} \quad (250)$$

Here $A, B_k, C_k, D_{kl} : \mathbb{R}^p \mapsto \mathbb{R}$ are products of simple univariate trigonometric functions. And the loss function can be constructed as

$$\begin{aligned} \mathcal{L}(\theta) = A(\theta)\mathcal{L}^{(A)} + \sum_{k=1}^p [B_k(\theta)\mathcal{L}_k^{(B)} + C_k(\theta)\mathcal{L}_k^{(C)}] \\ + \sum_{l>k}^p [D_{kl}(\theta)\mathcal{L}_{kl}^{(D)}] + O(\sin^3 \epsilon). \end{aligned} \quad (251)$$

Here $\mathcal{L}^{(A)}, \mathcal{L}_k^{(B)}, \mathcal{L}_k^{(C)}, \mathcal{L}_{kl}^{(D)} \in \mathbb{R}$ can be calculated from the hardware by measuring expectation value at $1 + 2p^2 - 2p$ different points where p is the number of parameters.

After the objective function landscape has been approximated, the author uses natural gradient descent to find the minimum conventionally. The natural gradient descent requires the Fubini–Study metric tensor to find the natural gradient, which can be approximated as

$$[F_Q]_{pq} = F_{BB}F_{BB}(\theta) + F_{AB}F_{AB}(\theta) + \cdots O(\sin^2 \epsilon), \quad (252)$$

where

$$F_{BB}(\theta) := 2 \frac{\partial B_p(\theta)}{\partial \theta_p} \frac{\partial B_q(\theta)}{\partial \theta_q} \quad (253)$$

$$F_{AB}(\theta) := 2 \frac{\partial B_p(\theta)}{\partial \theta_p} \frac{\partial A(\theta)}{\partial \theta_q} + 2 \frac{\partial A(\theta)}{\partial \theta_p} \frac{\partial B_n(\theta)}{\partial \theta_q}. \quad (254)$$

Once the local minima of the approximated landscape have been found, the algorithm updates the best-guessed parameters in this iteration to be the local minima. Then repeat the approximation and conventional optimization steps until the algorithm converges.

Jacobi diagonalization and Anderson acceleration An analytical method inspired by Jacobi diagonalization and Anderson acceleration has been introduced in Ref. [563]. This method is an improved version of the sequential sinusoidal fitting. Instead of fitting the landscape one dimension at a time, the landscape can be accurately reconstructed by only allowing a small subset of parameters to vary. This technique is similar to the Jacobi diagonalization algorithm for large matrices by iteratively optimizing a random subset of parameters [598]. The Anderson/Pulay DIIS sequence acceleration is then introduced to produce a better estimation for each iteration.

This algorithm first constructs the analytical landscape with a few parameters and then iteratively chooses a random subset of parameters to fit the analytical landscape for the optimal value. This iteration approach is similar to the Jacobian diagonalization method.

Parrish et al. [563] then introduce the DIIS (Direct inversion of the iterative subspace) method to better estimate and improve the convergence speed. Suppose the optimized parameter after i th iteration is θ^i , the error of the optimized parameter is ϵ^i , which is the difference between the optimal value and the optimized parameter θ^i . The DIIS algorithm gives a better estimation:

$$\theta^{i'} = \sum_i c_i \theta^i, \quad (255)$$

where c_i is a real coefficient that minimizes the square of the 2-norm of ϵ ,

$$O(c_i) = \sum_{ij} c_i c_j \epsilon^i \cdot \epsilon^j, \quad (256)$$

and subject to the normalization condition

$$\sum_i c_i = 1. \quad (257)$$

The DIIS algorithm utilizes the previous steps' historical value and extrapolates a better estimation for the next steps. Since the error ϵ^i is not accessible for each step, it can be approximated by

$$\epsilon^i \approx \delta\theta^i = \theta^i - \theta^{i-1} \quad (\text{Anderson style}), \quad (258)$$

or

$$\epsilon^i \approx g(\theta^i) \quad (\text{Pulay style}), \quad (259)$$

where $g(\theta^i)$ is the gradient of the cost function at θ^i . Here the Anderson style and Pulay style are different approaches to approximate the error value.

7.5. Engineering cost function

Collective optimization [599] In practice, when using VQE to solve eigenenergies of molecules, there are different Hamiltonians with varied bond lengths that can be studied. Since the two different solutions should be close when the Hamiltonian is only different in a small amount, all of the optimal solutions of the series of Hamiltonian should also be chained together in the parameter space, forming a snake-like shape. The entire optimization process can be considered optimizing the arrangement of the whole snake instead of optimizing points separately.

The collective optimization algorithm redefines the cost function into

$$\mathcal{L}(\theta(\lambda)) = \int_{\lambda_0}^{\lambda_1} [L(\theta(\lambda)) + E(\theta(\lambda))], \quad (260)$$

where $\theta(\lambda)$ is the optimized ansatz parameter for Hamiltonian of bond distance λ , the E term is the energy of the molecules, while the L term defines the internal energy of the snake.

$$L(\theta(\lambda)) = \alpha \left| \frac{\partial \theta(\lambda)}{\partial \lambda} \right|^2 + \beta \left| \frac{\partial^2 \theta(\lambda)}{\partial^2 \lambda} \right|^2. \quad (261)$$

The first term refers to the energy that depends on the snake's length, and the second term refers to the curvature of the snake. α and β are meta parameters.

In practice the snake is discretized into a sequence of parameters at different bond distance $r_i = (\theta_i(\lambda_1), \theta_i(\lambda_2), \dots, \theta_i(\lambda_K))$ where i is the i th component for each $\theta_i(\lambda_k)$. Then the r_i can be solved iteratively by

$$r_i^t = (\eta \mathbf{A} + \mathbb{1})^{-1} (r_i^{t-1} - \eta \frac{dE(r^{t-1})}{dr_i}), \quad (262)$$

where η is the learning rate, $E(r) = \sum_{k=1}^K E(\theta(\lambda_j))$ and A is a pentadiagonal banded matrix with $A_{i-2,i} = A_{i,i-2} = \beta$, $A_{i-1,i} = A_{i,i-1} = -\alpha - 4\beta$, $A_{i,i} = 2\alpha + 6\beta$.

The simulation result shows the collective optimization methods help to pull the parameters from local minimums.

Conditional Value-at-Risk as objective function Conditional Value-at-Risk (CVaR) is a measure that takes into account only the tail of the probability distribution. The CVaR of a random variable X for a confidence level $\alpha \in (0, 1]$ is defined as

$$\text{CVaR}_\alpha(X) = E[X|X \leq F_X^{-1}(\alpha)], \quad (263)$$

where F_X is the cumulative density function of X . To illustrate the idea, consider the random variable X has been sampled N times. The CVaR with confidence α can be calculated by selecting αN samples with the lowest value and evaluating the average. When $\alpha = 1$, CVaR equals the expectation value. For variational quantum algorithms, consider the result of each sampling to be a random variable. Then the CVaR can be used as an objective function and replace the expectation value [600].

$$\mathcal{L}(\theta) = \text{CVaR}_\alpha(X(\theta)), \quad (264)$$

where X is the random variable for each sample result.

The CVaR could give a reasonable benefit. Suppose $|\psi_0\rangle$ is the ground state, and $|\psi_1\rangle, |\psi_2\rangle, |\psi_3\rangle$ are first, second and third excited state. Define $|\psi_A\rangle = (|\psi_0\rangle + |\psi_3\rangle)\sqrt{2}$ and $|\psi_B\rangle = (|\psi_1\rangle + |\psi_2\rangle)\sqrt{2}$. Suppose the energy levels are equally separated, then $\langle\psi_A|H|\psi_A\rangle = \langle\psi_B|H|\psi_B\rangle$, therefore the optimizer will encounter a gradient plateau. However for our purpose, we would like to obtain the ground state, therefore $|A\rangle$ is better than $|B\rangle$ in practice. The CVaR could help with this problem by emphasizing the distribution $|A\rangle$. The CVaR emphasizes the best-observed samples and leads to a smooth objective function without introducing local minimum [600]. Also, the implementation of CVaR is relatively straightforward. Since CVaR throws away some of the samples, the accuracy of the estimation decreases. In order to get the same accuracy, the sampling number needs to be increased. The same amount of samples should be involved in the calculation.

Table 11

Comparison of optimization strategies mentioned in this section. C_M denotes the number of different measurement expectation values per iteration that need to be evaluated from the quantum computer. C_C denotes the complexity of the classical algorithm for each iteration. Since the gradient can be evaluated or approximated with different methods, one can use $g^{(1)}$ to denote the cost of evaluating first order gradients and $g^{(2)}$ to denote the cost of evaluation second order gradients. S denote the required sample shot number. k denote the number of Hamiltonians with different bond distance being optimized simultaneously. p denote the number of parameters in the ansatz.

Strategies	Type	Meta parameters	C_M	C_C	References
Simple gradient descent	First order	η : Learning rate.	$Sg^{(1)}$	$\mathcal{O}(p)$	[138,584]
RMSProp	First order	γ : Moving average parameter. η : Learning rate.	$Sg^{(1)}$	$\mathcal{O}(p)$	[602]
Adam	First order	β_1, β_2 : Moving average parameters. η : Learning rate.	$Sg^{(1)}$	$\mathcal{O}(p)$	[579]
BFGS	Second order	η : Learning rate.	$S(g^{(1)} + g^{(2)})$	$\mathcal{O}(p^3)$	[138,584, 593]
Quantum natural gradient decent	Second order	η : Learning rate.	$S(g^{(1)} + g^{(2)})$	$\mathcal{O}(p^3)$	[582–584]
Nelder–Mead	Gradient free	α_r : Reflection coefficient. α_e : Expansion coefficient. α_c : Contraction coefficient. α_s : Shrink coefficient.	S	$\mathcal{O}(p)$	[138,593]
Powell	Gradient free	α_k : Searching displacement. $\mathbf{d}_k$: Searching direction.	Depends on linear search method.	$\mathcal{O}(1)$	[138]
Rotosolve	Gradient free	None	$\mathcal{O}(3n)$	$\mathcal{O}(1)$	[133,134]
Fraxis	Gradient free	None	$\mathcal{O}(6n)$	$\mathcal{O}(1)$	[272,273]
Quantum analytical descent	Gradient free	η : Learning rate.	$S(1 + 2p^2 - 2p)$	NP	[135]
Anderson acceleration	Addon	None	No extra costs	$\mathcal{O}(p)$	[598]
Collective optimization	Addon	α : Coefficient for snake length β : Coefficient for snake curvature.	No extra costs	$\mathcal{O}(k)$	[599]
CVaR	Addon	α : Confident level	Multiply factor $1/\alpha$	$\mathcal{O}(1)$	[600]

Symmetry preserving cost function adjustments: One means to maintain electron number conservation in the output wavefunction of VQE is to impose a constraint on the cost function [38,518]. Namely, one can include a penalty term corresponding to violation of symmetries. The VQE cost function can therefore be re-written as

$$E(\theta, \mu) = \langle \psi(\theta) | \hat{H} | \psi(\theta) \rangle + \sum_i \mu \left[\langle \psi(\theta) | \hat{O}_i | \psi(\theta) \rangle - O_i \right]^2, \quad (265)$$

where $\hat{O}_i$ represents the symmetry operators that can be independently measured, for instance, the electron number operator, or the spin operator (square of total spin). O_i is the target expectation value for each of these operators, and μ is a Lagrangian parameter to determine the strength of the constraints. This method, referred to as Constrained VQE in Ref. [518], was shown to also eliminate ‘kinks’ appearing in the VQE implementation that had been shown to appear in Ref. [263]. It is worth pointing out (as done in Refs. [38,518]) that the Pauli strings used for operators representing the electron number or the total spin already need to be computed for the Hamiltonian thereby the method does not require additional Quantum costs with respect to the computation of the energy function. One caveat to this is that, as is the case for any constrained optimization problem, the optimization landscape becomes more complex with the addition of constraints. This, in addition to the management of the hyper-parameter μ , could result in additional optimization costs and risks of local minima. The use of this method, and further considerations regarding optimal application have been discussed in Ref. [601].

7.6. Discussion

In this Section, we have reviewed some of the latest optimization strategies adapted to optimize variational quantum ansätze. Some methods are adapted from the traditional numerical optimization, while some new methods are developed to provide some essential features specifically for variational quantum ansatz optimization (see Table 11).

The first key feature to consider is the speedup of the convergence for variational quantum ansatz. From the convergence speed per iteration, the analytical method is much faster than the gradient descent strategies. However, the analytical methods would require taking more measurement points to finish a single iteration, which is unfair to directly compare the convergence speed per iteration between optimization strategies. Numerical studies have nonetheless shown in multiple occasions that Rotosolve (and by association its extensions) indeed reach convergence significantly faster than other methods [133–135].

Several studies have been performed to compare the relative strength of convergence of different optimizers. Mihlikov et al. [603] show as part of a case study on Hydrogen that SPSA presents a clear advantage on the Nelder–Mead and the Powell optimizers. Bonet-Monroig et al. [604] test four different optimizers on a variety of small molecular systems and find that SPSA performs slightly better than all others. Beyond these studies, several methods allow for improved convergence on almost any optimizer: the covariance functions between the Hamiltonian and operator can also be used to increase the convergence speed when the optimization is almost converged [605]. Stochastic methods are also suggested as add-ons to reduce the impact of hardware noise, accelerating convergence speed [606–608].

Another key feature is the resilience of the barren plateau problem. The natural-gradient-based strategy is considered resilient to barren plateau, and its absolute value of gradient has a lower bound $1/(2^{2N+1})$ [131,132]. The analytical method is gradient-free and directly jumps into an optimal or approximately optimal position in each iteration, which can prevent entering a barren plateau through optimization dependent on a learning rate [133–135,271]. However, this is at the cost of increasing the measurement points and conventional computation.

From the discussion above, no optimization strategy outperforms other strategies in every aspect; the convergence needs to trade with the implementation complexity and measurement points. In fact, multiple strategies can be applied together to improve the overall performance of the optimizer. In practice, the sequential analytical fitting method has been mainly used when a non-traditional optimizer optimizes the variational ansatz. Quantum natural gradient-based methods are more popular in simulation-based studies.

A final comment is worth raising on the optimization of VQE ansätze: it does not need to be done entirely using samples from observable obtained on the quantum computer. In particular Okada et al. [305] have shown that one can use efficient circuit simulation to optimize local order parameters and subsequently use quantum devices post-optimization to measure the global quantities.

8. Error mitigation for VQE

When the noise of a quantum computer is below a certain threshold and a sufficient number of qubits are available, quantum error correction schemes can be applied to suppress the noise to arbitrarily small levels [76]. However, hardware demonstrations of simple quantum error correcting codes have been limited and have only demonstrated fault-tolerant universal quantum computation with limited error-correcting capabilities [609–616]. Error correction also brings different types of overheads, including large amounts of extra ancilla qubits, fast decoding and communication between quantum and conventional devices [76]. For example, assuming an optimistic error rate threshold of $\varepsilon = 10^{-3}$, the required number of physical qubits to start exploring interesting quantum chemistry problems could be of the order of 10^6 [233].

As an alternative, a series of techniques for mitigating the effects of noise in quantum algorithms running on NISQ hardware have been developed. These techniques have been shown to achieve a reduction in the noise levels of expectation value estimates, without requiring the large resources involved in error correction. Such methods will be critical for early implementations of the VQE algorithm in order to achieve the required precision for Quantum Chemistry computation. Although error mitigation can also benefit other algorithms such as the Quantum Approximate Optimization Algorithm [136], we focus our discussion on the relevance of these methods for VQE. In this section, we cover the most relevant error mitigation techniques for VQE algorithms. For a more complete review of error mitigation techniques, we refer readers to [617].

8.1. Symmetry verification

The computation of molecular Hamiltonians often comes with symmetry constraints. Formally, this results in the ground state of the Hamiltonian $|\psi_0\rangle$ being the eigenstate of the corresponding symmetry operator $\hat{S}$. One example is the particle number operator $\hat{N} = \sum_i \hat{a}_i^\dagger \hat{a}_i$, which is usually a symmetry of the system, and we are often sure that the corresponding ground state should have a certain number of particles. Symmetries can be used to ‘taper off qubits’ (Section 4.3) or as properties to design the ansatz (Section 6.2.3). Here, we discuss how symmetry can be used to mitigate errors on a quantum state ρ prepared by an ansatz.

The general idea is that we need to find a way to measure the expectation value $\langle \hat{S} \rangle$ of the symmetry operator $\hat{S}$. Then we can filter out the experiments ρ having the wrong number $\langle \hat{S} \rangle$, i.e. we post-select the experiment outcome based on $\langle \hat{S} \rangle$. Such a verification process has many error-mitigating benefits. The first is that verification filters out errors that violate the symmetry, which applies regardless of the error rates of the quantum computer, although the failure rate is higher when the error rates increase [274]. The potential for filtering noise on measurements has been demonstrated experimentally in a VQE algorithm [156]. At the same time, the measurement and the post-selection on the result can project the state into a subspace that preserves the symmetry, and hence increase the overlap of the state ρ prepared

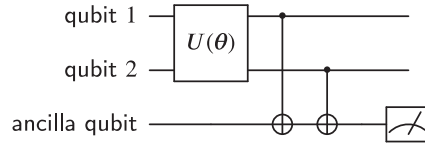


Fig. 14. Particle number parity verification circuit, adapted from Ref. [274]. Here the measurement outputs of the ancilla qubit give the qubit number parity, which is equivalent to the electron number parity in the Jordan–Wigner mapping.

using the ansatz with the symmetry preserving subspace [275]. Therefore, the symmetry verification method should be considered together with other symmetry preserving methods when preparing the target state in a VQE algorithm [156].

Here we briefly mention two specific techniques to measure $\hat{S}$. The first is called the *final symmetry verification*, in which we only verify that symmetries are respected at the end of the computation, i.e. after the ansatz circuit. This usually comes at no additional cost as the Pauli terms included in $\hat{S}$ are often measured as part of the Hamiltonian. The second is named *bulk symmetry-verification*, in which the symmetry verification step may be carried out during the computation, i.e. inside the ansatz circuit, without disrupting the computation.

Final symmetry verification: Verifying the symmetry at the end of the computation is relatively straightforward. For example, when using a Jordan–Wigner mapping, the electron number parity and the spin number parity of the state ρ produced by the ansatz can be directly measured at the end of the computation. We give an example of such measurement in Fig. 14. On the other hand, if the symmetry operator is more complex but can still be decomposed into a weighted sum of Pauli operators, it could then be measured using the same technique in VQE, even reusing data obtained already in measuring the energy of the state ρ .

The cost of this method can be characterized by [618]:

$$C = \frac{1}{\text{Tr}[\Pi_s \rho]}, \quad (266)$$

where Π_s projects into the eigenspace of the desired eigenvalue s . This is because only the $\text{Tr}[\Pi_s \rho]$ fraction of total circuit executions pass the symmetry verification test.

Bulk symmetry verification: Here we need to measure the observable $\hat{S}$ without disrupting the computation. At the moment, we can only achieve this when $\hat{S}$ can be decomposed into tensor products of one-qubit Pauli matrices

$$\hat{S} = \hat{s}_1 \otimes \hat{s}_2 \otimes \cdots, \quad (267)$$

where $\hat{s}_i \in \{I, X, Y, Z\}$. In Ref. [275], four different but equivalent circuits have been proposed using a similar construction, each having different advantages depending on the characteristics of the quantum computer. In general, a basis transformation is used to map $\hat{s}_i$ to the computational basis, using the pre-rotation gates already mentioned in Section 5.2. Information on all $\hat{s}_i$ in the computation basis is copied using CNOT gates to a qubit, which is then measured to extract the $\langle \hat{S} \rangle$ for filtering. We present all four circuits in Fig. 15.

Symmetry verification by post-processing: Verifying the symmetry $\hat{S}$ of a quantum state ρ projects the state into ρ_s , which lies within the eigenspace of the symmetry operator with the desired valued s . Subsequently, an observable $\hat{O}$ is measured on the projected state, giving a measured outcome $\text{Tr}[\hat{O} \rho_s]$. When both $\hat{S}$ and $\hat{O}$ are members of the Pauli group, it was shown [275] that

$$\text{Tr}[\hat{O} \rho_s] = \frac{\text{Tr}[\hat{O} \rho] + s \text{Tr}[\hat{O} \hat{S} \rho]}{1 + s \text{Tr}[\hat{S} \rho]}. \quad (268)$$

Therefore, the effect of symmetry verification, in this case, can be achieved with measurement outcomes of observables $\hat{O}$, $\hat{S}$, and $\hat{O} \hat{S}$. All three observables can be obtained as part of the measurements that in any case must be performed as part of the VQE optimization process. Therefore, we can achieve symmetry verification without measuring the symmetry operator $\hat{S}$ and post-selecting the measured number. This adds one more measured observable ($\hat{O} \hat{S}$) per Pauli observable, and one more measured observable $\hat{S}$ overall. This kind of symmetry verification is identical to a specific form of the quantum subspace expansion [275], where the excitation operators include the identity operator and the symmetry operator $\hat{S}$.

8.2. Extrapolation based methods

Extrapolation methods are based on the simple idea that the measurement result $\langle \hat{O} \rangle$ on a quantum computer is affected by the strength of the noise on the device, i.e. $\langle \hat{O} \rangle = \langle \hat{O} \rangle(\varepsilon)$, where ε characterizes the noise strength of the device. Therefore, using the data $\{\langle \hat{O} \rangle(\varepsilon_i)\}_{i \in \mathbb{N}}$ for the expectation value under several noise strength $\{\varepsilon_i\}$, we can construct a mathematical model $\langle \hat{O} \rangle(\varepsilon|\alpha)$ describing $\langle \hat{O} \rangle(\varepsilon)$, where $\alpha = \{\alpha_k\}$ are the set of parameters parameterizing

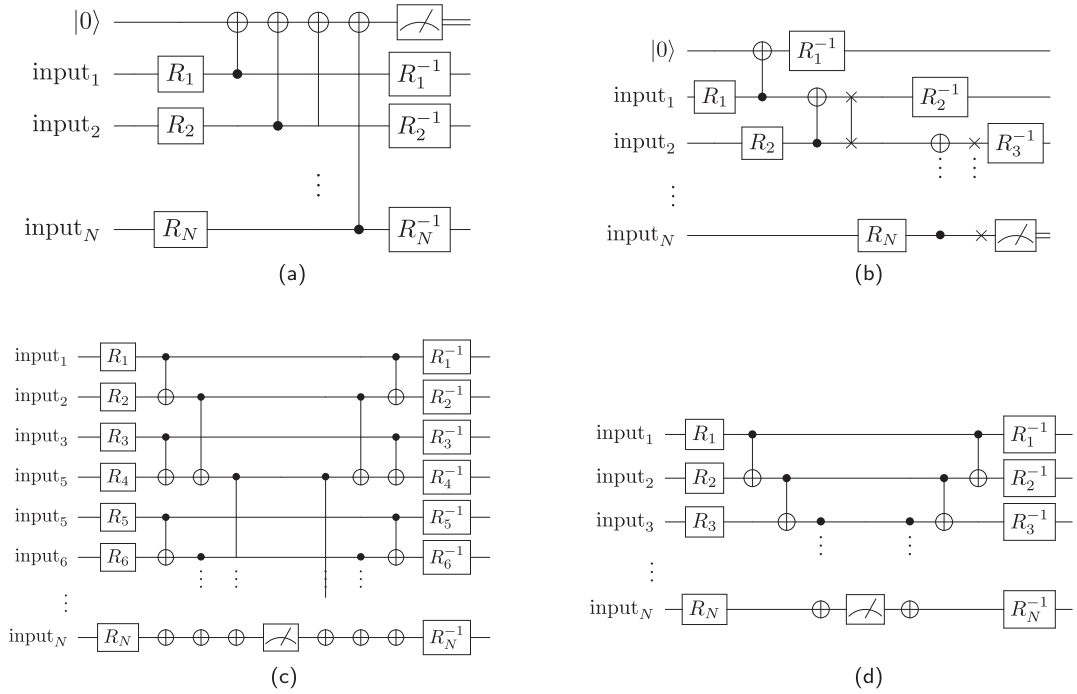


Fig. 15. Four different circuits that achieve the same bulk symmetry verification (Section 8.1) with different hardware requirements (adapted from Ref. [275]). In (a) and (b), an ancilla qubit initialized in state $|0\rangle$ is required for measuring the symmetry. In (c) and (d), the quantum computer should be able to perform a mid-circuit measurement. In all four circuits, the rotation gates $\{R_i\}$ perform a basis transformation such that the symmetry operator is equivalent to the qubit parity operator in the transformed basis.

the model used for extrapolation. We can then use the same mathematical model to predict the error-free expectation value $\langle \hat{O} \rangle_{\text{exact}} \equiv \langle \hat{O} \rangle(0) \approx \langle \hat{O} \rangle(\varepsilon = 0|\alpha)$. We denote by ε_0 the smallest noise strength of the quantum computer that can be achieved by hardware engineering, i.e. $\varepsilon_i \geq \varepsilon_0$ for all $i \in \mathbb{N}$. Of course, on a noiseless quantum computer we would have $\varepsilon_0 = 0$, which is not a realistic assumption. Hence, $\langle \hat{O} \rangle(\varepsilon_0)$ denotes the measurement outcome that would be achieved on the quantum computer without any error mitigation technique.

This method brings the additional overhead of the extra circuit executions required to obtain $\{\langle \hat{O} \rangle(\varepsilon_i)\}$ at different noise levels. It also potentially adds noise coming from the adjustment of the noise strength ε .

Care should be taken regarding implementing this method as part of a VQE algorithm: one may model the overall energy E as $E = E(\varepsilon|\alpha)$, but we argue here that it is more practical to model each Pauli observable separately. That is, each Pauli observable measured separately should be modeled with an independent $\langle \hat{O} \rangle(\varepsilon|\alpha)$. The reason is that the difference in noise from measuring different Pauli observables is a result of the different circuits required to rotate the state in the desired measurement basis. While this is likely insignificant for naive measurements of all Pauli observables, it becomes very relevant when grouping strategies relying on general commutativity of the Pauli observables are introduced (see Section 5.2), since the basis rotation circuit depth scales as $\mathcal{O}(N^2)$ [122]. It is worth pointing out that the parallel implementation for the VQE (as discussed in Section 2.7.2) also benefits from model fitting for each different hardware on which measurements are performed, especially when the noise characteristics vary across the hardware used. The effectiveness of extrapolation methods therefore depends on the ability to characterize and systematically increase the noise strength parameter ε (Section 8.2.1), and on the soundness of the mathematical model $\langle \hat{O} \rangle(\varepsilon|\alpha)$ (see Section 8.2.2).

8.2.1. Method to systematically increase the noise

Re-scaling method This method, proposed originally by Temme et al. [276], can achieve an effective scaling of the noise strength by rescaling the Hamiltonian H_{QPU} (here we are referring to the Hamiltonian driving the action of the quantum circuit on the initial state, rather than the Hamiltonian of the quantum chemistry problem we are trying to solve). The dynamics on a quantum computer can be described by the master equation

$$\partial_t \rho(t) = -i[\hat{H}_{\text{QPU}}(t), \rho(t)] + \varepsilon_0 \mathcal{L}[\rho(t)], \quad (269)$$

where the term $\varepsilon_0 \mathcal{L}[\rho(t)]$ represents the noise. An initial quantum state $\rho(0)$ is evolved under the Hamiltonian $\hat{H}_{\text{QPU}}(t)$ for a period of time T to the state ρ_{ε_0} where ε_0 characterizes the noise strength of the quantum computer. Another copy of the same initial state is evolved under a different Hamiltonian $\hat{H}_{\text{QPU}}(t/c)/c$ for a longer period of cT , resulting in another

state $\rho_{c\varepsilon_0}$. Then, assuming the noise $\mathcal{L}$ is independent of time t and of the Hamiltonian, one has

$$\rho_{c\varepsilon_0} - \rho_{\varepsilon_0} = c\varepsilon_0\mathcal{L}. \quad (270)$$

Therefore one can achieve an effective rescaling of the expectation value, resulting in $\langle\hat{O}\rangle(\varepsilon = c\varepsilon_0) - \langle\hat{O}\rangle(\varepsilon = \varepsilon_0) = c\varepsilon_0 \text{Tr}[\hat{O}\mathcal{L}]$. This scaling modifies the Hamiltonian on quantum computers directly and therefore requires precise control of the device, which has been demonstrated by the hardware team in Ref. [619]. However, the recently proposed OpenPulse specification [620] allows a more platform-independent way to control the underlying hardware which has been used to implement this error boosting method [621].

Pauli twirling Li and Benjamin [277] propose that one can boost the physical error rate to any desired value by randomly applying Pauli gates before and after a Clifford entangling two-qubit gate, and then randomly generating additional Pauli gates after the twirled two-qubit gates.

Inserting CNOT gates Refs. [622,623] showed that when every CNOT gate is replaced by j consecutive CNOT gates (j should be an odd number), the measurement outcome $\langle\hat{O}\rangle(j)$ depends on j by

$$\langle\hat{O}\rangle(j) = \langle\hat{O}\rangle_{\text{exact}} + kj + \mathcal{O}\left(\left(j \sum_{i=1}^{J_{\text{exact}}} \varepsilon_i\right)^2\right), \quad (271)$$

where $\langle\hat{O}\rangle_{\text{exact}}$ is the exact measurement outcome when there is no noise, ε_i characterizes the noise strength of the i th CNOT gate, k is the parameter characterizing the noise strength, and J_{exact} is the number of CNOT gates in the original circuit. In this model, it is assumed that the noise is dominated by the two-qubit depolarizing noise on the CNOT gates, and all other forms of error are considered negligible. The extrapolation method can therefore be used for extrapolating the case of $j = 0$.

This method becomes problematic when there are a large number of CNOT gates in the original circuit, each of which needs to be replaced by j CNOT gates. Eq. (271) means that errors in higher orders of j quickly overshadow lower-order contributions used for modeling the error.

As a solution, Ref. [623] provides an approach, which is appropriate for large numbers of CNOT gates. Instead of replacing each CNOT with r CNOT gates, the proposal is to replace the i th CNOT gate with j_i CNOT gates, where $(j_i - 1)/2 \sim \text{Poisson}(v)$, a Poisson distribution with mean v . That is, the number of additional CNOT gates is sampled each time from the Poisson distribution. As a result, Eq. (271) is slightly changed to

$$\langle\hat{O}\rangle(v) = \langle\hat{O}\rangle_{\text{exact}} + (1 + 2v)k + \mathcal{O}\left(\left((1 + 2v) \sum_{i=1}^{J_{\text{exact}}} \varepsilon_i\right)^2\right). \quad (272)$$

Using this adjustment, the circuit is run for different values of v to extrapolate to the case of $v = -1/2$. Errors in the quantum computer can be mitigated by extrapolating from measurement outcomes $\langle\hat{O}\rangle$ corresponding to different values of j (Eq. (271)) or v (Eq. (272)) using a linear model (Section 8.2.2). In particular, v can be any positive number chosen such that $1 + 2v$ is close to 1, making the second method more noise friendly than the first method, where j can only take integer values.

Unitary folding: This method is similar to the CNOT insertion method mentioned above. Instead of inserting CNOT gates, Ref. [624] proposes that one can insert only gates from the original circuit to boost the error rate. In this method, named *unitary folding*, one assumes a unitary circuit U is made from several layers $\{U_i\}_{i=1}^d$ of quantum gates

$$U = U_d \cdots U_2 U_1, \quad (273)$$

where d is the number of layers. To boost the error, U is transformed into an equivalent circuit

$$U_{\text{folded}} = U(U^\dagger U)^{L_1} U_{l_2}^\dagger U_{l_2-1}^\dagger \cdots U_1^\dagger U_1 \cdots U_{l_2-1} U_{l_2}, \quad (274)$$

where U has been *folded* $2L_2 + 1$ times, and an additional subset $\{l_1, l_2, \dots, l_2\}$ ($L_2 < d$) of the d layers is chosen to boost the error even further. There is no particular rule as to how this subset should be chosen. However, these additional gates give us additional freedom to control the noise strength. One can define λ as the ratio of layers in the folded circuit over the original circuit

$$\lambda = \frac{(2L_1 + 1)d + 2L_2}{d} = 2L_1 + 1 + \frac{2L_2}{d}. \quad (275)$$

Form this point, λ can be systematically increased with a step size of $2/d$, giving granular control of this noise parameter. Assuming a depolarizing noise model on all qubits, the measurement output of the folded circuit has the form

$$\langle\hat{O}\rangle(\lambda) = \langle\hat{O}\rangle_{\text{exact}} + b\varepsilon^\lambda, \quad (276)$$

where b and ε are coefficients that are determined from extrapolating from the measured data. Therefore, extrapolation to the case of $\lambda = 0$ can give the desired error-free measurement outcome.

Multi-parameter noise model: It is, in general, more realistic to characterize the noise of a quantum computer by multiple parameters, such as the times T_1 and T_2 (see [Appendix C.1.1](#)), which are usually used as figure of merits for quantum computers. In this case, the noise parameter ε becomes a vector $\varepsilon = (\varepsilon^{(1)}, \dots, \varepsilon^{(l)},)$ of l different noise parameters. Ref. [625] discusses an interesting scenario where T_1 and T_2 are estimated and used to build a model that recovers the population in state $|1\rangle$. Although we have not seen a method that systematically controls errors on multiple noise parameters, we note that all modeling methods discussed below may be generalized to this multi-parameter case.

8.2.2. Modeling the noise and extrapolating from measured data

Once a suitable technique has been developed to artificially increase the noise in the device, one can turn towards finding a model to extrapolate the data to obtain a better approximation of the true expectation value. A general formulation of the extrapolation can be made: given the model $\langle \hat{O} \rangle(\varepsilon|\alpha)$ and several measurement outcomes $\langle \hat{O} \rangle(\varepsilon_i)$, an optimal coefficient α^* can be found by optimizing function $f(\alpha)$, which measures the squared error between prediction $\langle \hat{O} \rangle(\varepsilon_i|\alpha)$ and the measurement outcome $\langle \hat{O} \rangle(\varepsilon_i)$

$$f(\alpha) = \sum_i |\langle \hat{O} \rangle(\varepsilon_i) - \langle \hat{O} \rangle(\varepsilon_i|\alpha)|^2. \quad (277)$$

This method is very common in data analysis, and is often used in regression techniques (for further details, see Ref. [626]).

Linear/polynomial model: In this model $\langle \hat{O} \rangle(\varepsilon|\alpha)$ can be generalized to

$$\langle \hat{O} \rangle(\varepsilon|\alpha) - \langle \hat{O} \rangle_{\text{exact}} = \sum_{k=0}^K \alpha_k \varepsilon^k + \mathcal{O}(\varepsilon^{K+1}), \quad (278)$$

which was shown to be valid in Ref. [276], when assuming that the noise on quantum computer is small enough, such that the residual noise $\langle \hat{O} \rangle(\varepsilon) - \langle \hat{O} \rangle(\varepsilon|\alpha) = \mathcal{O}(\varepsilon^{K+1})$. This function was also shown to characterize noise correctly when inserting additional identity gates into the quantum circuits, see Section 8.2.1.

Having obtained $\langle \hat{O} \rangle(\varepsilon_i)_{i=0}^L$, where $L + 1$ is the number of measurement outcomes we obtained for $L + 1$ different values of ε_i , we mention two methods to extrapolate an approximation to the exact measurement outcome $\langle \hat{O} \rangle_{\text{exact}}$. The first method is called Richardson extrapolation [276,277], and can be applied when $L = K$. Here, one rewrites Eq. (278) as

$$\langle \hat{O} \rangle(c_i) = \langle \hat{O} \rangle_{\text{exact}} + \sum_{k=0}^K \alpha_k (c_i \varepsilon_0)^k + \mathcal{O}(c_i^{K+1}), \quad (279)$$

where $c_i = \varepsilon_i/\varepsilon_0$ is the ratio of the noise strength of different cases relative to the reference value ε_0 . Next, one seeks a series of combination coefficients β_i such that

$$\sum_i \beta_i \langle \hat{O} \rangle(c_i) = \langle \hat{O} \rangle_{\text{exact}}, \quad (280)$$

for any ε_0 , ignoring the higher order terms $\mathcal{O}(\varepsilon_0^{N+1})$ in Eq. (279). That is, one considers Eq. (280) as an equivalence of two polynomials in ε_0 , leading to the conditions

$$\begin{aligned} \sum_i \beta_i &= 1, \\ \sum_i \beta_i c_i^k &= 0, \quad k = 1, 2, \dots, K, \end{aligned} \quad (281)$$

which admit the solution

$$\beta_i = \prod_{k \neq i} \frac{c_k}{c_k - c_i}. \quad (282)$$

It is worth noting that more explicit formulae for the constants c_i can be obtained assuming depolarizing noise in the CNOT-insertion based error boosting method [623], see Section 8.2.1.

The second method is a simple extension of the linear regression technique commonly found in machine learning [627], which we already mentioned above (Eq. (277)). Here, the optimization can be furthered simplified. One first rewrites

Eq. (278) into a matrix form by defining

$$A = \begin{pmatrix} 1 & \varepsilon_0 & \cdots & \varepsilon_0^K \\ 1 & \varepsilon_1 & \cdots & \varepsilon_1^K \\ \vdots & \vdots & & \vdots \\ 1 & \varepsilon_L & \cdots & \varepsilon_L^K \end{pmatrix}, \quad (283)$$

$$X = (\langle \hat{O} \rangle_{\text{exact}}, \alpha_1, \dots, \alpha_K)^T,$$

$$Y = (\langle \hat{O} \rangle(\varepsilon_0), \langle \hat{O} \rangle(\varepsilon_1), \dots, \langle \hat{O} \rangle(\varepsilon_L))^T,$$

$$f(X) = (AX - Y)^T(AX - Y),$$

where the cost function f measures the squared difference between the measurement outcome Y and the prediction AX (i.e. $\langle \hat{O} \rangle(\varepsilon_i|\alpha)$). The minima of f (which contains the desired $\langle \hat{O} \rangle_{\text{exact}}$) can be obtained using the equation [627]

$$X^* = \underset{X}{\operatorname{argmin}} f(X) = (A^T A)^{-1} A^T Y, \quad (284)$$

when $A^T A$ is not singular. Note that one often controls the ratio $c_i = \varepsilon_i/\varepsilon_0$, and the method works in the same manner by replacing ε_i with c_i in Eq. (283).

Here we compare the two methods. The Richardson extrapolation coincides with the second minimization method when $K = L$. This can be seen by noticing that A has the form of a Vandermonde matrix and its determinant when $K = L$ is $\prod_{0 \leq i < j \leq K} (\varepsilon_j - \varepsilon_i)$, which is always invertible when the ε_i have distinct values. However, although it may be less interesting to consider the case when $K > L$ when there are insufficient measurement samples to fix the free parameters in our model Eq. (278), we argue here that the $K < L$ case is interesting; this uses a polynomial of order K , which is smaller than L , the number of measurement samples. The argument is as follows: although when $K = L$ we may obtain a solution α such that each measurement outcome $M(\varepsilon_i)$ could be exactly predicted by our model $M(\varepsilon|\alpha)$, a model with higher-order polynomials is more sensitive to statistical fluctuations [624] and generally does not perform well for predicting new values. Similar arguments could be found in the machine learning literature under the topic of *over-fitting* [626]. In this case, the second method is also compatible with techniques addressing the over-fitting problem, such as the Lasso regression [627].

Finally, it is very important to note that extrapolation based on this model increases the cost of the VQE algorithm, possibly prohibitively. For each Pauli observable evaluation, the required circuit evaluation is multiplied by the size of data used to fit the model, and the variance of the observable is also increased by a factor γ . When the Richardson extrapolation is used, or when the matrix A in Eq. (283) is a square and non-singular matrix, γ can be derived analytically [276,628] as

$$\gamma = \sum_i \beta_i^2, \quad (285)$$

with β_i as in Eq. (282).

Exponential model: For a deep circuit, it is natural to expect that the error dominates the quantum computation exponentially fast. Ref. [629] demonstrated this in a simple example, where a quantum circuit consisting of J noisy gates is executed on a noisy quantum computer. It assumes that the noise on the i th identity gate ε_i can be modeled by a simple error model of the form

$$\mathcal{E}_i(\rho) = (1 - \varepsilon)\rho + \varepsilon \mathcal{E}'_i(\rho). \quad (286)$$

An example is the depolarizing noise, where $\mathcal{E}'_i(\rho) \propto \mathbb{1}$. With this assumption, the overall noise effect $\prod_i \mathcal{E}_i$, when J is large, can be approximated by

$$\prod_i \mathcal{E}_i \approx e^{-J\varepsilon} \sum_{k=0}^J \frac{(J\varepsilon)^k}{k!} \mathcal{X}_k, \quad (287)$$

where $\mathcal{X}_k$ contains contributions of $\mathcal{E}'$ applied k times towards the product. In order to fit an exponential noise extrapolation model, one can re-write $\langle \hat{O} \rangle(\varepsilon|\alpha)$ as

$$\langle \hat{O} \rangle(\varepsilon|\alpha) - \langle \hat{O} \rangle_{\text{exact}} = A \exp(-J\varepsilon). \quad (288)$$

Ref. [629] gives an explicit formula for $\langle \hat{O} \rangle(0)$, computable with prior knowledge of $\langle \hat{O} \rangle(\varepsilon_0)$ and $\langle \hat{O} \rangle(k\varepsilon_0)$,

$$\langle \hat{O} \rangle(0) = \frac{ke^{\varepsilon} \langle \hat{O} \rangle(\varepsilon) - e^{Jk\varepsilon} \langle \hat{O} \rangle(k\varepsilon)}{k - 1}. \quad (289)$$

Furthermore, one can combine the exponential model with the polynomial model into a poly-exponential extrapolation [624], where the noise model is

$$\langle \hat{O} \rangle(\varepsilon|\alpha) = A + B \exp(\text{poly}(\varepsilon)). \quad (290)$$

Here the degree of the polynomial in ε needs to be chosen by experience, and the coefficients can be obtained with the minimization method mentioned at the beginning of this section. Giurgica-Tiron et al. [624] demonstrate that this method successfully mitigates errors resulting from quantum circuit execution, however, they do not compare with other extrapolation methods in terms of cost and benefits.

Extrapolation based on this model similarly increases the cost of the VQE algorithm. For each Pauli string evaluation, the number of circuit repetitions is multiplied by the size of data, and as above, the variance of the observable is also increased by a factor γ . For a simple exponential model in Eq. (289), γ can be derived analytically as [628]

$$\gamma = \frac{k^2 \exp(2\varepsilon J) + \exp(2k\varepsilon J)}{(k-1)^2}, \quad (291)$$

assuming the variance of $\langle \hat{O} \rangle_{(\varepsilon_0)}$ and $\langle \hat{O} \rangle_{(k\varepsilon_0)}$ are equal. One can note that the variance in this exponential model tends to be significantly larger than for the polynomial model. However, it has been shown [629] that if enough sampling is affordable, this exponential model gives more accurate prediction of the exact measurement outcome than the polynomial model. The polynomial model can be seen as an approximation to this exponential model when the noise ε is not very high.

8.3. Probabilistic error cancellation

In Probabilistic error cancellation (PEC) [276,629], quantum circuits are actively modified with the overall goal of inverting the impact of noise on a quantum computer. Mathematically, the estimated ideal measurement value $\langle \hat{O} \rangle$ is obtained as

$$\langle \hat{O} \rangle_{\text{PEC}} = \sum_i q_i \langle \hat{O} \rangle_{\text{PEC},i} \quad (q_i \in \mathbb{R}), \quad (292)$$

where each measurement value $\langle \hat{O} \rangle_{\text{PEC},i}$ corresponds to a different, slightly modified, quantum circuit. Specifically, given a quantum circuit $U = \prod_k U_k$, one finds weights q_{i_k} and operations on quantum computer B_{i_k} such that

$$U_k = \sum_{i_k} q_{i_k} B_{i_k}, \quad (293)$$

and hence overall

$$U = \prod_k \left(\sum_{i_k} q_{i_k} B_{i_k} \right) = \sum_{i_1, i_2, \dots} \prod_k q_{i_k} B_{i_k}. \quad (294)$$

Here the weights $\prod_k q_{i_k}$ are used as the coefficient q_i in Eq. (292), where i is an abbreviation of the many indices $i_1, i_2, \dots$. We provide an example correcting single-qubit gates under the depolarizing noise in Appendix C.2.

One expects the terms in the summation to explode exponentially with respect to the depth of the circuit. Therefore, it helps to implement Eq. (292) probabilistically. To do this, one rewrites Eq. (292) as

$$\begin{aligned} \langle \hat{O} \rangle_{\text{PEC}} &= \gamma_{\text{PEC}} \sum_i \text{sgn}(q_i) P_{\text{PEC}}(\hat{O})_{\text{PEC},i}, \\ \gamma_{\text{PEC}} &= \sum_i |q_i|, \quad P_{\text{PEC}} = \frac{|q_i|}{\gamma_{\text{PEC}}}. \end{aligned} \quad (295)$$

Here $\text{sgn}(q_i)$ is the sign of q_i , and P_{PEC} is a proper probability distribution normalized by the constant γ_{PEC} . Eq. (295) provides a probabilistic implementation of the summation in Eq. (292). That is, we could sample from all the modified circuits according to the distribution P_{PEC} and execute them. The overall measurement $\langle \hat{O} \rangle_{\text{PEC}}$ can be estimated by adding measurement result $\langle \hat{O} \rangle_{\text{PEC},i}$ from each circuit weighted by the factor $\gamma_{\text{PEC}} \text{sgn}(q_i) P_{\text{PEC}}$. When the number of modified circuits is large, it is more practical to estimate $\langle \hat{O} \rangle_{\text{PEC}}$ using this probabilistic implementation. For this reason, this mitigation technique is called *probabilities error cancellation*. In the literature [276,629] pairs of weights q_{i_k} and operations B_{i_k} are considered as a *quasi-probability representation* of U_k , q_{i_k} or q_i . These are referred to as the *quasi-probability weights*, and incidentally the mitigation method is called the *quasi-probability method*.

Before discussing the quasi-probability representation and its construction, we first discuss its cost. Although the number of noisy circuits in a quasi-probability representation of the ideal circuit U grows exponentially with respect to the depth of the circuit (Eq. (294)), with the probabilistic implementation the cost of PEC only depends on how fast the probabilistic sampling converges, i.e. the variance of $\langle \hat{O} \rangle_{\text{PEC}}$. Assuming that the measurement outcomes of different circuits $\langle \hat{O} \rangle_{\text{PEC},i}$ are independent and that their variance can be bounded by 1 (which applies to the Pauli observable of interest in the VQE algorithm), the variance of $\langle \hat{O} \rangle_{\text{PEC}}$ could be bounded by γ_{PEC}^2 .

Now we detail how the modified circuits are constructed in the general case [629–631]. There are several approaches to this problem, and here we mention two important methods. The first constructs modified circuits using tomography data from quantum computers, the second constructs modified circuits based on data from a quantum computer and a given ansatz structure.

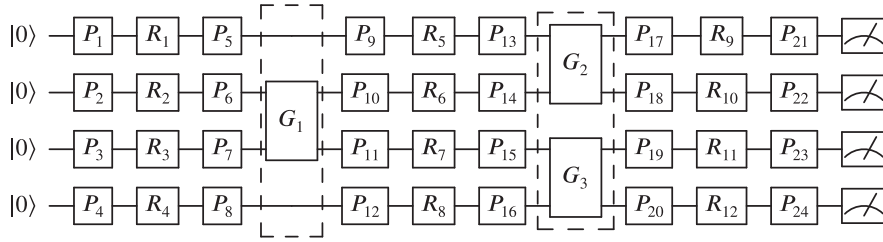


Fig. 16. Setup of the learning based probabilistic error cancellation method (adapted from Ref. [630]). A quantum circuit is partitioned into single qubit gates $\mathbf{R} = (R_1, R_2, \dots)$ followed by Clifford multi-qubit gates $\mathbf{G} = (G_1, G_2, \dots)$. Additional, single qubit Pauli gates $\mathbf{P} = (P_1, P_2, \dots)$ are inserted in between the single qubit gates and Clifford gates to mitigate errors.

Tomography based method [629]: Given a circuit U that consists of unitary gates $U = \prod_k U_k$, we assume that the effect of noise on each unitary gate can be characterized (by tomography) as $\mathcal{U}_{k,\text{noisy}} = \mathcal{N}_k \circ \mathcal{U}_k$, and that there is no correlation between errors on quantum gates acting on different qubits (spatial correlation) or different time steps (temporal correlation). Here $\mathcal{U}_k$ represents the quantum channel corresponding to the unitary gate U_k , and is typeset in an italic font to emphasize it being a quantum channel. From there, one must find pairs of quasi-probability weights $q_{i,k}$ and quantum operations $\mathcal{B}_{i,k}$ (which are not necessarily unitary gates) such that

$$\mathcal{U}_k = \sum_i q_{i,k} \mathcal{B}_{i,k} \mathcal{U}_{k,\text{noisy}}, \quad (296)$$

and overall the quantum channel $\mathcal{U}$ corresponding to the unitary gate U is expressed as

$$\mathcal{U} = \sum_{k,i} q_{i,k} \mathcal{B}_{i,k} \mathcal{U}_{k,\text{noisy}}. \quad (297)$$

To find such $\mathcal{U}$, one obtains the gate-wise noise model $\mathcal{N}_k$ for each k using tomography. Then, a different formalism to describe the quantum process $\mathcal{N}_k$ named Pauli transfer matrix formalism is used to obtain the (non-physical) channel $\mathcal{N}_k^{-1}$, described by a linear summation of several basis operations $\{\mathcal{B}'_i\}$. We refer the readers to Ref. [629] for implementation details.

Learning based method: This method, proposed by Strikis et al. [630] assumes that single-qubit errors are negligible, and can be applied to quantum circuits having a specific layer-wise structure. In each layer of these quantum circuits, single qubit gates are followed by Clifford multi-qubit gates. Pauli gates are inserted in between single qubits gates and Clifford multi-qubit gates to mitigate errors (see Fig. 16). Specifically, we denote all the Clifford multi-qubit gates collectively by $\mathbf{G}$ (called *frame gates* in the original paper), all single qubits in the unmodified circuit collectively by $\mathbf{R}$, and all the additional inserted Pauli gates by $\mathbf{P}$. To mitigate the error, different combinations of single qubit Pauli gates are inserted, labeled as $\mathbf{P} = (P_1, P_2, \dots)$, resulting in different quantum circuits. Each of these gives a measurement value labeled by $\langle \hat{O} \rangle(\mathbf{R}, \mathbf{P})$. The overall mitigation result $\langle \hat{O} \rangle^{\text{mit}}(\mathbf{R}, q)$ is

$$\langle \hat{O} \rangle^{\text{mit}}(\mathbf{R}, q) = \sum_{\mathbf{P}} q(\mathbf{P}) \langle \hat{O} \rangle(\mathbf{R}, \mathbf{P}), \quad (298)$$

where $q : \mathbf{P} \mapsto q(\mathbf{P}) \in \mathbb{R}$ associates a specific weight to a specific combination $\mathbf{P}$. Eq. (298) resembles the quasi-probability formula Eq. (271) closely, except for the structure of the circuit, and the procedure to modify the circuits. To find the appropriate weights $q(\mathbf{R})$, Strikis et al. [630] define a loss function

$$\text{Loss}(q) = \frac{1}{|\mathbb{T}|} \sum_{\mathbf{R} \in \mathbb{T}} |\langle \hat{O} \rangle^{\text{mit}, \text{qpu}}(\mathbf{R}, q) - \langle \hat{O} \rangle^{\text{mit}, \text{sim}}(\mathbf{R})|^2, \quad (299)$$

where the single qubit gates in the original circuit are substituted by single qubit gates in the set $\mathbb{T}$, called the training set. The substituted circuits are simulated to obtain an exact simulation result $\langle \hat{O} \rangle^{\text{mit}, \text{sim}}(\mathbf{R})$. Pauli gates are inserted in the manner specified in Fig. 16 and executed on a quantum computer to obtain the noisy result $\langle \hat{O} \rangle^{\text{qpu}, \text{noisy}}(\mathbf{R}, q) = \sum_{\mathbf{P}} q(\mathbf{P}) \langle \hat{O} \rangle(\mathbf{R}, \mathbf{P})$.

Overall, one aims to find the weights q such that the loss function in Eq. (299) is minimized. Strikis et al. [630] show that if the training set $\mathbb{T}$ is chosen as the set of all Clifford circuits, and if one assumes that there exists a q^* such that $\text{Loss}(q^*) = 0$, then

$$\langle \hat{O} \rangle^{\text{mit}, \text{qpu}}(\mathbf{R}, q^*) = \langle \hat{O} \rangle^{\text{exact}}(\mathbf{R}), \quad (300)$$

for any set of single qubit gates $\mathbf{R}$. Here it is also assumed that the single qubit errors (i.e. errors for $\mathbf{R}$) are negligible. The quantity $\langle \hat{O} \rangle^{\text{exact}}(\mathbf{R})$ is the exact result of the original circuit without error.

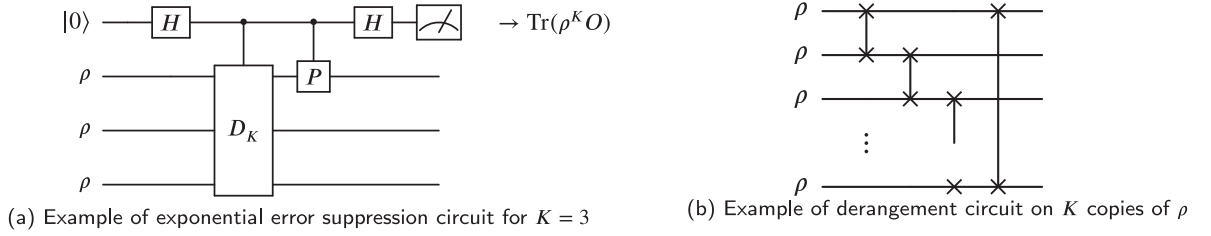


Fig. 17. General circuits to compute the numerator in Eq. (301) for exponential error suppression (adopted from Ref. [632]). Here for simplicity we only considered the case when $K = 3$, but the circuit could be easily generalized to arbitrary $K \geq 2$. D_K is the circuit which performs a derangement on all indices $\{i_1, \dots, i_K\}$. An example of D_K is shown to the right. It swaps all nearest neighbor qubits, as well as the first and the last qubit. Details on properties of D_K could be found in Appendix C.3.

Although Strikis et al. [630] have not discussed the optimization of the loss function $\text{Loss}(q)$, it discussed in detail the approach to sampling the training set $\mathbb{T}$ and the set of all Pauli operators $\mathbb{P}$ for estimating the loss function, which might be a concern since both sets grow exponentially with respect to the number of qubits.

8.4. Exponential error suppression

This method was independently proposed in Refs. [632,633], and further developed by Refs. [634–636]. It is called *exponential suppression by derangement* in Ref. [637], and *virtual distillation* in Ref. [633]. For fairness, we refer to this method as exponential error suppression to highlight its ability to suppress errors exponentially. Using different circuits, all variants of this method compute the value

$$\frac{\text{Tr}[\rho^K \hat{O}]}{\text{Tr}[\rho^K]}, \quad (301)$$

where a quantum state ρ is prepared by a circuit, e.g. the ansatz circuit of VQE. Here, $\hat{O}$ can be any observable, and K is a free parameter that can be adjusted. The effect of calculating Eq. (301) instead of directly $\text{Tr}[\rho O]$ is that the dominant eigenvector in ρ is amplified exponentially with respect to K . More precisely, from the spectral decomposition of ρ

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|, \quad (302)$$

where we adopt the convention that $p_i \geq p_{i+1}$, one can show [632] that

$$\frac{\text{Tr}(\rho^K \hat{O})}{\text{Tr}(\rho^K)} = \langle\psi_1|\hat{O}|\psi_1\rangle + \mathcal{O}(Q^K), \quad (303)$$

where $Q \equiv (p_1^{-1} - 1)p_2 < 1$.

Therefore, if one wants the difference between the contribution from the dominant eigenvector $\langle\psi_1|\hat{O}|\psi_1\rangle$ and the overall calculated value to be only ϵ , one needs K to be [632]:

$$K = \text{ceil}\left(\ln \frac{1}{\epsilon} + \frac{\ln(2p_2^{-1})}{\ln Q^{-1}}\right) \quad (304)$$

where $\text{ceil}(x)$ gives the ceiling of x , i.e. the smallest integer that is greater than or equal to x . The calculation of the numerator in Eq. (301) on a quantum computer can be achieved with the circuit outlined in Fig. 17, or by several alternative methods detailed in Appendix C.3.

To apply this method to the VQE algorithm, the mixed state ρ can be prepared by the ansatz circuit $U(\theta)$ in a noisy quantum computer, and then each Pauli observable P in the decomposed Hamiltonian is measured on ρ with this suppression technique. If the energy has been minimized, the dominant state $|\psi_1\rangle$ in ρ prepared by the ansatz circuit has energy exponentially close to the true ground state energy of the Hamiltonian. Therefore, exponential error suppression purifies the mixed state and helps to prepare a pure eigenstate of Hamiltonian when combined with VQE.

Meanwhile, the variance of the measured quantity, when $\hat{O}$ is again also a unitary operator, is comparable to the variance of measuring $\hat{O}$ directly. More precisely, the number of samples S required to achieve a small error ϵ_s^2 in variance of estimating either the numerator or the denominator is of the order [632]

$$\mathcal{O}[\epsilon_s^{-2(1+2f)}], \quad (305)$$

where $f = \ln(p_1^{-1})/\ln(Q^{-1}) > 0$, which converges to the limiting case without error mitigation method ($\mathcal{O}(\epsilon^{-2})$) as p_1 goes to 1 and p_2 goes to 0. Next, we note that $|\psi_1\rangle$ is not necessarily the ideal error-free computation result $U(\theta)|0\rangle$. Although

this does not affect the VQE algorithm itself as mentioned above, it affects some VQE-based algorithms. For example, the excited state VQE algorithms (see Section 9.1) that measure the fidelity between $|\psi_1\rangle$ and other quantum states, are affected by the fact that the inverse unitary $U(\theta)^\dagger$ no longer prepares the adjoint $\langle\psi_1|$ from $|0\rangle$. Although there is no known efficient solution to this problem, the infidelity between the ideal error-free computation result and $|\psi_1\rangle$, called the coherent mismatch, has been studied extensively in Ref. [637], where it is shown that the error from coherent mismatch can be guaranteed to be smaller than the error caused by the limitation in creating more copies (i.e. the limitation in K).

Finally, it is important to mention that there are several methods with different advantages to compute the quantities in Eq. (301), which we provide a high-level summary in Appendix C.3. Some of the methods require several copies of the same state ρ . In these methods, each copy $\rho^{(i)}$ is slightly different from ρ :

$$\rho^K \mapsto \rho^{(1)}\rho^{(2)}\dots\rho^{(K)}, \quad \rho^{(i)} \approx \rho. \quad (306)$$

However, as long as the dominant state (the state with the largest amplitude) in the spectral decomposition of $\rho^{(i)}$ remains $|\psi_i\rangle$, which is a reasonable assumption, the conclusions made above remain valid with minor modifications (see Ref. [632]).

8.5. Measurement readout error mitigation

State preparation and measurement (SPAM) errors are also important in NISQ quantum computation. It is often important to counter their influence with simple mitigation measures in order to achieve good accuracy on VQE experiments [263,619], which is particularly important for Pauli operators with large weight. In particular, the choice of encoding (see Section 4) has a direct implication for the level of readout errors affecting the VQE algorithm. For instance, the operators in the Jordan–Wigner mapping can have up to $\mathcal{O}(n)$ Pauli weight (non-identity operators), while the Bravyi–Kitaev mapping has up to $\mathcal{O}(\log(n))$ Pauli weight. This in turn means that less of the qubits are measured on each operator, thereby lowering the risk of readout error [124] (although once the operators are grouped, this advantage of low Pauli weight encoding disappears as the entire register must be measured). Similarly, grouping methods based on commutative groups of Pauli strings (see Section 5.2) transform the measurement basis such that multiple operators can be jointly measured by simply measuring one qubit at a time and therefore reduces measurement error probability (but could result in a much larger impact in case of error).

To mitigate this type of noise, typically one first performs basic calibration experiments on the quantum computer to obtain a measurement calibration matrix T (introduced below) which maps the probability of each state to be measured as each other, and use this information to mitigate SPAM errors on quantum computer outputs.

8.5.1. Measurement calibration matrix

The measurement calibration matrix is defined element-wise as

$$T_{i,j} = \text{Probability (measured state } i | \text{prepared in state } j), \quad (307)$$

where $i, j \in \{0, 1, \dots, 2^N - 1\}$, are all the computational basis states on the N qubits which we want to measure. T models measurement errors as a conventional Markov process, which can be justified using a specific quantum measurement mode [638]. T could be measured on a quantum computer by first preparing the state in the corresponding basis j using X gates, and measuring the resulting quantum computation in the computational basis. However, since the size of T grows exponentially with respect to the number of measured qubits, it can rapidly become too expensive to obtain. In this case, an approximation based on the assumption that crosstalk between different qubits during the measurement process is negligible helps alleviate the problem. Here, one approximates T by

$$T \approx T_1 \otimes T_2 \dots \otimes T_N, \quad (308)$$

where T_k is the measurement calibration matrix on the k th qubit among the N qubits to measure. $\{T_k\}_{k=1}^N$ could be obtained in a similar method, but it costs only $\mathcal{O}(N)$ number of measurements to obtain instead of $\mathcal{O}(2^N)$ for T . T_k is usually routinely reported by major quantum computing platforms such as IBM Quantum [639]. Although more elaborate schemes to efficiently estimate T while taking account of crosstalk effects on neighboring qubits have been studied in Ref. [640], they do not resolve the issue of exponential growth in the size of T matrix, and more research is needed to adapt this scheme for large n problems. At the moment, the approximation in Eq. (308) remains the best scalable approximation to T .

8.5.2. Correcting measurement outcomes

Having obtained T periodically during the experiment, the measurement outcome is corrected using T by the following procedure. One cannot simply apply the inverse of T to the measurement outcomes, since T may not be invertible, and the Moore–Penrose inverse of T may produce unphysical quantities, as even though T satisfies statistical properties (i.e. $T_{i,j} \geq 0$ and $\sum_i T_{i,j} = 1$), the inverse may not. Therefore, one approaches the inverse problem by solving the following optimization problem [641]

$$x = \underset{x}{\operatorname{argmin}} |Y - TX|^2, \quad (309)$$

subject to $\sum_i X_i = \sum_i Y_i$ and, $X_i \geq 0$, where Y is the vector of raw measurement outcome and x is the vector of error mitigated measurement outcomes. In the i th position of each vector is the number of occurrences of the measurement outcome in state i . The vector norm is defined as $|v| = \sqrt{v \cdot v}$. The condition ensures that the optimized value x is physical.

Obviously, the size of the matrix T and the vectors X and Y scales exponentially with N , the number of qubits measured, making it impractical for large N . To alleviate this problem, one can use the approximation scheme Eq. (308), and store both matrices and vectors in a sparse data format for processing. Meanwhile, correlated measurement errors between different qubits are ignored in this approximation, making it possible to derive a simple formula giving the mitigated measurement outcome from the noisy measurement outcome. This approach was first explored in Ref. [263], and later generalized [642,643] into the formula

$$\langle \hat{Z}_i \dots \hat{Z}_j \rangle = \sum_x p(x) \prod_{k=i, \dots, j} \frac{(-1)^{x_k} - p_k^-}{1 - p_k^+}, \quad (310)$$

where $\langle \hat{Z}_i \dots \hat{Z}_j \rangle$ is the mitigated measurement outcome of $\hat{Z}$ operators on qubits $i, \dots, j$, the summation $\sum_x$ is taken over all possible measurement outcomes x , $p(x)$ is the probability of each outcome x , x_k is the k th qubit component of x , $p_k^\pm = p_k(0|1) \pm p_k(1|0)$, $p_k(0|1)/p_k(1|0)$ is the probability of measuring 0/1 when the state is 1/0 on the k th qubit and can be easily obtain from T_k (Eq. (308)).

It is worth noting that when the measurement calibration matrix T is pathological; statistical uncertainties can be amplified and can result in oscillatory behavior of the mitigated measurement outcome [644]. Although an improved inversion method can alleviate this problem [644], when the readout errors are sufficiently small, the method mentioned above was shown experimentally to be good enough [644], see Ref. [644] for details of the improved method.

8.5.3. Exploiting state-dependent noise

In certain quantum computer architectures, the design of the quantum computer makes certain states more prone to measurement readout errors than other states. For example, state $|1\rangle$ is often prepared as a state having higher energy than $|0\rangle$, which may have a stronger tendency to drop back to $|0\rangle$ than for a state in $|0\rangle$ to be excited to state $|1\rangle$. In other words, the measurement error can be biased towards certain states. Ref. [645] examined this phenomenon on several IBMQ quantum computers and discovered a strong correlation of the measurement readout error and the hamming weight of the measured state. To decrease the readout error, Ref. [645] proposed splitting all the measurements into two parts, the standard part, and the inverted part. In the inverted part, additional X gates are inserted before all the measurements to invert the qubits, whereas in the standard part no modification is made. The measurement outcomes of the two parts are merged together with simple post-processing to invert the measurement outcomes on the inverted part. Therefore, the effect of the measurement bias on the readout can be averaged in the merged measurement outcomes and can lead to an improvement of performance. However, it was noted that the improvement in performance depends highly on the nature of the output state, and if the output state is favored by the measurement bias (such as the $|0 \dots 0\rangle$ state), this mitigation method may degrade the performance. Ref. [645] proposed another adaptive scheme to address this problem, which depends on our ability to predict the probability of certain states and its measurement bias relative to other states. The adaptive scheme therefore does not generalize well to experiments with more number of qubits. Therefore, we suggest accessing the performance of the non-adaptive approach on simplified quantum circuits where the exact quantum outcome can be calculated analytically before using this mitigation method.

8.6. Other error mitigation methods

There are a wide range of other error mitigation methods, which we mention briefly in what follows:

- Combining different error mitigation methods: Ref. [618] proposes combining exponential extrapolation techniques with quasi-probability and symmetry verification methods, showing by simulation that a lowered estimation bias and a reduced sampling cost could be achieved through mixing different methods. Ref. [646] combines the quasi-probability method (Section 8.3) with extrapolation methods (Section 8.2) to offer more advantages, including avoiding the need for a full gate-set tomography on the quantum computer, reducing the noise level to a “virtual” noise level that is below the hardware noise level, and potentially reducing the sampling cost of the quasi-probability method.
- Detecting errors in computation: Ref. [647] uses a simple 4-qubit error detecting code to filter out experiments where one detects an error on a quantum computer. Although it shows certain error mitigation capabilities, the limited set of quantum gates that can be executed with the error detecting code makes it difficult to apply this method to general VQE algorithms. Also, Ref. [648] uses different simplified circuits of the original VQE ansatz, constructed with Clifford gates, to detect a degradation in the quality of the output of the original circuit.
- Combining with reduced density matrix methods: Ref. [120] proposes using the physical constraint of reduced density matrices (RDMs) for error mitigation. This method has been experimentally tested in Ref. [50] for 2-RDMs where the McWeeny purification formula [649] is applied to constrain the RDMs as valid physical states.

- **Modeling the noise:** If the noise on the quantum computer can be described by an effective model, we can use data from the device to estimate the strength of noise in it and correct the measurement outcome based on this estimation. An example is discussed in the Appendix A of Ref. [159], where a simple probabilistic error model is used. In a similar spirit, Ref. [650] argues that a global depolarization model provides a good effective noise model on the quantum computer, which is used for error mitigation in the experiment to archive significant error reduction in the quantum algorithm. However, it remains unclear whether this method is scalable for algorithms requiring more qubits and deeper quantum circuits.
- **Incorporating machine learning:** we may take the advancement of machine learning technologies and apply them for mitigating errors. For example, one can learn the mapping from the noisy observables $\langle \hat{O} \rangle_{\text{noisy}}$ from the quantum computer to the exact one: $f : \langle \hat{O} \rangle_{\text{noisy}} \mapsto \langle \hat{O} \rangle_{\text{exact}}$. Although for a general quantum state $|\psi\rangle$ it is exponentially difficult to compute the expectation value $\langle \psi | \hat{O} | \psi \rangle$ on a conventional computer, such computation is efficient when $|\psi\rangle$ is prepared by certain class of quantum circuits: $|\psi'\rangle = U_{\text{simulable}}|0\rangle$. Therefore, one can use machine learning to learn a function f' mapping from $\langle \psi' | \hat{O} | \psi' \rangle_{\text{noisy}}$ to $\langle \psi' | \hat{O} | \psi' \rangle_{\text{exact}}$, and use same function f' mitigating errors on the expectation value on any state $\langle \psi | \hat{O} | \psi \rangle$, with assumption that f' is similar to f . This method has been explored experimentally for Clifford circuits [651] and for quantum circuits in the fermionic linear optics [652]. This approach has been further combined with the extrapolation based method mentioned in Section 8.2 [653]. Further, this idea has taken been further to develop artificial neural networks that predict the noise, or the correction needed, on given quantum circuits [654,655].
- **Noise aware circuit design:** Here, instead of trying to mitigate the noise from a given quantum circuit, it is asked whether one could design an ansatz that is better adapted to the noise on the quantum computer. For example, certain quantum circuits are naturally more resilient to quantum noise [656]. Then, machine learning can help produce more noise-resilient circuits for a specific task such as computing state overlap [657]. Recently, several works have been proposed applying machine learning approaches to design better ansätze (see Section 6 for a related discussion on ansatz design), and the resilience towards noise can be naturally incorporated in this design process [550,556].
- **Noise resilience from enhanced sampling:** Ref. [658] discusses a method to enhance the rate of information gain when sampling in quantum computers for estimating expectation values. It was found that by incorporating a well-calibrated noise model into their method, deeper quantum circuits can be utilized to obtain more accurate results in less time [659]. It remains to be seen whether this result holds for experiments involving more qubits.
- **Canceling noise using second derivatives:** Ref. [660] gives a different interpretation of noise on a quantum computer as a fluctuation of parameters in the cost function. This interpretation is valid for certain stochastic noise channels which include the depolarization noise channel. Based on this interpretation, the second derivatives of the cost function can be used to cancel noise in the cost function.

8.7. Impact of error mitigation on the scaling of VQE

Error mitigation procedures bring extra overheads to VQE algorithms and therefore affect the cost scaling. In this section, we describe how error mitigation procedures affect the scaling of VQE algorithms and discuss potential directions to alleviate the error mitigation overhead.

Firstly, error mitigation procedures tend to increase the cost of estimating expectation values of each Pauli term in the Hamiltonian. This can be quantified by a systematic increase of the variance of expectation values when more repeated experiments are required to ensure the statistical error of estimating the expectation value is smaller than a fixed threshold. Specific analysis of the increase in variance has been done for most error mitigation methods, which are summarized in the Table 12.

However, this analysis lacks a fair comparison between the cost of different error mitigation methods. Such a comparison is difficult due to the fact that the cost depends on the nature of noise on the quantum computer, the Hamiltonian considered, and sometimes on the desired accuracy of the method itself. Recently, Takagi et al. [96] attempt to set a general lower bound on the increase in variance in a broad class of error mitigation methods by showing that error mitigation methods tend to increase the difficulty of distinguishing different quantum states. It shows that

$$\gamma^2 \propto \frac{1 - 2b_{\max}}{K\sqrt{NQ}} \left(\frac{1}{1 - \varepsilon} \right)^{L/2}, \quad (311)$$

where γ^2 is the amplification ratio of the variance, $b_{\max}$ is the maximum error of the error mitigation method, also called the bias of the error mitigation method, N is the number of qubits, Q and K are constants that characterize the error mitigation method considered, L is the number of layers applied when using a layer-wise ansatz, and ε characterizes the noise present in the quantum computer. Hence, most error mitigation methods inevitably increase the variance exponentially with respect to both the number of qubits or the number of repeated layers in the ansatz circuit for VQE. This is expected, since, in comparison with the scalable error correction technology, where one actively interacts with the quantum system in order to suppress the errors in it, error mitigation methods are mostly “passive”, allowing errors to accumulate in the quantum system.

Table 12

The extra cost brought by different error mitigation methods is measured by a relative increase in the number of shots required to achieve certain statistical precision (relative shots amplification). Assuming a shot number of S_0 is required to achieve a statistical error of ϵ , with the introduction of error mitigation methods an increased number of shots λS_0 is required to achieve the same precision. Most of the increases in the methods presented in the table are due to the variances of observables being amplified by the error mitigation methods. An exception is the symmetry verification method where the extra cost comes from the probability that verification fails. Note that for the exponential error suppression method, the estimation can only be done approximately and depends on the factor f (which in turn depends on the amount of noise within the circuit), defined in Section 8.4. Although there is an exponential number of circuits that need to be evaluated in the probabilistic error cancellation method, in practice it only requires randomly sampling from these circuits, and the shot amplification is therefore only determined by its variance amplification factor γ_{PEC} . For the readout error mitigation method we expect the measurement calibration matrix T to increase the variance of the final observable, although an analytical result is currently not available.

Model	Relative shots amplification	Relevant section.
Symmetry verification [156,274,275]	$1/\text{Tr}[\Pi_s \rho]$	Section 8.1, Eq. (266).
Richardson extrapolation [276,277]	$\sum_i \beta_k^2$	Section 8.2.2, Eq. (285).
Exponential extrapolation [629]	$\frac{\sum_i \beta_k^2 \exp(2\epsilon f) + \exp(2k\epsilon f)}{(k-1)^2}$	Section 8.2.2, Eq. (289).
Probabilistic error cancellation [276]	γ_{PEC}^2	Section 8.3, Eq. (295).
Exponential error suppression [632,633]	(approx.) $\mathcal{O}(\frac{1}{\epsilon^{4f}})$	Section 8.4, Eq. (305).
Readout error mitigation	Analytical expression not available.	Section 8.5

Secondly, in the NISQ era, where the error correction technology is not applicable due to hardware limitations, there is substantial interest in studying and improving the efficiency of error mitigation methods. For example, the result of Ref. [96] indicates that the variance increase due to the error mitigation method may be offset by compromising on the accuracy of mitigation (see theorems 4 and 5 in Ref. [96]). Ref. [661] also provides an example where one trades off the accuracy of the mitigation method for a smaller variance (i.e. overhead) in the probabilistic error cancellation method (see Section 8.3). Another potential direction concerns the assumption made in Ref. [96] that the error mitigation method should apply to all possible quantum states. In a VQE algorithm this may not be necessary, since only the ground state is important, or more generally the eigenstates of a Hamiltonian. It remains to be researched whether the variance increase resulting from error mitigation methods can be adapted to specific subspaces of the whole Hilbert space in order to decrease its overhead.

Thirdly, it is important to note that it is unclear whether error mitigation is necessary during the initial training stage of the VQE algorithm. It is possible that error mitigation could be applied only when the optimization algorithm in VQE starts to converge. Choosing when to apply error mitigation methods can significantly reduce their cost, and therefore more research is required on this aspect.

Finally, we want to call for an investigation of how different error mitigation methods compare with each other. The increase in variance often depends on the mitigation parameter of each method, which often affects the accuracy of the method as well. Ref. [662] makes a fair comparison of different error mitigation methods based on the same noise model. There, it is shown that the symmetry verification method has a better extraction rate than several error mitigation methods analyzed (including the probabilistic error cancellation method and the exponential error suppression method mentioned in this section), which means that it can extract all components of the error mitigated state out of the noisy state. However, the size of the error mitigated state is measured by the *fidelity boost* metric in that paper, and is different for different error mitigation methods. Obviously, symmetry verification's fidelity boost depends on the number of symmetry elements available in the Hamiltonian. In general, although Cai [662] does not cover a few important error mitigation methods, notably the Richardson extrapolation methods, we believe that Cai provides an important step towards understanding the dependency of the computational cost on the noise in the computer, and a rigorous comparison of different error mitigation methods.

8.8. Noise robustness of VQE algorithms

It has been argued that the VQE algorithm is naturally robust to certain types of noise [38]. However, the reality is more complicated. For example, a systematic error such as a systematic deviation resulting from an over-rotation error can be mitigated in VQE by the optimization process. In general, if a coherent error results in a unitary $\tilde{U}_{\text{ansatz}}(\theta)$ that is different from the expected $U_{\text{ansatz}}(\theta)$, the optimization algorithm might be able to either correct it with a different angle θ_1 , satisfying $\tilde{U}_{\text{ansatz}}(\theta_1) = U_{\text{ansatz}}(\theta)$ [663], or the correction might not be needed as long as $\tilde{U}_{\text{ansatz}}(\theta)$ produced the desired ground state for VQE. However, this robustness also ties the results from a VQE to specific hardware, since the optimized angles can no-longer be adapted to another quantum computer. Furthermore, errors on the hardware affect the landscape of the cost function, potentially creating more local minima and therefore jeopardizing the optimization procedure. This has been experimentally investigated for the variational quantum factoring algorithm [664] (see Section 6.1 for relevant discussions) which is a close relative of the VQE algorithm discussed here. Finally, the variation in noise on different quantum computers poses a challenge to the parallelization of the VQE algorithm, see Section 2.7.2.

9. Beyond the ground state of isolated molecules: Extensions of VQE

Up until this point, we have provided details of the VQE in the context of finding the ground state of an isolated system. However, in reality, a molecule is generally coupled with a wider environment, as well as the physics being strongly influenced by its electronic excitations. In this section, we briefly review some of the modifications to VQE which enlarge the scope of applicability, including accessing beyond ground state properties of the system, as well as the use of VQE as a sub-component of other algorithms to access multi-resolution descriptions of larger systems.

9.1. Excited states VQE

The computation of excited states is key to many processes in quantum chemistry and materials science, governing the dominant optical, transport and reactive properties [665,666]. 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 optimization to avoid the variational collapse to the ground state. Conventional correlated quantum chemical approaches [667] include Equation of Motion (EOM) coupled-cluster [668], linear response theory [669], as well as multi-reference approaches for stronger correlation [216,670]. 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 diagonalized 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 [239,642,671], and low-rank excitations of the ground state motivated by an equation-of-motion formalism [672,673]. These approaches can also be used to yield improved ground state estimates [239,674].

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. [672], and subsequently implemented on a quantum device [675]. 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 [676,677].

Variational approaches: An alternative approach relies on directly optimizing 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 [518] (for details about this method, please refer back to Section 7.5). 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 [678]: $\hat{H}' = (\hat{H} - \gamma \mathbb{1})^2$, 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. [679] as a mean to probe many-body localization 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) [680] 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 minimize 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 optimization of the ansatz becomes increasingly more constrained with the number of excited states desired.

Higgott et al. [459] 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 minimize both the expectation value of the Hamiltonian and maximize 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 quantum cost that could be significant for a NISQ device (i.e. a large number of SWAP gates), or possibly 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. [681] propose to compute the overlap term with a low depth SWAP test, and uses variational time evolution [291]. Chan et al. [682] extend this excited state method by merging it with ADAPT-VQE [217] (see Section 6.3.1). Kottmann et al. [576] 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 (see Section 7). Wakaura and Suksomo [683] propose an adaptation of the VQE cost function to minimize 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 [683].

The discriminative VQE (DVQE) [460] is a further alternative method, and relies on training of a generative adversarial network to enforce orthogonality between the ground state and a trial excited state. Generative Adversarial Networks (GANs) are machine learning tools, which are composed of two neural networks competing against each other: a generator, which is trained to produce a specific data pattern (e.g. an image), and a discriminator, which is trained to distinguish between true instances of this data pattern, and generated instances. When the training is successful, the generator learns to generate data patterns, which are indistinguishable from true ones. This concept was ported to quantum computing with the Quantum GAN [684,685], a method, which can be used to learn an approximation of an unknown pure state. The DVQE proposed in [460] inverts the logic of the QGAN, forcing generator and discriminator to collaborate for the generator to generate a state, which is as easy to distinguish as possible from the ground state: an orthogonal state. In order to ensure that the generated state is the first excited state, one must at the same time minimize the expectation value of the Hamiltonian. Subsequent excited states can be found by repeating the procedure iteratively. The scaling of the depth required for the discriminator remains unknown and could become an impediment for the method.

An alternative approach is the Variance VQE method [686], which replaces the usual cost function of VQE by minimizing 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 Refs. [687–692]). 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. [686] by combining both energy and variance minimization in order to allow for computation of low lying excited states. Zhang et al. [693] propose an adaptive variant of this method to compute highly excited states of Hamiltonians. The ansatz is grown by choosing operators from a pool of Pauli operators, akin to the methods listed in Section 6.3.

Dynamical correlation functions Equilibrium dynamical correlation functions are the key quantities governing the linear response behavior of quantum systems, encoding the information of the excitation spectrum over all energy scales. These functions can either be represented in the time or frequency (energy) domain, with a key dynamic response function being the single-particle Green's function, describing the charged excitation spectrum of the system. Any method to systematically calculate individual excited state energies (e.g. via a quantum subspace expansion) and the relevant transition amplitudes coupling them to the ground state, can in principle compute these dynamical correlation functions via its spectral representation [694–698]. However, other VQE approaches exist which directly target these correlation functions in either the (imaginary) time or frequency domains [671,694,699]. These include VQE-based variational approaches to directly solve the linear equations for the response of a system at a given frequency [700], which can be cast as a modified cost function, with a similarly parameterized VQE ansatz [461,701,702]. While these approaches can describe the correlation functions to high accuracy over the whole energy range without restricting to a low-energy subspace, their challenge arises chiefly from the substantially more difficult optimization problem for the ansatz, originating from the larger condition number of the cost function, as well as the necessity for Hadamard tests to compute the transition amplitudes between the excited (or response) and ground states (for more details, see Appendix B).

9.2. 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 exploit 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. We have provided below three examples of embedding methods adapted in this way:

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 Section 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 [101,703]. Furthermore, the CAS-CI can be variationally optimized, 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) [478,704], 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:

$$|\Psi_{\text{CASSCF}}\rangle = |\mathbf{R}, \mathbf{c}\rangle = e^{-\mathbf{R}} \sum_{\mu} c_{\mu} |\mu\rangle, \quad (312)$$

where $\mathbf{R}$ parameterizes the single-particle anti-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 optimized on a classical device, while the parameterized 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 [705–707]. These CAS-based approaches were initially proposed in combination with VQE as a solver for the active space in Ref. [233] and were subsequently successfully demonstrated practically on quantum computers in Refs. [159,492,708], including self-consistent optimization of the active space.

It should be noted that in order to achieve this optimization 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 [159]. 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) [709]. 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 Rossmann et al. [710] within a VQE description of a correlated active space. Shade et al. [711] also extend these ideas to the reduced density matrix 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 [712,713] 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 [714], but cast as a static wavefunction theory (see below). In order to optimize 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 [715–719], and has been implemented on quantum computers with proof of principles for relevant applications such as protein–ligand interactions for drug design [152] (with an alternative method based on perturbation theory proposed in Ref. [151]). 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) [720–722] was also tested and implemented on a quantum device [159], 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 [714]. 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 for 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. [331] 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 realizations of correlated material systems via VQE-type parameterized algorithms in Refs. [694,695,697,723]. An alternative method to compute the Green's function over the whole energy range is based on the quantum subspace expansion [698].

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.

10. Conclusion and outlook

The VQE is among the most promising near-term applications for quantum computers. While it is supported by strong theoretical arguments, there are still numerous open questions about the future applicability of the method. In this review, we outline the most relevant research that has been produced on the different components of the VQE, and how these are connected. To begin with, the choice of the Hamiltonian has direct implications for the number of qubits required, and for the capacity of VQE to achieve accurate results. A canonical orbital basis in second quantization has been the dominant choice in the community due to limited number of qubits required, although alternatives such as plane wave bases [148,235] and frozen natural orbitals basis [85] have been proposed. Once the problem Hamiltonian has been constructed (and provided one decides to use second quantization), it is critical to decide on a mapping from fermionic operators to spin operators that best suits the system studied. While the Jordan–Wigner mapping [119] has been most commonly used due to its convenient implementation, low-weight mappings (in particular the ternary tree mapping [256] for molecular Hamiltonians, and the generalized superfast encoding [257] for lattice models) have been shown to bring significant benefits with respect to circuit depth [281] for fermionic-based ansätze and resilience to the barren plateau problem [87,91,258].

One of the most discussed drawbacks of VQE is the large number of measurements required to perform the optimization process [82,83,85]. Some savings can be accomplished by managing the distribution of measurements, by weighting shots allocated to each operators by their respective weights in the Hamiltonian [82,120,121] or by truncating terms with significant impact on the energy [38]. However, to significantly reduce the scaling of VQE one should turn towards methods which jointly measure Hamiltonian terms. At present for molecular Hamiltonian this can most efficiently be done using a decomposed interactions method [124,262]. For lattice models, however, because low-weight encoding are more easily achieved, a simple qubit-wise commutativity grouping scheme may be the most cost effective [38,120,122,123,157,263–267].

The core of the VQE remains the ansatz, or parameterized quantum circuit, chosen to model the trial wavefunction. Among fixed structured ansätze the k-Unitary paired Generalized Coupled Cluster (k-UpCCGSD) [80] and its extension [311] stand out due to their ability to achieve excellent accuracy (evidenced numerous times [80,219,294,311,483]), while maintaining a scaling linear in the number of qubits. For lattice models, while Unitary Coupled Cluster ansätze have been tested (for instance [294]), one may prefer to use an ansatz for which the physically-motivated construction is more adapted. One such ansatz is the Hamiltonian Variational Ansatz [82,150], and its extensions [235,297], which have been shown to be a promising methods for the study of many-body physics through VQE [150,235,297]. Adaptive methods (all stemming from ADAPT-VQE [217]), are promising avenues with demonstrated benefits [289], but their overall cost and scaling remains unclear.

The choice of optimizer directly impacts the ability of the VQE to converge and the overall cost of the method. Comparative studies remain sparse, and it is challenging to infer the ability of specific optimizer to navigate the complex optimization landscape of VQE [93]. However, two optimizers have received at least some level of numerical supports regarding their convergence rate and ability to avoid optimization pitfalls: the quantum natural gradient [582–584], and RotoSolve [133,134]. Finally, appropriate mitigation of quantum noise is likely required for VQE. Variational algorithms have been shown to exhibit at least some degree of ability to learn the biases created by quantum noise [38,94,95]. However, further methods, such as symmetry verification and extrapolation based methods, despite coming at a significant computational cost, could be required to achieve the target accuracy with VQE.

Despite numerical simulations being presented in the vast majority of papers proposing new methods, comprehensive, comparative and applied studies remain rare. This is most likely due to the rapid development of the field and the amount of different interacting components the VQE requires. Some notable exceptions (in particular [85]) have attempted to focus on best practices to study applicability of VQE to practical electronic structure problems. Further numerical studies of these aspects can undoubtedly help guiding future research in the field. We identify four broad areas of suggested

further research: (1) optimal measurement strategies to reduce the substantial number of shots required to estimate the expectation value of a Hamiltonian using a VQE ansatz; (2) the real cost and benefits of the massive parallelization potential of VQE, which is likely required for the method to ever be relevant; (3) the extent to which the method is subject to the barren plateau problem and whether current mitigation techniques are sufficient for VQE to scale, this would also involve further studying the complexity of the VQE optimization landscape, the expected convergence rate of the method and the relevance of various optimizers in avoiding local minima; and (4) the extent of the noise resilience of VQE, and the extent to which noise can be mitigated in a tractable manner.

One interesting aspect of VQE that our review has not covered is that numerous algorithms have been inspired by it. For example, the Full Quantum Eigensolver [724] (FQE) includes a quantum gradient descent to perform an optimization similar to VQE, but directly on the quantum computer to speed up convergence; the Projective Quantum Eigensolver [725] optimizes energy using residuals from projection of the Schrödinger equation, rather than gradients. The Quantum Assisted Simulator (QAS) [726,727] and its extensions [728] aims to avoid the barren plateau problem by replacing the usual optimization loop of VQE by a three-step process: selection of an ansatz, measurement of overlap matrices on a quantum computer, conventional computing post-processing based on the data from the second step. The Quantum Sampling Regression [729] also bypasses the VQE optimization loop and focuses on sampling the Fourier basis of the energy landscape using an ansatz to perform post-processing on a conventional computer, thereby bypassing the quantum–classical communication costs such as latency. The Variational Quantum State Eigensolver (VQSE) [730] adapts the VQE to cases in which one searches eigenvalues and eigenvectors of a quantum state density matrix. The Permutation VQE (PermVQE) [731] nests the VQE in an additional optimization loop with the objective to permute the order of the qubits at each iteration and minimize long-range correlations in ground state energy. The development of such algorithms is based on a deep knowledge of both the advantages and limitations of the VQE. The Filtering VQE [732] uses filtering operators in a VQE process to find ground states more efficiently and accurately, showing the method outperforms standard VQE and Quantum Approximate Optimization Algorithm in the case of a quadratic unconstrained binary optimization Hamiltonian [733]. Deep VQE [734] can be used to break down a given Hamiltonian into sub-units which can be solved separately to reduce the qubit requirements on very large systems. This method was also extended to excited states computation by Mizuta et al. [735]. Meitei et al. [736] presented a method, later optimized by Asthana et al. [737], to perform VQE directly at pulse level (ctrl-VQE), removing the need for a gate based ansatz. Finally, variational methods have also been used to model real and imaginary time evolution of quantum systems for ground and excited states discovery [291,738–744].

To assess the future worth of the VQE as an electronic structure method, one must also consider how it can be paired with other algorithms, such as those described in the paragraph above or with QPE, to help accelerate computation where needed. The algorithm is no doubt among the most promising methods for near-term applications of quantum computers. As is the case for many emerging technologies, it faces daunting obstacles that will need to be overcome before it can be used in practical calculations, but the VQE also promises to deliver outstanding applications in fields such as drug discovery, chemical engineering, or material sciences.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

J. Tilly is supported by an industrial CASE (iCASE) studentship, funded by and UK EPSRC [EP/R513143/1], in collaboration with University College London and Odyssey Therapeutics. H.C. is supported through a Teaching Fellowship from UCL. D.P. is supported by an industrial CASE (iCASE) studentship, funded by and UK EPSRC [EP/T517793/1], in collaboration with University College London and Odyssey Therapeutics. Y.L. is supported by the National Natural Science Foundation of China (Grant No. 11875050 and 12088101) and NSAF, China (Grant No. U1930403). E.G. is supported by the UK EPSRC [EP/P510270/1]. L.W. is an honorary research fellow at the Department of Computer Science, University College London. I.R. acknowledges the support of the UK government department for Business, Energy and Industrial Strategy through the UK National Quantum Technologies Programme. G.H.B. gratefully acknowledges support from the Royal Society via a University Research Fellowship, United Kingdom, as well as funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 759063).

Appendix A. Qubit encodings and fenwick trees

The qubit encodings that have been presented in Section 4.1 can be translated into the language of graph theory and be visualized in a simple way as Fenwick trees [293]. The tree corresponding to the Bravyi–Kitaev (like the one shown in Fig. 18 and in Fig 19) can be built recursively according to a procedure that mirrors the construction of the change of basis matrix of Eq. (52): the tree corresponding to the single-qubit encoding is the trivial graph, and the tree corresponding to the 2^x -qubit encoding is built from two copies of the tree corresponding to the 2^{x-1} -qubit encoding, where vertex k

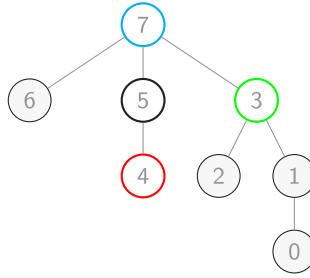


Fig. 18. Fenwick tree for the Bravyi–Kitaev encoding of 8 qubits showing the update set $U(5) = \{7\}$ (cyan), flip set $F(5) = \{4\}$ (red) and remainder set $R(5) = \{3\}$ (green) for qubit 5, from which we can read off the Bravyi–Kitaev operators (see Section 4.1.3).

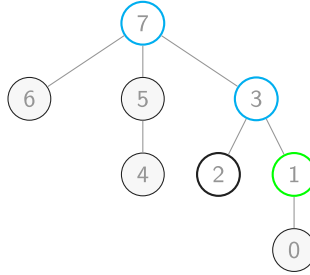


Fig. 19. The same Fenwick tree as in Fig. 18, this time showing the update set $U(2) = \{3, 7\}$ (cyan) and remainder set $R(2) = P(2) = \{1\}$ (green) for qubit 2 (because of how the Bravyi–Kitaev tree is constructed the leaves are the even vertices and so we see again that the flip set of an even qubit is always empty $F(2) = \emptyset$).

on the first tree corresponds to vertex $k + 2^{x-1}$ on the second, and the two trees are joined together by making vertex $2^{x-1} - 1$ a child of vertex $2^x - 1$.

The definitions of the qubit sets given when describing the Bravyi–Kitaev encoding (see Section 4.1.3) now translate into the following simple statements about Fenwick trees [293]:

- The update set of the j th qubit $U(j)$ corresponds then to the set of ancestors of vertex j on the tree.
- The flip set $F(j)$ is the set of children of the j th vertex.
- The remainder set $R(j)$ is the set of children of the ancestors of vertex j whose values is less than j .

This construction allows us to read off qubits sets for each qubit from the tree corresponding to our encoding (as shown in Fig. 18 and in Fig. 19 for our usual examples of qubits 2 and 5) and hence the representation of the creation and annihilation operators using Eq. (58). In the case where the number of qubits n is not a power of 2 we construct the Fenwick tree for the next power of 2 and determine the qubit sets, and then we discard from the qubit sets all the qubits greater or equal to n .

The parity encoding can also be represented this way, giving rise to the linear graph in Fig. 20. We can further generalize this to other encodings by considering disconnected Fenwick trees. We can then change our definition of $R(j)$ to include both the children of vertex j 's ancestors and the roots whose value is less than j . We then have that the Jordan–Wigner encoding corresponds to a totally disconnected graph (a graph with no edges). In the case of both encodings inspection of the corresponding Fenwick tree together with Eq. (58) recovers the same representations of $\hat{a}_j^\dagger$ and $\hat{a}_j$ operators as in Eq. (51) and (45). We can turn this argument upside down, and define new fermionic encodings starting from collections of Fenwick trees [293].

Appendix B. Hadamard test

Here we briefly introduce a common quantum subroutine called the Hadamard test [76]. Hadamard test is frequently used to compute the amplitude $\langle \psi | U | \psi \rangle$ (both its real and imaginary part) of an initial state $|\psi\rangle$ and a unitary gate U . This algorithm is summarized in the circuit diagram in Fig. 21. Here we explain the circuit in detail. To measure the real part of the amplitude, the quantum computer is initialized in a product state $|\psi\rangle \otimes |0\rangle$ with one ancilla qubit. The Hadamard gate Had converts the ancilla from $|0\rangle$ to $(|0\rangle + |1\rangle)/\sqrt{2}$. Then the controlled- U gate with control qubit on the ancilla results in the state $(U|\psi\rangle \otimes |1\rangle + |\psi\rangle \otimes |0\rangle)/\sqrt{2}$. The second Hadamard gate gives the state $((I + U)|\psi\rangle \otimes |0\rangle + (I - U)|\psi\rangle \otimes |1\rangle)/2$. In the final measurement, the probability of obtaining 0 is $(1 + \text{Re}\langle \psi | U | \psi \rangle)/2$, and the probability of obtaining 1 is $(1 - \text{Re}\langle \psi | U | \psi \rangle)/2$. Therefore, the difference between the two probabilities is $\text{Re}\langle \psi | U | \psi \rangle$. Similarly calculation shows that in the second circuit (Fig. 21(b)), the probability of obtaining 0 minus the probability of obtaining 1 is $\text{Im}\langle \psi | U | \psi \rangle$.

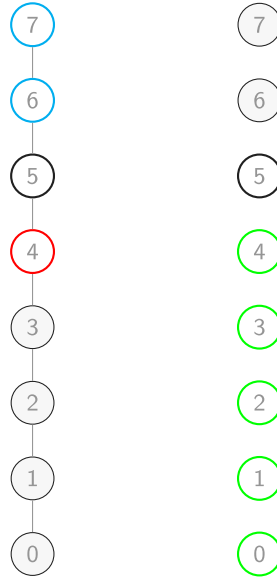


Fig. 20. Fenwick trees for the parity (left) and Jordan–Wigner (right) encoding of 8 qubits showing update $U(5)$ (cyan), flip $F(5)$ (red) and remainder sets $R(5)$ (green) for qubit 5.

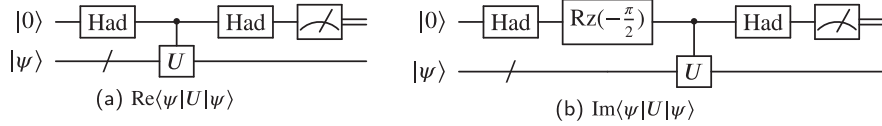


Fig. 21. Hadamard test circuits. The probability of measuring 0 minus the probability of measuring 1 in (a) and (b) gives respectively the real the imaginary part of the amplitude of $\langle\psi|U|\psi\rangle$.

Appendix C. Error mitigation appendix

C.1. Common noise models

Here we mention a few noise models that may be common in the literature, as for most relevant concepts in quantum computing, readers can refer to Ref. [76] for further details.

C.1.1. Relaxation rates

We start with the T_1 and T_2 relaxation time of a quantum computer. They are commonly used as a figure of merit for the noise robustness of a quantum computer. T_1 and T_2 capture the error on a single qubit quantum state ρ due to thermalization to an equilibrium state with the environment. Specifically, assume the initial quantum state ρ is

$$\rho = \begin{pmatrix} a & b \\ b^* & 1-a \end{pmatrix}. \quad (313)$$

Then, after a period of time t , the quantum state is changed by the noise $\mathcal{N}$ into [76]

$$\mathcal{N}(\rho) = \begin{pmatrix} (a - a_0)e^{-t/T_1} + a_0 & be^{-t/T_2} \\ b^*e^{-t/T_2} & (a_0 - a)e^{-t/T_1} + 1 - a_0 \end{pmatrix}, \quad (314)$$

where a_0 represents the thermal equilibrium state, and usually $a_0 = 1$ on superconducting qubits. The parameters T_1 and T_2 are also known as spin–lattice (or ‘longitudinal’) and spin–spin (or ‘transverse’) relaxation rates. They can be obtained experimentally and are widely available metrics for most quantum computers today.

C.1.2. Over-rotation

This type of error happens when a physical pulse used to generate a particular gate are mis-calibrated [745]. In the case of Pauli rotation gates $R_i(\theta)$, $i \in \{x, y, z\}$, this error manifests a small deviation δ in the angle θ , resulting in an actual

gate of modified rotation gate $R_i(\theta + \delta)$. The deviation may be systematic [746], but may also be stochastic and usually varies differently from different quantum computer.

C.1.3. Depolarizing noise model

The depolarizing model $\mathcal{N}_{\text{depolar}}$ is a common noise model. On a quantum state ρ of n qubits, it changes ρ according to

$$\mathcal{N}_{\text{depolar}}(\rho) = (1 - \varepsilon)\rho + \varepsilon \frac{\mathbb{1}}{2^n}, \quad (315)$$

where $\mathbb{1}$ is the identity matrix on n qubits, and ε characterizes the strength of the depolarizing noise.

C.1.4. Dephasing noise model

The dephasing noise model $\mathcal{N}_{\text{dephasing}}$ is related to the disappearance of off-diagonal part of a density matrix, hence closely related to the T_2 relaxation rate of a quantum computer. Its action on a single qubit state ρ is described by

$$\begin{aligned} \mathcal{N}_{\text{dephasing}}(\rho) &= (1 - \varepsilon)\rho + \varepsilon Z \rho Z \\ &= \begin{pmatrix} a & b(1 - 2\varepsilon) \\ b^*(1 - 2\varepsilon) & 1 - a \end{pmatrix}, \end{aligned} \quad (316)$$

where a and b characterize the initial state of ρ (Eq. (313)), ε characterizes the strength of this noise model. It is clear that the depolarizing noise does not affect the diagonal part of a single qubit state, therefore it commutes with most symmetry operator and could not be detected by the symmetry verification technique mentioned in Section 8.1.

C.1.5. Damping error

Here as suggested by the name, the damping error decreases the population of $|1\rangle$ in a single qubit state, and is closely related to the T_1 relaxation time. This error model can be described by the Kraus operators

$$K_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1 - \varepsilon} \end{pmatrix}, \quad K_1 = \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{\varepsilon} \end{pmatrix}, \quad (317)$$

where ε characterizes the strength of this noise model. Its action on the initial state ρ in Eq. (313) is

$$\begin{aligned} \mathcal{N}_{\text{damping}}(\rho) &= \sum_{i=0}^1 K_i \rho K_i^\dagger \\ &= \begin{pmatrix} 1 + (a - 1)(1 - \varepsilon) & b\sqrt{1 - \varepsilon} \\ b^*\sqrt{1 - \varepsilon} & (1 - a)(1 - \varepsilon) \end{pmatrix}. \end{aligned} \quad (318)$$

C.2. Example for probabilistic error cancellation

Here we present an illustrative example of using probabilistic error cancellation to correct the depolarizing error on a one-qubit quantum computer. The depolarizing noise $\mathcal{N}_{\text{dep}}$ changes an initial state ρ to

$$\mathcal{N}_{\text{dep}}(\rho) = \left(1 - \frac{3}{4}p\right)\rho + \frac{p}{4} \sum_{\sigma \in \{X, Y, Z\}} \sigma \rho \sigma. \quad (319)$$

Here $p \in [0, \frac{4}{3}]$ characterizes the strength of the depolarizing noise, and $\{X, Y, Z\}$ are the Pauli matrices. One can show that $\mathcal{N}_{\text{dep}}^{-1}$ can be written as [276]

$$\begin{aligned} \mathcal{N}_{\text{dep}}^{-1}(\rho) &= \gamma_{\text{PEC, dep}} \left(q_{\text{dep,1}} \rho + q_{\text{dep,2}} \sum_{\sigma \in \{X, Y, Z\}} \sigma \rho \sigma \right), \\ \gamma_{\text{PEC, dep}} &= \frac{p + 2}{2 - 2p}, \\ q_{\text{dep,1}} &= \frac{4 - p}{2p + 4}, \\ q_{\text{dep,2}} &= -\frac{p}{2p + 4}. \end{aligned} \quad (320)$$

Suppose the effective quantum channel of an ideal unitary U on quantum computer could be characterized as $\mathcal{N}_1 \circ U$, where the circle $\circ$ means composition of quantum channels. One could define three additional quantum circuits, each

formed by appending one of the three Pauli gates after the unitary U . One obtains the measurement outcome from the four circuits $\{U, X \circ U, Y \circ U, Z \circ U\}$ containing the original one, and calculates according to the weights in Eq. (320)

$$\begin{aligned}
 & \langle \hat{O} \rangle_{\text{PEC, dep}} \\
 &= \mathcal{V}_{\text{PEC, dep}} \left(q_{\text{dep},1} \underbrace{\text{Tr}[\hat{O} \mathcal{N}_1 \circ U(\rho)]}_{\text{original circuit}} + q_{\text{dep},2} \sum_{\sigma \in \{X,Y,Z\}} \underbrace{\text{Tr}[\hat{O} \sigma \circ \mathcal{N}_1 \circ U(\rho)]}_{\text{appended by a Pauli gate}} \right) \\
 &= \text{Tr} \left[\hat{O} \underbrace{\mathcal{N}_1^{-1}(\mathcal{N}_1 \circ U(\rho))}_{\text{quasi-probability representation of } U \text{ acting on } \rho} \right] \\
 &= \text{Tr}[\hat{O} U(\rho)].
 \end{aligned} \tag{321}$$

Therefore, the noise effect of $\mathcal{N}_1$ can be canceled with the PEC method.

C.3. Implementation of exponential error suppression

Here we will mention several methods to actually compute the value in Eq. (301) using quantum computers. For simplicity, we mostly consider the computation of the numerator

$$\text{Tr}(\rho^K \hat{O}), \tag{322}$$

since the denominator is a special case of the numerator with $\hat{O}$ being the identity observable.

C.3.1. Ancilla assisted method

The first method needs the assistance of an extra ancilla qubits. Considering n copies of the same state $\rho^{\otimes K}$. We want to measure the observable O only on one system, but in as an overlap of the original state, and a permuted state. Specifically, we want to compute the quantity

$$\begin{aligned}
 & \langle \psi_{i_1} | \otimes \cdots \otimes \langle \psi_{i_n} | O^{(1)} | \psi_{s(i_1)} \rangle \otimes \cdots \otimes | \psi_{s(i_K)} \rangle \\
 &= \langle \psi_{i_1} | O | \psi_{i_1} \rangle \delta_{i_1, s(i_1)} \cdots \delta_{i_n, s(i_K)},
 \end{aligned} \tag{323}$$

where for simplicity we consider only one possibility $|\psi_{i_1}\rangle \otimes \cdots \otimes |\psi_{i_K}\rangle$ in the statistical mixture $\rho^{\otimes K}$, $O^{(1)}$ is the same observable O on any one of the n copies of the same system, s is called the K -cycle derangement in the Group theory. Explicitly, s is a permutation in the indices $\{1, 2, \dots, K\}$, and can be written as mappings

$$\begin{aligned}
 & s(i_1) = i_2 \\
 & s(i_2) = i_3 \\
 & \dots \\
 & s(i_{n-1}) = i_K \\
 & s(i_n) = i_1,
 \end{aligned} \tag{324}$$

where all indices $\{i_1, \dots, i_K\}$ are distinct numbers. An example is

$$s(i) = i + 1, (i < K), s(K) = 1. \tag{325}$$

s being a permutation of all indices (i.e. $\{i_1, \dots, i_K\}$ are distinct numbers above) enforces the delta functions $\delta_{i_1, s(i_1)} \cdots \delta_{i_n, s(i_K)}$ in Eq. (323). For the statistical mixture $\rho^{\otimes K}$, we are effectively computing $\text{Tr}[\rho^K O]$ since

$$\begin{aligned}
 & \sum_{i_1 \cdots i_n} p_{i_1} \cdots p_{i_K} \\
 & \times \langle \psi_{i_1} | \otimes \cdots \otimes \langle \psi_{i_K} | O^{(1)} | \psi_{s(i_1)} \rangle \otimes \cdots \otimes | \psi_{s(i_K)} \rangle \\
 &= \sum_{i_1} p_{i_1}^K \langle \psi_{i_1} | O | \psi_{i_1} \rangle = \text{Tr}[\rho^K O].
 \end{aligned} \tag{326}$$

Given an explicit expression for s , it is easy to construct a circuit D_K which performs the unitary mapping

$$|\psi_{i_1}\rangle \otimes \cdots \otimes |\psi_{i_K}\rangle \mapsto |\psi_{s(i_1)}\rangle \otimes \cdots \otimes |\psi_{s(i_K)}\rangle. \tag{327}$$

For example, for the s in Eq. (325), D_m can be constructed by performing CNOT operations on pairs of qubits $(1, 2), \dots, (K-1, K), (K, 1)$. Having constructed $|\psi_{s(i_1)}\rangle \otimes \cdots \otimes |\psi_{s(i_K)}\rangle$, measuring the overlap in Eq. (323) could be achieved easily with a modified Hadamard test, which is illustrated in the circuit diagram in Fig. 17(a).

We note that several improvements of the derangement operation D_K exist in the literature [633,747]. In particular, Ref. [747] utilizes a deep circuit to achieve derangement on a single copy, avoiding creating K copies of ρ and applying the derangement operation on all K copies. Although D_K seems to be prone to errors on actual quantum computer due

to its long-range nature, as long as the error does not affect the orthogonal relations in Eq. (323), the noisy D_K is still a valid derangement operation, and it is shown in Ref. [632] that the error induced by D_K could be mitigated effectively with extrapolation based mitigation techniques.

C.3.2. Diagonalization method

In the second method, we reformulate the derangement operation as a matrix S , and implement a unitary version of S and the observable after the derangement operation OS in order to estimate the denominator and the numerator of Eq. (301). Compared with the previous method, this method does not require additional ancilla qubit and the long range controlled- D_K operation (Fig. 17(a)). This method only performs local operations B_i which connects the same set of qubits in each copy of the state ρ in $\rho^{\otimes K}$. We will first describe a general method, and then specialize on a specific case ($K = 2$, O acts only one qubit only) to illustrate the potential of this method.

We can define a matrix S which swaps quantum systems as defined by the derangement permutation s in Eq. (324)

$$S|\psi_{i_1}\rangle \otimes \cdots \otimes |\psi_{i_K}\rangle = |\psi_{s(i_1)}\rangle \otimes \cdots \otimes |\psi_{s(i_K)}\rangle. \quad (328)$$

It should be easy to verify the equalities (see for example Eq. (323))

$$\text{Tr}(O^{(1)}S\rho^{\otimes K}) = \text{Tr}(O\rho^K), \quad \text{Tr}(S\rho^{\otimes K}) = \text{Tr}(\rho^K), \quad (329)$$

where the observable $O^{(1)}$ is the same observable O on the first one of the n copies of the system. It should be noted that S in Eq. (329) is not intended to define a unitary operation, since it acts on the density matrix ρ as $S\rho$ instead of $S\rho S^\dagger$. Although S and $O^{(1)}S$ are not Hermitian in general, in many cases they may be diagonalized by unitary matrices. That is, we could find unitary matrices B_S and B_{OS} such that

$$B_S \Lambda_S B_S^\dagger = S, \quad B_{OS} \Lambda_{OS} B_{OS}^\dagger = O^{(1)}S, \quad (330)$$

where Λ_S and Λ_{OS} are diagonal matrices. Then, in principle we may perform the unitaries B_S and B_{OS} on the state $\rho^{\otimes K}$, and measure the K copies of ρ each in their computation basis to obtain the numerator and denominator of Eq. (301). Naively, it may seem difficult to calculate the diagonalization since both S and $O^{(1)}S$ operate on the large Hilbert space of $\rho^{\otimes K}$. However, since S permutes the K copies of ρ , S can be easily factorized into tensor products

$$S = S_1 \otimes S_2 \otimes \cdots \otimes S_N, \quad (331)$$

where N is the number of qubits of state ρ , S_i permutes all the i th qubit on each copy. For example, when $K = 2$, S_i interchanges the qubits in the same position on the two copies of ρ , i.e.

$$S_i = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (332)$$

Therefore, for the qubits which are not related to the observable O , diagonalizing S is simplified into diagonalizing S_i , which would not be difficult assuming K is not large. Meanwhile, in many cases of VQE, the observable O may be limited to act only on a few qubits. In this case, the diagonalization of $O^{(1)}S$ could be simplified into the diagonalization of S_i above, and the diagonalization of

$$O \otimes_{i \in \mathbb{S}_O} S_i, \quad (333)$$

where $\mathbb{S}_O$ is the subset of qubits on which O acts. Furthermore, when O is a one-qubit observable (e.g. $O = Z_1$), we can define the symmetrized version of O as

$$O_{\text{sym}} = \sum_{j=1}^K O^{(j)}. \quad (334)$$

where $O^{(j)}$ is the same observable O on the j th copy among n copies of ρ . Obviously O_{sym} and S commute. Additionally, O_{sym} and S can be factorized into local unitaries acting on the same qubit in the ρ among the K copies. Hence, the two diagonalization unitaries coincide ($B_O = B_{OS}$) and can also be factorized as tensor products of unitaries acting only on the same qubit in the ρ among the K copies. In particular, we illustrate an explicit example when $K = 2$ of this method in Fig. 22. Sec. II.A of Ref. [633] also gives the explicit formula for $\Lambda_{O/OS}$ and diagonalizing unitary $B_{O/OS}$ when $K = 2$.

Finally, using grouping methods (see Section 5), we can turn a complex multiple qubit measurement into a simple one-qubit measurement, making this implementation more appealing since it only needs local qubit connections.

C.3.3. Dual state purification method

It was shown in Ref. [635] that for the special case $K = 2$, exponential error suppression could be achieved without creating multiple copies of the state ρ , and without any controlled unitary gates. The disadvantage is that this method requires measuring all qubits as a final step, and it doubles the circuit depth. The core idea is to “measure” ρ on the state

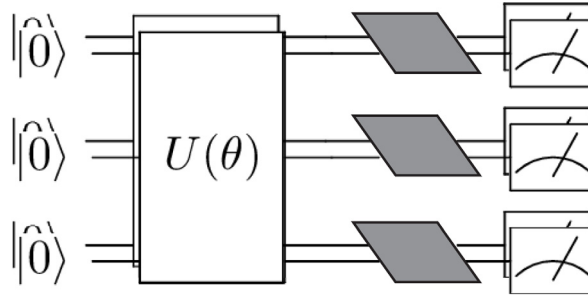


Fig. 22. An illustration of the exponential error suppression mitigation implemented with the simultaneous diagonalization method mentioned in Appendix C.3 (adopted from Ref. [633]). We prepare two copies of the same quantum state using the same ansatz circuit $U(\theta)$, and then apply the diagonalizing gate on each pair of the qubit as specified by Eq. (330), and measure the observables in the computation basis (Eq. (330)) in the computation basis.

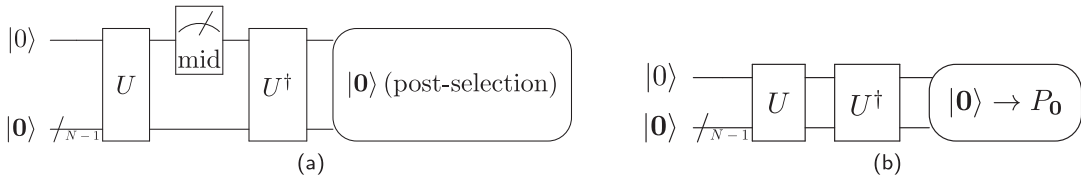


Fig. 23. Quantum circuit to measure Pauli Z on the first qubit using the dual state purification method in exponential error suppression (Appendix C.3.3). In (a), a mid-circuit measurement labeled by mid is carried out in addition to the final measurement on all qubits, and only experiments where all the final qubits are measured to be 0 are considered. In (b), no mid-circuit measurement is carried out, and it should be noted that dual to the noise present in the quantum computer, the combined effect of applying U and $U^\dagger$ is not identity in (b).

$O\rho$, therefore giving the observable $\text{Tr}(\rho O\rho) = \text{Tr}(O\rho^2)$. It should be noted that, although that same spirit can be applied to the general case where m is an even number $K = 2, 4, \text{etc.}$ [636], only in the case of $K = 2$ can we achieve exponential error suppression without the over-head of controlled unitary gates.

The circuit diagram in Fig. 23 shows an example of dual state purification method when the observable to be measured is Pauli-Z on the first qubit (denoted by Z_1). In the diagram, the circuit U which prepares a state ρ from $|0\rangle$ is applied twice: U in the first time and its inverse $U^\dagger$ in the second time. In the first diagram (Fig. 23(a)), a mid-circuit measurement is carried out, and one also post-select on the experiments where the final measurement on all qubits gives the same outcome 0. In this case, one denotes the probability that the final measurement obtaining 0 in all qubits when the mid-circuit measurement gives an outcome of $b \in \{0, 1\}$ as $P_{0,b}$. Also, in a separate experiment illustrated in Fig. 23(b), the mid-circuit measurement K is not carried out, and one denotes the probability that the final measurement obtaining 0 in all qubits as P_0 . Then, the two circuits outcome can be combined to the following value, which closely approximate $\text{Tr}(Z_1\rho^2)/\text{Tr}(\rho^2)$:

$$\frac{P_{0,0} - P_{0,1}}{P_0}. \quad (335)$$

This implementation can be adapted to measure any quantum observable O by expanding O into a linear sum of Pauli observables. Details could be found in Ref. [635] and an explicit extension of the circuit in Fig. 23 to arbitrary observable O which squares to the identity (e.g. Pauli observables) is available in the Appendix A of Ref. [636].

C.3.4. Shadow tomography based method

This method has been proposed in Ref. [748,749], and uses the shadow tomography technique (see Section 5.2.1) to estimate the quantity in Eq. (301), or more simply the numerator $\text{Tr}(\rho^K \hat{O})$ (Eq. (322)). The method proposed in Ref. [749] is more complicated and involves error correction codes, but shares the same spirit with Ref. [748] in using shadow tomography to estimate the numerator $\text{Tr}(\rho^K \hat{O})$. By applying a random unitary U generated from a pool on ρ , and measuring it on the computational basis, we can obtain a classical shadow $\tilde{\rho}$ of ρ (see Section 5.2.1 for details), and compute the numerator using the shadow $\tilde{\rho}$ as a surrogate for ρ . An explicit formula is available in Ref. [748] when the pool of random unitaries satisfy the 3-design property. One significant advantage of the shadow tomography based method over other methods presented above, is that shadow tomography does not require additional copies of the same state (Appendices C.3.1 and C.3.2) or doubling of the depth of the circuit (Appendix C.3.3). On the other hand, the number of required random unitaries (N_U) to sample in order to achieve an estimate of $\text{Tr}(\rho^K \hat{O})$ within certain precision, might scale exponentially with respect to the number of qubits N . A preliminary numerical study in Ref. [748] suggests that $N_U \approx 2^{0.82N}$ for $K = 2$, which is better than the cost of performing state tomography on ρ , but is out of reach for large scale quantum experiments.

References

- [1] P. Benioff, The computer as a physical system: A microscopic quantum mechanical Hamiltonian model of computers as represented by Turing machines, *J. Stat. Phys.* 22 (5) (1980) 563–591, <http://dx.doi.org/10.1007/bf01011339>.
- [2] R.P. Feynman, Simulating physics with computers, *Internat. J. Theoret. Phys.* 21 (6–7) (1982) 467–488, <http://dx.doi.org/10.1007/bf02650179>.
- [3] I.L. Chuang, N. Gershenfeld, M. Kubinec, Experimental implementation of fast quantum searching, *Phys. Rev. Lett.* 80 (15) (1998) 3408–3411, <http://dx.doi.org/10.1103/physrevlett.80.3408>.
- [4] J.A. Jones, M. Mosca, Approximate quantum counting on an NMR ensemble quantum computer, *Phys. Rev. Lett.* 83 (5) (1999) 1050–1053, <http://dx.doi.org/10.1103/physrevlett.83.1050>.
- [5] D.W. Leung, I.L. Chuang, F. Yamaguchi, Y. Yamamoto, Efficient implementation of coupled logic gates for quantum computation, *Phys. Rev. A* 61 (4) (2000) <http://dx.doi.org/10.1103/physreva.61.042310>.
- [6] L.M.K. Vandersypen, M. Steffen, G. Breyta, C.S. Yannoni, M.H. Sherwood, I.L. Chuang, Experimental realization of Shor's quantum factoring algorithm using nuclear magnetic resonance, *Nature* 414 (6866) (2001) 883–887, <http://dx.doi.org/10.1038/414883a>.
- [7] H. Häffner, W. Hänsel, C.F. Roos, J. Benhelm, D. Chek-al kar, M. Chwalla, T. Körber, U.D. Rapol, M. Riebe, P.O. Schmidt, C. Becher, O. Gühne, W. Dür, R. Blatt, Scalable multiparticle entanglement of trapped ions, *Nature* 438 (7068) (2005) 643–646, <http://dx.doi.org/10.1038/nature04279>.
- [8] C. Negrevergne, T.S. Mahesh, C.A. Ryan, M. Ditty, F. Cyr-Racine, W. Power, N. Boulant, T. Havel, D.G. Cory, R. Laflamme, Benchmarking quantum control methods on a 12-qubit system, *Phys. Rev. Lett.* 96 (17) (2006) <http://dx.doi.org/10.1103/physrevlett.96.170501>.
- [9] J.H. Plantenberg, P.C. de Groot, C.J.P.M. Harmans, J.E. Mooij, Demonstration of controlled-NOT quantum gates on a pair of superconducting quantum bits, *Nature* 447 (7146) (2007) 836–839, <http://dx.doi.org/10.1038/nature05896>.
- [10] D. Hanneke, J.P. Home, J.D. Jost, J.M. Amini, D. Leibfried, D.J. Wineland, Realization of a programmable two-qubit quantum processor, *Nat. Phys.* 6 (1) (2009) 13–16, <http://dx.doi.org/10.1038/nphys1453>.
- [11] T. Monz, P. Schindler, J.T. Barreiro, M. Chwalla, D. Nigg, W.A. Coish, M. Harlander, W. Hänsel, M. Hennrich, R. Blatt, 14-Qubit entanglement: creation and coherence, *Phys. Rev. Lett.* 106 (13) (2011) <http://dx.doi.org/10.1103/physrevlett.106.130506>.
- [12] S.J. Devitt, A.M. Stephens, W.J. Munro, K. Nemoto, Requirements for fault-tolerant factoring on an atom-optics quantum computer, *Nature Commun.* 4 (1) (2013) <http://dx.doi.org/10.1038/ncomms3524>.
- [13] S.J. Devitt, Performing quantum computing experiments in the cloud, *Phys. Rev. A* 94 (3) (2016) <http://dx.doi.org/10.1103/physreva.94.032329>.
- [14] T. Monz, D. Nigg, E.A. Martinez, M.F. Brandl, P. Schindler, R. Rines, S.X. Wang, I.L. Chuang, R. Blatt, Realization of a scalable Shor algorithm, *Science* 351 (6277) (2016) 1068–1070, <http://dx.doi.org/10.1126/science.aad9480>.
- [15] P. Jurcevic, A. Javadi-Abhari, L.S. Bishop, I. Lauer, D.F. Bogorin, M. Brink, L. Capelluto, O. Günlük, T. Itoko, N. Kanazawa, A. Kandala, G.A. Keefe, K. Krsulich, W. Landers, E.P. Lewandowski, D.T. McClure, G. Nannicini, A. Narasgond, H.M. Nayfeh, E. Pritchett, M.B. Rothwell, S. Srinivasan, N. Sundaresan, C. Wang, K.X. Wei, C.J. Wood, J.-B. Yau, E.J. Zhang, O.E. Dial, J.M. Chow, J.M. Gambetta, Demonstration of quantum volume 64 on a superconducting quantum computing system, *Quantum Sci. Technol.* 6 (2) (2021) 025020, <http://dx.doi.org/10.1088/2058-9565/abe519>.
- [16] J.M. Pino, J.M. Dreiling, C. Figgatt, J.P. Gaebler, S.A. Moses, M.S. Allman, C.H. Baldwin, M. Foss-Feig, D. Hayes, K. Mayer, C. Ryan-Anderson, B. Neyenhuis, Demonstration of the trapped-ion quantum CCD computer architecture, *Nature* 592 (7853) (2021) 209–213, <http://dx.doi.org/10.1038/s41586-021-03318-4>.
- [17] S. Ebadi, T.T. Wang, H. Levine, A. Keesling, G. Semeghini, A. Omran, D. Bluvstein, R. Samajdar, H. Pichler, W.W. Ho, S. Choi, S. Sachdev, M. Greiner, V. Vuletić, M.D. Lukin, Quantum phases of matter on a 256-atom programmable quantum simulator, *Nature* 595 (7866) (2021) 227–232, <http://dx.doi.org/10.1038/s41586-021-03582-4>.
- [18] J. Preskill, Quantum computing in the NISQ era and beyond, *Quantum* 2 (2018) 79, <http://dx.doi.org/10.22331/q-2018-08-06-79>.
- [19] M. Brooks, Beyond quantum supremacy: the hunt for useful quantum computers, *Nature* 574 (7776) (2019) 19–21, <http://dx.doi.org/10.1038/d41586-019-02936-3>.
- [20] F. Arute, K. Arya, R. Babbush, D. Bacon, J.C. Bardin, R. Barends, R. Biswas, S. Boixo, F.G.S.L. Brandao, D.A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, K. Guerin, S. Habegger, M.P. Harrigan, M.J. Hartmann, A. Ho, M. Hoffmann, T. Huang, T.S. Humble, S.V. Isakov, E. Jeffrey, Z. Jiang, D. Kafri, K. Kechedzhi, J. Kelly, P.V. Klimov, S. Knysh, A. Korotkov, F. Kostritsa, D. Landhuis, M. Lindmark, E. Lucero, D. Lyakh, S. Mandrà, J.R. McClean, M. McEwen, A. Megrant, X. Mi, K. Michielsen, M. Mohseni, J. Mutus, O. Naaman, M. Neeley, C. Neill, M.Y. Niu, E. Ostby, A. Petukhov, J.C. Platt, C. Quintana, E.G. Rieffel, P. Roushan, N.C. Rubin, D. Sank, K.J. Satzinger, V. Smelyanskiy, K.J. Sung, M.D. Trevithick, A. Vainsencher, B. Villalonga, T. White, Z.J. Yao, P. Yeh, A. Zalcman, H. Neven, J.M. Martinis, Quantum supremacy using a programmable superconducting processor, *Nature* 574 (7779) (2019) 505–510, <http://dx.doi.org/10.1038/s41586-019-1666-5>.
- [21] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu, P. Hu, X.-Y. Yang, W.-J. Zhang, H. Li, Y. Li, X. Jiang, L. Gan, G. Yang, L. You, Z. Wang, L. Li, N.-L. Liu, C.-Y. Lu, J.-W. Pan, Quantum computational advantage using photons, *Science* 370 (6523) (2020) 1460–1463, <http://dx.doi.org/10.1126/science.abe8770>.
- [22] Y. Wu, W.-S. Bao, S. Cao, F. Chen, M.-C. Chen, X. Chen, T.-H. Chung, H. Deng, Y. Du, D. Fan, M. Gong, C. Guo, C. Guo, S. Guo, L. Han, L. Hong, H.-L. Huang, Y.-H. Huo, L. Li, N. Li, S. Li, Y. Li, F. Liang, C. Lin, J. Lin, H. Qian, D. Qiao, H. Rong, H. Su, L. Sun, L. Wang, S. Wang, D. Wu, Y. Xu, K. Yan, W. Yang, Y. Yang, Y. Ye, J. Yin, C. Ying, J. Yu, C. Zha, C. Zhang, H. Zhang, K. Zhang, Y. Zhang, H. Zhao, Y. Zhao, L. Zhou, Q. Zhu, C.-Y. Lu, C.-Z. Peng, X. Zhu, J.-W. Pan, Strong quantum computational advantage using a superconducting quantum processor, 2021, [arXiv:2106.14734v1](https://arxiv.org/abs/2106.14734v1).
- [23] L.S. Madsen, F. Laudenbach, M.F. Askarani, F. Rortais, T. Vincent, J.F.F. Bulmer, F.M. Miatto, L. Neuhaus, L.G. Helt, M.J. Collins, A.E. Lita, T. Gerrits, S.W. Nam, V.D. Vaidya, M. Menotti, I. Dhand, Z. Vernon, N. Quesada, J. Lavoie, Quantum computational advantage with a programmable photonic processor, *Nature* 606 (7912) (2022) 75–81, <http://dx.doi.org/10.1038/s41586-022-04725-x>.
- [24] P. Shor, 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, pp. 124–134, <http://dx.doi.org/10.1109/sfcs.1994.365700>.
- [25] C.-Y. Lu, D.E. Browne, T. Yang, J.-W. Pan, Demonstration of a compiled version of Shor's quantum factoring algorithm using photonic qubits, *Phys. Rev. Lett.* 99 (25) (2007) <http://dx.doi.org/10.1103/physrevlett.99.250504>.
- [26] B.P. Lanyon, T.J. Weinhold, N.K. Langford, M. Barbieri, D.F.V. James, A. Gilchrist, A.G. White, Experimental demonstration of a compiled version of Shor's algorithm with quantum entanglement, *Phys. Rev. Lett.* 99 (25) (2007) <http://dx.doi.org/10.1103/physrevlett.99.250505>.
- [27] E. Lucero, R. Barends, Y. Chen, J. Kelly, M. Mariantoni, A. Megrant, P. O'Malley, D. Sank, A. Vainsencher, J. Wenner, T. White, Y. Yin, A.N. Cleland, J.M. Martinis, Computing prime factors with a Josephson phase qubit quantum processor, *Nat. Phys.* 8 (10) (2012) 719–723, <http://dx.doi.org/10.1038/nphys2385>.
- [28] E. Martín-López, A. Laing, T. Lawson, R. Alvarez, X.-Q. Zhou, J.L. O'Brien, Experimental realization of Shor's quantum factoring algorithm using qubit recycling, *Nat. Photonics* 6 (11) (2012) 773–776, <http://dx.doi.org/10.1038/nphoton.2012.259>.
- [29] I.L. Markov, M. Saeedi, Faster quantum number factoring via circuit synthesis, *Phys. Rev. A* 87 (1) (2013) <http://dx.doi.org/10.1103/physreva.87.012310>.

- [30] M. Amico, Z.H. Saleem, M. Mumph, Experimental study of Shor's factoring algorithm using the IBM Q experience, *Phys. Rev. A* 100 (1) (2019) <http://dx.doi.org/10.1103/physreva.100.012305>.
- [31] L.K. Grover, A fast quantum mechanical algorithm for database search, in: *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing, STOC '96*, Association for Computing Machinery, New York, NY, USA, ISBN: 0897917855, 1996, pp. 212–219, <http://dx.doi.org/10.1145/237814.237866>.
- [32] C.H. Bennett, E. Bernstein, G. Brassard, U. Vazirani, Strengths and weaknesses of quantum computing, *SIAM J. Comput.* 26 (5) (1997) 1510–1523, <http://dx.doi.org/10.1137/s0097539796300933>.
- [33] N.J. Cerf, L.K. Grover, C.P. Williams, Nested quantum search and NP-hard problems, *Appl. Algebra Eng. Commun. Comput.* 10 (4–5) (2000) 311–338, <http://dx.doi.org/10.1007/s002000050134>.
- [34] A. Ambainis, Quantum search algorithms, *ACM SIGACT News* 35 (2) (2004) 22–35, <http://dx.doi.org/10.1145/992287.992296>.
- [35] A. Ambainis, Quantum walk algorithm for element distinctness, *SIAM J. Comput.* 37 (1) (2007) 210–239, <http://dx.doi.org/10.1137/s0097539705447311>.
- [36] D.J. Bernstein, Grover vs. McEliece, in: *Lecture Notes in Computer Science*, Vol. 6061, Springer Berlin Heidelberg, 2010, pp. 73–80, http://dx.doi.org/10.1007/978-3-642-12929-2_6.
- [37] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P.J. Love, A. Aspuru-Guzik, J.L. O'Brien, A variational eigenvalue solver on a photonic quantum processor, *Nature Commun.* 5 (1) (2014) <http://dx.doi.org/10.1038/ncomms5213>.
- [38] J.R. McClean, J. Romero, R. Babbush, A. Aspuru-Guzik, The theory of variational hybrid quantum-classical algorithms, *New J. Phys.* 18 (2) (2016) 023023, <http://dx.doi.org/10.1088/1367-2630/18/2/023023>.
- [39] P. Deglmann, A. Schäfer, C. Lennartz, Application of quantum calculations in the chemical industry—An overview, *Int. J. Quantum Chem.* 115 (3) (2014) 107–136, <http://dx.doi.org/10.1002/qua.24811>.
- [40] B.J. Williams-Noonan, E. Yuriev, D.K. Chalmers, Free energy methods in drug design: Prospects of “alchemical perturbation” in medicinal chemistry, *J. Med. Chem.* 61 (3) (2017) 638–649, <http://dx.doi.org/10.1021/acs.jmedchem.7b00681>.
- [41] A. Heifetz (Ed.), *Quantum Mechanics in Drug Discovery*, Springer US, ISBN: 9781071602829, 2020, <http://dx.doi.org/10.1007/978-1-0716-0282-9>.
- [42] M.A. Continentino, *Key Methods and Concepts in Condensed Matter Physics*, IOP Publishing, ISBN: 9780750333955, 2021, <http://dx.doi.org/10.1088/978-0-7503-3395-5>.
- [43] A.V. der Ven, Z. Deng, S. Banerjee, S.P. Ong, Rechargeable Alkali-ion battery materials: Theory and computation, *Chem. Rev.* 120 (14) (2020) 6977–7019, <http://dx.doi.org/10.1021/acs.chemrev.9b00601>.
- [44] Y. Cao, J. Romero, A. Aspuru-Guzik, Potential of quantum computing for drug discovery, *IBM J. Res. Dev.* 62 (6) (2018) 6:1–6:20, <http://dx.doi.org/10.1147/jrd.2018.2888987>.
- [45] N.S. Blunt, J. Camps, O. Crawford, R. Izsák, S. Leontica, A. Mirani, A.E. Moylett, S.A. Scivier, C. Sünderhauf, P. Schopf, J.M. Taylor, N. Holzmann, A perspective on the current state-of-the-art of quantum computing for drug discovery applications, 2022, <http://dx.doi.org/10.48550/ARXIV.2206.00551>, URL: <https://arxiv.org/abs/2206.00551>.
- [46] V. Lordi, J.M. Nichol, Advances and opportunities in materials science for scalable quantum computing, *MRS Bull.* 46 (7) (2021) 589–595, <http://dx.doi.org/10.1557/s43577-021-00133-0>.
- [47] Y. Cao, J. Romero, J.P. Olson, M. Degroote, P.D. Johnson, M. Kieferová, I.D. Kivlichan, T. Menke, B. Peropadre, N.P.D. Sawaya, S. Sim, L. Veis, A. Aspuru-Guzik, Quantum chemistry in the age of quantum computing, *Chem. Rev.* 119 (19) (2019) 10856–10915, <http://dx.doi.org/10.1021/acs.chemrev.8b00803>.
- [48] Y. Zhou, E.M. Stoudenmire, X. Waintal, What limits the simulation of quantum computers? *Phys. Rev. X* 10 (4) (2020) <http://dx.doi.org/10.1103/physrevx.10.041038>.
- [49] S. Boixo, S.V. Isakov, V.N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M.J. Bremner, J.M. Martinis, H. Neven, Characterizing quantum supremacy in near-term devices, *Nat. Phys.* 14 (6) (2018) 595–600, <http://dx.doi.org/10.1038/s41567-018-0124-x>.
- [50] A.J. McCaskey, Z.P. Parks, J. Jakowski, S.V. Moore, T.D. Morris, T.S. Humble, R.C. Pooser, Quantum chemistry as a benchmark for near-term quantum computers, *Npj Quantum Inf.* 5 (1) (2019) <http://dx.doi.org/10.1038/s41534-019-0209-0>.
- [51] M. Silverman, *Quantum Superposition : Counterintuitive Consequences of Coherence, Entanglement, and Interference*, Springer, Berlin, ISBN: 3540718834, 2008.
- [52] L.E. Ballentine, *Quantum superposition: Counterintuitive consequences of coherence, entanglement and interference* quantum superposition: Counterintuitive consequences of coherence, entanglement and interference mark P.Silverman. 378 pp. Springer, Berlin, 2008. Price: \$89.95 ISBN 978-3-540-71883-3, *Amer. J. Phys.* 76 (11) (2008) 1078–1079, <http://dx.doi.org/10.1119/1.2973810>.
- [53] T. Helgaker, P. Jørgensen, J. Olsen, *Molecular Electronic-Structure Theory*, John Wiley & Sons, Ltd, ISBN: 9781118531471, 2000, <http://dx.doi.org/10.1002/9781119019572>.
- [54] P. Kratzer, J. Neugebauer, The basics of electronic structure theory for periodic systems, *Front. Chem.* 7 (2019) <http://dx.doi.org/10.3389/fchem.2019.00106>.
- [55] X. Li, N. Govind, C. Isborn, A.E. DePrince, K. Lopata, Real-time time-dependent electronic structure theory, *Chem. Rev.* 120 (18) (2020) 9951–9993, <http://dx.doi.org/10.1021/acs.chemrev.0c00223>.
- [56] R. Miceli, M. McGuigan, Effective matrix model for nuclear physics on a quantum computer, in: *2019 New York Scientific Data Summit*, NYSDS, 2019, pp. 1–4, <http://dx.doi.org/10.1109/NYSDS.2019.8909693>.
- [57] O.D. Matteo, A. McCoy, P. Gysbers, T. Miyagi, R.M. Woloshyn, P. Navrátil, Improving Hamiltonian encodings with the Gray code, *Phys. Rev. A* 103 (4) (2021) <http://dx.doi.org/10.1103/physreva.103.042405>.
- [58] O. Kiss, M. Grossi, P. Lougovski, F. Sanchez, S. Vallecorsa, T. Papenbrock, Quantum computing of the ${}^6\text{Li}$ nucleus via ordered unitary coupled cluster, 2022, <http://dx.doi.org/10.48550/ARXIV.2205.00864>, URL: <https://arxiv.org/abs/2205.00864>.
- [59] A.M. Romero, J. Engel, H.L. Tang, S.E. Economou, Solving nuclear structure problems with the adaptive variational quantum algorithm, 2022, <http://dx.doi.org/10.48550/ARXIV.2203.01619>, URL: <https://arxiv.org/abs/2203.01619>.
- [60] M.C. Bañuls, R. Blatt, J. Catani, A. Celi, J.I. Cirac, M. Dalmonte, L. Fallani, K. Jansen, M. Lewenstein, S. Montangero, C.A. Muschik, B. Reznik, E. Rico, L. Tagliacozzo, K.V. Acoleyen, F. Verstraete, U.-J. Wiese, M. Wingate, J. Zakrzewski, P. Zoller, Simulating lattice gauge theories within quantum technologies, *Eur. Phys. J. D* 74 (8) (2020) <http://dx.doi.org/10.1140/epjd/e2020-100571-8>.
- [61] S.D. Bass, E. Zohar, Quantum technologies in particle physics, *Phil. Trans. R. Soc. A* 380 (2216) (2021) <http://dx.doi.org/10.1098/rsta.2021.0072>.
- [62] C.W. Bauer, Z. Davoudi, A.B. Balantekin, T. Bhattacharya, M. Carena, W.A. de Jong, P. Draper, A. El-Khadra, N. Gemelke, M. Hanada, D. Kharzeev, H. Lamm, Y.-Y. Li, J. Liu, M. Lukin, Y. Meurice, C. Monroe, B. Nachman, G. Pagano, J. Preskill, E. Rinaldi, A. Roggero, D.I. Santiago, M.J. Savage, I. Siddiqi, G. Siopsis, D. Van Zanten, N. Wiebe, Y. Yamauchi, K. Yeter-Aydeniz, S. Zorzetti, Quantum simulation for high energy physics, 2022, <http://dx.doi.org/10.48550/ARXIV.2204.03381>, URL: <https://arxiv.org/abs/2204.03381>.
- [63] S. McArdle, A. Mayorov, X. Shan, S. Benjamin, X. Yuan, Digital quantum simulation of molecular vibrations, *Chem. Sci.* 10 (22) (2019) 5725–5735, <http://dx.doi.org/10.1039/c9sc01313j>.
- [64] N.P.D. Sawaya, J. Huh, Quantum algorithm for calculating molecular vibronic spectra, *J. Phys. Chem. Lett.* 10 (13) (2019) 3586–3591, <http://dx.doi.org/10.1021/acs.jpclett.9b01117>.

- [65] P.J. Ollitrault, A. Baiardi, M. Reiher, I. Tavernelli, Hardware efficient quantum algorithms for vibrational structure calculations, *Chem. Sci.* 11 (26) (2020) 6842–6855, <http://dx.doi.org/10.1039/d0sc01908a>.
- [66] S. Jahangiri, J.M. Arrazola, N. Quesada, A. Delgado, Quantum algorithm for simulating molecular vibrational excitations, *Phys. Chem. Chem. Phys.* 22 (44) (2020) 25528–25537, <http://dx.doi.org/10.1039/d0cp03593a>.
- [67] E. Löststedt, K. Yamanouchi, T. Tsuchiya, Y. Tachikawa, Calculation of vibrational eigenenergies on a quantum computer: Application to the Fermi resonance in CO₂, *Phys. Rev. A* 103 (6) (2021) <http://dx.doi.org/10.1103/physreva.103.062609>.
- [68] N.P.D. Sawaya, F. Paesani, D.P. Tabor, Near- and long-term quantum algorithmic approaches for vibrational spectroscopy, *Phys. Rev. A* 104 (6) (2021) <http://dx.doi.org/10.1103/physreva.104.062419>.
- [69] K. Mitarai, Y.O. Nakagawa, W. Mizukami, Theory of analytical energy derivatives for the variational quantum eigensolver, *Phys. Rev. Res.* 2 (1) (2020) <http://dx.doi.org/10.1103/physrevresearch.2.013129>.
- [70] K. Omiya, Y.O. Nakagawa, S. Koh, W. Mizukami, Q. Gao, T. Kobayashi, Analytical energy gradient for state-averaged orbital-optimized variational quantum eigensolvers and its application to a photochemical reaction, *J. Chem. Theory Comput.* 18 (2) (2022) 741–748, <http://dx.doi.org/10.1021/acs.jctc.1c00877>.
- [71] J. Liu, L. Wan, Z. Li, J. Yang, Simulating periodic systems on a quantum computer using molecular orbitals, *J. Chem. Theory Comput.* 16 (11) (2020) 6904–6914, <http://dx.doi.org/10.1021/acs.jctc.0c00881>.
- [72] N. Yoshioka, T. Sato, Y.O. Nakagawa, Y. ya Ohnishi, W. Mizukami, Variational quantum simulation for periodic materials, *Phys. Rev. Res.* 4 (1) (2022) <http://dx.doi.org/10.1103/physrevresearch.4.013052>.
- [73] D.Z. Manrique, I.T. Khan, K. Yamamoto, V. Wichitwechkarn, D.M. Ramo, Momentum-space unitary coupled cluster and translational quantum subspace expansion for periodic systems on quantum computers, 2020, <http://dx.doi.org/10.48550/ARXIV.2008.08694>, URL: <https://arxiv.org/abs/2008.08694>.
- [74] M. Lubasch, J. Joo, P. Moinier, M. Kiffner, D. Jaksch, Variational quantum algorithms for nonlinear problems, *Phys. Rev. A* 101 (1) (2020) <http://dx.doi.org/10.1103/physreva.101.010301>.
- [75] A. Joseph, T. White, V. Chandra, M. McGuigan, Quantum computing of Schwarzschild-de sitter black holes and kantowski-sachs cosmology, 2022, <http://dx.doi.org/10.48550/ARXIV.2202.09906>, URL: <https://arxiv.org/abs/2202.09906>.
- [76] M.A. Nielsen, I.L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, Cambridge, ISBN: 9780511976667, 2009, <http://dx.doi.org/10.1017/cbo9780511976667>.
- [77] J.W. Rayleigh, In finding the correction for the open end of an organ-pipe, *Phil. Trans.* 161 (77) (1870).
- [78] W. Ritz, Über eine neue methode zur lösung gewisser variationsprobleme der mathematischen physik, *J. Reine Angew. Math.* 135 (1908) 1–61.
- [79] G.B. Arfken, H.J. Weber, Rayleigh-ritz variational technique, in: *Mathematical Methods for Physicists*, third ed., Academic Press, Orlando, FL, ISBN: 9781483277820, 1985, pp. 957–961.
- [80] J. Lee, W.J. Huggins, M. Head-Gordon, K.B. Whaley, Generalized unitary coupled cluster wave functions for quantum computation, *J. Chem. Theory Comput.* 15 (1) (2018) 311–324, <http://dx.doi.org/10.1021/acs.jctc.8b01004>.
- [81] A. Szabo, N. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, in: *Dover Books on Chemistry*, Dover Publications, ISBN: 9780486691862, 1996.
- [82] D. Wecker, M.B. Hastings, M. Troyer, Progress towards practical quantum variational algorithms, *Phys. Rev. A* 92 (4) (2015) 042303, <http://dx.doi.org/10.1103/physreva.92.042303>.
- [83] V.E. Elfving, B.W. Broer, M. Webber, J. Gavartin, M.D. Halls, K.P. Lorton, A. Bochevarov, How will quantum computers provide an industrially relevant computational advantage in quantum chemistry? 2020, [arXiv:2009.12472v1](https://arxiv.org/abs/2009.12472v1).
- [84] M. Kühn, S. Zanker, P. Deglmann, M. Marthaler, H. Weiß, Accuracy and resource estimations for quantum chemistry on a near-term quantum computer, *J. Chem. Theory Comput.* 15 (9) (2019) 4764–4780, <http://dx.doi.org/10.1021/acs.jctc.9b00236>.
- [85] J.F. Gonthier, M.D. Radin, C. Buda, E.J. Daskoćil, C.M. Abuan, J. Romero, Identifying challenges towards practical quantum advantage through resource estimation: the measurement roadblock in the variational quantum eigensolver, 2020, [arXiv:2012.04001v1](https://arxiv.org/abs/2012.04001v1).
- [86] J.R. McClean, S. Boixo, V.N. Smelyanskiy, R. Babbush, H. Neven, Barren plateaus in quantum neural network training landscapes, *Nature Commun.* 9 (1) (2018) <http://dx.doi.org/10.1038/s41467-018-07090-4>.
- [87] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, P.J. Coles, Cost function dependent barren plateaus in shallow parameterized quantum circuits, *Nature Commun.* 12 (1) (2021) 1791, <http://dx.doi.org/10.1038/s41467-021-21728-w>.
- [88] A.A. Holmes, H.J. Changlani, C.J. Umrigar, Efficient heat-bath sampling in fock space, *J. Chem. Theory Comput.* 12 (4) (2016) 1561–1571, <http://dx.doi.org/10.1021/acs.jctc.5b01170>.
- [89] C.O. Marrero, M. Kieferová, N. Wiebe, Entanglement induced barren plateaus, 2020, [arXiv:2010.15968v2](https://arxiv.org/abs/2010.15968v2).
- [90] T.L. Patti, K. Najafi, X. Gao, S.F. Yelin, Entanglement devised barren plateau mitigation, *Phys. Rev. Res.* 3 (3) (2021) <http://dx.doi.org/10.1103/physrevresearch.3.033090>.
- [91] A.V. Uvarov, J.D. Biamonte, On barren plateaus and cost function locality in variational quantum algorithms, *J. Phys. A* 54 (24) (2021) 245301, <http://dx.doi.org/10.1088/1751-8121/abfac7>, [arXiv:2011.10530v1](https://arxiv.org/abs/2011.10530v1); *J. Phys. A: Math. Theor.* 54 (24) (2021) 245301.
- [92] K. Sharma, M. Cerezo, L. Cincio, P.J. Coles, Trainability of dissipative perceptron-based quantum neural networks, 2020, [arXiv:2005.12458v1](https://arxiv.org/abs/2005.12458v1).
- [93] L. Bittel, M. Kliesch, Training variational quantum algorithms is NP-hard, *Phys. Rev. Lett.* 127 (12) (2021) <http://dx.doi.org/10.1103/physrevlett.127.120502>, [arXiv:2101.07267v1](https://arxiv.org/abs/2101.07267v1); *Phys. Rev. Lett.* 127 (2021) 120502.
- [94] E. Fontana, N. Fitzpatrick, D.M. Ramo, R. Duncan, I. Rungger, Evaluating the noise resilience of variational quantum algorithms, *Phys. Rev. A* 104 (2) (2021) <http://dx.doi.org/10.1103/physreva.104.022403>.
- [95] E. Fontana, M. Cerezo, A. Arrasmith, I. Rungger, P.J. Coles, Optimizing parameterized quantum circuits via noise-induced breaking of symmetries, 2020, [arXiv:2011.08763](https://arxiv.org/abs/2011.08763).
- [96] R. Takagi, S. Endo, S. Minagawa, M. Gu, Fundamental limits of quantum error mitigation, 2021, [arXiv:2109.04457v2](https://arxiv.org/abs/2109.04457v2).
- [97] D.R. Hartree, The wave mechanics of an atom with a non-Coulomb central field. Part II. Some results and discussion, *Math. Proc. Cambridge Philos. Soc.* 24 (1) (1928) 111–132, <http://dx.doi.org/10.1017/s0305004100011920>.
- [98] J.C. Slater, The self consistent field and the structure of atoms, *Phys. Rev.* 32 (3) (1928) 339–348, <http://dx.doi.org/10.1103/physrev.32.339>.
- [99] J.A. Gaunt, A theory of Hartree's atomic fields, *Math. Proc. Cambridge Philos. Soc.* 24 (2) (1928) 328–342, <http://dx.doi.org/10.1017/s0305004100015851>.
- [100] D.R. Hartree, Self-consistent field, with exchange, for beryllium, *Proc. R. Soc. Lond. A* 1 150 (869) (1935) 9–33, <http://dx.doi.org/10.1098/rspa.1935.0085>.
- [101] F. Jensen, *Introduction to Computational Chemistry*, Wiley, Chichester, UK Hoboken, NJ, ISBN: 978-1-118-82595-2, 2017.
- [102] E. Grant, L. Wossnig, M. Ostaszewski, M. Benedetti, An initialization strategy for addressing barren plateaus in parameterized quantum circuits, *Quantum* 3 (2019) 214, <http://dx.doi.org/10.22331/q-2019-12-09-214>.
- [103] J.D. Whitfield, J. Yan, W. Wang, J.T. Heath, B. Harrison, Quantum computing 2022, 2022, <http://dx.doi.org/10.48550/ARXIV.2201.09877>, URL: <https://arxiv.org/abs/2201.09877>.
- [104] S. McArdle, S. Endo, A. Aspuru-Guzik, S.C. Benjamin, X. Yuan, Quantum computational chemistry, *Rev. Modern Phys.* 92 (1) (2020) 15003, <http://dx.doi.org/10.1103/revmodphys.92.015003>.

- [105] B. Bauer, S. Bravyi, M. Motta, G.K.-L. Chan, Quantum algorithms for quantum chemistry and quantum materials science, *Chem. Rev.* 120 (22) (2020) 12685–12717, <http://dx.doi.org/10.1021/acs.chemrev.9b00829>.
- [106] M. Motta, J. Rice, Emerging quantum computing algorithms for quantum chemistry, 2021, [arXiv:2109.02873v1](https://arxiv.org/abs/2109.02873v1).
- [107] M. Cerezo, A. Arrasmith, R. Babbush, S.C. Benjamin, S. Endo, K. Fujii, J.R. McClean, K. Mitarai, X. Yuan, L. Cincio, P.J. Coles, Variational quantum algorithms, *Nat. Rev. Phys.* 3 (9) (2021) 625–644, <http://dx.doi.org/10.1038/s42254-021-00348-9>.
- [108] K. Bharti, A. Cervera-Lierta, T.H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J.S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, A. Aspuru-Guzik, Noisy intermediate-scale quantum (NISQ) algorithms, 2021, [arXiv:2101.08448v2](https://arxiv.org/abs/2101.08448v2).
- [109] D.A. Fedorov, B. Peng, N. Govind, Y. Alexeev, VQE method: A short survey and recent developments, 2021, [arXiv:2103.08505v1](https://arxiv.org/abs/2103.08505v1).
- [110] A. Anand, P. Schleich, S. Alperin-Lea, P.W.K. Jensen, S. Sim, M. Díaz-Tinoco, J.S. Kottmann, M. Degroote, A.F. Izmaylov, A. Aspuru-Guzik, A quantum computing view on unitary coupled cluster theory, 2021, [arXiv:2109.15176v1](https://arxiv.org/abs/2109.15176v1).
- [111] R.G. Parr, On the genesis of a theory, *Int. J. Quantum Chem.* 37 (4) (1990) 327–347, <http://dx.doi.org/10.1002/qua.560370407>.
- [112] R.A. Friesner, Ab initio quantum chemistry: Methodology and applications, *Proc. Natl. Acad. Sci. USA* 102 (19) (2005) 6648–6653, <http://dx.doi.org/10.1073/pnas.0408036102>.
- [113] V. Némoshkalenko, *Computational Methods in Solid State Physics*, Gordon and Breach Science Publishers, Amsterdam, ISBN: 9789056990947, 1998.
- [114] R.M. Martin, *Electronic Structure*, Cambridge University Press, 2004, <http://dx.doi.org/10.1017/cbo9780511805769>.
- [115] M.P. Marder, *Condensed Matter Physics*, John Wiley & Sons, Inc., 2010, <http://dx.doi.org/10.1002/9780470949955>.
- [116] J.W. Clark, M.L. Ristig (Eds.), *Theory of Spin Lattices and Lattice Gauge Models*, Springer Berlin Heidelberg, 1997, <http://dx.doi.org/10.1007/bfb0104298>.
- [117] W. Pauli, Über den zusammenhang des abschlusses der elektronengruppen im atom mit der komplexstruktur der spektren, *Z. Phys.* 31 (1) (1925) 765–783, <http://dx.doi.org/10.1007/bf02980631>.
- [118] D. Griffiths, P. Griffiths, *Introduction to Quantum Mechanics*, in: Pearson International Edition, Pearson Prentice Hall, ISBN: 978-0-13-111892-8, 2005, URL: <https://books.google.co.uk/books?id=BsvAQAAIAAJ>.
- [119] P. Jordan, E. Wigner, Über das Paulische äquivalenzverbot, *Z. Phys.* 47 (9) (1928) 631–651, <http://dx.doi.org/10.1007/BF01331938>.
- [120] N.C. Rubin, R. Babbush, J. McClean, Application of fermionic marginal constraints to hybrid quantum algorithms, *New J. Phys.* 20 (5) (2018) 053020, <http://dx.doi.org/10.1088/1367-2630/aab919>.
- [121] A. Arrasmith, L. Cincio, R.D. Somma, P.J. Coles, Operator sampling for shot-frugal optimization in variational algorithms, 2020, [arXiv:2004.06252v1](https://arxiv.org/abs/2004.06252v1).
- [122] P. Gokhale, O. Angiuli, Y. Ding, K. Gui, T. Tomesh, M. Suchara, M. Martonosi, F.T. Chong, Minimizing state preparations in variational quantum eigensolver by partitioning into commuting families, 2019, [arXiv:1907.13623v1](https://arxiv.org/abs/1907.13623v1).
- [123] I. Hamamura, T. Imamichi, Efficient evaluation of quantum observables using entangled measurements, *Npj Quantum Inf.* 6 (1) (2020) 56, <http://dx.doi.org/10.1038/s41534-020-0284-2>.
- [124] W.J. Huggins, J.R. McClean, N.C. Rubin, Z. Jiang, N. Wiebe, K.B. Whaley, R. Babbush, Efficient and noise resilient measurements for quantum chemistry on near-term quantum computers, *Npj Quantum Inf.* 7 (1) (2021) 23, <http://dx.doi.org/10.1038/s41534-020-00341-7>.
- [125] G. Torlai, G. Mazzola, G. Carleo, A. Mezzacapo, Precise measurement of quantum observables with neural-network estimators, *Phys. Rev. Res.* 2 (2) (2020) 022060(R), <http://dx.doi.org/10.1103/physrevresearch.2.022060>.
- [126] H.-Y. Huang, R. Kueng, J. Preskill, Predicting many properties of a quantum system from very few measurements, *Nat. Phys.* 16 (10) (2020) 1050–1057, <http://dx.doi.org/10.1038/s41567-020-0932-7>.
- [127] Z. Holmes, K. Sharma, M. Cerezo, P.J. Coles, Connecting ansatz expressibility to gradient magnitudes and barren plateaus, 2021, [arXiv:2101.02138v1](https://arxiv.org/abs/2101.02138v1).
- [128] K. Nakaji, N. Yamamoto, Expressibility of the alternating layered ansatz for quantum computation, *Quantum* 5 (2021) 434, <http://dx.doi.org/10.22331/q-2021-04-19-434>.
- [129] G.E. Crooks, Gradients of parameterized quantum gates using the parameter-shift rule and gate decomposition, 2019, [arXiv:1905.13311v1](https://arxiv.org/abs/1905.13311v1).
- [130] A.F. Izmaylov, R.A. Lang, T.-C. Yen, Analytic gradients in variational quantum algorithms: Algebraic extensions of the parameter-shift rule to general unitary transformations, 2021, [arXiv:2107.08131v2](https://arxiv.org/abs/2107.08131v2).
- [131] T. Haug, K. Bharti, M.S. Kim, Capacity and quantum geometry of parameterized quantum circuits, 2021, [arXiv:2102.01659v2](https://arxiv.org/abs/2102.01659v2).
- [132] T. Haug, M.S. Kim, Optimal training of variational quantum algorithms without barren plateaus, 2021, [arXiv:2104.14543v3](https://arxiv.org/abs/2104.14543v3).
- [133] K.M. Nakanishi, K. Fujii, S. Todo, Sequential minimal optimization for quantum-classical hybrid algorithms, *Phys. Rev. Res.* 2 (4) (2020) <http://dx.doi.org/10.1103/physrevresearch.2.043158>.
- [134] M. Ostaszewski, E. Grant, M. Benedetti, Structure optimization for parameterized quantum circuits, *Quantum* 5 (2021) 391, <http://dx.doi.org/10.22331/q-2021-01-28-391>, [arXiv:1905.09692](https://arxiv.org/abs/1905.09692).
- [135] B. Koczor, S.C. Benjamin, Quantum natural gradient generalised to non-unitary circuits, 2019, [arXiv:1912.08660v4](https://arxiv.org/abs/1912.08660v4).
- [136] E. Farhi, J. Goldstone, S. Gutmann, A quantum approximate optimization algorithm, 2014, [arXiv:1411.4028v1](https://arxiv.org/abs/1411.4028v1).
- [137] C. Bravo-Prieto, R. LaRose, M. Cerezo, Y. Subasi, L. Cincio, P.J. Coles, Variational quantum linear solver, 2019, [http://dx.doi.org/10.48550/ARXIV.1909.05820](https://arxiv.org/abs/1909.05820), URL: <https://arxiv.org/abs/1909.05820>.
- [138] A. Pellow-Jarman, I. Sinayskiy, A. Pillay, F. Petruccione, A comparison of various classical optimizers for a variational quantum linear solver, *Quantum Inf. Process.* 20 (6) (2021) 202, <http://dx.doi.org/10.1007/s11128-021-03140-x>.
- [139] Y. Kwak, W.J. Yun, S. Jung, J. Kim, Quantum neural networks: Concepts, applications, and challenges, 2021, [http://dx.doi.org/10.48550/ARXIV.2108.01468](https://arxiv.org/abs/2108.01468), URL: <https://arxiv.org/abs/2108.01468>.
- [140] D.P. García, J. Cruz-Benito, F.J. García-Peñalvo, Systematic literature review: Quantum machine learning and its applications, 2022, [http://dx.doi.org/10.48550/ARXIV.2201.04093](https://arxiv.org/abs/2201.04093), URL: <https://arxiv.org/abs/2201.04093>.
- [141] A. Abbas, D. Sutter, C. Zoufal, A. Lucchi, A. Figalli, S. Woerner, The power of quantum neural networks, *Nat. Comput. Sci.* 1 (6) (2021) 403–409, <http://dx.doi.org/10.1038/s43588-021-00084-1>.
- [142] M. Schuld, N. Killoran, Quantum machine learning in feature Hilbert spaces, *Phys. Rev. Lett.* 122 (4) (2019) [http://dx.doi.org/10.1103/physrevlett.122.040504](https://arxiv.org/abs/1808.07588v1).
- [143] H.-Y. Huang, M. Broughton, M. Mohseni, R. Babbush, S. Boixo, H. Neven, J.R. McClean, Power of data in quantum machine learning, *Nature Commun.* 12 (1) (2021) [http://dx.doi.org/10.1038/s41467-021-22539-9](https://arxiv.org/abs/2103.14033v1).
- [144] J. Preskill, Quantum computing and the entanglement frontier, 2012, [arXiv:1203.5813v3](https://arxiv.org/abs/1203.5813v3).
- [145] E. Pednault, J.A. Gunnels, G. Nannicini, L. Horeh, R. Wisniewski, Leveraging secondary storage to simulate deep 54-qubit sycamore circuits, 2019, [arXiv:1910.09534v2](https://arxiv.org/abs/1910.09534v2).
- [146] Y.A. Liu, X.L. Liu, F.N. Li, H. Fu, Y. Yang, J. Song, P. Zhao, Z. Wang, D. Peng, H. Chen, et al., Closing the “quantum supremacy” gap, in: *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, ACM, 2021, [http://dx.doi.org/10.1145/3458817.3487399](https://arxiv.org/abs/2108.01468).
- [147] D. Jaschke, S. Montangero, Is quantum computing green? An estimate for an energy-efficiency quantum advantage, 2022, [http://dx.doi.org/10.48550/ARXIV.2205.12092](https://arxiv.org/abs/2205.12092), URL: <https://arxiv.org/abs/2205.12092>.

- [148] P.K. Barkoutsos, J.F. Gonthier, I. Sokolov, N. Moll, G. Salis, A. Fuhrer, M. Ganzhorn, D.J. Egger, M. Troyer, A. Mezzacapo, S. Filipp, I. Tavernelli, Quantum algorithms for electronic structure calculations: Particle-hole Hamiltonian and optimized wave-function expansions, *Phys. Rev. A* 98 (2) (2018) 022322, <http://dx.doi.org/10.1103/physreva.98.022322>.
- [149] P.J.J. O'Malley, R. Babbush, I.D. Kivlichan, J. Romero, J.R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A.G. Fowler, E. Jeffrey, E. Lucero, A. Megrant, J.Y. Mutus, M. Neeley, C. Neill, C. Quintana, D. Sank, A. Vainsencher, J. Wenner, T.C. White, P.V. Coveney, P.J. Love, H. Neven, A. Aspuru-Guzik, J.M. Martinis, Scalable quantum simulation of molecular energies, *Phys. Rev. X* 6 (3) (2016) 031007, <http://dx.doi.org/10.1103/physrevx.6.031007>.
- [150] R. Wiersema, C. Zhou, Y. de Sereville, J.F. Carrasquilla, Y.B. Kim, H. Yuen, Exploring entanglement and optimization within the Hamiltonian variational ansatz, *PRX Quantum* 1 (2) (2020) <http://dx.doi.org/10.1103/prxquantum.1.020319>.
- [151] F.D. Malone, R.M. Parrish, A.R. Welden, T. Fox, M. Degroote, E. Kyoseva, N. Moll, R. Santagati, M. Streif, Towards the simulation of large scale protein-ligand interactions on NISQ-era quantum computers, 2021, [arXiv:2110.01589v1](https://arxiv.org/abs/2110.01589v1).
- [152] J.J.M. Kirsopp, C.D. Paola, D.Z. Manrique, M. Krompiec, G. Greene-Diniz, W. Guba, A. Meyder, D. Wolf, M. Strahm, D.M. Ramo, Quantum computational quantification of protein-ligand interactions, 2021, [arXiv:2110.08163v1](https://arxiv.org/abs/2110.08163v1).
- [153] E.R. Anschuetz, B.T. Kiani, Beyond barren plateaus: Quantum variational algorithms are swamped with traps, 2022, <http://dx.doi.org/10.48550/ARXIV.2205.05786>, URL: <https://arxiv.org/abs/2205.05786>.
- [154] M.W. Krentel, The complexity of optimization problems, *J. Comput. System Sci.* 36 (3) (1988) 490–509, [http://dx.doi.org/10.1016/0022-0000\(88\)90039-6](http://dx.doi.org/10.1016/0022-0000(88)90039-6).
- [155] G.S. Barron, C.J. Wood, Measurement error mitigation for variational quantum algorithms, 2020, [arXiv:2010.08520v1](https://arxiv.org/abs/2010.08520v1).
- [156] R. Sagastizabal, X. Bonet-Monroig, M. Singh, M.A. Rol, C.C. Bultink, X. Fu, C.H. Price, V.P. Ostroukh, N. Muthusubramanian, A. Bruno, M. Beekman, N. Haider, T.E. O'Brien, L. DiCarlo, Experimental error mitigation via symmetry verification in a variational quantum eigensolver, *Phys. Rev. A* 100 (1) (2019) 010302(R), <http://dx.doi.org/10.1103/physreva.100.010302>.
- [157] Y. Nam, J.-S. Chen, N.C. Pisenti, K. Wright, K. Delaney, D. Maslov, K.R. Brown, S. Allen, J.M. Amini, J. Apisdorf, K.M. Beck, A. Blinov, V. Chaplin, M. Chmielewski, C. Collins, S. Debnath, K.M. Hudek, A.M. Ducore, M. Keenan, S.M. Kreikemeier, J. Mizrahi, P. Solomon, M. Williams, J.D. Wong-Campos, D. Moehring, C. Monroe, J. Kim, Ground-state energy estimation of the water molecule on a trapped-ion quantum computer, *Npj Quantum Inf.* 6 (1) (2020) <http://dx.doi.org/10.1038/s41534-020-0259-3>.
- [158] F. Arute, K. Arya, R. Babbush, D. Bacon, J.C. Bardin, R. Barends, S. Boixo, M. Broughton, B.B. Buckley, D.A. Buell, B. Burkett, N. Bushnell, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, S. Demura, A. Dunsworth, E. Farhi, A. Fowler, B. Foxen, C. Gidney, M. Giustina, R. Graff, S. Habegger, M.P. Harrigan, A. Ho, S. Hong, T. Huang, W.J. Huggins, L. Ioffe, S.V. Isakov, E. Jeffrey, Z. Jiang, C. Jones, D. Kafri, K. Kechedzhi, J. Kelly, S. Kim, P.V. Klimov, A. Korotkov, F. Kostritsa, D. Landhuis, P. Laptev, M. Lindmark, E. Lucero, O. Martin, J.M. Martinis, J.R. McClean, M. McEwen, A. Megrant, X. Mi, M. Mohseni, W. Mruczkiewicz, J. Mutus, O. Naaman, M. Neeley, C. Neill, H. Neven, M.Y. Niu, T.E. O'Brien, E. Ostby, A. Petukhov, H. Putterman, C. Quintana, P. Roushan, N.C. Rubin, D. Sank, K.J. Satzinger, V. Smelyanskiy, D. Strain, K.J. Sung, M. Szalay, T.Y. Takeshita, A. Vainsencher, T. White, N. Wiebe, Z.J. Yao, P. Yeh, A. Zalcman, Hartree-fock on a superconducting qubit quantum computer, *Science* 369 (6507) (2020) 1084–1089, <http://dx.doi.org/10.1126/science.abb9811>.
- [159] J. Tilly, P.V. Sriluckshmy, A. Patel, E. Fontana, I. Rungger, E. Grant, R. Anderson, J. Tennyson, G.H. Booth, Reduced density matrix sampling: Self-consistent embedding and multiscale electronic structure on current generation quantum computers, *Phys. Rev. Res.* 3 (3) (2021) <http://dx.doi.org/10.1103/physrevresearch.3.033230>, [arXiv:2104.05531v1](https://arxiv.org/abs/2104.05531v1); *Phys. Rev. Res.* 3 (2021) 033230.
- [160] F. Benfenati, G. Mazzola, C. Capecci, P.K. Barkoutsos, P.J. Ollitrault, I. Tavernelli, L. Guidoni, Improved accuracy on noisy devices by nonunitary variational quantum eigensolver for chemistry applications, *J. Chem. Theory Comput.* 17 (7) (2021) 3946–3954, <http://dx.doi.org/10.1021/acs.jctc.1c00091>.
- [161] F.A. Evangelista, G.K.-L. Chan, G.E. Scuseria, Exact parameterization of fermionic wave functions via unitary coupled cluster theory, *J. Chem. Phys.* 151 (24) (2019) 244112, <http://dx.doi.org/10.1063/1.5133059>.
- [162] P. Hohenberg, W. Kohn, Inhomogeneous electron gas, *Phys. Rev.* 136 (3B) (1964) B864–B871, <http://dx.doi.org/10.1103/physrev.136.b864>.
- [163] M. Levy, Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the v-representability problem, *Proc. Natl. Acad. Sci.* 76 (12) (1979) 6062–6065, <http://dx.doi.org/10.1073/pnas.76.12.6062>.
- [164] G. Vignale, M. Rasolt, Density-functional theory in strong magnetic fields, *Phys. Rev. Lett.* 59 (20) (1987) 2360–2363, <http://dx.doi.org/10.1103/physrevlett.59.2360>.
- [165] W. Kohn, L.J. Sham, Self-consistent equations including exchange and correlation effects, *Phys. Rev.* 140 (4A) (1965) A1133–A1138, <http://dx.doi.org/10.1103/physrev.140.a1133>.
- [166] R.G. Parr, Y. Weitao, *Density-Functional Theory of Atoms and Molecules*, Oxford University Press, ISBN: 9780197560709, 1995, <http://dx.doi.org/10.1093/oso/9780195092769.001.0001>.
- [167] D. Bagayoko, Understanding density functional theory (DFT) and completing it in practice, *AIP Adv.* 4 (12) (2014) 127104, <http://dx.doi.org/10.1063/1.4903408>.
- [168] J.J. Eriksen, T.A. Anderson, J.E. Deustua, K. Ghanem, D. Hait, M.R. Hoffmann, S. Lee, D.S. Levine, I. Magoulas, J. Shen, N.M. Tubman, K.B. Whaley, E. Xu, Y. Yao, N. Zhang, A. Alavi, G.K.-L. Chan, M. Head-Gordon, W. Liu, P. Piecuch, S. Sharma, S.L. Ten-no, C.J. Umrigar, J. Gauss, The ground state electronic energy of benzene, *J. Phys. Chem. Lett.* 11 (20) (2020) 8922–8929, <http://dx.doi.org/10.1021/acs.jpclett.0c02621>.
- [169] K.T. Williams, Y. Yao, J. Li, L. Chen, H. Shi, M. Motta, C. Niu, U. Ray, S. Guo, R.J. Anderson, J. Li, L.N. Tran, C.-N. Yeh, B. Mussard, S. Sharma, F. Bruneval, M. van Schilfgaarde, G.H. Booth, G.K.-L. Chan, S. Zhang, E. Gull, D. Zgid, A. Millis, C.J. Umrigar, L.K.W. and, Direct comparison of many-body methods for realistic electronic Hamiltonians, *Phys. Rev. X* 10 (1) (2020) <http://dx.doi.org/10.1103/physrevx.10.011041>.
- [170] R. Babbush, D.W. Berry, J.R. McClean, H. Neven, Quantum simulation of chemistry with sublinear scaling in basis size, *Npj Quantum Inf.* 5 (1) (2019) <http://dx.doi.org/10.1038/s41534-019-0199-y>.
- [171] C.D. Sherrill, H.F. Schaefer, The configuration interaction method: Advances in highly correlated approaches, in: *Advances in Quantum Chemistry*, Elsevier, 1999, pp. 143–269, [http://dx.doi.org/10.1016/s0065-3276\(08\)60532-8](http://dx.doi.org/10.1016/s0065-3276(08)60532-8).
- [172] K.D. Vogiatzis, D. Ma, J. Olsen, L. Gagliardi, W.A. de Jong, Pushing configuration-interaction to the limit: Towards massively parallel MCSCF calculations, *J. Chem. Phys.* 147 (18) (2017) 184111, <http://dx.doi.org/10.1063/1.4989858>, [arXiv:https://doi.org/10.1063/1.4989858](https://arxiv.org/abs/https://doi.org/10.1063/1.4989858).
- [173] I.G. Ross, Calculations of the energy levels of acetylene by the method of antisymmetric molecular orbitals, including σ - π interaction, *Trans. Faraday Soc.* 48 (1952) 973, <http://dx.doi.org/10.1039/tf9524800973>.
- [174] J. Foresman, *Exploring Chemistry with Electronic Structure Methods*, Gaussian, Inc, Pittsburgh, PA, ISBN: 978-0963676931, 1996.
- [175] S.R. White, Density matrix formulation for quantum renormalization groups, *Phys. Rev. Lett.* 69 (19) (1992) 2863–2866, <http://dx.doi.org/10.1103/physrevlett.69.2863>.
- [176] S. Sharma, G.K.-L. Chan, Spin-adapted density matrix renormalization group algorithms for quantum chemistry, *J. Chem. Phys.* 136 (12) (2012) 124121, <http://dx.doi.org/10.1063/1.3695642>.
- [177] N.M. Tubman, J. Lee, T.Y. Takeshita, M. Head-Gordon, K.B. Whaley, A deterministic alternative to the full configuration interaction quantum Monte Carlo method, *J. Chem. Phys.* 145 (4) (2016) 044112, <http://dx.doi.org/10.1063/1.4955109>.
- [178] G.H. Booth, A. Alavi, Approaching chemical accuracy using full configuration-interaction quantum Monte Carlo: A study of ionization potentials, *J. Chem. Phys.* 132 (17) (2010) 174104, <http://dx.doi.org/10.1063/1.3407895>.

- [179] R.E. Thomas, G.H. Booth, A. Alavi, Accurate *Ab Initio* Calculation of ionization potentials of the first-row transition metals with the configuration-interaction quantum Monte Carlo technique, *Phys. Rev. Lett.* 114 (3) (2015) <http://dx.doi.org/10.1103/physrevlett.114.033001>.
- [180] J. Li, M. Otten, A.A. Holmes, S. Sharma, C.J. Umrigar, Fast semistochastic heat-bath configuration interaction, *J. Chem. Phys.* 149 (21) (2018) 214110, <http://dx.doi.org/10.1063/1.5055390>.
- [181] K.A. Peterson, D. Feller, D.A. Dixon, Chemical accuracy in *ab initio* thermochemistry and spectroscopy: current strategies and future challenges, *Theor. Chem. Acc.* 131 (1) (2012) <http://dx.doi.org/10.1007/s00214-011-1079-5>.
- [182] N.M. Tubman, C.D. Freeman, D.S. Levine, D. Hait, M. Head-Gordon, K.B. Whaley, Modern approaches to exact diagonalization and selected configuration interaction with the adaptive sampling CI method, *J. Chem. Theory Comput.* 16 (4) (2020) 2139–2159, <http://dx.doi.org/10.1021/acs.jctc.8b00536>.
- [183] D. Hait, N.M. Tubman, D.S. Levine, K.B. Whaley, M. Head-Gordon, 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, *J. Chem. Theory Comput.* 15 (10) (2019) 5370–5385, <http://dx.doi.org/10.1021/acs.jctc.9b00674>.
- [184] D.S. Levine, D. Hait, N.M. Tubman, S. Lehtola, K.B. Whaley, M. Head-Gordon, CASSCF with extremely large active spaces using the adaptive sampling configuration interaction method, *J. Chem. Theory Comput.* 16 (4) (2020) 2340–2354, <http://dx.doi.org/10.1021/acs.jctc.9b01255>.
- [185] F.R. Petruzielo, A.A. Holmes, H.J. Changlani, M.P. Nightingale, C.J. Umrigar, Semistochastic projector Monte Carlo method, *Phys. Rev. Lett.* 109 (23) (2012) <http://dx.doi.org/10.1103/physrevlett.109.230201>.
- [186] A.A. Holmes, N.M. Tubman, C.J. Umrigar, Heat-bath configuration interaction: An efficient selected configuration interaction algorithm inspired by heat-bath sampling, *J. Chem. Theory Comput.* 12 (8) (2016) 3674–3680, <http://dx.doi.org/10.1021/acs.jctc.6b00407>.
- [187] S. Sharma, A.A. Holmes, G. Jeanmairat, A. Alavi, C.J. Umrigar, Semistochastic heat-bath configuration interaction method: Selected configuration interaction with semistochastic perturbation theory, *J. Chem. Theory Comput.* 13 (4) (2017) 1595–1604, <http://dx.doi.org/10.1021/acs.jctc.6b01028>.
- [188] J.E.T. Smith, B. Mussard, A.A. Holmes, S. Sharma, Cheap and near exact CASSCF with large active spaces, *J. Chem. Theory Comput.* 13 (11) (2017) 5468–5478, <http://dx.doi.org/10.1021/acs.jctc.7b00900>.
- [189] A.A. Holmes, C.J. Umrigar, S. Sharma, Excited states using semistochastic heat-bath configuration interaction, *J. Chem. Phys.* 147 (16) (2017) 164111, <http://dx.doi.org/10.1063/1.4998614>.
- [190] D. Cleland, G.H. Booth, C. Overly, A. Alavi, Taming the first-row diatomics: A full configuration interaction quantum Monte Carlo study, *J. Chem. Theory Comput.* 8 (11) (2012) 4138–4152, <http://dx.doi.org/10.1021/ct300504f>.
- [191] R.J. Anderson, G.H. Booth, Four-component full configuration interaction quantum Monte Carlo for relativistic correlated electron problems, *J. Chem. Phys.* 153 (18) (2020) 184103, <http://dx.doi.org/10.1063/5.0029863>.
- [192] G.H. Booth, D. Cleland, A.J.W. Thom, A. Alavi, Breaking the carbon dimer: The challenges of multiple bond dissociation with full configuration interaction quantum Monte Carlo methods, *J. Chem. Phys.* 135 (8) (2011) 084104, <http://dx.doi.org/10.1063/1.3624383>.
- [193] N.S. Blunt, A.J.W. Thom, C.J.C. Scott, Preconditioning and perturbative estimators in full configuration interaction quantum Monte Carlo, *J. Chem. Theory Comput.* 15 (6) (2019) 3537–3551, <http://dx.doi.org/10.1021/acs.jctc.9b00049>.
- [194] S.R. White, Density-matrix algorithms for quantum renormalization groups, *Phys. Rev. B* 48 (14) (1993) 10345–10356, <http://dx.doi.org/10.1103/physrevb.48.10345>.
- [195] S.R. White, R.L. Martin, *Ab initio* quantum chemistry using the density matrix renormalization group, *J. Chem. Phys.* 110 (9) (1999) 4127–4130, <http://dx.doi.org/10.1063/1.478295>.
- [196] A.O. Mitrushenkov, G. Fano, F. Ortolani, R. Linguerri, P. Palmieri, Quantum chemistry using the density matrix renormalization group, *J. Chem. Phys.* 115 (15) (2001) 6815–6821, <http://dx.doi.org/10.1063/1.1389475>.
- [197] Ö. Legeza, J. Röder, B.A. Hess, Controlling the accuracy of the density-matrix renormalization-group method: The dynamical block state selection approach, *Phys. Rev. B* 67 (2003) 125114, <http://dx.doi.org/10.1103/PhysRevB.67.125114>.
- [198] G.K.-L. Chan, M. Head-Gordon, Highly correlated calculations with a polynomial cost algorithm: A study of the density matrix renormalization group, *J. Chem. Phys.* 116 (11) (2002) 4462–4476, <http://dx.doi.org/10.1063/1.1449459>.
- [199] G.K.-L. Chan, S. Sharma, The density matrix renormalization group in quantum chemistry, *Annu. Rev. Phys. Chem.* 62 (1) (2011) 465–481, <http://dx.doi.org/10.1146/annurev-physchem-032210-103338>.
- [200] R. Olivares-Amaya, W. Hu, N. Nakatani, S. Sharma, J. Yang, G.K.-L. Chan, Theab-initiodensity matrix renormalization group in practice, *J. Chem. Phys.* 142 (3) (2015) 034102, <http://dx.doi.org/10.1063/1.4905329>.
- [201] S. Wouters, D.V. Neck, The density matrix renormalization group for *ab initio* quantum chemistry, *Eur. Phys. J. D* 68 (9) (2014) <http://dx.doi.org/10.1140/epjd/e2014-50500-1>.
- [202] T. Yanai, Y. Kurashige, W. Mizukami, J. Chalupský, T.N. Lan, M. Saitow, Density matrix renormalization group for *ab initio* Calculations and associated dynamic correlation methods: A review of theory and applications, *Int. J. Quantum Chem.* 115 (5) (2014) 283–299, <http://dx.doi.org/10.1002/qua.24808>.
- [203] S. Knecht, E.D. Hedegård, S. Keller, A. Kovyshin, Y. Ma, A. Muolo, C.J. Stein, M. Reiher, New approaches for *ab initio* calculations of molecules with strong electron correlation, *CHIMIA Int. J. Chem.* 70 (4) (2016) 244–251, <http://dx.doi.org/10.2533/chimia.2016.244>.
- [204] J. Biamonte, V. Bergholm, Tensor networks in a nutshell, 2017, <http://dx.doi.org/10.48550/ARXIV.1708.00006>, URL: <https://arxiv.org/abs/1708.00006>.
- [205] R. Orús, Tensor networks for complex quantum systems, *Nat. Rev. Phys.* 1 (9) (2019) 538–550, <http://dx.doi.org/10.1038/s42254-019-0086-7>.
- [206] M. Schneider, J. Ostmeier, K. Jansen, T. Luu, C. Urbach, The hubbard model with fermionic tensor networks, 2021, <http://dx.doi.org/10.48550/ARXIV.2110.15340>, URL: <https://arxiv.org/abs/2110.15340>.
- [207] R. Kaneko, I. Danshita, Tensor-network study of correlation-spreading dynamics in the two-dimensional Bose-Hubbard model, *Commun. Phys.* 5 (1) (2022) <http://dx.doi.org/10.1038/s42005-022-00848-9>.
- [208] J. Toulouse, R. Assaraf, C.J. Umrigar, Chapter fifteen - introduction to the variational and diffusion Monte Carlo methods, in: P.E. Hoggan, T. Ozdogan (Eds.), *Electron Correlation in Molecules – Ab Initio beyond Gaussian Quantum Chemistry*, in: *Advances in Quantum Chemistry*, vol. 73, Academic Press, 2016, pp. 285–314, <http://dx.doi.org/10.1016/bs.aiq.2015.07.003>.
- [209] R.J. Needs, M.D. Towler, N.D. Drummond, P. López Ríos, J.R. Trail, Variational and diffusion quantum Monte Carlo calculations with the CASINO code, *J. Chem. Phys.* 152 (15) (2020) 154106, <http://dx.doi.org/10.1063/1.5144288>, [arXiv:https://arxiv.org/abs/10.1063/1.5144288](https://arxiv.org/abs/10.1063/1.5144288).
- [210] D. Pfau, J.S. Spencer, A.G.D.G. Matthews, W.M.C. Foulkes, *Ab initio* solution of the many-electron schrödinger equation with deep neural networks, *Phys. Rev. Res.* 2 (3) (2020) <http://dx.doi.org/10.1103/physrevresearch.2.033429>.
- [211] J.J. Eriksen, D.A. Matthews, P. Jørgensen, J. Gauss, Assessment of the accuracy of coupled cluster perturbation theory for open-shell systems. I. Triples expansions, *J. Chem. Phys.* 144 (19) (2016) 194102, <http://dx.doi.org/10.1063/1.4948780>.
- [212] J.J. Eriksen, D.A. Matthews, P. Jørgensen, J. Gauss, Assessment of the accuracy of coupled cluster perturbation theory for open-shell systems. II. Quadruples expansions, *J. Chem. Phys.* 144 (19) (2016) 194103, <http://dx.doi.org/10.1063/1.4948781>.
- [213] I.Y. Zhang, A. Grüneis, Coupled cluster theory in materials science, *Front. Mater.* 6 (2019) <http://dx.doi.org/10.3389/fmats.2019.00123>.
- [214] R.J. Bartlett, M. Musiał, Coupled-cluster theory in quantum chemistry, *Rev. Modern Phys.* 79 (1) (2007) 291–352, <http://dx.doi.org/10.1103/revmodphys.79.291>.

- [215] Q. Wang, M. Duan, E. Xu, J. Zou, S. Li, Describing strong correlation with block-correlated coupled cluster theory, *J. Phys. Chem. Lett.* 11 (18) (2020) 7536–7543, <http://dx.doi.org/10.1021/acs.jpclett.0c02117>.
- [216] D.I. Lyakh, M. Musiał, V.F. Lotrich, R.J. Bartlett, Multireference nature of chemistry: The coupled-cluster view, *Chem. Rev.* 112 (1) (2011) 182–243, <http://dx.doi.org/10.1021/cr2001417>.
- [217] H.R. Grimsley, S.E. Economou, E. Barnes, N.J. Mayhall, An adaptive variational algorithm for exact molecular simulations on a quantum computer, *Nature Commun.* 10 (1) (2019) <http://dx.doi.org/10.1038/s41467-019-10988-2>.
- [218] B.T. Gard, L. Zhu, G.S. Barron, N.J. Mayhall, S.E. Economou, E. Barnes, Efficient symmetry-preserving state preparation circuits for the variational quantum eigensolver algorithm, *Npj Quantum Inf.* 6 (1) (2020) 10, <http://dx.doi.org/10.1038/s41534-019-0240-1>.
- [219] H.R. Grimsley, D. Claudino, S.E. Economou, E. Barnes, N.J. Mayhall, Is the trotterized UCCSD ansatz chemically well-defined? *J. Chem. Theory Comput.* 16 (1) (2019) 1–6, <http://dx.doi.org/10.1021/acs.jctc.9b01083>.
- [220] G. Carleo, M. Troyer, Solving the quantum many-body problem with artificial neural networks, *Science* 355 (6325) (2017) 602–606, <http://dx.doi.org/10.1126/science.aag2302>.
- [221] Y. Nomura, A.S. Darmawan, Y. Yamaji, M. Imada, Restricted Boltzmann machine learning for solving strongly correlated quantum systems, *Phys. Rev. B* 96 (20) (2017) <http://dx.doi.org/10.1103/physrevb.96.205152>.
- [222] K. Choo, G. Carleo, N. Regnault, T. Neupert, Symmetries and many-body excitations with neural-network quantum states, *Phys. Rev. Lett.* 121 (16) (2018) <http://dx.doi.org/10.1103/physrevlett.121.167204>.
- [223] A. Glielmo, Y. Rath, G. Csányi, A.D. Vita, G.H. Booth, Gaussian process states: A data-driven representation of quantum many-body physics, *Phys. Rev. X* 10 (4) (2020) <http://dx.doi.org/10.1103/physrevx.10.041026>.
- [224] D. Luo, B.K. Clark, Backflow transformations via neural networks for quantum many-body wave functions, *Phys. Rev. Lett.* 122 (22) (2019) <http://dx.doi.org/10.1103/physrevlett.122.226401>.
- [225] K. Choo, A. Mezzacapo, G. Carleo, Fermionic neural-network states for ab-initio electronic structure, *Nature Commun.* 11 (1) (2020) <http://dx.doi.org/10.1038/s41467-020-15724-9>.
- [226] J. Hermann, Z. Schätzle, F. Noé, Deep-neural-network solution of the electronic Schrödinger equation, *Nat. Chem.* 12 (10) (2020) 891–897, <http://dx.doi.org/10.1038/s41557-020-0544-y>.
- [227] A.Y. Kitaev, Quantum measurements and the abelian stabilizer problem, 1995, [arXiv:quant-ph/9511026v1](https://arxiv.org/abs/quant-ph/9511026v1).
- [228] D.S. Abrams, S. Lloyd, Simulation of many-body Fermi systems on a universal quantum computer, *Phys. Rev. Lett.* 79 (13) (1997) 2586–2589, <http://dx.doi.org/10.1103/physrevlett.79.2586>.
- [229] D.S. Abrams, S. Lloyd, Quantum algorithm providing exponential speed increase for finding eigenvalues and eigenvectors, *Phys. Rev. Lett.* 83 (24) (1999) 5162–5165, <http://dx.doi.org/10.1103/physrevlett.83.5162>.
- [230] R. Cleve, A. Ekert, C. Macchiavello, M. Mosca, Quantum algorithms revisited, *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* 454 (1969) (1998) 339–354, <http://dx.doi.org/10.1098/rspa.1998.0164>.
- [231] A. Aspuru-Guzik, A.D. Dutoi, P.J. Love, M. Head-Gordon, Simulated quantum computation of molecular energies, *Science* 309 (5741) (2005) 1704–1707, <http://dx.doi.org/10.1126/science.1113479>.
- [232] J.R. McClean, R. Babbush, P.J. Love, A. Aspuru-Guzik, Exploiting locality in quantum computation for quantum chemistry, *J. Phys. Chem. Lett.* 5 (24) (2014) 4368–4380, <http://dx.doi.org/10.1021/jz501649m>.
- [233] M. Reiher, N. Wiebe, K.M. Svore, D. Wecker, M. Troyer, Elucidating reaction mechanisms on quantum computers, *Proc. Natl. Acad. Sci.* 114 (29) (2017) 7555–7560, <http://dx.doi.org/10.1073/pnas.1619152114>.
- [234] A.M. Childs, D. Maslov, Y. Nam, N.J. Ross, Y. Su, Toward the first quantum simulation with quantum speedup, *Proc. Natl. Acad. Sci.* 115 (38) (2018) 9456–9461, <http://dx.doi.org/10.1073/pnas.1801723115>.
- [235] R. Babbush, N. Wiebe, J. McClean, J. McClain, H. Neven, G.K.-L. Chan, Low-depth quantum simulation of materials, *Phys. Rev. X* 8 (1) (2018) 011044, <http://dx.doi.org/10.1103/physrevx.8.011044>.
- [236] N.M. Tubman, C. Mejuto-Zaera, J.M. Epstein, D. Hait, D.S. Levine, W. Huggins, Z. Jiang, J.R. McClean, R. Babbush, M. Head-Gordon, K.B. Whaley, Postponing the orthogonality catastrophe: efficient state preparation for electronic structure simulations on quantum devices, 2018, [arXiv:1809.05523](https://arxiv.org/abs/1809.05523).
- [237] K. Sugisaki, S. Nakazawa, K. Toyota, K. Sato, D. Shiomi, T. Takui, 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.* 5 (1) (2018) 167–175, <http://dx.doi.org/10.1021/acscentsci.8b00788>.
- [238] K. Sugisaki, S. Yamamoto, S. Nakazawa, K. Toyota, K. Sato, D. Shiomi, T. Takui, Open shell electronic state calculations on quantum computers: A quantum circuit for the preparation of configuration state functions based on Serber construction, *Chem. Phys. Lett.* 737 (2019) 100002, <http://dx.doi.org/10.1016/j.cpletx.2018.100002>.
- [239] M. Motta, C. Sun, A.T.K. Tan, M.J. O’Rourke, E. Ye, A.J. Minnich, F.G.S.L. Brandão, G.K.-L. Chan, Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution, *Nat. Phys.* 16 (2) (2019) 205–210, <http://dx.doi.org/10.1038/s41567-019-0704-4>.
- [240] B. Murta, J. Fernández-Rossier, Gutzwiller wave function on a digital quantum computer, *Phys. Rev. B* 103 (24) (2021) <http://dx.doi.org/10.1103/physrevb.103.241113>.
- [241] T. Albash, D.A. Lidar, Adiabatic quantum computation, *Rev. Modern Phys.* 90 (1) (2018) <http://dx.doi.org/10.1103/revmodphys.90.015002>.
- [242] S. Matsuura, T. Yamazaki, V. Senicourt, L. Huntington, A. Zaribafian, VanQver: the variational and adiabatically navigated quantum eigensolver, *New J. Phys.* 22 (5) (2020) 053023, <http://dx.doi.org/10.1088/1367-2630/ab8080>.
- [243] J. Du, N. Xu, X. Peng, P. Wang, S. Wu, D. Lu, NMR implementation of a molecular hydrogen quantum simulation with adiabatic state preparation, *Phys. Rev. Lett.* 104 (3) (2010) <http://dx.doi.org/10.1103/physrevlett.104.030502>.
- [244] B.P. Lanyon, J.D. Whitfield, G.G. Gillett, M.E. Goggin, M.P. Almeida, J.D. Biamonte, M. Mohseni, B.J. Powell, M. Barbieri, A. Aspuru-Guzik, A.G. White, Towards quantum chemistry on a quantum computer, *Nature Chem.* 2 (2) (2010) 106–111, <http://dx.doi.org/10.1038/nchem.483>.
- [245] Z. Li, M.-H. Yung, H. Chen, D. Lu, J.D. Whitfield, X. Peng, A. Aspuru-Guzik, J. Du, Solving quantum ground-state problems with nuclear magnetic resonance, *Sci. Rep.* 1 (1) (2011) <http://dx.doi.org/10.1038/srep00088>.
- [246] Y. Wang, F. Dolde, J. Biamonte, R. Babbush, V. Bergholm, S. Yang, I. Jakobi, P. Neumann, A. Aspuru-Guzik, J.D. Whitfield, J. Wrachtrup, Quantum simulation of helium hydride cation in a solid-state spin register, *ACS Nano* 9 (8) (2015) 7769–7774, <http://dx.doi.org/10.1021/acsnano.5b01651>.
- [247] S. Paesani, A.A. Gentile, R. Santagati, J. Wang, N. Wiebe, D.P. Tew, J.L. O’Brien, M.G. Thompson, Experimental Bayesian quantum phase estimation on a silicon photonic chip, *Phys. Rev. Lett.* 118 (10) (2017) <http://dx.doi.org/10.1103/physrevlett.118.100503>.
- [248] R. Santagati, J. Wang, A.A. Gentile, S. Paesani, N. Wiebe, J.R. McClean, S. Morley-Short, P.J. Shadbolt, D. Bonneau, J.W. Silverstone, D.P. Tew, X. Zhou, J.L. O’Brien, M.G. Thompson, Witnessing eigenstates for quantum simulation of Hamiltonian spectra, *Sci. Adv.* 4 (1) (2018) <http://dx.doi.org/10.1126/sciadv.aap9646>.
- [249] H. Beinert, R.H. Holm, E. Münck, Iron-sulfur clusters: Nature’s modular, multipurpose structures, *Science* 277 (5326) (1997) 653–659, <http://dx.doi.org/10.1126/science.277.5326.653>.
- [250] G.H. Low, I.L. Chuang, Hamiltonian simulation by qubitization, *Quantum* 3 (2019) 163, <http://dx.doi.org/10.22331/q-2019-07-12-163>.
- [251] Z. Li, J. Li, N.S. Dattani, C.J. Umrigar, G.K.-L. Chan, The electronic complexity of the ground-state of the FeMo cofactor of nitrogenase as relevant to quantum simulations, *J. Chem. Phys.* 150 (2) (2019) 024302, <http://dx.doi.org/10.1063/1.5063376>.

- [252] J. Lee, D.W. Berry, C. Gidney, W.J. Huggins, J.R. McClean, N. Wiebe, R. Babbush, Even more efficient quantum computations of chemistry through tensor hypercontraction, *PRX Quantum* 2 (3) (2021) <http://dx.doi.org/10.1103/prxquantum.2.030305>.
- [253] R. Babbush, C. Gidney, D.W. Berry, N. Wiebe, J. McClean, A. Paler, A. Fowler, H. Neven, Encoding electronic spectra in quantum circuits with linear t complexity, *Phys. Rev. X* 8 (4) (2018) <http://dx.doi.org/10.1103/physrevx.8.041015>.
- [254] I.D. Kivlichan, C. Gidney, D.W. Berry, N. Wiebe, J. McClean, W. Sun, Z. Jiang, N. Rubin, A. Fowler, A. Aspuru-Guzik, H. Neven, R. Babbush, Improved fault-tolerant quantum simulation of condensed-phase correlated electrons via trotterization, *Quantum* 4 (2020) 296, <http://dx.doi.org/10.22331/q-2020-07-16-296>.
- [255] D. Wang, O. Higgott, S. Brierley, Accelerated variational quantum eigensolver, *Phys. Rev. Lett.* 122 (14) (2019) <http://dx.doi.org/10.1103/physrevlett.122.140504>.
- [256] Z. Jiang, A. Kalev, W. Mruczkiewicz, H. Neven, Optimal fermion-to-qubit mapping via ternary trees with applications to reduced quantum states learning, *Quantum* 4 (2020) 276, <http://dx.doi.org/10.22331/q-2020-06-04-276>.
- [257] K. Setia, S. Bravyi, A. Mezzacapo, J.D. Whitfield, Superfast encodings for fermionic quantum simulation, *Phys. Rev. Res.* 1 (3) (2019) <http://dx.doi.org/10.1103/physrevresearch.1.033033>.
- [258] A. Uvarov, J.D. Biamonte, D. Yudin, Variational quantum eigensolver for frustrated quantum systems, *Phys. Rev. B* 102 (7) (2020) <http://dx.doi.org/10.1103/physrevb.102.075104>.
- [259] N.P.D. Sawaya, M. Smelyanskiy, J.R. McClean, A. Aspuru-Guzik, Error sensitivity to environmental noise in quantum circuits for chemical state preparation, *J. Chem. Theory Comput.* 12 (7) (2016) 3097–3108, <http://dx.doi.org/10.1021/acs.jctc.6b00220>.
- [260] C. Derby, J. Klassen, J. Bausch, T. Cubitt, Compact fermion to qubit mappings, *Phys. Rev. B* 104 (3) (2021) <http://dx.doi.org/10.1103/physrevb.104.035118>.
- [261] C. Derby, J. Klassen, A compact Fermion to qubit mapping part 2: Alternative lattice geometries, 2021, [arXiv:2101.10735v1](https://arxiv.org/abs/2101.10735v1).
- [262] T.-C. Yen, A.F. Izmaylov, Cartan sub-algebra approach to efficient measurements of quantum observables, 2021, [arXiv:2007.01234](https://arxiv.org/abs/2007.01234).
- [263] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J.M. Chow, J.M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, *Nature* 549 (7671) (2017) 242–246, <http://dx.doi.org/10.1038/nature23879>.
- [264] C. Hempel, C. Maier, J. Romero, J. McClean, T. Monz, H. Shen, P. Jurcevic, B.P. Lanyon, P. Love, R. Babbush, A. Aspuru-Guzik, R. Blatt, C.F. Roos, Quantum chemistry calculations on a trapped-ion quantum simulator, *Phys. Rev. X* 8 (3) (2018) <http://dx.doi.org/10.1103/physrevx.8.031022>.
- [265] C. Kokail, C. Maier, R. van Bijnen, T. Brydges, M.K. Joshi, P. Jurcevic, C.A. Muschik, P. Silvi, R. Blatt, C.F. Roos, P. Zoller, Self-verifying variational quantum simulation of lattice models, *Nature* 569 (7756) (2019) 355–360, <http://dx.doi.org/10.1038/s41586-019-1177-4>.
- [266] A.F. Izmaylov, T.-C. Yen, I.G. Ryabinkin, Revising the measurement process in the variational quantum eigensolver: is it possible to reduce the number of separately measured operators? *Chem. Sci.* 10 (13) (2019) 3746–3755, <http://dx.doi.org/10.1039/c8sc05592k>.
- [267] V. Verteletskyi, T.-C. Yen, A.F. Izmaylov, Measurement optimization in the variational quantum eigensolver using a minimum clique cover, *J. Chem. Phys.* 152 (12) (2020) 124114, <http://dx.doi.org/10.1063/1.5141458>.
- [268] T.E. O'Brien, M. Streif, N.C. Rubin, R. Santagati, Y. Su, W.J. Huggins, J.J. Goings, N. Moll, E. Kyoseva, M. Degroote, C.S. Tautermann, J. Lee, D.W. Berry, N. Wiebe, R. Babbush, Efficient quantum computation of molecular forces and other energy gradients, 2021, <http://dx.doi.org/10.48550/ARXIV.2111.12437>, URL: <https://arxiv.org/abs/2111.12437>.
- [269] Y.S. Yordanov, V. Armaos, C.H.W. Barnes, D.R.M. Arvidsson-Shukur, Iterative qubit-excitation based variational quantum eigensolver, 2020, [arXiv:2011.10540v1](https://arxiv.org/abs/2011.10540v1).
- [270] H.L. Tang, V. Shkolnikov, G.S. Barron, H.R. Grimsley, N.J. Mayhall, E. Barnes, S.E. Economou, Qubit-ADAPT-VQE: An adaptive algorithm for constructing hardware-efficient ansätze on a quantum processor, *PRX Quantum* 2 (2021) 020310, <http://dx.doi.org/10.1103/prxquantum.2.020310>.
- [271] J.G. Vidal, D.O. Theis, Calculus on parameterized quantum circuits, 2018, <http://dx.doi.org/10.48550/ARXIV.1812.06323>, URL: <https://arxiv.org/abs/1812.06323>.
- [272] H.C. Watanabe, R. Raymond, Y.-y. Ohnishi, E. Kaminishi, M. Sugawara, Optimizing parameterized quantum circuits with free-axis selection, 2021, <http://dx.doi.org/10.48550/ARXIV.2104.14875>, URL: <https://arxiv.org/abs/2104.14875>.
- [273] K. Wada, R. Raymond, Y.-y. Ohnishi, E. Kaminishi, M. Sugawara, N. Yamamoto, H.C. Watanabe, Simulating time evolution with fully optimized single-qubit gates on parameterized quantum circuits, 2021, <http://dx.doi.org/10.48550/ARXIV.2111.05538>, URL: <https://arxiv.org/abs/2111.05538>.
- [274] S. McArdle, X. Yuan, S. Benjamin, Error-mitigated digital quantum simulation, *Phys. Rev. Lett.* 122 (18) (2019) 1–17, <http://dx.doi.org/10.1103/physrevlett.122.180501>.
- [275] X. Bonet-Monroig, R. Sagastizabal, M. Singh, T.E. O'Brien, Low-cost error mitigation by symmetry verification, *Phys. Rev. A* 98 (6) (2018) 062339, <http://dx.doi.org/10.1103/physreva.98.062339>.
- [276] K. Temme, S. Bravyi, J.M. Gambetta, Error mitigation for short-depth quantum circuits, *Phys. Rev. Lett.* 119 (18) (2017) 1–15, <http://dx.doi.org/10.1103/physrevlett.119.180509>.
- [277] Y. Li, S.C. Benjamin, Efficient variational quantum simulator incorporating active error minimization, *Phys. Rev. X* 7 (2) (2017) 021050, <http://dx.doi.org/10.1103/physrevx.7.021050>.
- [278] D.W. Berry, M. Kieferová, A. Scherer, Y.R. Sanders, G.H. Low, N. Wiebe, C. Gidney, R. Babbush, Improved techniques for preparing eigenstates of fermionic Hamiltonians, *Npj Quantum Information* 4 (1) (2018) <http://dx.doi.org/10.1038/s41534-018-0071-5>.
- [279] J.T. Seeley, M.J. Richard, P.J. Love, The Bravyi-Kitaev transformation for quantum computation of electronic structure, *J. Chem. Phys.* 137 (22) (2012) 224109, <http://dx.doi.org/10.1063/1.4768229>.
- [280] S.B. Bravyi, A.Y. Kitaev, Fermionic quantum computation, *Ann. Physics* 298 (1) (2002) 210–226, <http://dx.doi.org/10.1006/aphy.2002.6254>.
- [281] A. Tranter, S. Sofia, J. Seeley, M. Kaicher, J. McClean, R. Babbush, P.V. Coveney, F. Mintert, F. Wilhelm, P.J. Love, The Bravyi-Kitaev transformation: Properties and applications, *Int. J. Quantum Chem.* 115 (19) (2015) 1431–1441, <http://dx.doi.org/10.1002/qua.24969>.
- [282] S. Bravyi, J.M. Gambetta, A. Mezzacapo, K. Temme, Tapering off qubits to simulate fermionic Hamiltonians, 2017, [arXiv:1701.08213v1](https://arxiv.org/abs/1701.08213v1).
- [283] K. Setia, R. Chen, J.E. Rice, A. Mezzacapo, M. Pistoia, J.D. Whitfield, Reducing qubit requirements for quantum simulations using molecular point group symmetries, *J. Chem. Theory Comput.* 16 (10) (2020) 6091–6097, <http://dx.doi.org/10.1021/acs.jctc.0c00113>.
- [284] W.M. Kirby, A. Tranter, P.J. Love, Contextual subspace variational quantum eigensolver, *Quantum* 5 (2021) 456, <http://dx.doi.org/10.22331/q-2021-05-14-456>.
- [285] P. Gokhale, F.T. Chong, $O(N^3)$ measurement cost for variational quantum eigensolver on molecular Hamiltonians, 2019, [arXiv:1908.11857v1](https://arxiv.org/abs/1908.11857v1).
- [286] O. Crawford, B. van Straaten, D. Wang, T. Parks, E. Campbell, S. Brierley, Efficient quantum measurement of Pauli operators in the presence of finite sampling error, *Quantum* 5 (2021) 385, <http://dx.doi.org/10.22331/q-2021-01-20-385>.
- [287] I.G. Ryabinkin, R.A. Lang, S.N. Genin, A.F. Izmaylov, Iterative qubit coupled cluster approach with efficient screening of generators, *J. Chem. Theory Comput.* 16 (2) (2020) 1055–1063, <http://dx.doi.org/10.1021/acs.jctc.9b01084>.
- [288] K. Zhang, M.-H. Hsieh, L. Liu, D. Tao, Toward trainability of quantum neural networks, 2020, [arXiv:2011.06258v2](https://arxiv.org/abs/2011.06258v2).
- [289] D. Claudino, J. Wright, A.J. McCaskey, T.S. Humble, Benchmarking adaptive variational quantum eigensolvers, *Front. Chem.* 8 (2020) <http://dx.doi.org/10.3389/fchem.2020.606863>.

- [290] J. Stokes, J. Isaac, N. Killoran, G. Carleo, Quantum natural gradient, *Quantum* 4 (2020) 269, <http://dx.doi.org/10.22331/q-2020-05-25-269>, [arXiv:1909.02108](https://arxiv.org/abs/1909.02108).
- [291] S. McArdle, T. Jones, S. Endo, Y. Li, S.C. Benjamin, X. Yuan, Variational ansatz-based quantum simulation of imaginary time evolution, *Npj Quantum Inf.* 5 (1) (2019) 75, <http://dx.doi.org/10.1038/s41534-019-0187-2>, [arXiv:1804.03023](https://arxiv.org/abs/1804.03023).
- [292] C. Köppl, H.-J. Werner, Parallel and low-order scaling implementation of Hartree-Fock exchange using local density fitting, *J. Chem. Theory Comput.* 12 (7) (2016) 3122–3134, <http://dx.doi.org/10.1021/acs.jctc.6b00251>.
- [293] V. Havlíček, M. Troyer, J.D. Whitfield, Operator locality in the quantum simulation of fermionic models, *Phys. Rev. A* 95 (3) (2017) <http://dx.doi.org/10.1103/physreva.95.032332>.
- [294] I.O. Sokolov, P.K. Barkoutsos, P.J. Ollitrault, D. Greenberg, J. Rice, M. Pistoia, I. Tavernelli, Quantum orbital-optimized unitary coupled cluster methods in the strongly correlated regime: Can quantum algorithms outperform their classical equivalents? *J. Chem. Phys.* 152 (12) (2020) 124107, <http://dx.doi.org/10.1063/1.5141835>.
- [295] K. Setia, J.D. Whitfield, Bravyi-Kitaev Superfast simulation of electronic structure on a quantum computer, *J. Chem. Phys.* 148 (16) (2018) 164104, <http://dx.doi.org/10.1063/1.5019371>.
- [296] N. Vogt, S. Zanker, J.-M. Reiner, T. Eckl, A. Maruszczyk, M. Marthaler, Preparing symmetry broken ground states with variational quantum algorithms, 2020, [arXiv:2007.01582v2](https://arxiv.org/abs/2007.01582v2).
- [297] A. Choquette, A.D. Paolo, P.K. Barkoutsos, D. Sénéchal, I. Tavernelli, A. Blais, Quantum-optimal-control-inspired ansatz for variational quantum algorithms, *Phys. Rev. Res.* 3 (2) (2021) <http://dx.doi.org/10.1103/physrevresearch.3.023092>.
- [298] C. Fan, F.Y. Wu, General lattice model of phase transitions, *Phys. Rev. B* 2 (3) (1970) 723–733, <http://dx.doi.org/10.1103/physrevb.2.723>.
- [299] J. Marro, R. Dickman, *Nonequilibrium Phase Transitions in Lattice Models*, Cambridge University Press, 1999, <http://dx.doi.org/10.1017/cbo9780511524288>.
- [300] T.N. Stanislavchuk, A.P. Litvinchuk, R. Hu, Y.H. Jeon, S.D. Ji, S.-W. Cheong, A.A. Sirenko, Optical properties, lattice dynamics, and structural phase transition in hexagonal 2H-BaMnO₃ single crystals, *Phys. Rev. B* 92 (13) (2015) <http://dx.doi.org/10.1103/physrevb.92.134308>.
- [301] J.-M. Choi, F. Dar, R.V. Pappu, LASSI: A lattice model for simulating phase transitions of multivalent proteins, *PLoS Comput. Biol.* 15 (10) (2019) e1007028, <http://dx.doi.org/10.1371/journal.pcbi.1007028>.
- [302] Y.-N. Li, S. Wang, T. Wang, R. Gao, C.-Y. Geng, Y.-W. Li, J. Wang, H. Jiao, Energies and spin states of FeS^{0/-}, FeS₂^{0/-}, Fe₂S₂^{0/-}, Fe₃S₄^{0/-}, and Fe₄S₄^{0/-} clusters, *ChemPhysChem* 14 (6) (2013) 1182–1189, <http://dx.doi.org/10.1002/cphc.201201043>.
- [303] M. Kjaergaard, M.E. Schwartz, J. Braumüller, P. Krantz, J.I.-J. Wang, S. Gustavsson, W.D. Oliver, Superconducting qubits: Current state of play, *Annu. Rev. Condens. Matter Phys.* 11 (1) (2020) 369–395, <http://dx.doi.org/10.1146/annurev-conmatphys-031119-050605>.
- [304] J.P.T. Stenger, D. Gunlyck, C.S. Hellberg, Expanding variational quantum eigensolvers to larger systems by dividing the calculations between classical and quantum hardware, *Phys. Rev. A* 105 (2) (2022) <http://dx.doi.org/10.1103/physreva.105.022438>.
- [305] K.N. Okada, K. Osaki, K. Mitarai, K. Fujii, Identification of topological phases using classically-optimized variational quantum eigensolver, 2022, <http://dx.doi.org/10.48550/ARXIV.2202.02909>, URL: <https://arxiv.org/abs/2202.02909>.
- [306] Y. Zhang, L. Cincio, C.F.A. Negre, P. Czarnik, P. Coles, P.M. Anisimov, S.M. Mniszewski, S. Tretiak, P.A. Dub, Variational quantum eigensolver with reduced circuit complexity, 2021, [arXiv:2106.07619v1](https://arxiv.org/abs/2106.07619v1).
- [307] W.J. Hehre, R.F. Stewart, J.A. Pople, Self-consistent molecular-orbital methods. I. Use of Gaussian expansions of Slater-type atomic orbitals, *J. Chem. Phys.* 51 (6) (1969) 2657–2664, <http://dx.doi.org/10.1063/1.1672392>.
- [308] R.F. Stewart, Small Gaussian expansions of Slater-type orbitals, *J. Chem. Phys.* 52 (1) (1970) 431–438, <http://dx.doi.org/10.1063/1.1672702>.
- [309] T.H. Dunning, Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen, *J. Chem. Phys.* 90 (2) (1989) 1007–1023, <http://dx.doi.org/10.1063/1.456153>.
- [310] A.J.C. Varandas, Basis-set extrapolation of the correlation energy, *J. Chem. Phys.* 113 (20) (2000) 8880–8887, <http://dx.doi.org/10.1063/1.1319644>, [arXiv:https://doi.org/10.1063/1.1319644](https://arxiv.org/abs/https://doi.org/10.1063/1.1319644).
- [311] J.S. Kottmann, P. Schleich, T. Tamayo-Mendoza, A. Aspuru-Guzik, Reducing qubit requirements while maintaining numerical precision for the variational quantum eigensolver: A basis-set-free approach, *J. Phys. Chem. Lett.* 12 (1) (2021) 663–673, <http://dx.doi.org/10.1021/acs.jpclett.0c03410>.
- [312] S. Wiesner, Simulations of many-body quantum systems by a quantum computer, 1996, [arXiv:quant-ph/9603028v1](https://arxiv.org/abs/quant-ph/9603028v1).
- [313] C. Zalka, Simulating quantum systems on a quantum computer, *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* 454 (1969) (1998) 313–322, <http://dx.doi.org/10.1098/rspa.1998.0162>.
- [314] D.A. Lidar, H. Wang, Calculating the thermal rate constant with exponential speedup on a quantum computer, *Phys. Rev. E* 59 (2) (1999) 2429–2438, <http://dx.doi.org/10.1103/physreve.59.2429>.
- [315] I. Kassal, S.P. Jordan, P.J. Love, M. Mohseni, A. Aspuru-Guzik, Polynomial-time quantum algorithm for the simulation of chemical dynamics, *Proc. Natl. Acad. Sci.* 105 (48) (2008) 18681–18686, <http://dx.doi.org/10.1073/pnas.0808245105>.
- [316] N.J. Ward, I. Kassal, A. Aspuru-Guzik, Preparation of many-body states for quantum simulation, *J. Chem. Phys.* 130 (19) (2009) 194105, <http://dx.doi.org/10.1063/1.3115177>.
- [317] N.C. Jones, J.D. Whitfield, P.L. McMahon, M.-H. Yung, R.V. Meter, A. Aspuru-Guzik, Y. Yamamoto, Faster quantum chemistry simulation on fault-tolerant quantum computers, *New J. Phys.* 14 (11) (2012) 115023, <http://dx.doi.org/10.1088/1367-2630/14/11/115023>.
- [318] I.D. Kivlichan, N. Wiebe, R. Babbush, A. Aspuru-Guzik, Bounding the costs of quantum simulation of many-body physics in real space, *J. Phys. A* 50 (30) (2017) 305301, <http://dx.doi.org/10.1088/1751-8121/aa77b8>.
- [319] P. Verma, L. Huntington, M.P. Coons, Y. Kawashima, T. Yamazaki, A. Zaribafiyani, Scaling up electronic structure calculations on quantum computers: The frozen natural orbital based method of increments, *J. Chem. Phys.* 155 (3) (2021) 034110, <http://dx.doi.org/10.1063/5.0054647>.
- [320] Y. Mochizuki, K. Okuwaki, T. Kato, Y. Minato, Reduction of orbital space for molecular orbital calculations with quantum computation simulator for educators, *ChemRxiv* (2019) <http://dx.doi.org/10.26434/chemrxiv.9863810.v1>.
- [321] S. Lloyd, Universal quantum simulators, *Science* 273 (5278) (1996) 1073–1078, <http://dx.doi.org/10.1126/science.273.5278.1073>.
- [322] B.M. Boghosian, W. Taylor, Simulating quantum mechanics on a quantum computer, *Physica D* 120 (1–2) (1998) 30–42, [http://dx.doi.org/10.1016/s0167-2789\(98\)00042-6](http://dx.doi.org/10.1016/s0167-2789(98)00042-6).
- [323] Y. Su, D.W. Berry, N. Wiebe, N. Rubin, R. Babbush, Fault-tolerant quantum simulations of chemistry in first quantization, 2021, [arXiv:2105.12767v3](https://arxiv.org/abs/2105.12767v3).
- [324] J.C. Slater, The theory of complex spectra, *Phys. Rev.* 34 (10) (1929) 1293–1322, <http://dx.doi.org/10.1103/physrev.34.1293>.
- [325] E.U. Condon, The theory of complex spectra, *Phys. Rev.* 36 (7) (1930) 1121–1133, <http://dx.doi.org/10.1103/physrev.36.1121>.
- [326] N. Moll, P. Barkoutsos, L.S. Bishop, J.M. Chow, A. Cross, D.J. Egger, S. Filipp, A. Fuhrer, J.M. Gambetta, M. Ganzhorn, A. Kandala, A. Mezzacapo, P. Müller, W. Riess, G. Salis, J. Smolin, I. Tavernelli, K. Temme, Quantum optimization using variational algorithms on near-term quantum devices, *Quantum Sci. Technol.* 3 (3) (2018) 030503, <http://dx.doi.org/10.1088/2058-9565/aab822>.
- [327] W.A. Harrison, *Elementary Electronic Structure*, WSPC, 2004, <http://dx.doi.org/10.1142/5432>.
- [328] M. Vojta, Quantum phase transitions, *Rep. Progr. Phys.* 66 (12) (2003) 2069–2110, <http://dx.doi.org/10.1088/0034-4885/66/12/r01>.
- [329] S. Sachdev, *Quantum Phase Transitions*, Cambridge University Press, 2009, <http://dx.doi.org/10.1017/cbo9780511973765>.
- [330] S. Simon, *The Oxford Solid State Basics*, Oxford University Press, Oxford, ISBN: 9780199680771, 2013.

- [331] B. Bauer, D. Wecker, A.J. Millis, M.B. Hastings, M. Troyer, Hybrid quantum-classical approach to correlated materials, *Phys. Rev. X* 6 (3) (2016) <http://dx.doi.org/10.1103/physrevx.6.031045>.
- [332] J.L. Bosse, A. Montanaro, Probing ground state properties of the kagome antiferromagnetic heisenberg model using the variational quantum eigensolver, 2021, [arXiv:2108.08086](https://arxiv.org/abs/2108.08086).
- [333] J. Kattemölle, J. van Wezel, Variational quantum eigensolver for the Heisenberg antiferromagnet on the kagome lattice, 2021, [arXiv:2108.02175](https://arxiv.org/abs/2108.02175).
- [334] J. Pesonen, L. Halonen, Recent advances in the theory of vibration–rotation Hamiltonians, in: *Advances in Chemical Physics*, John Wiley & Sons, Ltd, ISBN: 9780471428022, 2003, pp. 269–349, <http://dx.doi.org/10.1002/0471428027.ch4>, chap. 4.
- [335] J. McClain, Q. Sun, G.K.-L. Chan, T.C. Berkelbach, Gaussian-based coupled-cluster theory for the ground-state and band structure of solids, *J. Chem. Theory Comput.* 13 (3) (2017) 1209–1218, <http://dx.doi.org/10.1021/acs.jctc.7b00049>.
- [336] N. Moll, A. Fuhrer, P. Staar, I. Tavernelli, Optimizing qubit resources for quantum chemistry simulations in second quantization on a quantum computer, *J. Phys. A* 49 (29) (2016) 295301, <http://dx.doi.org/10.1088/1751-8113/49/29/295301>.
- [337] M. Steudtner, S. Wehner, Fermion-to-qubit mappings with varying resource requirements for quantum simulation, *New J. Phys.* 20 (6) (2018) 063010, <http://dx.doi.org/10.1088/1367-2630/aac54f>.
- [338] L. Clinton, J. Bausch, T. Cubitt, Hamiltonian simulation algorithms for near-term quantum hardware, *Nature Commun.* 12 (1) (2021) <http://dx.doi.org/10.1038/s41467-021-25196-0>, [arXiv:2003.06886v3](https://arxiv.org/abs/2003.06886v3); *Nat. Commun.* 12 (2021) 4989.
- [339] C. Cade, L. Mineh, A. Montanaro, S. Stanisic, Strategies for solving the Fermi-Hubbard model on near-term quantum computers, *Phys. Rev. B* 102 (23) (2020) <http://dx.doi.org/10.1103/physrevb.102.235122>.
- [340] A. Tranter, P.J. Love, F. Mintert, P.V. Coveney, A comparison of the Bravyi–Kitaev and Jordan–Wigner transformations for the quantum simulation of quantum chemistry, *J. Chem. Theory Comput.* 14 (11) (2018) 5617–5630, <http://dx.doi.org/10.1021/acs.jctc.8b00450>.
- [341] G. Ortiz, J.E. Gubernatis, E. Knill, R. Laflamme, Quantum algorithms for fermionic simulations, *Phys. Rev. A* 64 (2) (2001) <http://dx.doi.org/10.1103/physreva.64.022319>.
- [342] R. Somma, G. Ortiz, J.E. Gubernatis, E. Knill, R. Laflamme, Simulating physical phenomena by quantum networks, *Phys. Rev. A* 65 (4) (2002) <http://dx.doi.org/10.1103/physreva.65.042323>.
- [343] R. Somma, G. Ortiz, E. Knill, J. Gubernatis, Quantum simulations of physics problems, *Int. J. Quantum Inf.* 01 (02) (2003) 189–206, <http://dx.doi.org/10.1142/s0219749903000140>.
- [344] Y.S. Yordanov, V. Armaos, C.H.W. Barnes, D.R.M. Arvidsson-Shukur, Qubit-excitation-based adaptive variational quantum eigensolver, *Commun. Phys.* 4 (1) (2021) <http://dx.doi.org/10.1038/s42005-021-00730-0>.
- [345] M. Consiglio, W.J. Chetcuti, C. Bravo-Prieto, S. Ramos-Calderer, A. Minguzzi, J.I. Latorre, L. Amico, T.J.G. Apollaro, Variational quantum eigensolver for SU(N) Fermions, 2021, <http://dx.doi.org/10.48550/ARXIV.2106.15552>, URL: <https://arxiv.org/abs/2106.15552>.
- [346] A.Y. Vlasov, Clifford algebras, spin groups and qubit trees, 2019, [arXiv:1904.09912v5](https://arxiv.org/abs/1904.09912v5).
- [347] E. Majorana, Teoria simmetrica dell’elettrone e del positrone, *Il Nuovo Cimento* 14 (4) (1937) 171–184, <http://dx.doi.org/10.1007/bf02961314>.
- [348] E. Majorana, L. Maiani, A symmetric theory of electrons and positrons, in: *Ettore Majorana Scientific Papers*, Springer Berlin Heidelberg, 2006, pp. 201–233, http://dx.doi.org/10.1007/978-3-540-48095-2_10.
- [349] B. Kayser, Are neutrinos their own antiparticles? *J. Phys. Conf. Ser.* 173 (2009) 012013, <http://dx.doi.org/10.1088/1742-6596/173/1/012013>.
- [350] A.B. Balantekin, A. de Gouvêa, B. Kayser, Addressing the majorana vs. Dirac question with neutrino decays, *Phys. Lett. B* 789 (2019) 488–495, <http://dx.doi.org/10.1016/j.physletb.2018.11.068>.
- [351] M. Hirsch, R. Srivastava, J.W. Valle, Can one ever prove that neutrinos are Dirac particles? *Phys. Lett. B* 781 (2018) 302–305, <http://dx.doi.org/10.1016/j.physletb.2018.03.073>.
- [352] S. Bilenky, Neutrinos: Majorana or Dirac?, 2020, [arXiv:2008.02110](https://arxiv.org/abs/2008.02110).
- [353] F. Wilczek, Majorana returns, *Nat. Phys.* 5 (9) (2009) 614–618, <http://dx.doi.org/10.1038/nphys1380>.
- [354] F. Verstraete, J.I. Cirac, Mapping local Hamiltonians of fermions to local Hamiltonians of spins, *J. Stat. Mech. Theory Exp.* 2005 (09) (2005) P09012, <http://dx.doi.org/10.1088/1742-5468/2005/09/p09012>.
- [355] R.C. Ball, Fermions without Fermion fields, *Phys. Rev. Lett.* 95 (17) (2005) <http://dx.doi.org/10.1103/physrevlett.95.176407>.
- [356] M. Levin, X.-G. Wen, Fermions, strings, and gauge fields in lattice spin models, *Phys. Rev. B* 67 (24) (2003) <http://dx.doi.org/10.1103/physrevb.67.245316>.
- [357] X.-G. Wen, Quantum order from string-net condensations and the origin of light and massless fermions, *Phys. Rev. D* 68 (6) (2003) <http://dx.doi.org/10.1103/physrevd.68.065003>.
- [358] J.D. Whitfield, V. Havlíček, M. Troyer, Local spin operators for fermion simulations, *Phys. Rev. A* 94 (3) (2016) <http://dx.doi.org/10.1103/physreva.94.030301>.
- [359] M. Steudtner, S. Wehner, Quantum codes for quantum simulation of fermions on a square lattice of qubits, *Phys. Rev. A* 99 (2) (2019) <http://dx.doi.org/10.1103/physreva.99.022308>.
- [360] Z. Jiang, J. McClean, R. Babbush, H. Neven, Majorana loop stabilizer codes for error mitigation in Fermionic quantum simulations, *Phys. Rev. A* 12 (6) (2019) <http://dx.doi.org/10.1103/physrevapplied.12.064041>.
- [361] R.W. Chien, J.D. Whitfield, Custom fermionic codes for quantum simulation, 2020, [arXiv:2009.11860v1](https://arxiv.org/abs/2009.11860v1).
- [362] J. Bausch, T. Cubitt, C. Derby, J. Klassen, Mitigating errors in local Fermionic encodings, 2020, [arXiv:2003.07125v2](https://arxiv.org/abs/2003.07125v2).
- [363] R.W. Chien, S. Xue, T.S. Hardikar, K. Setia, J.D. Whitfield, Analysis of superfast encoding performance for electronic structure simulations, *Phys. Rev. A* 100 (3) (2019) <http://dx.doi.org/10.1103/physreva.100.032337>.
- [364] A. Messiah, *Quantum Mechanics*, Dover Publications, City, ISBN: 9780486784557, 2014.
- [365] F. Cotton, *Chemical Applications of Group Theory*, Wiley, New York, ISBN: 978-0-471-51094-9, 1990.
- [366] F. Zhang, N. Gomes, N.F. Berthussen, P.P. Orth, C.-Z. Wang, K.-M. Ho, Y.-X. Yao, Shallow-circuit variational quantum eigensolver based on symmetry-inspired Hilbert space partitioning for quantum chemical calculations, *Phys. Rev. Res.* 3 (1) (2021) <http://dx.doi.org/10.1103/physrevresearch.3.013039>.
- [367] R. Raussendorf, Contextuality in measurement-based quantum computation, *Phys. Rev. A* 88 (2) (2013) <http://dx.doi.org/10.1103/physreva.88.022322>.
- [368] M. Howard, J. Wallman, V. Veitch, J. Emerson, Contextuality supplies the ‘magic’ for quantum computation, *Nature* 510 (7505) (2014) 351–355, <http://dx.doi.org/10.1038/nature13460>.
- [369] A. Cabello, S. Severini, A. Winter, Graph-theoretic approach to quantum correlations, *Phys. Rev. Lett.* 112 (4) (2014) <http://dx.doi.org/10.1103/physrevlett.112.040401>.
- [370] A. Cabello, M. Kleinmann, C. Budroni, Necessary and sufficient condition for quantum state-independent contextuality, *Phys. Rev. Lett.* 114 (25) (2015) <http://dx.doi.org/10.1103/physrevlett.114.250402>.
- [371] R. Ramanathan, P. Horodecki, Necessary and sufficient condition for state-independent contextual measurement scenarios, *Phys. Rev. Lett.* 112 (4) (2014) <http://dx.doi.org/10.1103/physrevlett.112.040404>.
- [372] W.M. Kirby, P.J. Love, Contextuality test of the nonclassicality of variational quantum eigensolvers, *Phys. Rev. Lett.* 123 (20) (2019) <http://dx.doi.org/10.1103/physrevlett.123.200501>.

- [373] W.M. Kirby, P.J. Love, Classical simulation of noncontextual Pauli Hamiltonians, *Phys. Rev. A* 102 (3) (2020) <http://dx.doi.org/10.1103/physreva.102.032418>.
- [374] M. Dekking, *A Modern Introduction to Probability and Statistics : Understanding Why and how*, Springer, London, ISBN: 978-1-84628-168-6, 2005.
- [375] J.M. Kübler, A. Arrasmith, L. Cincio, P.J. Coles, An adaptive optimizer for measurement-frugal variational algorithms, *Quantum* 4 (2020) 263, <http://dx.doi.org/10.22331/q-2020-05-11-263>.
- [376] T.-C. Yen, V. Verteletskyi, A.F. Izmaylov, Measuring all compatible operators in one series of single-qubit measurements using unitary transformations, *J. Chem. Theory Comput.* 16 (4) (2020) 2400–2409, <http://dx.doi.org/10.1021/acs.jctc.0c00008>.
- [377] A.F. Izmaylov, T.-C. Yen, R.A. Lang, V. Verteletskyi, Unitary partitioning approach to the measurement problem in the variational quantum eigensolver method, *J. Chem. Theory Comput.* 16 (1) (2019) 190–195, <http://dx.doi.org/10.1021/acs.jctc.9b00791>.
- [378] G.M. D'Ariano, M.G. Paris, M.F. Sacchi, Quantum tomography, in: *Advances in Imaging and Electron Physics*, Elsevier, 2003, pp. 205–308, [http://dx.doi.org/10.1016/s1076-5670\(03\)80065-4](http://dx.doi.org/10.1016/s1076-5670(03)80065-4).
- [379] M. Cramer, M.B. Plenio, S.T. Flammia, R. Somma, D. Gross, S.D. Bartlett, O. Landon-Cardinal, D. Poulin, Y.-K. Liu, Efficient quantum state tomography, *Nature Commun.* 1 (1) (2010) <http://dx.doi.org/10.1038/ncomms1147>.
- [380] M. Christandl, R. Renner, Reliable quantum state tomography, *Phys. Rev. Lett.* 109 (12) (2012) <http://dx.doi.org/10.1103/physrevlett.109.120403>.
- [381] A. Bisio, G. Chiribella, G. D'Ariano, S. Facchini, P. Perinotti, Optimal quantum tomography, *IEEE J. Sel. Top. Quantum Electron.* 15 (6) (2009) 1646–1660, <http://dx.doi.org/10.1109/jstqe.2009.2029243>.
- [382] R. O'Donnell, J. Wright, Efficient quantum tomography, in: *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing*, ACM, 2016, pp. 899–912, <http://dx.doi.org/10.1145/2897518.2897544>.
- [383] R. O'Donnell, J. Wright, Efficient quantum tomography II, in: *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, ACM, 2017, pp. 962–974, <http://dx.doi.org/10.1145/3055399.3055454>.
- [384] J. Haah, A.W. Harrow, Z. Ji, X. Wu, N. Yu, Sample-optimal tomography of quantum states, *IEEE Trans. Inform. Theory* (2017) 1, <http://dx.doi.org/10.1109/tit.2017.2719044>.
- [385] B.C. Hall, *Lie Groups, Lie Algebras, and Representations*, Springer International Publishing, 2015, <http://dx.doi.org/10.1007/978-3-319-13467-3>.
- [386] J. Cotler, F. Wilczek, Quantum overlapping tomography, *Phys. Rev. Lett.* 124 (10) (2020) 100401, <http://dx.doi.org/10.1103/physrevlett.124.100401>.
- [387] T.J. Evans, R. Harper, S.T. Flammia, Scalable Bayesian Hamiltonian learning, 2019, [arXiv:1912.07636v1](https://arxiv.org/abs/1912.07636v1).
- [388] S. Aaronson, Shadow tomography of quantum states, 2017, [arXiv:1711.01053v2](https://arxiv.org/abs/1711.01053v2).
- [389] S. Aaronson, G.N. Rothblum, Gentle measurement of quantum states and differential privacy, in: *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, in: *STOC 2019*, Association for Computing Machinery, New York, NY, USA, ISBN: 9781450367059, 2019, pp. 322–333, <http://dx.doi.org/10.1145/3313276.3316378>.
- [390] M.R. Jerrum, L.G. Valiant, V.V. Vazirani, Random generation of combinatorial structures from a uniform distribution, *Theoret. Comput. Sci.* 43 (1986) 169–188, [http://dx.doi.org/10.1016/0304-3975\(86\)90174-x](http://dx.doi.org/10.1016/0304-3975(86)90174-x).
- [391] C. Blair, *Problem Complexity and Method Efficiency in Optimization* (A. S. Nemirovsky and D. B. Yudin), Vol. 27, Society for Industrial & Applied Mathematics (SIAM), 1985, pp. 264–265, <http://dx.doi.org/10.1137/1027074>.
- [392] C. Hadfield, S. Bravyi, R. Raymond, A. Mezzacapo, Measurements of quantum Hamiltonians with locally-biased classical shadows, 2020, [arXiv:2006.15788v1](https://arxiv.org/abs/2006.15788v1).
- [393] H.-Y. Huang, R. Kueng, J. Preskill, Efficient estimation of Pauli observables by derandomization, *Phys. Rev. Lett.* 127 (3) (2021) <http://dx.doi.org/10.1103/physrevlett.127.030503>, [arXiv:2103.07510v1](https://arxiv.org/abs/2103.07510v1); *Phys. Rev. Lett.* 127 (2021) 030503.
- [394] A. Acharya, S. Saha, A.M. Sengupta, Informationally complete POVM-based shadow tomography, 2021, [arXiv:2105.05992](https://arxiv.org/abs/2105.05992).
- [395] C. Hadfield, Adaptive Pauli shadows for energy estimation, 2021, [http://dx.doi.org/10.48550/ARXIV.2105.12207](https://arxiv.org/abs/2105.12207), URL: <https://arxiv.org/abs/2105.12207>.
- [396] H.-Y. Hu, Y.-Z. You, Hamiltonian-driven shadow tomography of quantum states, *Phys. Rev. Res.* 4 (1) (2022) <http://dx.doi.org/10.1103/physrevresearch.4.013054>.
- [397] H.-Y. Hu, S. Choi, Y.-Z. You, Classical shadow tomography with locally scrambled quantum dynamics, 2021, [http://dx.doi.org/10.48550/ARXIV.2107.04817](https://arxiv.org/abs/2107.04817), URL: <https://arxiv.org/abs/2107.04817>.
- [398] A. Zhao, N.C. Rubin, A. Miyake, Fermionic partial tomography via classical shadows, *Phys. Rev. Lett.* 127 (11) (2021) <http://dx.doi.org/10.1103/physrevlett.127.110504>.
- [399] K. Bu, D.E. Koh, R.J. Garcia, A. Jaffe, Classical shadows with Pauli-invariant unitary ensembles, 2022, [http://dx.doi.org/10.48550/ARXIV.2202.03272](https://arxiv.org/abs/2202.03272), URL: <https://arxiv.org/abs/2202.03272>.
- [400] J.M. Lukens, K.J.H. Law, R.S. Bennink, A Bayesian analysis of classical shadows, *Npj Quantum Inf.* 7 (1) (2021) <http://dx.doi.org/10.1038/s41534-021-00447-6>.
- [401] S. Chen, W. Yu, P. Zeng, S.T. Flammia, Robust shadow estimation, *PRX Quantum* 2 (3) (2021) <http://dx.doi.org/10.1103/prxquantum.2.030348>.
- [402] D.E. Koh, S. Grewal, Classical shadows with noise, 2020, [http://dx.doi.org/10.48550/ARXIV.2011.11580](https://arxiv.org/abs/2011.11580), URL: <https://arxiv.org/abs/2011.11580>.
- [403] S.H. Sack, R.A. Medina, A.A. Michailidis, R. Kueng, M. Serbyn, Avoiding barren plateaus using classical shadows, 2022, [http://dx.doi.org/10.48550/ARXIV.2201.08194](https://arxiv.org/abs/2201.08194), URL: <https://arxiv.org/abs/2201.08194>.
- [404] D.H. Ackley, G.E. Hinton, T.J. Sejnowski, A learning algorithm for Boltzmann machines, *Cogn. Sci.* 9 (1) (1985) 147–169, http://dx.doi.org/10.1207/s15516709cog0901_7.
- [405] R.G. Melko, G. Carleo, J. Carrasquilla, J.I. Cirac, Restricted Boltzmann machines in quantum physics, *Nat. Phys.* 15 (9) (2019) 887–892, <http://dx.doi.org/10.1038/s41567-019-0545-1>.
- [406] G. Torlai, R.G. Melko, Machine-learning quantum states in the NISQ era, *Annu. Rev. Condens. Matter Phys.* 11 (1) (2020) 325–344, <http://dx.doi.org/10.1146/annurev-conmatphys-031119-050651>.
- [407] D. Gottesman, Stabilizer codes and quantum error correction, 1997, [arXiv:quant-ph/9705052v1](https://arxiv.org/abs/quant-ph/9705052v1).
- [408] A. Jena, S. Genin, M. Mosca, Pauli partitioning with respect to gate sets, 2019, [arXiv:1907.07859v1](https://arxiv.org/abs/1907.07859v1).
- [409] M.V. den Nest, J. Dehaene, B.D. Moor, Graphical description of the action of local Clifford transformations on graph states, *Phys. Rev. A* 69 (2) (2004) <http://dx.doi.org/10.1103/physreva.69.022316>.
- [410] S. Aaronson, D. Gottesman, Improved simulation of stabilizer circuits, *Phys. Rev. A* 70 (5) (2004) <http://dx.doi.org/10.1103/physreva.70.052328>.
- [411] K.N. Patel, I.L. Markov, J.P. Hayes, Optimal synthesis of linear reversible circuits, *Quantum Inf. Comput.* 8 (3) (2008) 282–294, <http://dx.doi.org/10.5555/2011763.2011767>.
- [412] A. Zhao, A. Tranter, W.M. Kirby, S.F. Ung, A. Miyake, P.J. Love, Measurement reduction in variational quantum algorithms, *Phys. Rev. A* 101 (6) (2020) 062322, <http://dx.doi.org/10.1103/physreva.101.062322>.
- [413] A. Ralli, P.J. Love, A. Tranter, P.V. Coveney, Implementation of measurement reduction for the variational quantum eigensolver, *Phys. Rev. Res.* 3 (3) (2021) <http://dx.doi.org/10.1103/physrevresearch.3.033195>.
- [414] M. Motta, E. Ye, J.R. McClean, Z. Li, A.J. Minnich, R. Babbush, G.K.-L. Chan, Low rank representations for quantum simulation of electronic structure, *Npj Quantum Inf.* 7 (1) (2021) 83, <http://dx.doi.org/10.1038/s41534-021-00416-z>.

- [415] B. Peng, K. Kowalski, Highly efficient and scalable compound decomposition of two-electron integral tensor and its application in coupled cluster calculations, *J. Chem. Theory Comput.* 13 (9) (2017) 4179–4192, <http://dx.doi.org/10.1021/acs.jctc.7b00605>.
- [416] D.W. Berry, C. Gidney, M. Motta, J.R. McClean, R. Babbush, Qubitization of arbitrary basis quantum chemistry leveraging sparsity and low rank factorization, *Quantum* 3 (2019) 208, <http://dx.doi.org/10.22331/q-2019-12-02-208>.
- [417] T.B. Pedersen, F. Aquilante, R. Lindh, Density fitting with auxiliary basis sets from Cholesky decompositions, *Theor. Chem. Acc.* 124 (1) (2009) 1–10, <http://dx.doi.org/10.1007/s00214-009-0608-y>.
- [418] N.H.F. Beebe, J. Linderberg, Simplifications in the generation and transformation of two-electron integrals in molecular calculations, *Int. J. Quantum Chem.* 12 (4) (1977) 683–705, <http://dx.doi.org/10.1002/qua.560120408>.
- [419] H. Koch, A.S. de Merás, T.B. Pedersen, Reduced scaling in electronic structure calculations using Cholesky decompositions, *J. Chem. Phys.* 118 (21) (2003) 9481–9484, <http://dx.doi.org/10.1063/1.1578621>.
- [420] W. Purwanto, H. Krakauer, Y. Virgus, S. Zhang, Assessing weak hydrogen binding on Ca⁺ centers: An accurate many-body study with large basis sets, *J. Chem. Phys.* 135 (16) (2011) 164105, <http://dx.doi.org/10.1063/1.3654002>.
- [421] N. Mardirossian, J.D. McClain, G.K.-L. Chan, Lowering of the complexity of quantum chemistry methods by choice of representation, *J. Chem. Phys.* 148 (4) (2018) 044106, <http://dx.doi.org/10.1063/1.5007779>.
- [422] C.G. Broyden, The convergence of a class of double-rank minimization algorithms 1. general considerations, *IMA J. Appl. Math.* 6 (1) (1970) 76–90, <http://dx.doi.org/10.1093/imamat/6.1.76>.
- [423] R. Fletcher, A new approach to variable metric algorithms, *Comput. J.* 13 (3) (1970) 317–322, <http://dx.doi.org/10.1093/comjnl/13.3.317>.
- [424] D. Goldfarb, A family of variable-metric methods derived by variational means, *Math. Comp.* 24 (109) (1970) 23, <http://dx.doi.org/10.1090/s0025-5718-1970-0258249-6>.
- [425] D.F. Shanno, Conditioning of quasi-Newton methods for function minimization, *Math. Comp.* 24 (111) (1970) 647, <http://dx.doi.org/10.1090/s0025-5718-1970-0274029-x>.
- [426] M. Garey, *Computers and Intractability : A Guide to the Theory of NP-Completeness*, W.H. Freeman, San Francisco, ISBN: 978-0-7167-1045-5, 1979.
- [427] R.M. Karp, Reducibility among combinatorial problems, in: R.E. Miller, J.W. Thatcher, J.D. Bohlinger (Eds.), *Complexity of Computer Computations*, in: The IBM Research Symposia Series, Springer US, Boston, MA, ISBN: 978-1-4684-2001-2, 1972, pp. 85–103, http://dx.doi.org/10.1007/978-1-4684-2001-2_9.
- [428] D.J.A. Welsh, An upper bound for the chromatic number of a graph and its application to timetabling problems, *Comput. J.* 10 (1) (1967) 85–86, <http://dx.doi.org/10.1093/comjnl/10.1.85>.
- [429] D.W. Matula, G. Marble, J.D. Isaacson, GRAPH COLORING ALGORITHMS††This research was supported in part by the Advanced Research Projects Agency of the Department of Defense under contract SD-302 and by the National Science Foundation under contract GJ-446, *Graph Theory Comput.* (1972) 109–122, <http://dx.doi.org/10.1016/b978-1-4832-3187-7.50015-5>.
- [430] D. Brélaz, New methods to color the vertices of a graph, *Commun. ACM* 22 (4) (1979) 251–256, <http://dx.doi.org/10.1145/359094.359101>.
- [431] H.J. Vallury, M.A. Jones, C.D. Hill, L.C.L. Hollenberg, Quantum computed moments correction to variational estimates, *Quantum* 4 (2020) 373, <http://dx.doi.org/10.22331/q-2020-12-15-373>.
- [432] R.D. Dutton, A new graph colouring algorithm, *Comput. J.* 24 (1) (1981) 85–86, <http://dx.doi.org/10.1093/comjnl/24.1.85>.
- [433] A. Hertz, A fast algorithm for coloring Meyniel graphs, *J. Combin. Theory Ser. B* 50 (2) (1990) 231–240, [http://dx.doi.org/10.1016/0095-8956\(90\)90078-e](http://dx.doi.org/10.1016/0095-8956(90)90078-e).
- [434] R. Boppana, M.M. Halldórsson, Approximating maximum independent sets by excluding subgraphs, *BIT* 32 (2) (1992) 180–196, <http://dx.doi.org/10.1007/bf01994876>.
- [435] E. Tomita, A. Tanaka, H. Takahashi, The worst-case time complexity for generating all maximal cliques and computational experiments, *Theoret. Comput. Sci.* 363 (1) (2006) 28–42, <http://dx.doi.org/10.1016/j.tcs.2006.06.015>.
- [436] M. Kubale (Ed.), *Graph Colorings*, American Mathematical Society, ISBN: 9780821879429, 2004, <http://dx.doi.org/10.1090/conm/352>.
- [437] F.T. Leighton, A graph coloring algorithm for large scheduling problems, *J. Res. Natl. Inst. Stand. Technol.* 84 (1979) 489–506, URL: https://nvlpubs.nist.gov/nistpubs/jres/84/jresv84n6p489_a1b.pdf.
- [438] S. Sim, P.D. Johnson, A. Aspuru-Guzik, Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms, *Adv. Quantum Technol.* 2 (12) (2019) 1900070, <http://dx.doi.org/10.1002/qute.201900070>.
- [439] T. Hubregtsen, J. Pichlmeier, P. Stecher, K. Bertels, Evaluation of parameterized quantum circuits: on the relation between classification accuracy, expressibility, and entangling capability, *Quantum Mach. Intell.* 3 (1) (2021) <http://dx.doi.org/10.1007/s42484-021-00038-w>.
- [440] S.E. Rasmussen, N.J.S. Loft, T. Bækkegaard, M. Kues, N.T. Zinner, Reducing the amount of single-qubit rotations in VQE and related algorithms, *Adv. Quantum Technol.* 3 (12) (2020) 2000063, <http://dx.doi.org/10.1002/qute.202000063>.
- [441] L. Funcke, T. Hartung, K. Jansen, S. Kühn, P. Stornati, Dimensional expressivity analysis of parametric quantum circuits, *Quantum* 5 (2021) 422, <http://dx.doi.org/10.22331/q-2021-03-29-422>.
- [442] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, P.J. Coles, Noise-induced barren plateaus in variational quantum algorithms, 2020, [arXiv:2007.14384v3](https://arxiv.org/abs/2007.14384v3).
- [443] A. Arrasmith, Z. Holmes, M. Cerezo, P.J. Coles, Equivalence of quantum barren plateaus to cost concentration and narrow gorges, 2021, [arXiv:2104.05868v1](https://arxiv.org/abs/2104.05868v1).
- [444] A. Arrasmith, M. Cerezo, P. Czarnik, L. Cincio, P.J. Coles, Effect of barren plateaus on gradient-free optimization, 2020, [arXiv:2011.12245v1](https://arxiv.org/abs/2011.12245v1).
- [445] J. Napp, Quantifying the barren plateau phenomenon for a model of unstructured variational ansätze, 2022, <https://arxiv.org/abs/2203.06174>, URL: <https://arxiv.org/abs/2203.06174>.
- [446] M. Cerezo, P.J. Coles, Higher order derivatives of quantum neural networks with barren plateaus, *Quantum Sci. Technol.* 6 (3) (2021) 035006, <http://dx.doi.org/10.1088/2058-9565/abf51a>.
- [447] F. Sauvage, S. Sim, A.A. Kunitsa, W.A. Simon, M. Mauri, A. Perdomo-Ortiz, FLIP: A flexible initializer for arbitrarily-sized parameterized quantum circuits, 2021, [arXiv:2103.08572v2](https://arxiv.org/abs/2103.08572v2).
- [448] A. Kulshrestha, I. Safran, BEINIT: Avoiding barren plateaus in variational quantum algorithms, 2022, <https://arxiv.org/abs/2204.13751>, URL: <https://arxiv.org/abs/2204.13751>.
- [449] T. Volkoff, P.J. Coles, Large gradients via correlation in random parameterized quantum circuits, *Quantum Sci. Technol.* 6 (2) (2021) 025008, <http://dx.doi.org/10.1088/2058-9565/abd891>.
- [450] A. Skolik, J.R. McClean, M. Mohseni, P. van der Smagt, M. Leib, Layerwise learning for quantum neural networks, *Quantum Mach. Intell.* 3 (1) (2021) <http://dx.doi.org/10.1007/s42484-020-00036-4>.
- [451] E. Campos, A. Nasrallah, J. Biamonte, Abrupt transitions in variational quantum circuit training, *Phys. Rev. A* 103 (3) (2021) <https://arxiv.org/abs/2103.032607>.
- [452] A. Wu, G. Li, Y. Ding, Y. Xie, Mitigating noise-induced gradient vanishing in variational quantum algorithm training, 2021, <https://arxiv.org/abs/2111.13209>, URL: <https://arxiv.org/abs/2111.13209>.
- [453] A. Pesah, M. Cerezo, S. Wang, T. Volkoff, A.T. Sornborger, P.J. Coles, Absence of barren plateaus in quantum convolutional neural networks, 2020, [arXiv:2011.02966v1](https://arxiv.org/abs/2011.02966v1).

- [454] I.G. Ryabinkin, T.-C. Yen, S.N. Genin, A.F. Izmaylov, Qubit coupled cluster method: A systematic approach to quantum chemistry on a quantum computer, *J. Chem. Theory Comput.* 14 (12) (2018) 6317–6326, <http://dx.doi.org/10.1021/acs.jctc.8b00932>.
- [455] R. Xia, S. Kais, Qubit coupled cluster singles and doubles variational quantum eigensolver ansatz for electronic structure calculations, *Quantum Sci. Technol.* 6 (1) (2020) 015001, <http://dx.doi.org/10.1088/2058-9565/abb74>.
- [456] J. Romero, R. Babbush, J.R. McClean, C. Hempel, P.J. Love, A. Aspuru-Guzik, Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz, *Quantum Sci. Technol.* 4 (1) (2018) 014008, <http://dx.doi.org/10.1088/2058-9565/aad3e4>.
- [457] W. Mizukami, K. Mitarai, Y.O. Nakagawa, T. Yamamoto, T. Yan, Y.-y. Ohnishi, Orbital optimized unitary coupled cluster theory for quantum computer, *Phys. Rev. Res.* 2 (3) (2020) 033421, <http://dx.doi.org/10.1103/physrevresearch.2.033421>.
- [458] K. Mitarai, T. Yan, K. Fujii, Generalization of the output of a variational quantum eigensolver by parameter interpolation with a low-depth ansatz, *Phys. Rev. A* 11 (4) (2019) <http://dx.doi.org/10.1103/physrevapplied.11.044087>.
- [459] O. Higgott, D. Wang, S. Brierley, Variational quantum computation of excited states, *Quantum* 3 (2019) 156, <http://dx.doi.org/10.22331/q-2019-07-01-156>.
- [460] J. Tilly, G. Jones, H. Chen, L. Wossnig, E. Grant, Computation of molecular excited states on IBM quantum computers using a discriminative variational quantum eigensolver, *Phys. Rev. A* 102 (6) (2020) <http://dx.doi.org/10.1103/physreva.102.062425>.
- [461] H. Chen, M. Nusspickel, J. Tilly, G.H. Booth, Variational quantum eigensolver for dynamic correlation functions, *Phys. Rev. A* 104 (3) (2021) <http://dx.doi.org/10.1103/physreva.104.032405>, [arXiv:2105.01703v2](https://arxiv.org/abs/2105.01703v2); *Phys. Rev. A* 104 (2021) 032405.
- [462] A. Wu, G. Li, Y. Wang, B. Feng, Y. Ding, Y. Xie, Towards efficient ansatz architecture for variational quantum algorithms, 2021, <http://dx.doi.org/10.48550/ARXIV.2111.13730>, URL: <https://arxiv.org/abs/2111.13730>.
- [463] C. Bravo-Prieto, J. Lumbrales-Zarapico, L. Tagliacozzo, J.I. Latorre, Scaling of variational quantum circuit depth for condensed matter systems, *Quantum* 4 (2020) 272, <http://dx.doi.org/10.22331/q-2020-05-28-272>.
- [464] R.J. Bartlett, S.A. Kucharski, J. Noga, Alternative coupled-cluster ansätze II. The unitary coupled-cluster method, *Chem. Phys. Lett.* 155 (1) (1989) 133–140, [http://dx.doi.org/10.1016/s0009-2614\(89\)87372-5](http://dx.doi.org/10.1016/s0009-2614(89)87372-5).
- [465] A.G. Taube, R.J. Bartlett, New perspectives on unitary coupled-cluster theory, *Int. J. Quantum Chem.* 106 (15) (2006) 3393–3401, <http://dx.doi.org/10.1002/qua.21198>.
- [466] M.-A. Filip, A.J.W. Thom, A stochastic approach to unitary coupled cluster, *J. Chem. Phys.* 153 (21) (2020) 214106, <http://dx.doi.org/10.1063/5.0026141>.
- [467] F. Coester, H. Kümmel, Short-range correlations in nuclear wave functions, *Nucl. Phys.* 17 (1960) 477–485, [http://dx.doi.org/10.1016/0029-5582\(60\)90140-1](http://dx.doi.org/10.1016/0029-5582(60)90140-1).
- [468] J. Čížek, On the correlation problem in atomic and molecular systems. Calculation of wavefunction components in urself-type expansion using quantum-field theoretical methods, *J. Chem. Phys.* 45 (11) (1966) 4256–4266, <http://dx.doi.org/10.1063/1.1727484>.
- [469] J. Paldus, J. Čížek, I. Shavitt, Correlation problems in atomic and molecular systems. IV. Extended coupled-pair many-electron theory and its application to the BH3Molecule, *Phys. Rev. A* 5 (1) (1972) 50–67, <http://dx.doi.org/10.1103/physreva.5.50>.
- [470] J. Paldus, Correlation problems in atomic and molecular systems. V. Spin-adapted coupled cluster many-electron theory, *J. Chem. Phys.* 67 (1) (1977) 303–318, <http://dx.doi.org/10.1063/1.434526>.
- [471] J. Cizek, J. Paldus, Coupled cluster approach, *Phys. Scr.* 21 (3–4) (1980) 251–254, <http://dx.doi.org/10.1088/0031-8949/21/3-4/006>.
- [472] J. Paldus, J. Čížek, M. Takahashi, Approximate account of the connected quadruply excited clusters in the coupled-pair many-electron theory, *Phys. Rev. A* 30 (5) (1984) 2193–2209, <http://dx.doi.org/10.1103/physreva.30.2193>.
- [473] P. Piecuch, R. Toboza, J. Paldus, Coupled-cluster approaches to an approximate account of triply and quadruply excited clusters: Implementation of the orthogonally spin-adapted CCD + ST(CCD), CCSD + T(CCD), and ACPQ + ST(ACPQ) formalisms, *Int. J. Quantum Chem.* 55 (2) (1995) 133–146, <http://dx.doi.org/10.1002/qua.560550208>.
- [474] D.W. Small, M. Head-Gordon, A fusion of the closed-shell coupled cluster singles and doubles method and valence-bond theory for bond breaking, *J. Chem. Phys.* 137 (11) (2012) 114103, <http://dx.doi.org/10.1063/1.4751485>.
- [475] D.W. Small, K.V. Lawler, M. Head-Gordon, Coupled cluster valence bond method: Efficient computer implementation and application to multiple bond dissociations and strong correlations in the acenes, *J. Chem. Theory Comput.* 10 (5) (2014) 2027–2040, <http://dx.doi.org/10.1021/ct500112y>.
- [476] J. Lee, D.W. Small, E. Epifanovsky, M. Head-Gordon, Coupled-cluster valence-bond singles and doubles for strongly correlated systems: Block-tensor based implementation and application to oligoacenes, *J. Chem. Theory Comput.* 13 (2) (2017) 602–615, <http://dx.doi.org/10.1021/acs.jctc.6b01092>.
- [477] J. Lee, D.W. Small, M. Head-Gordon, Open-shell coupled-cluster valence-bond theory augmented with an independent amplitude approximation for three-pair correlations: Application to a model oxygen-evolving complex and single molecular magnet, *J. Chem. Phys.* 149 (24) (2018) 244121, <http://dx.doi.org/10.1063/1.5052667>.
- [478] B.O. Roos, P.R. Taylor, P.E. Sigbahn, A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach, *Chem. Phys.* 48 (2) (1980) 157–173, [http://dx.doi.org/10.1016/0301-0104\(80\)80045-0](http://dx.doi.org/10.1016/0301-0104(80)80045-0).
- [479] M. Nooijen, Can the eigenstates of a many-body Hamiltonian be represented exactly using a general two-body cluster expansion? *Phys. Rev. Lett.* 84 (10) (2000) 2108–2111, <http://dx.doi.org/10.1103/physrevlett.84.2108>.
- [480] T. Stein, T.M. Henderson, G.E. Scuseria, Seniority zero pair coupled cluster doubles theory, *J. Chem. Phys.* 140 (21) (2014) 214113, <http://dx.doi.org/10.1063/1.4880819>.
- [481] P.A. Limacher, P.W. Ayers, P.A. Johnson, S.D. Baerdemacker, D.V. Neck, P. Bultinck, A new mean-field method suitable for strongly correlated electrons: Computationally facile antisymmetric products of nonorthogonal geminals, *J. Chem. Theory Comput.* 9 (3) (2013) 1394–1401, <http://dx.doi.org/10.1021/ct300902c>.
- [482] W.J. Huggins, J. Lee, U. Baek, B. O’Gorman, K.B. Whaley, A non-orthogonal variational quantum eigensolver, *New J. Phys.* 22 (7) (2020) 073009, <http://dx.doi.org/10.1088/1367-2630/ab867b>.
- [483] G. Greene-Diniz, D.M. Ramo, Generalized unitary coupled cluster excitations for multireference molecular states optimized by the variational quantum eigensolver, *Int. J. Quantum Chem.* 121 (4) (2020) <http://dx.doi.org/10.1002/qua.26352>.
- [484] J.S. Kottmann, A. Aspuru-Guzik, Optimized low-depth quantum circuits for molecular electronic structure using a separable pair approximation, 2021, [arXiv:2105.03836v2](https://arxiv.org/abs/2105.03836v2).
- [485] C. Edmiston, M. Krauss, Pseudonatural orbitals as a basis for the superposition of configurations. II. Energy surface for linear H3, *J. Chem. Phys.* 49 (1) (1968) 192–205, <http://dx.doi.org/10.1063/1.1669809>.
- [486] R. Ahlrichs, H. Lischka, V. Staemmler, W. Kutzelnigg, PNO-CI (pair natural orbital configuration interaction) and CEPA-PNO (coupled electron pair approximation with pair natural orbitals) calculations of molecular systems. I. Outline of the method for closed-shell states, *J. Chem. Phys.* 62 (4) (1975) 1225–1234, <http://dx.doi.org/10.1063/1.430637>.
- [487] W. Meyer, PNO-CI studies of electron correlation effects. I. Configuration expansion by means of nonorthogonal orbitals, and application to the ground state and ionized states of methane, *J. Chem. Phys.* 58 (3) (1973) 1017–1035, <http://dx.doi.org/10.1063/1.1679283>.
- [488] W. Meyer, Ionization energies of water from PNO-CI calculations, *Int. J. Quantum Chem.* 5 (S5) (2009) 341–348, <http://dx.doi.org/10.1002/qua.560050839>.

- [489] F. Neese, F. Wennmohs, A. Hansen, Efficient and accurate local approximations to coupled-electron pair approaches: An attempt to revive the pair natural orbital method, *J. Chem. Phys.* 130 (11) (2009) 114108, <http://dx.doi.org/10.1063/1.3086717>.
- [490] F. Neese, A. Hansen, D.G. Liakos, Efficient and accurate approximations to the local coupled cluster singles doubles method using a truncated pair natural orbital basis, *J. Chem. Phys.* 131 (6) (2009) 064103, <http://dx.doi.org/10.1063/1.3173827>.
- [491] P.R. Nagy, M. Kállay, Approaching the basis set limit of CCSD(T) energies for large molecules with local natural orbital coupled-cluster methods, *J. Chem. Theory Comput.* 15 (10) (2019) 5275–5298, <http://dx.doi.org/10.1021/acs.jctc.9b00511>.
- [492] S. Yalouz, B. Senjean, J. Günther, F. Buda, T.E. O'Brien, L. Visscher, A state-averaged orbital-optimized hybrid quantum–classical algorithm for a democratic description of ground and excited states, *Quantum Sci. Technol.* 6 (2) (2021) 024004, <http://dx.doi.org/10.1088/2058-9565/abd334>.
- [493] U. Bozkaya, J.M. Turney, Y. Yamaguchi, H.F. Schaefer, C.D. Sherrill, Quadratically convergent algorithm for orbital optimization in the orbital-optimized coupled-cluster doubles method and in orbital-optimized second-order Møller-Plesset perturbation theory, *J. Chem. Phys.* 135 (10) (2011) 104103, <http://dx.doi.org/10.1063/1.3631129>.
- [494] U. Bozkaya, C.D. Sherrill, Orbital-optimized coupled-electron pair theory and its analytic gradients: Accurate equilibrium geometries, harmonic vibrational frequencies, and hydrogen transfer reactions, *J. Chem. Phys.* 139 (5) (2013) 054104, <http://dx.doi.org/10.1063/1.4816628>.
- [495] I.W. Bulik, T.M. Henderson, G.E. Scuseria, Can single-reference coupled cluster theory describe static correlation? *J. Chem. Theory Comput.* 11 (7) (2015) 3171–3179, <http://dx.doi.org/10.1021/acs.jctc.5b00422>.
- [496] K. Kowalski, Properties of coupled-cluster equations originating in excitation sub-algebras, *J. Chem. Phys.* 148 (9) (2018) 094104, <http://dx.doi.org/10.1063/1.5010693>.
- [497] N.P. Bauman, E.J. Bylaska, S. Krishnamoorthy, G.H. Low, N. Wiebe, C.E. Granade, M. Roetteler, M. Troyer, K. Kowalski, Downfolding of many-body Hamiltonians using active-space models: Extension of the sub-system embedding sub-algebras approach to unitary coupled cluster formalisms, *J. Chem. Phys.* 151 (1) (2019) 014107, <http://dx.doi.org/10.1063/1.5094643>.
- [498] N.P. Bauman, G.H. Low, K. Kowalski, Quantum simulations of excited states with active-space downfolded Hamiltonians, *J. Chem. Phys.* 151 (23) (2019) 234114, <http://dx.doi.org/10.1063/1.5128103>.
- [499] K. Kowalski, N.P. Bauman, Sub-system quantum dynamics using coupled cluster downfolding techniques, *J. Chem. Phys.* 152 (24) (2020) 244127, <http://dx.doi.org/10.1063/5.0008436>.
- [500] K. Kowalski, Dimensionality reduction of the many-body problem using coupled-cluster subsystem flow equations: Classical and quantum computing perspective, *Phys. Rev. A* 104 (3) (2021) <http://dx.doi.org/10.1103/physreva.104.032804>.
- [501] B.N. P, J. Chládek, L. Veis, J. Pittner, K. Karol, Variational quantum eigensolver for approximate diagonalization of downfolded Hamiltonians using generalized unitary coupled cluster ansatz, *Quantum Sci. Technol.* 6 (3) (2021) 034008, <http://dx.doi.org/10.1088/2058-9565/abf602>.
- [502] M. Metcalf, N.P. Bauman, K. Kowalski, W.A. de Jong, Resource-efficient chemistry on quantum computers with the variational quantum eigensolver and the double unitary coupled-cluster approach, *J. Chem. Theory Comput.* 16 (10) (2020) 6165–6175, <http://dx.doi.org/10.1021/acs.jctc.0c00421>.
- [503] Y. Matsuzawa, Y. Kurashige, Jastrow-type decomposition in quantum chemistry for low-depth quantum circuits, *J. Chem. Theory Comput.* 16 (2) (2020) 944–952, <http://dx.doi.org/10.1021/acs.jctc.9b00963>.
- [504] W.M.C. Foulkes, L. Mitars, R.J. Needs, G. Rajagopal, Quantum Monte Carlo simulations of solids, *Rev. Modern Phys.* 73 (1) (2001) 33–83, <http://dx.doi.org/10.1103/revmodphys.73.33>.
- [505] E. Neuscamman, Communication: A Jastrow factor coupled cluster theory for weak and strong electron correlation, *J. Chem. Phys.* 139 (18) (2013) 181101, <http://dx.doi.org/10.1063/1.4829536>.
- [506] F. Pavošević, T. Culpitt, S. Hammes-Schiffer, Multicomponent coupled cluster singles and doubles theory within the nuclear-electronic orbital framework, *J. Chem. Theory Comput.* 15 (1) (2018) 338–347, <http://dx.doi.org/10.1021/acs.jctc.8b01120>.
- [507] F. Pavošević, J. Flick, Polaritonic unitary coupled cluster for quantum computations, 2021, [arXiv:2106.09842v1](https://arxiv.org/abs/2106.09842v1).
- [508] P.-L. Dallaire-Demers, J. Romero, L. Veis, S. Sim, A. Aspuru-Guzik, Low-depth circuit ansatz for preparing correlated fermionic states on a quantum computer, *Quantum Sci. Technol.* 4 (4) (2019) 045005, <http://dx.doi.org/10.1088/2058-9565/ab3951>.
- [509] L.Z. Stolarczyk, H.J. Monkhorst, Quasiparticle Fock-space coupled-cluster theory, *Mol. Phys.* 108 (21–23) (2010) 3067–3089, <http://dx.doi.org/10.1080/00268976.2010.518981>.
- [510] Z. Rolik, M. Kállay, A quasiparticle-based multi-reference coupled-cluster method, *J. Chem. Phys.* 141 (13) (2014) 134112, <http://dx.doi.org/10.1063/1.4896703>.
- [511] A. Signoracci, T. Duguet, G. Hagen, G.R. Jansen, Ab initio Bogoliubov coupled cluster theory for open-shell nuclei, *Phys. Rev. C* 91 (6) (2015) <http://dx.doi.org/10.1103/physrevc.91.064320>.
- [512] S. Sim, J. Romero, J.F. Gonthier, A.A. Kunita, Adaptive pruning-based optimization of parameterized quantum circuits, *Quantum Sci. Technol.* 6 (2) (2021) 025019, <http://dx.doi.org/10.1088/2058-9565/abe107>.
- [513] D.A. Fedorov, Y. Alexeev, S.K. Gray, M. Otten, Unitary selective coupled-cluster method, *Quantum* 6 (2022) 703, <http://dx.doi.org/10.22331/q-2022-05-02-703>.
- [514] B. O’Gorman, W.J. Huggins, E.G. Rieffel, K.B. Whaley, Generalized swap networks for near-term quantum computing, 2019, [arXiv:1905.05118v1](https://arxiv.org/abs/1905.05118v1).
- [515] G.-L.R. Anselmetti, D. Wierichs, C. Gogolin, R.M. Parrish, Local, expressive, quantum-number-preserving VQE ansätze for fermionic systems, 2021, [arXiv:2104.05695v2](https://arxiv.org/abs/2104.05695v2).
- [516] Y.S. Yordanov, D.R.M. Arvidsson-Shukur, C.H.W. Barnes, Efficient quantum circuits for quantum computational chemistry, *Phys. Rev. A* 102 (6) (2020) <http://dx.doi.org/10.1103/physreva.102.062612>.
- [517] A.F. Izmaylov, M. Díaz-Tinoco, R.A. Lang, On the order problem in construction of unitary operators for the variational quantum eigensolver, *Phys. Chem. Chem. Phys.* 22 (23) (2020) 12980–12986, <http://dx.doi.org/10.1039/d0cp01707h>.
- [518] I.G. Ryabinkin, S.N. Genin, A.F. Izmaylov, Constrained variational quantum eigensolver: Quantum computer search engine in the fock space, *J. Chem. Theory Comput.* 15 (1) (2018) 249–255, <http://dx.doi.org/10.1021/acs.jctc.8b00943>.
- [519] M. Roth, M. Ganzhorn, N. Moll, S. Filipp, G. Salis, S. Schmidt, Analysis of a parametrically driven exchange-type gate and a two-photon excitation gate between superconducting qubits, *Phys. Rev. A* 96 (6) (2017) 062323, <http://dx.doi.org/10.1103/physreva.96.062323>.
- [520] D. Egger, M. Ganzhorn, G. Salis, A. Fuhrer, P. Müller, P. Barkoutsos, N. Moll, I. Tavernelli, S. Filipp, Entanglement generation in superconducting qubits using holonomic operations, *Phys. Rev. Appl.* 11 (2019) 014017, <http://dx.doi.org/10.1103/PhysRevApplied.11.014017>.
- [521] D.C. McKay, S. Filipp, A. Mezzacapo, E. Magesan, J.M. Chow, J.M. Gambetta, Universal gate for fixed-frequency qubits via a tunable bus, *Phys. Rev. A* 6 (6) (2016) 064007, <http://dx.doi.org/10.1103/physrevapplied.6.064007>.
- [522] M. Ganzhorn, D. Egger, P. Barkoutsos, P. Ollitrault, G. Salis, N. Moll, M. Roth, A. Fuhrer, P. Mueller, S. Woerner, I. Tavernelli, S. Filipp, Gate-efficient simulation of molecular eigenstates on a quantum computer, *Phys. Rev. A* 11 (4) (2019) 044092, <http://dx.doi.org/10.1103/physrevapplied.11.044092>.
- [523] E. Farhi, J. Goldstone, S. Gutmann, M. Sipser, Quantum computation by adiabatic evolution, 2000, [arXiv:quant-ph/0001106v1](https://arxiv.org/abs/quant-ph/0001106v1).
- [524] M.S. Jattana, F. Jin, H. De Raedt, K. Michielsen, Assessment of the variational quantum eigensolver: application to the heisenberg model, 2022, [http://dx.doi.org/10.48550/ARXIV.2201.05065](https://arxiv.org/abs/2201.05065), URL: <https://arxiv.org/abs/2201.05065>.

- [525] S. Stanisic, J.L. Bosse, F.M. Gambetta, R.A. Santos, W. Mroczkiewicz, T.E. O'Brien, E. Ostby, A. Montanaro, Observing ground-state properties of the Fermi-Hubbard model using a scalable algorithm on a quantum computer, 2021, <http://dx.doi.org/10.48550/ARXIV.2112.02025>, URL: <https://arxiv.org/abs/2112.02025>.
- [526] Z. Jiang, K.J. Sung, K. Kechedzhi, V.N. Smelyanskiy, S. Boixo, Quantum algorithms to simulate many-body physics of correlated Fermions, *Phys. Rev. A* 9 (4) (2018) <http://dx.doi.org/10.1103/physrevapplied.9.044036>.
- [527] F. Verstraete, J.I. Cirac, J.I. Latorre, Quantum circuits for strongly correlated quantum systems, *Phys. Rev. A* 79 (3) (2009) <http://dx.doi.org/10.1103/physreva.79.032316>.
- [528] A.J. Ferris, Fourier transform for Fermionic systems and the spectral tensor network, *Phys. Rev. Lett.* 113 (1) (2014) <http://dx.doi.org/10.1103/physrevlett.113.010401>.
- [529] J. Werschnik, E.K.U. Gross, Quantum optimal control theory, *J. Phys. B: At. Mol. Opt. Phys.* 40 (18) (2007) R175–R211, <http://dx.doi.org/10.1088/0953-4075/40/18/r01>.
- [530] D. D'Alessandro, Introduction to Quantum Control and Dynamics, Chapman and Hall/CRC, ISBN: 9780429150425, 2007, <http://dx.doi.org/10.1201/9781584888833>.
- [531] M.M. Müller, R.S. Said, F. Jelezko, T. Calarco, S. Montangero, One decade of quantum optimal control in the chopped random basis, 2021, <http://dx.doi.org/10.48550/ARXIV.2104.07687>, URL: <https://arxiv.org/abs/2104.07687>.
- [532] J. Liu, Z. Li, J. Yang, An efficient adaptive variational quantum solver of the Schrödinger equation based on reduced density matrices, *J. Chem. Phys.* 154 (24) (2021) 244112, <http://dx.doi.org/10.1063/5.0054822>.
- [533] H.R. Grimsley, G.S. Barron, E. Barnes, S.E. Economou, N.J. Mayhall, ADAPT-VQE is insensitive to rough parameter landscapes and barren plateaus, 2022, <http://dx.doi.org/10.48550/ARXIV.2204.07179>, URL: <https://arxiv.org/abs/2204.07179>.
- [534] M. Bilkis, M. Cerezo, G. Verdon, P.J. Coles, L. Cincio, A semi-agnostic ansatz with variable structure for quantum machine learning, 2021, [arXiv:2103.06712v1](https://arxiv.org/abs/2103.06712v1).
- [535] Y. Fan, C. Cao, X. Xu, Z. Li, D. Lv, M.-H. Yung, Circuit-depth reduction of unitary-coupled-cluster ansatz by energy sorting, 2021, <http://dx.doi.org/10.48550/ARXIV.2106.15210>, URL: <https://arxiv.org/abs/2106.15210>.
- [536] H. Lipkin, N. Meshkov, A. Glick, Validity of many-body approximation methods for a solvable model, *Nucl. Phys.* 62 (2) (1965) 188–198, [http://dx.doi.org/10.1016/0029-5582\(65\)90862-x](http://dx.doi.org/10.1016/0029-5582(65)90862-x).
- [537] E. Caurier, G. Martínez-Pinedo, F. Nowacki, A. Poves, A.P. Zuker, The shell model as a unified view of nuclear structure, *Rev. Modern Phys.* 77 (2) (2005) 427–488, <http://dx.doi.org/10.1103/revmodphys.77.427>.
- [538] K.L.G. Heyde, The nuclear shell model, in: *The Nuclear Shell Model*, Springer Berlin Heidelberg, 1994, pp. 58–154, http://dx.doi.org/10.1007/978-3-642-79052-2_4.
- [539] V.O. Shkolnikov, N.J. Mayhall, S.E. Economou, E. Barnes, Avoiding symmetry roadblocks and minimizing the measurement overhead of adaptive variational quantum eigensolvers, 2021, <http://dx.doi.org/10.48550/ARXIV.2109.05340>, URL: <https://arxiv.org/abs/2109.05340>.
- [540] A. Mukherjee, N.F. Berthussen, J.C. Getelina, P.P. Orth, Y.-X. Yao, Comparative study of adaptive variational quantum eigensolvers for multi-orbital impurity models, 2022, <http://dx.doi.org/10.48550/ARXIV.2203.06745>, URL: <https://arxiv.org/abs/2203.06745>.
- [541] Z.-J. Zhang, T.H. Kyaw, J.S. Kottmann, M. Degroote, A. Aspuru-Guzik, Mutual information-assisted adaptive variational quantum eigensolver, *Quantum Sci. Technol.* 6 (3) (2021) 035001, <http://dx.doi.org/10.1088/2058-9565/abdca4>.
- [542] L. Amico, R. Fazio, A. Osterloh, V. Vedral, Entanglement in many-body systems, *Rev. Modern Phys.* 80 (2) (2008) 517–576, <http://dx.doi.org/10.1103/revmodphys.80.517>.
- [543] Z. Huang, S. Kais, Entanglement as measure of electron–electron correlation in quantum chemistry calculations, *Chem. Phys. Lett.* 413 (1–3) (2005) 1–5, <http://dx.doi.org/10.1016/j.cplett.2005.07.045>.
- [544] J. Rissler, R.M. Noack, R.M. White, Measuring orbital interaction using quantum information theory, *Chem. Phys.* 323 (2–3) (2006) 519–531, <http://dx.doi.org/10.1016/j.chemphys.2005.10.018>.
- [545] R.A. Lang, I.G. Ryabinkin, A.F. Izmaylov, Unitary transformation of the electronic Hamiltonian with an exact quadratic truncation of the Baker–Campbell–Hausdorff expansion, *J. Chem. Theory Comput.* 17 (1) (2020) 66–78, <http://dx.doi.org/10.1021/acs.jctc.0c00170>.
- [546] S.N. Genin, I.G. Ryabinkin, N.R. Paisley, S.O. Whelan, M.G. Helander, Z.M. Hudson, Estimating phosphorescent emission energies in Ir^{III} complexes using large-scale quantum computing simulations, *Angew. Chem.* 134 (19) (2022) <http://dx.doi.org/10.1002/ange.202116175>.
- [547] I.G. Ryabinkin, A.F. Izmaylov, S.N. Genin, A posteriori corrections to the iterative qubit coupled cluster method to minimize the use of quantum resources in large-scale calculations, *Quantum Sci. Technol.* 6 (2) (2021) 024012, <http://dx.doi.org/10.1088/2058-9565/abdb8e>.
- [548] A.G. Rattew, S. Hu, M. Pistoia, R. Chen, S. Wood, A domain-agnostic, noise-resistant, hardware-efficient evolutionary variational quantum eigensolver, 2019, [arXiv:1910.09694v4](https://arxiv.org/abs/1910.09694v4).
- [549] D. Chivilikhin, A. Samarin, V. Ulyantsev, I. Iorsh, A.R. Oganov, O. Kyriienko, Mog-VQE: Multiobjective genetic variational quantum eigensolver, 2020, [arXiv:2007.04424v1](https://arxiv.org/abs/2007.04424v1).
- [550] Y. Du, T. Huang, S. You, M.-H. Hsieh, D. Tao, Quantum circuit architecture search: error mitigation and trainability enhancement for variational quantum solvers, 2020, [arXiv:2010.10217v2](https://arxiv.org/abs/2010.10217v2).
- [551] S.-X. Zhang, C.-Y. Hsieh, S. Zhang, H. Yao, Differentiable quantum architecture search, 2020, [arXiv:2010.08561v2](https://arxiv.org/abs/2010.08561v2).
- [552] H. Liu, K. Simonyan, Y. Yang, DARTS: Differentiable architecture search, 2018, [arXiv:1806.09055v2](https://arxiv.org/abs/1806.09055v2).
- [553] S.-X. Zhang, C.-Y. Hsieh, S. Zhang, H. Yao, Neural predictor based quantum architecture search, 2021, [arXiv:2103.06524v1](https://arxiv.org/abs/2103.06524v1).
- [554] E.-J. Kuo, Y.-L.L. Fang, S.Y.-C. Chen, Quantum architecture search via deep reinforcement learning, 2021, [arXiv:2104.07715v1](https://arxiv.org/abs/2104.07715v1).
- [555] C. Chen, Z. He, L. Li, S. Zheng, H. Situ, Quantum architecture search with meta-learning, 2021, [arXiv:2106.06248v1](https://arxiv.org/abs/2106.06248v1).
- [556] H. Wang, Y. Ding, J. Gu, Y. Lin, D.Z. Pan, F.T. Chong, S. Han, QuantumNAS: Noise-adaptive search for robust quantum circuits, 2021, [arXiv:2107.10845v2](https://arxiv.org/abs/2107.10845v2).
- [557] M.C. Gutzwiller, Effect of correlation on the ferromagnetism of transition metals, *Phys. Rev. Lett.* 10 (5) (1963) 159–162, <http://dx.doi.org/10.1103/physrevlett.10.159>.
- [558] J. Hubbard, Electron correlations in narrow energy bands, *Proc. R. Soc. A* 276 (1365) (1963) 238–257, <http://dx.doi.org/10.1098/rspa.1963.0204>.
- [559] K. Seki, Y. Otsuka, S. Yunoki, Gutzwiller wave function on a quantum computer using a discrete Hubbard–Stratonovich transformation, *Phys. Rev. B* 105 (15) (2022) <http://dx.doi.org/10.1103/physrevb.105.155119>.
- [560] L. Gentini, A. Cuccoli, S. Pirandola, P. Verrucchi, L. Banchi, Noise-resilient variational hybrid quantum-classical optimization, *Phys. Rev. A* 102 (5) (2020) 052414, <http://dx.doi.org/10.1103/physreva.102.052414>.
- [561] D.S. França, R. García-Patrón, Limitations of optimization algorithms on noisy quantum devices, *Nat. Phys.* (2021) <http://dx.doi.org/10.1038/s41567-021-01356-3>.
- [562] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, N. Killoran, Evaluating analytic gradients on quantum hardware, *Phys. Rev. A* 99 (3) (2019) 032331, <http://dx.doi.org/10.1103/physreva.99.032331>, [arXiv:1811.11184](https://arxiv.org/abs/1811.11184).
- [563] R.M. Parrish, J.T. Iosue, A. Ozaeta, P.L. McMahon, A Jacobi diagonalization and Anderson acceleration algorithm for variational quantum algorithm parameter optimization, 2019, [arXiv:1904.03206v1](https://arxiv.org/abs/1904.03206v1).
- [564] S. Cao, L. Wossnig, B. Vlastakis, P. Leek, E. Grant, Cost-function embedding and dataset encoding for machine learning with parameterized quantum circuits, *Phys. Rev. A* 101 (2020) 052309, <http://dx.doi.org/10.1103/PhysRevA.101.052309>.

- [565] J. Ebel, T. Joas, M. Schalk, P. Weinbrenner, A. Angerer, J. Majer, F. Reinhard, Dispersive readout of room-temperature ensemble spin sensors, *Quantum Sci. Technol.* 6 (3) (2021) 03LT01, <http://dx.doi.org/10.1088/2058-9565/abfaaf>.
- [566] D. Lu, A. Brodutch, J. Park, H. Katiyar, T. Jochym-O'Connor, R. Laflamme, NMR quantum information processing, 2015, [arXiv:1501.01353v1](https://arxiv.org/abs/1501.01353v1).
- [567] J. Kiefer, J. Wolfowitz, Stochastic estimation of the maximum of a regression function, *Ann. Math. Stat.* 23 (3) (1952) 462–466, <http://dx.doi.org/10.1214/aoms/1177729392>, Publisher: Institute of Mathematical Statistics.
- [568] S. Finck, H.-G. Beyer, Performance analysis of the simultaneous perturbation stochastic approximation algorithm on the noisy sphere model, *Theoret. Comput. Sci.* 419 (C) (2012) 50–72, <http://dx.doi.org/10.1016/j.tcs.2011.11.015>, Publisher: Theor Comput Sci.
- [569] T. Durrani, G. Morison, SPSSA for noisy non-stationary blind source separation, in: 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). Vol. 5, 2003, pp. V–285, <http://dx.doi.org/10.1109/ICASSP.2003.1199925>.
- [570] L. Wang, J. Zhu, J.C. Spall, Mixed simultaneous perturbation stochastic approximation for gradient-free optimization with noisy measurements, in: 2018 Annual American Control Conference (ACC), IEEE, 2018, pp. 3774–3779, <http://dx.doi.org/10.23919/acc.2018.8430974>, ISSN: 2378-5861.
- [571] J. Spall, Accelerated second-order stochastic optimization using only function measurements, in: Proceedings of the 36th IEEE Conference on Decision and Control, Vol. 2, 1997, pp. 1417–1424, <http://dx.doi.org/10.1109/CDC.1997.657661>.
- [572] K. Mitarai, K. Fujii, Methodology for replacing indirect measurements with direct measurements, *Phys. Rev. Res.* 1 (1) (2019) 013006, <http://dx.doi.org/10.1103/physrevresearch.1.013006>.
- [573] T. Hubregtsen, F. Wilde, S. Qasim, J. Eisert, Single-component gradient rules for variational quantum algorithms, *Quantum Sci. Technol.* 7 (3) (2022) 035008, <http://dx.doi.org/10.1088/2058-9565/ac6824>.
- [574] D. Wierichs, J. Izaac, C. Wang, C.Y.-Y. Lin, General parameter-shift rules for quantum gradients, *Quantum* 6 (2022) 677, <http://dx.doi.org/10.22331/q-2022-03-30-677>.
- [575] D.O. Theis, Optimality of finite-support parameter shift rules for derivatives of variational quantum circuits, 2021, <http://dx.doi.org/10.48550/ARXIV.2112.14669>, URL: <https://arxiv.org/abs/2112.14669>.
- [576] J.S. Kottmann, A. Anand, A. Aspuru-Guzik, A feasible approach for automatically differentiable unitary coupled-cluster on quantum computers, *Chem. Sci.* 12 (10) (2021) 3497–3508, <http://dx.doi.org/10.1039/d0sc06627c>.
- [577] C. Lemaréchal, Cauchy and the gradient method, 2012, pp. 251–254, URL: https://www.math.uni-bielefeld.de/documenta/vol-ismf/40_lemarechal-claude.html.
- [578] R. Courant, Variational methods for the solution of problems of equilibrium and vibrations, *Bull. Amer. Math. Soc.* 49 (1) (1943) 1–23, [bams/1183504922](https://doi.org/10.2307/2300492).
- [579] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2014, [arXiv:1412.6980v9](https://arxiv.org/abs/1412.6980v9).
- [580] R.H. Byrd, P. Lu, J. Nocedal, C. Zhu, A limited memory algorithm for bound constrained optimization, *SIAM J. Sci. Comput.* 16 (5) (1995) 1190–1208, <http://dx.doi.org/10.1137/0916069>.
- [581] P. Huembeli, A. Dauphin, Characterizing the loss landscape of variational quantum circuits, *Quantum Sci. Technol.* 6 (2) (2021) 025011, <http://dx.doi.org/10.1088/2058-9565/abdb9c>.
- [582] S.-i. Amari, Natural gradient works efficiently in learning, *Neural Comput.* 10 (2) (1998) 251–276, <http://dx.doi.org/10.1162/089976698300017746>, Conference Name: Neural Computation.
- [583] J. Martens, New insights and perspectives on the natural gradient method, 2014, [arXiv:1412.1193v11](https://arxiv.org/abs/1412.1193v11).
- [584] D. Wierichs, C. Gogolin, M. Kastoryano, Avoiding local minima in variational quantum eigensolvers with the natural gradient optimizer, *Phys. Rev. Res.* 2 (4) (2020) 043246, <http://dx.doi.org/10.1103/physrevresearch.2.043246>.
- [585] R. Wiersema, N. Killoran, Optimizing quantum circuits with Riemannian gradient flow, 2022, <http://dx.doi.org/10.48550/ARXIV.2202.06976>, URL: <https://arxiv.org/abs/2202.06976>.
- [586] S. Kullback, R.A. Leibler, On information and sufficiency, *Ann. Math. Stat.* 22 (1) (1951) 79–86, <http://dx.doi.org/10.1214/aoms/1177729694>.
- [587] D. Šafránek, Simple expression for the quantum Fisher information matrix, *Phys. Rev. A* 97 (4) (2018) 042322, <http://dx.doi.org/10.1103/physreva.97.042322>, [arXiv: 1801.00945](https://arxiv.org/abs/1801.00945).
- [588] J. Liu, H. Yuan, X.-M. Lu, X. Wang, Quantum Fisher information matrix and multiparameter estimation, *J. Phys. A* 53 (2) (2019) 023001, <http://dx.doi.org/10.1088/1751-8121/ab5d4d>, [arXiv:1907.08037](https://arxiv.org/abs/1907.08037).
- [589] N. Yamamoto, On the natural gradient for variational quantum eigensolver, 2019, [arXiv:1909.05074v1](https://arxiv.org/abs/1909.05074v1).
- [590] B. van Straaten, B. Koczor, Measurement cost of metric-aware variational quantum algorithms, *PRX Quantum* 2 (3) (2021) <http://dx.doi.org/10.1103/prxquantum.2.030324>, [arXiv:2005.05172v2](https://arxiv.org/abs/2005.05172v2); *PRX Quantum* 2 (2021) 030324.
- [591] J. Gacon, C. Zoufal, G. Carleo, S. Woerner, Simultaneous perturbation stochastic approximation of the quantum Fisher information, *Quantum* 5 (2021) 567, <http://dx.doi.org/10.22331/q-2021-10-20-567>.
- [592] J.A. Nelder, R. Mead, A simplex method for function minimization, *Comput. J.* 7 (4) (1965) 308–313, <http://dx.doi.org/10.1093/comjnl/7.4.308>.
- [593] W. Lavrijsen, A. Tudor, J. Müller, C. Iancu, W. de Jong, Classical optimizers for noisy intermediate-scale quantum devices, in: 2020 IEEE International Conference on Quantum Computing and Engineering, QCE, 2020, pp. 267–277, <http://dx.doi.org/10.1109/QCE49297.2020.00041>.
- [594] M.J.D. Powell, An efficient method for finding the minimum of a function of several variables without calculating derivatives, *Comput. J.* 7 (2) (1964) 155–162, <http://dx.doi.org/10.1093/comjnl/7.2.155>.
- [595] J.G. Vidal, D.O. Theis, Calculus on parameterized quantum circuits, 2018, <http://dx.doi.org/10.48550/ARXIV.1812.06323>, URL: <https://arxiv.org/abs/1812.06323>.
- [596] D. Wierichs, J. Izaac, C. Wang, C.Y.-Y. Lin, General parameter-shift rules for quantum gradients, *Quantum* 6 (2022) 677, <http://dx.doi.org/10.22331/q-2022-03-30-677>.
- [597] K.J. Sung, J. Yao, M.P. Harrigan, N.C. Rubin, Z. Jiang, L. Lin, R. Babbush, J.R. McClean, Using models to improve optimizers for variational quantum algorithms, *Quantum Sci. Technol.* 5 (4) (2020) 044008, <http://dx.doi.org/10.1088/2058-9565/abb6d9>.
- [598] G.H. Golub, H.A. van der Vorst, Eigenvalue computation in the 20th century, *J. Comput. Appl. Math.* 123 (1–2) (2000) 35–65, [http://dx.doi.org/10.1016/S0377-0427\(00\)00413-1](http://dx.doi.org/10.1016/S0377-0427(00)00413-1).
- [599] D.-B. Zhang, T. Yin, Collective optimization for variational quantum eigensolvers, *Phys. Rev. A* 101 (3) (2020) 032311, <http://dx.doi.org/10.1103/physreva.101.032311>.
- [600] P.K. Barkoutsos, G. Nannicini, A. Robert, I. Tavernelli, S. Woerner, Improving variational quantum optimization using CVaR, *Quantum* 4 (2020) 256, <http://dx.doi.org/10.22331/q-2020-04-20-256>, [arXiv: 1907.04769](https://arxiv.org/abs/1907.04769).
- [601] K. Kuroiwa, Y.O. Nakagawa, Penalty methods for a variational quantum eigensolver, *Phys. Rev. Res.* 3 (2021) 013197, <http://dx.doi.org/10.1103/PhysRevResearch.3.013197>, URL: <https://link.aps.org/doi/10.1103/PhysRevResearch.3.013197>.
- [602] X.W. Xuchen You, Exponentially many local minima in quantum neural networks, 2021, [arXiv:2110.02479v1](https://arxiv.org/abs/2110.02479v1).
- [603] I. Miháliková, M. Friák, M. Pivluska, M. Plesch, M. Saip, M. Šob, Best-practice aspects of quantum-computer calculations: A case study of the hydrogen molecule, *Molecules* 27 (3) (2022) 597, <http://dx.doi.org/10.3390/molecules27030597>.
- [604] X. Bonet-Monroig, H. Wang, D. Vermetten, B. Senjean, C. Moussa, T. Bäck, V. Dunjko, T.E. O'Brien, Performance comparison of optimization methods on variational quantum algorithms, 2021, <http://dx.doi.org/10.48550/ARXIV.2111.13454>, URL: <https://arxiv.org/abs/2111.13454>.
- [605] G. Boyd, B. Koczor, Training variational quantum circuits with CoVar: covariance root finding with classical shadows, 2022, <http://dx.doi.org/10.48550/ARXIV.2204.08494>, URL: <https://arxiv.org/abs/2204.08494>.

- [606] J. Mueller, W. Lavrijsen, C. Iancu, W. de Jong, Accelerating noisy VQE optimization with Gaussian processes, 2022, <http://dx.doi.org/10.48550/ARXIV.2204.07331>, URL: <https://arxiv.org/abs/2204.07331>.
- [607] J.A. Gidi, B. Candia, A.D. Muñoz-Moller, A. Rojas, L. Pereira, M. Muñoz, L. Zambrano, A. Delgado, Stochastic optimization algorithms for quantum applications, 2022, <http://dx.doi.org/10.48550/ARXIV.2203.06044>, URL: <https://arxiv.org/abs/2203.06044>.
- [608] T.L. Patti, O. Shehab, K. Najafi, S.F. Yelin, Markov chain Monte-Carlo enhanced variational quantum algorithms, 2021, <http://dx.doi.org/10.48550/ARXIV.2112.02190>, URL: <https://arxiv.org/abs/2112.02190>.
- [609] C.K. Andersen, A. Remm, S. Lazar, S. Krinner, N. Lacroix, G.J. Norris, M. Gabureac, C. Eichler, A. Wallraff, Repeated quantum error detection in a surface code, *Nat. Phys.* 16 (8) (2020) 875–880, <http://dx.doi.org/10.1038/s41567-020-0920-y>.
- [610] R. Barends, J. Kelly, A. Megrant, A. Veitia, D. Sank, E. Jeffrey, T.C. White, J. Mutus, A.G. Fowler, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, C. Neill, P. O'Malley, P. Roushan, A. Vainsencher, J. Wenner, A.N. Korotkov, A.N. Cleland, J.M. Martinis, Superconducting quantum circuits at the surface code threshold for fault tolerance, *Nature* 508 (7497) (2014) 500–503, <http://dx.doi.org/10.1038/nature13171>.
- [611] P. Campagne-Ibarcq, A. Eickbusch, S. Touzard, E. Zalusky-Geller, N.E. Frattini, V.V. Sivak, P. Reinhold, S. Puri, S. Shankar, R.J. Schoelkopf, L. Frunzio, M. Mirrahimi, M.H. Devoret, Quantum error correction of a qubit encoded in grid states of an oscillator, *Nature* 584 (7821) (2020) 368–372, <http://dx.doi.org/10.1038/s41586-020-2603-3>.
- [612] D. Nigg, M. Müller, E.A. Martinez, P. Schindler, M. Hennrich, T. Monz, M.A. Martin-Delgado, R. Blatt, Quantum computations on a topologically encoded qubit, *Science* 345 (6194) (2014) 302–305, <http://dx.doi.org/10.1126/science.1253742>.
- [613] N. Ofek, A. Petrenko, R. Heeres, P. Reinhold, Z. Leghtas, B. Vlastakis, Y. Liu, L. Frunzio, S.M. Girvin, L. Jiang, M. Mirrahimi, M.H. Devoret, R.J. Schoelkopf, Extending the lifetime of a quantum bit with error correction in superconducting circuits, *Nature* 536 (7617) (2016) 441–445, <http://dx.doi.org/10.1038/nature18949>.
- [614] G. Waldherr, Y. Wang, S. Zaiser, M. Jamali, T. Schulte-Herbrüggen, H. Abe, T. Ohshima, J. Isoya, J.F. Du, P. Neumann, J. Wrachtrup, Quantum error correction in a solid-state hybrid spin register, *Nature* 506 (7487) (2014) 204–207, <http://dx.doi.org/10.1038/nature12919>.
- [615] S. Krinner, N. Lacroix, A. Remm, A. Di Paolo, E. Genois, C. Leroux, C. Hellings, S. Lazar, F. Swiadek, J. Herrmann, G.J. Norris, C.K. Andersen, M. Müller, A. Blais, C. Eichler, A. Wallraff, Demonstration of fault-tolerant universal quantum gate operations, *Nature* 605 (7911) (2022) 675–680, <http://dx.doi.org/10.1038/s41586-022-04721-1>.
- [616] S. Krinner, N. Lacroix, A. Remm, A.D. Paolo, E. Genois, C. Leroux, C. Hellings, S. Lazar, F. Swiadek, J. Herrmann, G.J. Norris, C.K. Andersen, M. Müller, A. Blais, C. Eichler, A. Wallraff, Realizing repeated quantum error correction in a distance-three surface code, *Nature* 605 (7911) (2022) 669–674, <http://dx.doi.org/10.1038/s41586-022-04566-8>, URL: <https://www.nature.com/articles/s41586-022-04566-8>.
- [617] S. Endo, Z. Cai, S.C. Benjamin, X. Yuan, Hybrid quantum-classical algorithms and quantum error mitigation, *J. Phys. Soc. Japan* 90 (3) (2021) 032001, <http://dx.doi.org/10.7566/jpsj.90.032001>.
- [618] Z. Cai, Multi-exponential error extrapolation and combining error mitigation techniques for NISQ applications, *Npj Quantum Inf.* 7 (1) (2021) <http://dx.doi.org/10.1038/s41534-021-00404-3>, arXiv:2007.01265v2; *npj Quantum Inf.* 7 (2021) 80.
- [619] A. Kandala, K. Temme, A.D. Córcoles, A. Mezzacapo, J.M. Chow, J.M. Gambetta, Error mitigation extends the computational reach of a noisy quantum processor, *Nature* 567 (7749) (2019) 491–495, <http://dx.doi.org/10.1038/s41586-019-1040-7>.
- [620] D.C. McKay, T. Alexander, L. Bello, M.J. Biercuk, L. Bishop, J. Chen, J.M. Chow, A.D. Córcoles, D. Egger, S. Filipp, J. Gomez, M. Hush, A. Javadi-Abhari, D. Moreda, P. Nation, B. Paulovicks, E. Winston, C.J. Wood, J. Wootton, J.M. Gambetta, Qiskit backend specifications for OpenQASM and OpenPulse experiments, 2018, arXiv:1809.03452v1.
- [621] J.W.O. Garmon, R.C. Pooser, E.F. Dumitrescu, Benchmarking noise extrapolation with the OpenPulse control framework, *Phys. Rev. A* 101 (4) (2020) 042308, <http://dx.doi.org/10.1103/physreva.101.042308>.
- [622] E.F. Dumitrescu, A.J. McCaskey, G. Hagen, G.R. Jansen, T.D. Morris, T. Papenbrock, R.C. Pooser, D.J. Dean, P. Lougovski, Cloud quantum computing of an atomic nucleus, *Phys. Rev. Lett.* 120 (21) (2018) 210501, <http://dx.doi.org/10.1103/physrevlett.120.210501>.
- [623] A. He, B. Nachman, W.A. de Jong, C.W. Bauer, Zero-noise extrapolation for quantum-gate error mitigation with identity insertions, *Phys. Rev. A* 102 (1) (2020) 012426, <http://dx.doi.org/10.1103/physreva.102.012426>, arXiv:2003.04941.
- [624] T. Giurgica-Tiron, Y. Hindy, R. LaRose, A. Mari, W.J. Zeng, Digital zero noise extrapolation for quantum error mitigation, in: 2020 IEEE International Conference on Quantum Computing and Engineering (QCE), IEEE, Denver, CO, USA, ISBN: 978-1-72818-969-7, 2020, pp. 306–316, <http://dx.doi.org/10.1109/qce49297.2020.00045>.
- [625] M. Otten, S.K. Gray, Recovering noise-free quantum observables, *Phys. Rev. A* 99 (1) (2019) 012338, <http://dx.doi.org/10.1103/physreva.99.012338>.
- [626] I. Goodfellow, *Deep Learning*, The MIT Press, Cambridge, Massachusetts, ISBN: 0262035618, 2016.
- [627] T. Hastie, R. Tibshirani, M. Wainwright, *Statistical Learning with Sparsity: The Lasso and Generalizations*, in: Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Taylor & Francis, ISBN: 9781498712163, 2015, URL: <https://books.google.co.uk/books?id=LnUlrGEACAAJ>.
- [628] S. Endo, Hybrid Quantum-Classical Algorithms and Error Mitigation (Ph.D. thesis), University of Oxford, 2019, URL: <https://ora.ox.ac.uk/objects/uuid:6733c0f6-1b19-4d12-a899-18946aa5df85>.
- [629] S. Endo, S.C. Benjamin, Y. Li, Practical quantum error mitigation for near-future applications, *Phys. Rev. X* 8 (3) (2018) <http://dx.doi.org/10.1103/physrevx.8.031027>.
- [630] A. Strikis, D. Qin, Y. Chen, S.C. Benjamin, Y. Li, Learning-based quantum error mitigation, 2020, arXiv:2005.07601v2.
- [631] R. Takagi, Optimal resource cost for error mitigation, *Phys. Rev. Res.* 3 (2021) 033178, <http://dx.doi.org/10.1103/PhysRevResearch.3.033178>.
- [632] B. Koczor, Exponential error suppression for near-term quantum devices, *Phys. Rev. X* 11 (3) (2021) 031057, <http://dx.doi.org/10.1103/physrevx.11.031057>.
- [633] W.J. Huggins, S. McArdle, T.E. O'Brien, J. Lee, N.C. Rubin, S. Boixo, K.B. Whaley, R. Babbush, J.R. McClean, Virtual distillation for quantum error mitigation, 2020, arXiv:2011.07064v3.
- [634] T.E. O'Brien, S. Polla, N.C. Rubin, W.J. Huggins, S. McArdle, S. Boixo, J.R. McClean, R. Babbush, Error mitigation via verified phase estimation, *PRX Quantum* 2 (2) (2021) 020317, <http://dx.doi.org/10.1103/prxquantum.2.020317>.
- [635] M. Huo, Y. Li, Dual-state purification for practical quantum error mitigation, 2021, arXiv:2105.01239v1.
- [636] Z. Cai, Resource-efficient purification-based quantum error mitigation, 2021, arXiv:2107.07279v1.
- [637] B. Koczor, The dominant eigenvector of a noisy quantum state, 2021, arXiv:2104.00608v2.
- [638] M.R. Geller, Rigorous measurement error correction, *Quantum Sci. Technol.* 5 (3) (2020) 03LT01, <http://dx.doi.org/10.1088/2058-9565/ab9591>.
- [639] IBM quantum, 2021, <https://quantum-computing.ibm.com/>. Accessed: 2021-10-20.
- [640] M.R. Geller, M. Sun, Toward efficient correction of multiqubit measurement errors: pair correlation method, *Quantum Sci. Technol.* 6 (2) (2021) 025009, <http://dx.doi.org/10.1088/2058-9565/abd5c9>, arXiv:2001.09980v2; *Quantum Sci. Technol.* 6 (2021) 025009.
- [641] J. Gambetta, M. Treinish, P. Nation, P. Kassebaum, D.M. Rodríguez, Qiskit-Bot, S. De La Puente González, Shaohan Hu, K. Krsulich, L. Zdanski, J. Yu, D. McKay, J. Gomez, L. Capelluto, Travis-S-IBM, J. Gacon, Lerongil, Rafey Iqbal Rahman, S. Wood, Drew, J. Schwarm, L. Bello, MELVIN GEORGE, M. Marques, O.C. Hamido, RohitMidha23, S. Dague, S. Garion, Tigerjack, Y. Kobayashi, Qiskit/qiskit: Qiskit 0.25.0, 2021, <http://dx.doi.org/10.5281/ZENODO.4660156>.
- [642] K. Yeter-Aydeniz, R.C. Pooser, G. Siopsis, Practical quantum computation of chemical and nuclear energy levels using quantum imaginary time evolution and Lanczos algorithms, *Npj Quantum Inf.* 6 (1) (2020) 63, <http://dx.doi.org/10.1038/s41534-020-00290-1>.

- [643] K. Yeter-Aydeniz, E.F. Dumitrescu, A.J. McCaskey, R.S. Bennink, R.C. Pooser, G. Siopsis, Scalar quantum field theories as a benchmark for near-term quantum computers, *Phys. Rev. A* 99 (3) (2019) 032306, <http://dx.doi.org/10.1103/physreva.99.032306>.
- [644] B. Nachman, M. Urbanek, W.A. de Jong, C.W. Bauer, Unfolding quantum computer readout noise, *Npj Quantum Inf.* 6 (1) (2020) 84, <http://dx.doi.org/10.1038/s41534-020-00309-7>.
- [645] S.S. Tannu, M.K. Qureshi, Mitigating measurement errors in quantum computers by exploiting state-dependent bias, in: *Proceedings of the 52nd Annual IEEE/ACM International Symposium on Microarchitecture*, ACM, Columbus OH USA, ISBN: 978-1-4503-6938-1, 2019, pp. 279–290, <http://dx.doi.org/10.1145/3352460.3358265>.
- [646] A. Mari, N. Shammah, W.J. Zeng, Extending quantum probabilistic error cancellation by noise scaling, 2021, [arXiv:2108.02237v1](https://arxiv.org/abs/2108.02237v1).
- [647] M. Urbanek, B. Nachman, W.A. de Jong, Error detection on quantum computers improving the accuracy of chemical calculations, *Phys. Rev. A* 102 (2) (2020) <http://dx.doi.org/10.1103/physreva.102.022427>.
- [648] R. Mezher, J. Mills, E. Kashefi, Mitigating errors by quantum verification and post-selection, 2021, [arXiv:2109.14329v1](https://arxiv.org/abs/2109.14329v1).
- [649] L.A. Truflandier, R.M. Dianzinga, D.R. Bowler, Communication: Generalized canonical purification for density matrix minimization, *J. Chem. Phys.* 144 (9) (2016) 091102, <http://dx.doi.org/10.1063/1.4943213>.
- [650] J. Vovrosh, K.E. Khosla, S. Greenaway, C. Self, M.S. Kim, J. Knolle, Simple mitigation of global depolarizing errors in quantum simulations, *Phys. Rev. E* 104 (3) (2021) <http://dx.doi.org/10.1103/physreve.104.035309>.
- [651] P. Czarnik, A. Arrasmith, P.J. Coles, L. Cincio, Error mitigation with clifford quantum-circuit data, 2020, [arXiv:2005.10189v2](https://arxiv.org/abs/2005.10189v2).
- [652] A. Montanaro, S. Stanisic, Error mitigation by training with fermionic linear optics, 2021, [arXiv:2102.02120v1](https://arxiv.org/abs/2102.02120v1).
- [653] A. Lowe, M.H. Gordon, P. Czarnik, A. Arrasmith, P.J. Coles, L. Cincio, Unified approach to data-driven quantum error mitigation, *Phys. Rev. Res.* 3 (3) (2021) <http://dx.doi.org/10.1103/physrevresearch.3.033098>, [arXiv:2011.01157v2](https://arxiv.org/abs/2011.01157v2); *Phys. Rev. Res.* 3 (2021) 033098.
- [654] C. Kim, K.D. Park, J.-K. Rhee, Quantum error mitigation with artificial neural network, *IEEE Access* 8 (2020) 188853–188860, <http://dx.doi.org/10.1109/access.2020.3031607>.
- [655] A. Zlokap, A. Gheorghiu, A deep learning model for noise prediction on near-term quantum devices, 2020, [arXiv:2005.10811v1](https://arxiv.org/abs/2005.10811v1).
- [656] I.H. Kim, Noise-resilient preparation of quantum many-body ground states, 2017, [arXiv:1703.00032](https://arxiv.org/abs/1703.00032).
- [657] L. Cincio, K. Rudinger, M. Sarovar, P.J. Coles, Machine learning of noise-resilient quantum circuits, *PRX Quantum* 2 (1) (2021) 010324, <http://dx.doi.org/10.1103/prxquantum.2.010324>.
- [658] G. Wang, D.E. Koh, P.D. Johnson, Y. Cao, Minimizing estimation runtime on noisy quantum computers, *PRX Quantum* 2 (2021) 010346, <http://dx.doi.org/10.1103/PRXQuantum.2.010346>.
- [659] A. Katabarwa, A. Kunitsa, B. Peropadre, P. Johnson, Reducing runtime and error in VQE using deeper and noisier quantum circuits, 2021, [arXiv:2110.10664](https://arxiv.org/abs/2110.10664).
- [660] K. Ito, W. Mizukami, K. Fujii, Universal noise-precision relations in variational quantum algorithms, 2021, [arXiv:2106.03390](https://arxiv.org/abs/2106.03390).
- [661] C. Piveteau, D. Sutter, S. Woerner, Quasiprobability decompositions with reduced sampling overhead, 2021, [arXiv:2101.09290v1](https://arxiv.org/abs/2101.09290v1).
- [662] Z. Cai, A practical framework for quantum error mitigation, 2021, [arXiv:2110.05389](https://arxiv.org/abs/2110.05389).
- [663] X. Yuan, S. Endo, Q. Zhao, Y. Li, S.C. Benjamin, Theory of variational quantum simulation, *Quantum* 3 (2019) 191, <http://dx.doi.org/10.22331/q-2019-10-07-191>, [arXiv:1812.08767](https://arxiv.org/abs/1812.08767).
- [664] A.H. Karamlou, W.A. Simon, A. Katabarwa, T.L. Scholten, B. Peropadre, Y. Cao, Analyzing the performance of variational quantum factoring on a superconducting quantum processor, 2020, [arXiv:2012.07825v2](https://arxiv.org/abs/2012.07825v2).
- [665] K.M.R. Karim, H.R. Ong, H. Abdullah, A. Yousuf, C.K. Cheng, M.M.R. Khan, Photoelectrochemical reduction of carbon dioxide to methanol on p-type CuFe2O4 under visible light irradiation, *Int. J. Hydrogen Energy* 43 (39) (2018) 18185–18193, <http://dx.doi.org/10.1016/j.ijhydene.2018.07.174>.
- [666] C.D. Zeinalipour-Yazdi, J.S.J. Hargreaves, C.R.A. Catlow, Low-t mechanisms of ammonia synthesis on Co3Mo3N, *J. Phys. Chem. C* 122 (11) (2018) 6078–6082, <http://dx.doi.org/10.1021/acs.jpcc.7b12364>.
- [667] S. Matsika, A.I. Krylov, Introduction: Theoretical modeling of excited state processes, *Chem. Rev.* 118 (15) (2018) 6925–6926, <http://dx.doi.org/10.1021/acs.chemrev.8b00436>.
- [668] J.F. Stanton, R.J. Bartlett, The equation of motion coupled-cluster method. a systematic biorthogonal approach to molecular excitation energies, transition probabilities, and excited state properties, *J. Chem. Phys.* 98 (9) (1993) 7029–7039, <http://dx.doi.org/10.1063/1.464746>.
- [669] H.J. Monkhorst, Calculation of properties with the coupled-cluster method, *Int. J. Quantum Chem.* 12 (S11) (2009) 421–432, <http://dx.doi.org/10.1002/qua.560120850>.
- [670] B. Jeziorski, H.J. Monkhorst, Coupled-cluster method for multideterminantal reference states, *Phys. Rev. A* 24 (4) (1981) 1668–1681, <http://dx.doi.org/10.1103/physreva.24.1668>.
- [671] S.-N. Sun, M. Motta, R.N. Tazhigulov, A.T. Tan, G.K.-L. Chan, A.J. Minnich, Quantum computation of finite-temperature static and dynamical properties of spin systems using quantum imaginary time evolution, *PRX Quantum* 2 (1) (2021) 010317, <http://dx.doi.org/10.1103/prxquantum.2.010317>.
- [672] J.R. McClean, M.E. Kimchi-Schwartz, J. Carter, W.A. de Jong, Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states, *Phys. Rev. A* 95 (4) (2017) 042308, <http://dx.doi.org/10.1103/physreva.95.042308>.
- [673] P.J. Ollitrault, A. Kandala, C.-F. Chen, P.K. Barkoutsos, A. Mezzacapo, M. Pistoia, S. Sheldon, S. Woerner, J.M. Gambetta, I. Tavernelli, Quantum equation of motion for computing molecular excitation energies on a noisy quantum processor, *Phys. Rev. Res.* 2 (4) (2020) <http://dx.doi.org/10.1103/physrevresearch.2.043140>.
- [674] R.M. Parrish, E.G. Hohenstein, P.L. McMahon, T.J. Martínez, Quantum computation of electronic transitions using a variational quantum eigensolver, *Phys. Lett.* 122 (23) (2019) 230401, <http://dx.doi.org/10.1103/physrevlett.122.230401>.
- [675] J.I. Colless, V.V. Ramasesh, D. Dahlen, M.S. Blok, M.E. Kimchi-Schwartz, J.R. McClean, J. Carter, W.A. de Jong, I. Siddiqi, Computation of molecular spectra on a quantum processor with an error-resilient algorithm, *Phys. Rev. X* 8 (1) (2018) 011021, <http://dx.doi.org/10.1103/physrevx.8.011021>.
- [676] N.S. Blunt, A. Alavi, G.H. Booth, Nonlinear biases, stochastically sampled effective Hamiltonians, and spectral functions in quantum Monte Carlo methods, *Phys. Rev. B* 98 (8) (2018) <http://dx.doi.org/10.1103/physrevb.98.085118>.
- [677] E.N. Epperly, L. Lin, Y. Nakatsukasa, A theory of quantum subspace diagonalization, 2021, [arXiv:2110.07492](https://arxiv.org/abs/2110.07492).
- [678] L.-W. Wang, A. Zunger, Solving Schrödinger's equation around a desired energy: Application to silicon quantum dots, *J. Chem. Phys.* 100 (3) (1994) 2394–2397, <http://dx.doi.org/10.1063/1.466486>.
- [679] S. Liu, S.-X. Zhang, C.-Y. Hsieh, S. Zhang, H. Yao, Probing many-body localization by excited-state VQE, 2021, [http://dx.doi.org/10.48550/ARXIV.2111.13719](https://arxiv.org/abs/2111.13719), URL: <https://arxiv.org/abs/2111.13719>.
- [680] K.M. Nakanishi, K. Mitarai, K. Fujii, Subspace-search variational quantum eigensolver for excited states, *Phys. Rev. Res.* 1 (3) (2019) 1–7, <http://dx.doi.org/10.1103/physrevresearch.1.033062>.
- [681] T. Jones, S. Endo, S. McArdle, X. Yuan, S.C. Benjamin, Variational quantum algorithms for discovering Hamiltonian spectra, *Phys. Rev. A* 99 (6) (2019) <http://dx.doi.org/10.1103/physreva.99.062304>.
- [682] H.H.S. Chan, N. Fitzpatrick, J. Segarra-Martí, M.J. Bearpark, D.P. Tew, Molecular excited state calculations with adaptive wavefunctions on a quantum eigensolver emulation: Reducing circuit depth and separating spin states, 2021, [arXiv:2105.10275](https://arxiv.org/abs/2105.10275).

- [683] H. Wakaura, A.B. Suksmono, Tangent vector variational quantum eigensolver: A robust variational quantum eigensolver against the inaccuracy of derivative, 2021, [arXiv:2105.01141](https://arxiv.org/abs/2105.01141).
- [684] S. Lloyd, C. Weedbrook, Quantum generative adversarial learning, *Phys. Rev. Lett.* 121 (4) (2018) [http://dx.doi.org/10.1103/physrevlett.121.040502](https://doi.org/10.1103/physrevlett.121.040502).
- [685] M. Benedetti, E. Grant, L. Wossnig, S. Severini, Adversarial quantum circuit learning for pure state approximation, *New J. Phys.* 21 (4) (2019) 043023, [http://dx.doi.org/10.1088/1367-2630/ab14b5](https://doi.org/10.1088/1367-2630/ab14b5).
- [686] D.-B. Zhang, Z.-H. Yuan, T. Yin, Variational quantum eigensolvers by variance minimization, 2020, [arXiv:2006.15781](https://arxiv.org/abs/2006.15781).
- [687] J.H. Bartlett, J.J. Gibbons, C.G. Dunn, The normal helium atom, *Phys. Rev.* 47 (9) (1935) 679–680, [http://dx.doi.org/10.1103/physrev.47.679](https://doi.org/10.1103/physrev.47.679).
- [688] F. Siringo, L. Marotta, A variational method from the variance of energy, *Eur. Phys. J. C* 44 (2) (2005) 293–298, [http://dx.doi.org/10.1140/epjc/s2005-02358-x](https://doi.org/10.1140/epjc/s2005-02358-x).
- [689] C.J. Umrigar, C. Filippi, Energy and variance optimization of many-body wave functions, *Phys. Rev. Lett.* 94 (15) (2005) [http://dx.doi.org/10.1103/physrevlett.94.150201](https://doi.org/10.1103/physrevlett.94.150201).
- [690] V. Khemani, F. Pollmann, S.L. Sondhi, Obtaining highly excited eigenstates of many-body localized Hamiltonians by the density matrix renormalization group approach, *Phys. Rev. Lett.* 116 (24) (2016) [http://dx.doi.org/10.1103/physrevlett.116.247204](https://doi.org/10.1103/physrevlett.116.247204).
- [691] F. Pollmann, V. Khemani, J.I. Cirac, S.L. Sondhi, Efficient variational diagonalization of fully many-body localized Hamiltonians, *Phys. Rev. B* 94 (4) (2016) [http://dx.doi.org/10.1103/physrevb.94.041116](https://doi.org/10.1103/physrevb.94.041116).
- [692] F. Vicentini, A. Biella, N. Regnault, C. Ciuti, Variational neural-network ansatz for steady states in open quantum systems, *Phys. Rev. Lett.* 122 (25) (2019) [http://dx.doi.org/10.1103/physrevlett.122.250503](https://doi.org/10.1103/physrevlett.122.250503).
- [693] F. Zhang, N. Gomes, Y. Yao, P.P. Orth, T. Iadecola, Adaptive variational quantum eigensolvers for highly excited states, *Phys. Rev. B* 104 (7) (2021) [http://dx.doi.org/10.1103/physrevb.104.075159](https://doi.org/10.1103/physrevb.104.075159), [arXiv:2104.12636v1](https://arxiv.org/abs/2104.12636v1); *Phys. Rev. B* 104 (2021) 075159.
- [694] S. Endo, I. Kurata, Y.O. Nakagawa, Calculation of the Green's function on near-term quantum computers, *Phys. Rev. Res.* 2 (3) (2020) 033281, [http://dx.doi.org/10.1103/physrevresearch.2.033281](https://doi.org/10.1103/physrevresearch.2.033281).
- [695] I. Rungger, N. Fitzpatrick, H. Chen, C.H. Alderete, H. Apel, A. Cowtan, A. Patterson, D.M. Ramo, Y. Zhu, N.H. Nguyen, E. Grant, S. Chretien, L. Wossnig, N.M. Linke, R. Duncan, Dynamical mean field theory algorithm and experiment on quantum computers, 2019, [arXiv:1910.04735v2](https://arxiv.org/abs/1910.04735v2).
- [696] J. Zhu, Y.O. Nakagawa, C.-F. Li, G.-C. Guo, Y.-S. Zhang, Calculating the Green's function of two-site Fermionic hubbard model in a photonic system, 2021, [arXiv:2104.12361v1](https://arxiv.org/abs/2104.12361v1).
- [697] F. Jamet, A. Agarwal, C. Lupo, D.E. Browne, C. Weber, I. Rungger, Krylov variational quantum algorithm for first principles materials simulations, 2021, [arXiv:2105.13298v2](https://arxiv.org/abs/2105.13298v2).
- [698] F. Jamet, A. Agarwal, I. Rungger, Quantum subspace expansion algorithm for Green's functions, 2022, [arXiv:2205.00094](https://arxiv.org/abs/2205.00094).
- [699] D. Wecker, M.B. Hastings, N. Wiebe, B.K. Clark, C. Nayak, M. Troyer, Solving strongly correlated electron models on a quantum computer, *Phys. Rev. A* 92 (6) (2015) [http://dx.doi.org/10.1103/physreva.92.062318](https://doi.org/10.1103/physreva.92.062318).
- [700] X. Xu, J. Sun, S. Endo, Y. Li, S.C. Benjamin, X. Yuan, Variational algorithms for linear algebra, *Sci. Bull.* 66 (21) (2021) 2181–2188, [http://dx.doi.org/10.1016/j.scib.2021.06.023](https://doi.org/10.1016/j.scib.2021.06.023), [arXiv:1909.03898](https://arxiv.org/abs/1909.03898).
- [701] X. Cai, W.-H. Fang, H. Fan, Z. Li, Quantum computation of molecular response properties, *Phys. Rev. Res.* 2 (3) (2020) 033324, [http://dx.doi.org/10.1103/physrevresearch.2.033324](https://doi.org/10.1103/physrevresearch.2.033324).
- [702] Y. Tong, D. An, N. Wiebe, L. Lin, Fast inversion, preconditioned quantum linear system solvers, fast Green's-function computation, and fast evaluation of matrix functions, *Phys. Rev. A* 104 (3) (2021) [http://dx.doi.org/10.1103/physreva.104.032422](https://doi.org/10.1103/physreva.104.032422).
- [703] B.G. Levine, A.S. Durden, M.P. Esch, F. Liang, Y. Shu, CAS without SCF—Why to use CASCI and where to get the orbitals, *J. Chem. Phys.* 154 (9) (2021) 090902, [http://dx.doi.org/10.1063/5.0042147](https://doi.org/10.1063/5.0042147).
- [704] Q. Sun, J. Yang, G.K.-L. Chan, A general second order complete active space self-consistent-field solver for large-scale systems, *Chem. Phys. Lett.* 683 (2017) 291–299, [http://dx.doi.org/10.1016/j.cplett.2017.03.004](https://doi.org/10.1016/j.cplett.2017.03.004).
- [705] D.R. Mañeru, A.K. Pal, I. de P. R. Moreira, S.N. Datta, F. Illas, The Triplet–Singlet gap in the m-xylylene radical: A not so simple one, *J. Chem. Theory Comput.* 10 (1) (2013) 335–345, [http://dx.doi.org/10.1021/ct400883m](https://doi.org/10.1021/ct400883m).
- [706] Y. Li, L. Wang, X. Guo, J. Zhang, A CASSCF/CASPT2 insight into excited-state intramolecular proton transfer of four imidazole derivatives, *J. Comput. Chem.* 36 (32) (2015) 2374–2380, [http://dx.doi.org/10.1002/jcc.24216](https://doi.org/10.1002/jcc.24216).
- [707] J. Olsen, The CASSCF method: A perspective and commentary, *Int. J. Quantum Chem.* 111 (13) (2011) 3267–3272, [http://dx.doi.org/10.1002/qua.23107](https://doi.org/10.1002/qua.23107).
- [708] T. Takeshita, N.C. Rubin, Z. Jiang, E. Lee, R. Babbush, J.R. McClean, Increasing the representation accuracy of quantum simulations of chemistry without extra quantum resources, *Phys. Rev. X* 10 (1) (2020) [http://dx.doi.org/10.1103/physrevx.10.011004](https://doi.org/10.1103/physrevx.10.011004).
- [709] M. Abe, G. Gopakumar, T. Nakajima, K. Hirao, 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, pp. 157–177, [http://dx.doi.org/10.1007/978-1-4020-8184-2_6](https://doi.org/10.1007/978-1-4020-8184-2_6).
- [710] M. Rossmanek, P.K. Barkoutsos, P.J. Ollitrault, I. Tavernelli, Quantum HF/DFT-embedding algorithms for electronic structure calculations: Scaling up to complex molecular systems, *J. Chem. Phys.* 154 (11) (2021) 114105, [http://dx.doi.org/10.1063/5.0029536](https://doi.org/10.1063/5.0029536).
- [711] R. Schade, C. Bauer, K. Tamoev, L. Mazur, C. Plessl, T.D. Kühne, Parallel quantum chemistry on noisy intermediate-scale quantum computers, 2022, [http://dx.doi.org/10.48550/ARXIV.2202.02417](https://doi.org/10.48550/ARXIV.2202.02417), URL: <https://arxiv.org/abs/2202.02417>.
- [712] G. Knizia, G.K.-L. Chan, Density matrix embedding: A strong-coupling quantum embedding theory, *J. Chem. Theory Comput.* 9 (3) (2013) 1428–1432, [http://dx.doi.org/10.1021/ct301044e](https://doi.org/10.1021/ct301044e).
- [713] S. Wouters, C.A. Jiménez-Hoyos, Q. Sun, G.K.-L. Chan, A practical guide to density matrix embedding theory in quantum chemistry, *J. Chem. Theory Comput.* 12 (6) (2016) 2706–2719, [http://dx.doi.org/10.1021/acs.jctc.6b00316](https://doi.org/10.1021/acs.jctc.6b00316).
- [714] A. Georges, G. Kotliar, W. Krauth, M.J. Rozenberg, Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions, *Rev. Modern Phys.* 68 (1) (1996) 13–125, [http://dx.doi.org/10.1103/revmodphys.68.13](https://doi.org/10.1103/revmodphys.68.13).
- [715] N.C. Rubin, A hybrid classical/quantum approach for large-scale studies of quantum systems with density matrix embedding theory, 2016, [arXiv:1610.06910v2](https://arxiv.org/abs/1610.06910v2).
- [716] T. Yamazaki, S. Matsuura, A. Narimani, A. Saidmuradov, A. Zaribafian, Towards the practical application of near-term quantum computers in quantum chemistry simulations: A problem decomposition approach, 2018, [arXiv:1806.01305v1](https://arxiv.org/abs/1806.01305v1).
- [717] H. Ma, M. Govoni, G. Galli, Quantum simulations of materials on near-term quantum computers, *Npj Comput. Mater.* 6 (1) (2020) [http://dx.doi.org/10.1038/s41524-020-00353-z](https://doi.org/10.1038/s41524-020-00353-z).
- [718] L. Mineev, A. Montanaro, Solving the hubbard model using density matrix embedding theory and the variational quantum eigensolver, *Phys. Rev. B* 105 (12) (2022) [http://dx.doi.org/10.1103/physrevb.105.125117](https://doi.org/10.1103/physrevb.105.125117).
- [719] W. Li, Z. Huang, C. Cao, Y. Huang, Z. Shuai, X. Sun, J. Sun, X. Yuan, D. Lv, Toward practical quantum embedding simulation of realistic chemical systems on near-term quantum computers, 2021, [arXiv:2109.08062v1](https://arxiv.org/abs/2109.08062v1).
- [720] E. Fertitta, G.H. Booth, Rigorous wave function embedding with dynamical fluctuations, *Phys. Rev. B* 98 (23) (2018) [http://dx.doi.org/10.1103/physrevb.98.235132](https://doi.org/10.1103/physrevb.98.235132).

- [721] E. Fertitta, G.H. Booth, Energy-weighted density matrix embedding of open correlated chemical fragments, *J. Chem. Phys.* 151 (1) (2019) 014115, <http://dx.doi.org/10.1063/1.5100290>.
- [722] P.V. Sriluckshmy, M. Nusspickel, E. Fertitta, G.H. Booth, Fully algebraic and self-consistent effective dynamics in a static quantum embedding, *Phys. Rev. B* 103 (8) (2021) <http://dx.doi.org/10.1103/physrevb.103.085131>.
- [723] J.M. Kreula, L. García-Álvarez, L. Lamata, S.R. Clark, E. Solano, D. Jaksch, Few-qubit quantum-classical simulation of strongly correlated lattice fermions, *EPJ Quantum Technol.* 3 (1) (2016) <http://dx.doi.org/10.1140/epjqt/s40507-016-0049-1>.
- [724] S. Wei, H. Li, G. Long, A full quantum eigensolver for quantum chemistry simulations, *Research* 2020 (2020) 1–11, <http://dx.doi.org/10.34133/2020/1486935>.
- [725] N.H. Stair, F.A. Evangelista, Simulating many-body systems with a projective quantum eigensolver, *PRX Quantum* 2 (3) (2021) <http://dx.doi.org/10.1103/prxquantum.2.030301>.
- [726] K. Bharti, T. Haug, Quantum assisted simulator, 2020, [arXiv:2011.06911v1](https://arxiv.org/abs/2011.06911v1).
- [727] T. Haug, K. Bharti, Generalized quantum assisted simulator, *Quantum Sci. Technol.* 7 (4) (2022) 045019, <http://dx.doi.org/10.1088/2058-9565/ac83e7>.
- [728] K. Bharti, T. Haug, Iterative quantum assisted eigensolver, 2020, [arXiv:2010.05638v1](https://arxiv.org/abs/2010.05638v1).
- [729] P. Rivero, I.C. Cloët, Z. Sullivan, An optimal quantum sampling regression algorithm for variational eigensolving in the low qubit number regime, 2020, [arXiv:2012.02338](https://arxiv.org/abs/2012.02338).
- [730] M. Cerezo, K. Sharma, A. Arrasmith, P.J. Coles, Variational quantum state eigensolver, 2020, [arXiv:2004.01372v1](https://arxiv.org/abs/2004.01372v1).
- [731] N.V. Tkachenko, J. Sud, Y. Zhang, S. Tretiak, P.M. Anisimov, A.T. Arrasmith, P.J. Coles, L. Cincio, P.A. Dub, Correlation-informed permutation of qubits for reducing ansatz depth in the variational quantum eigensolver, *PRX Quantum* 2 (2) (2021) <http://dx.doi.org/10.1103/prxquantum.2.020337>.
- [732] D. Amaro, C. Modica, M. Rosenkranz, M. Fiorentini, M. Benedetti, M. Lubasch, Filtering variational quantum algorithms for combinatorial optimization, *Quantum Sci. Technol.* 7 (1) (2022) 015021, <http://dx.doi.org/10.1088/2058-9565/ac3e54>.
- [733] G. Kochenberger, J.-K. Hao, F. Glover, M. Lewis, Z. Lü, H. Wang, Y. Wang, The unconstrained binary quadratic programming problem: a survey, *J. Comb. Optim.* 28 (1) (2014) 58–81, <http://dx.doi.org/10.1007/s10878-014-9734-0>.
- [734] K. Fujii, K. Mizuta, H. Ueda, K. Mitarai, W. Mizukami, Y.O. Nakagawa, Deep Variational Quantum Eigensolver: a divide-and-conquer method for solving a larger problem with smaller size quantum computers, 2020, <http://dx.doi.org/10.48550/ARXIV.2007.10917>, URL: <https://arxiv.org/abs/2007.10917>.
- [735] K. Mizuta, M. Fujii, S. Fujii, K. Ichikawa, Y. Imamura, Y. Okuno, Y.O. Nakagawa, Deep variational quantum eigensolver for excited states and its application to quantum chemistry calculation of periodic materials, *Phys. Rev. Res.* 3 (4) (2021) <http://dx.doi.org/10.1103/physrevresearch.3.043121>.
- [736] O.R. Meitei, B.T. Gard, G.S. Barron, D.P. Pappas, S.E. Economou, E. Barnes, N.J. Mayhall, Gate-free state preparation for fast variational quantum eigensolver simulations, *Npj Quantum Inf.* 7 (1) (2021) <http://dx.doi.org/10.1038/s41534-021-00493-0>.
- [737] A. Asthana, C. Liu, O.R. Meitei, S.E. Economou, E. Barnes, N.J. Mayhall, Minimizing state preparation times in pulse-level variational molecular simulations, 2022, <http://dx.doi.org/10.48550/ARXIV.2203.06818>, URL: <https://arxiv.org/abs/2203.06818>.
- [738] X. Yuan, S. Endo, Q. Zhao, Y. Li, S.C. Benjamin, Theory of variational quantum simulation, *Quantum* 3 (2019) 191, <http://dx.doi.org/10.22331/q-2019-10-07-191>.
- [739] E. Gustafson, P. Dreher, Z. Hang, Y. Meurice, Indexed improvements for real-time trotter evolution of a $(1 + 1)$ field theory using NISQ quantum computers, *Quantum Sci. Technol.* 6 (4) (2021) 045020, <http://dx.doi.org/10.1088/2058-9565/ac1dff>.
- [740] S.-H. Lin, R. Dilip, A.G. Green, A. Smith, F. Pollmann, Real- and imaginary-time evolution with compressed quantum circuits, *PRX Quantum* 2 (1) (2021) <http://dx.doi.org/10.1103/prxquantum.2.010342>.
- [741] R. Mansuroglu, T. Eckstein, L. Nützel, S.A. Wilkinson, M.J. Hartmann, Variational Hamiltonian simulation for translational invariant systems via classical pre-processing, 2021, <http://dx.doi.org/10.48550/ARXIV.2106.03680>, URL: <https://arxiv.org/abs/2106.03680>.
- [742] H. Nishi, T. Kosugi, Y. ichiro Matsushita, Implementation of quantum imaginary-time evolution method on NISQ devices by introducing nonlocal approximation, *Npj Quantum Inf.* 7 (1) (2021) <http://dx.doi.org/10.1038/s41534-021-00409-y>.
- [743] M. Benedetti, M. Fiorentini, M. Lubasch, Hardware-efficient variational quantum algorithms for time evolution, *Phys. Rev. Res.* 3 (3) (2021) <http://dx.doi.org/10.1103/physrevresearch.3.033083>.
- [744] N. Gomes, A. Mukherjee, F. Zhang, T. Iadecola, C.-Z. Wang, K.-M. Ho, P.P. Orth, Y.-X. Yao, Adaptive variational quantum imaginary time evolution approach for ground state preparation, *Adv. Quantum Technol.* 4 (12) (2021) 2100114, <http://dx.doi.org/10.1002/qute.202100114>.
- [745] J.T. Merrill, K.R. Brown, Progress in compensating pulse sequences for quantum computation, 2012, [arXiv:1203.6392v1](https://arxiv.org/abs/1203.6392v1).
- [746] D. Bultrini, M.H. Gordon, E. López, G. Sierra, Simple mitigation strategy for a systematic gate error in IBMQ, 2020, [arXiv:2012.00831v3](https://arxiv.org/abs/2012.00831v3).
- [747] P. Czarnik, A. Arrasmith, L. Cincio, P.J. Coles, Qubit-efficient exponential suppression of errors, 2021, [arXiv:2102.06056v2](https://arxiv.org/abs/2102.06056v2).
- [748] A. Seif, Z.-P. Cui, S. Zhou, S. Chen, L. Jiang, Shadow distillation: quantum error mitigation with classical shadows for near-term quantum processors, 2022, [arXiv:2203.07309](https://arxiv.org/abs/2203.07309).
- [749] H.-Y. Hu, R. LaRose, Y.-Z. You, E. Rieffel, Z. Wang, Logical shadow tomography: efficient estimation of error-mitigated observables, 2022, [arXiv:2203.07263](https://arxiv.org/abs/2203.07263).