

# Variational Quantum Algorithms

M. Cerezo,<sup>1,2,3</sup> Andrew Arrasmith,<sup>1</sup> Ryan Babbush,<sup>4</sup> Simon C. Benjamin,<sup>5</sup> Suguru Endo,<sup>6</sup> Keisuke Fujii,<sup>7,8,9</sup> Jarrod R. McClean,<sup>4</sup> Kosuke Mitarai,<sup>7,10,11</sup> Xiao Yuan,<sup>12,13</sup> Lukasz Cincio,<sup>1,3</sup> and Patrick J. Coles<sup>1,3</sup>

<sup>1</sup>*Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*

<sup>2</sup>*Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM, USA*

<sup>3</sup>*Quantum Science Center, Oak Ridge, TN 37931, USA*

<sup>4</sup>*Google Quantum AI Team, Venice, CA 90291, United States of America*

<sup>5</sup>*Department of Materials, University of Oxford, Parks Road, Oxford OX1 3PH, United Kingdom*

<sup>6</sup>*NTT Secure Platform Laboratories, NTT Corporation, Musashino, Tokyo 180-8585, Japan*

<sup>7</sup>*Graduate School of Engineering Science, Osaka University, Osaka 560-8531, Japan*

<sup>8</sup>*Center for Quantum Information and Quantum Biology,*

*Institute for Open and Transdisciplinary Research Initiatives, Osaka University, Osaka 560-8531, Japan*

<sup>9</sup>*Center for Emergent Matter Science, RIKEN, Saitama 351-0198, Japan*

<sup>10</sup>*Center for Quantum Information and Quantum Biology,*

*Institute for Open and Transdisciplinary Research Initiatives, Osaka 560-8531, Japan*

<sup>11</sup>*JST, PRESTO, Saitama 332-0012, Japan*

<sup>12</sup>*Center on Frontiers of Computing Studies, Department of Computer Science, Peking University, Beijing 100871, China*

<sup>13</sup>*Stanford Institute for Theoretical Physics, Stanford University, Stanford California 94305, USA*

Applications such as simulating large quantum systems or solving large-scale linear algebra problems are immensely challenging for classical computers due to their extremely high computational cost. Quantum computers promise to unlock these applications, although fault-tolerant quantum computers will likely not be available for several years. Currently available quantum devices have serious constraints, including limited qubit numbers and noise processes that limit circuit depth. Variational Quantum Algorithms (VQAs), which employ a classical optimizer to train a parametrized quantum circuit, have emerged as a leading strategy to address these constraints. VQAs have now been proposed for essentially all applications that researchers have envisioned for quantum computers, and they appear to be the best hope for obtaining quantum advantage. Nevertheless, challenges remain including the trainability, accuracy, and efficiency of VQAs. In this review article we present an overview of the field of VQAs. Furthermore, we discuss strategies to overcome their challenges as well as the exciting prospects for using them as a means to obtain quantum advantage.

## I. INTRODUCTION

The field of quantum computing owes its inception to the realization that the classical simulation of quantum systems is a difficult task [1]. The computational resources needed to simulate a generic quantum system grow exponentially with the system size, making the problem quickly intractable even with state-of-the-art algorithms and supercomputers. Exponential scaling acts as a roadblock to applications such as pharmaceutical design based on first principles, discovering new high-temperature superconductors, and simulating high-energy particle collisions.

In the quantum computing field a string of promising theoretical results have motivated the immense, decades-long challenge of realising effective physical hardware. Shor's quantum algorithm was proposed to factor numbers with exponential speedup over classical methods [2], dynamical quantum simulation algorithms were developed [3], and quantum algorithms for solving linear systems of equations were identified [4]. The focus was on the asymptotic scaling of quantum algorithms, with the typical assumption that a large-scale fault-tolerant quantum computer would someday be available. Scaling pre-factors, qubit requirements, and quantum circuit depth were not a significant concern in this paradigm.

In 2016, users were provided access to the first cloud-

based quantum computer [5]. Noise and qubit limitations prevented serious implementations of the aforementioned quantum algorithms [6]. But excitement grew as to what could be done with these new devices, which have been called Noisy Intermediate-Scale Quantum (NISQ) computers [7]. Current state-of-the-art device size ranges from 50 to 100 qubits. This size actually allows one to achieve quantum supremacy, outperforming the best classical supercomputer, for certain contrived mathematical tasks [8, 9].

Nevertheless, the true promise of quantum computers, speedup for practical applications, which is often called quantum advantage, has yet to be realized. Moreover, the availability of fault-tolerant quantum computers appears to still be many years, or even decades, away. The key technological question is therefore how to make the best use of today's NISQ devices to achieve quantum advantage. Any such strategy must account for: (1) Limited numbers of qubits, (2) Limited connectivity of qubits, and (3) Coherent and incoherent errors that limit quantum circuit depth.

Variational Quantum Algorithms (VQAs) have emerged as the leading strategy to obtain quantum advantage on NISQ devices. Accounting for all of the constraints imposed by the NISQ era with a single strategy requires an optimization-based or learning-based approach, and indeed VQAs employ such an approach.

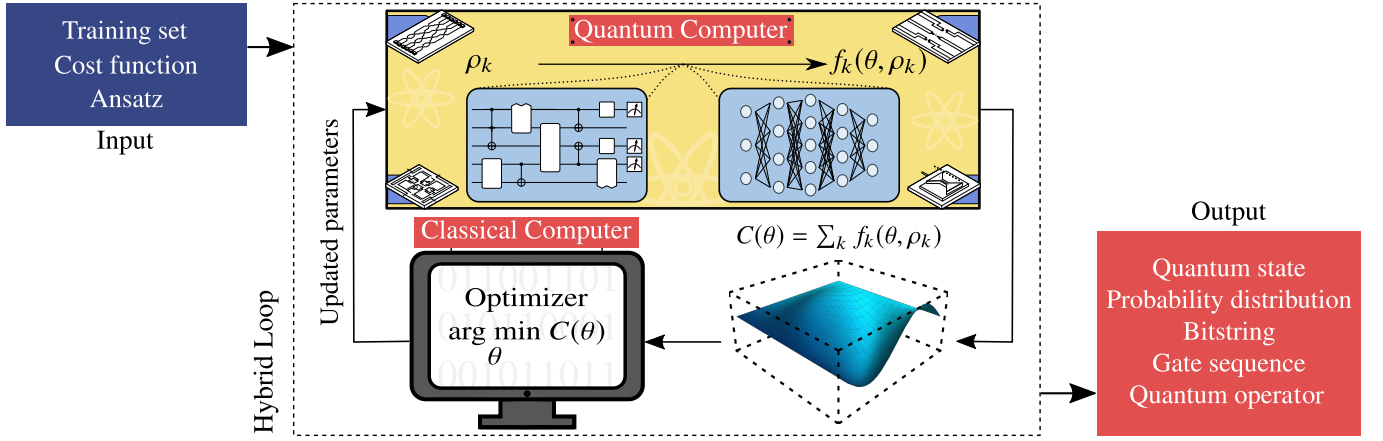


FIG. 1. **Schematic diagram of a Variational Quantum Algorithm (VQA).** The inputs to a VQA are: a cost function  $C(\theta)$  which encodes the solution to the problem, an ansatz whose parameters are trained to minimize the cost, and (possibly) a set of training data used during the optimization. At each iteration of the loop one employs a quantum computer to efficiently estimate the cost (or its gradients). This information is fed into a classical computer that leverages the power of optimizers to navigate the cost landscape and solve the optimization problem in Eq. (1). Once a termination condition is met, the VQA outputs an estimate of the solution to the problem. The form of the output depends on the precise task at hand. In the figure are indicated some of the most common types of output.

VQAs are arguably the natural quantum analog of the highly successful machine-learning methods employed in classical computing, such as neural networks. Moreover, VQAs leverage the toolbox of classical optimization, as VQAs employ parametrized quantum circuits to be run on the quantum computer, and then outsource the parameter optimization to a classical optimizer. This has the added advantage of keeping the quantum circuit depth shallow and hence mitigating noise, in contrast to quantum algorithms developed for the fault-tolerant era.

VQAs have now been proposed for a plethora of applications, covering essentially all of the applications that researchers have envisioned for quantum computers. While they may be the key to obtaining near-term quantum advantage, VQAs still face important challenges, including their trainability, accuracy, and efficiency. In this review, we discuss the exciting prospects for VQAs, and we highlight the challenges that must be overcome to obtain the ultimate goal of quantum advantage.

The structure of this review is as follows. In Section II we introduce the basic framework behind VQAs. Therein we present a general description of cost functions, and we discuss some of the most widely used ansatzes and optimizers. Then, in Section III we present different applications for VQAs ranging from finding ground and excited states, quantum simulation, optimization, and machine learning, among others. Section IV contains a discussion of the main challenges (and their potential solutions) for VQAs. These encompass trainability problems such as barren plateaus, the effect of hardware noise, and error mitigation techniques. In Section V we discuss opportunities where VQAs could be employed for obtaining quantum advantage in the near-term. Such applications include using VQAs to solve problems in quantum chemistry, nuclear and particle physics, and for optimization

and machine learning. Finally, Section VI contains our final discussions and outlooks.

## II. BASIC CONCEPTS AND TOOLS

One of the main advantages of Variational Quantum Algorithms (VQAs) is that they provide a general framework that can be used to solve a wide array of problems. While this versatility translates into different algorithmic structures with varying levels of complexity, there are basic elements that most (if not all) VQAs share in common. In this section we review the building blocks of VQAs in the hope that these can be used as blueprints for the development of novel algorithms.

Let us start by considering a task one wishes to solve. This implies having access to a description of the problem, and also possibly to a set of training data. As schematically shown in Fig. 1, the first step to developing a VQA is to define a cost (or loss) function  $C$  which encodes the solution to the problem. One then proposes an ansatz, i.e., a quantum operation depending on a set of continuous or discrete parameters  $\theta$  that can be optimized (see below for a more in-depth discussion of ansatzes). This ansatz is then trained (with data from the training set) in a hybrid quantum-classical loop to solve the optimization task

$$\theta^* = \arg \min_{\theta} C(\theta). \quad (1)$$

The trademark of VQAs is that they employ a quantum computer to estimate the cost function  $C(\theta)$  (or its gradient) while leveraging the power of classical optimizers to train the parameters  $\theta$ . In what follows, we provide

additional details for each step of the VQA architecture shown in Fig. 1.

### A. Cost function

A crucial aspect of a VQA is encoding the problem into a cost function. Similar to classical machine learning, the cost function maps values of the trainable parameters  $\theta$  to real numbers. More abstractly, the cost defines a hyper-surface usually called the cost landscape (see Fig. 1) such that the task of the optimizer is to navigate through the landscape and find the global minima. Without loss of generality, we consider that the cost can be expressed as

$$C(\theta) = \sum_k f_k (\text{Tr}[O_k U(\theta) \rho_k U^\dagger(\theta)]) . \quad (2)$$

Here  $U(\theta)$  is a parametrized unitary,  $\theta$  is composed of discrete and continuous parameters,  $\{\rho_k\}$  are input states from a training set,  $\{O_k\}$  are a set of observables, and  $f_k$  are functions which encode the task at hand. We remark that for a given problem, there can be different choices of  $f_k$  which lead to useful costs.

Let us now discuss desirable criteria that the cost function should meet. First, the cost must be *faithful* in that the minimum of  $C(\theta)$  corresponds to the solution of the problem. Second, one must be able to *efficiently estimate*  $C(\theta)$  by performing measurements on a quantum computer and possibly performing classical post-processing. An implicit assumption here is that the cost should not be efficiently computable with a classical computer, as this would imply that no quantum advantage can be achieved with the VQA. In addition, it is also useful for  $C(\theta)$  to be *operationally meaningful*, so that smaller cost values indicate a better solution quality. Finally, the cost must be *trainable*, which means that it must be possible to efficiently optimize the parameters  $\theta$ . In Section IV A we discuss in more detail the issue of trainability for VQAs.

For a given VQA to be implementable in near-term quantum hardware, the quantum circuits employed to estimate  $C(\theta)$  must keep the circuit depth and ancilla requirements small. This is due to the fact that NISQ devices are prone to gate errors, have limited qubit counts, and that these qubits have short decoherence times. Hence the construction of efficient cost evaluation circuits is an important aspect of VQA research.

### B. Ansatz

Another important aspect of a VQA is its ansatz. Generically speaking the form of the ansatz dictates what the parameters  $\theta$  are, and hence, how they can be trained to minimize the cost. The specific structure of an ansatz will generally depend on the task at hand, as in many cases one can employ information about the problem to

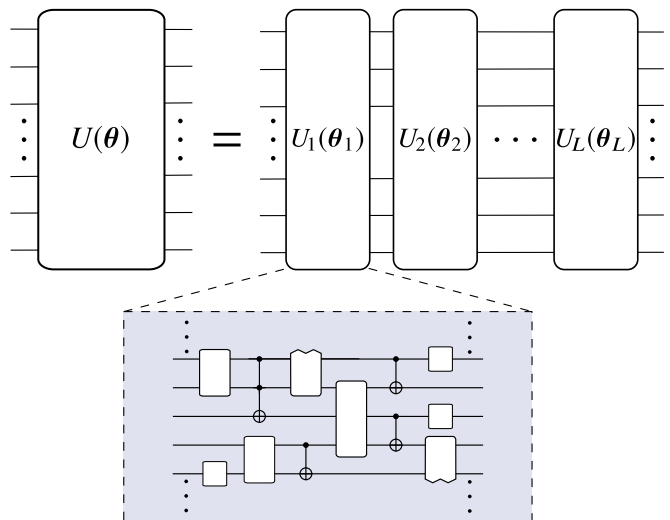


FIG. 2. **Schematic diagram of an ansatz.** The unitary  $U(\theta)$  can be expressed as a product of  $L$  unitaries  $U_l(\theta_l)$  sequentially acting on an input state. As indicated, each unitary  $U_l(\theta_l)$  can in turn be decomposed into a sequence of parametrized and unparametrized gates.

tailor an ansatz. These are the so-called *problem-inspired ansatzes*. On the other hand, some ansatz architectures are generic and *problem-agnostic*, meaning that they can be used even when no relevant information is readily available. For the cost function in (2), the parameters  $\theta$  can be encoded in a unitary  $U(\theta)$  that is applied to the input states to the quantum circuit. As shown in Fig. 2,  $U(\theta)$  can be generically expressed as the product of  $L$  sequentially applied unitaries

$$U(\theta) = U_L(\theta_L) \cdots U_2(\theta_2) U_1(\theta_1), \quad (3)$$

with

$$U_l(\theta_l) = \prod_m e^{-i\theta_m H_m} W_m. \quad (4)$$

Here  $W_m$  is an unparametrized unitary and  $H_m$  is a Hermitian operator. Below we describe some of the most widely used ansatzes in the literature, starting with those that can be expressed as (3), and then presenting more general architectures.

**Hardware efficient ansatz.** The hardware efficient ansatz [10] is a generic name used for ansatzes that are aimed at reducing the circuit depth needed to implement  $U(\theta)$  when using a given quantum hardware. Here one employs unitaries  $W_m$  and  $e^{-i\theta_m H_m}$  that are taken from a gate alphabet determined from interactions specific to a quantum hardware. The latter avoids the circuit depth overhead arising from translating an arbitrary unitary into a sequence of gates easily implementable in a device. One of the main advantages of the hardware efficient ansatz is its versatility, as it can accommodate encoding symmetries [11, 12] and bringing correlated qubits closer for depth reduction [13], as well as being especially

useful to study Hamiltonians that are similar to the device’s interactions [14]. Additionally, “layered” hardware efficient ansatz, where gates act on alternating pairs of qubits in a brick-like structure, have been prominently used as problem-agnostic architectures. However, as discussed in Section IV A, this ansatz can lead to trainability problems when randomly initialized.

*Unitary coupled clustered ansatz.* The Unitary Coupled (UCC) ansatz is a problem-inspired ansatz widely used for quantum chemistry problems where the goal is to obtain the ground state energy of a fermionic molecular Hamiltonian  $H$  (see Section III A for a more detailed description of using VQAs to find ground states). The UCC ansatz proposes a candidate for such ground state based on exciting some reference state  $|\psi_0\rangle$  (usually the Hartree-Fock state of  $H$ ) as  $e^{T(\boldsymbol{\theta})-T(\boldsymbol{\theta})^\dagger}|\psi_0\rangle$ . Here,  $T = \sum_k T_k$  is the cluster operator [15, 16] and  $T_k$  are excitation operators. In the so-called UCCSD ansatz the summation is truncated to contain single excitations  $T_1 = \sum_{i,j} \theta_{ij}^j a_i^\dagger a_j$ , and double excitations  $T_2 = \sum_{i,j,k,l} \theta_{ij,kl}^{k,l} a_i^\dagger a_j^\dagger a_k a_l$ , where  $\{a_i^\dagger\}$  ( $\{a_i\}$ ) are fermionic creation (annihilation) operators. To implement this ansatz in a quantum computer one employs the Jordan-Wigner or the Bravyi-Kitaev transformations to map the fermionic operators to spin operators, resulting in an ansatz of the form (3). We remark that there are many variants of the UCC ansatz [17], with some of them reducing the circuit depth by considering more efficient methods for compiling the fermionic operators [18–21].

*Quantum alternating operator ansatz.* The Quantum Alternating Operator Ansatz (QAOA) was originally introduced to obtain approximate solutions for combinatorial optimization problems [22], and later generalized as a standalone ansatz [23] shown to be computationally universal [24] (see Section III C for a more detailed description of using VQAs for optimization tasks). The QAOA can be understood as a Trotterized adiabatic transformation where the order  $p$  of the Trotterization determines the precision of the solution. The goal of the QAOA is to map an input state  $|\psi_0\rangle$ , to the ground-state of a given problem Hamiltonian  $H_P$  by sequentially applying a problem unitary  $e^{-i\gamma H_P}$  and a mixer unitary  $e^{-i\beta H_M}$  [25]. That is,  $U(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \prod_{l=1}^p e^{-i\beta_l H_M} e^{-i\gamma_l H_P}$ , where  $\boldsymbol{\theta} = (\boldsymbol{\gamma}, \boldsymbol{\beta})$ . Note that this ansatz is naturally of the form (3), although decomposing these unitaries into native gates may result in a lengthy circuit due to many-body terms in  $H_C$  as well as limited device connectivity. One of the strengths of QAOA is the fact that the feasible subspace for certain problems is smaller than the full Hilbert space, and this restriction may result in a better-performing algorithm.

*Variational Hamiltonian ansatz.* Inspired by the QAOA, the variational Hamiltonian ansatz also aims to prepare a trial ground states for a given Hamiltonian  $H = \sum_k H_k$  by Trotterizing an adiabatic state preparation process [26]. Here, each Trotter step corresponds to a variational ansatz so that the unitary is

given by  $U(\boldsymbol{\theta}) = \prod_l (\prod_k e^{-\theta_{l,k} H_k})$ , and again is of the form (3). Due to its versatility, the variational Hamiltonian ansatz has been implemented for quantum chemistry [20, 27], optimization [23], and for quantum simulation problems [28].

*ADAPT-VQE.* In many ansatz one optimizes over continuous parameters (such as rotation angles), while the structure of the circuit is kept fixed. While this allows one to control the overall circuit complexity, it may miss refinements attained by optimizing the circuit structure itself, including the addition or removal of unnecessary circuit elements. This possibility was explored in a framework called ADAPT-VQE [29, 30], which seeks to adaptively add specific elements to the ansatz to maximize the benefit while minimizing the number of circuit elements. One can then view the problem as a sparse model problem, and while such an optimization is known to be hard, heuristic or greedy approximations that seek to add one term at a time have been shown to be helpful [29, 30]. Complementary approaches based on exploring a plurality of ansatz variants simultaneously as an evolving cohort have also shown promising performance [31].

*Sub-logical ansatz and quantum optimal control.* The parameters  $\boldsymbol{\theta}$  are often specified at the logical circuit level (e.g. rotation angle), however sometimes they have a direct translation to device-level parameters below the logical level. Hence, one can make these device-level parameters the definition of the ansatz, as this can lead to additional flexibilities [32]. This approach also establishes a connection to the idea of quantum optimal control, which is often used to determine the translation from logical to physical device parameters, and which is especially applicable for quantum simulations [33, 34]. Refs. [35, 36] have explored using variational quantum algorithms to construct optimal control sequences. While this can increase the number of parameters, the additional flexibility may allow for on-the-fly calibration effects that have been seen to reduce the effects of coherent noise [35–37].

*Hybrid ansatz.* In some cases, one can combine quantum ansatz with classical strategies to push some of the complexity onto a classical device. For instance, in quantum chemistry one can exploit the classical simulability of free fermion dynamics to apply quantum operations via classical post-processing [38–43]. A different approach is to use as ansatz a trainable linear combination of parametrized states  $|\psi(\{c_\mu\}, \boldsymbol{\theta})\rangle = \sum_\mu c_\mu |\psi_\mu(\boldsymbol{\theta}_\mu)\rangle$  with  $\{c_\mu\}$  classically optimizable coefficient [44–50]. Moreover, as quantum circuits can be viewed as tensor networks [51], it is quite natural to combine the existing tensor network techniques with a quantum ansatz [52–54]. An alternative hybrid approach was proposed via the deep variational quantum eigensolver, where the algorithm divides the whole system into small subsystems and sequentially solves each subsystem and the interaction between the subsystems [55]. Finally, there is also a hybrid method that combines variational Monte Carlo techniques with a quantum ansatz to classically apply the so-called Jastrow operator  $e^{(\sum_{i,j} J_{ij} n_i n_j)}$  (for  $J$  a sym-



metric matrix) to a parametrized quantum state  $|\psi(\boldsymbol{\theta})\rangle$  with the goal of obtaining a more accurate result by optimizing together  $J$  and  $\boldsymbol{\theta}$  [56].

*Ansatz for mixed states.* Since mixed states play an important role in many applications, e.g. systems at finite temperature, several ansätze have been developed to construct a mixed state  $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$  of  $n$  qubits. A first approach (which comes at the cost of requiring up to  $2n$  qubits) is based on preparing a pure state that has  $\rho$  as a reduced state in some subsystem of qubits. Refs. [52, 57] have proposed a method to variationally obtain a purification  $|\psi\rangle = \sum_i \sqrt{p_i} |\psi_i\rangle |\phi_i\rangle$  of  $\rho$ , while [58] introduced a method to construct a state  $|\rho\rangle = \frac{1}{c} \sum_i p_i |\psi_i\rangle |\psi_i\rangle$  with normalization  $c = \sum_i p_i^2$ . Alternatively, one can also train a probability distribution  $\{p_i(\boldsymbol{\phi})\}$  and a set of states  $\{|\psi_i(\boldsymbol{\theta}_i)\rangle\}$  to construct  $\rho$  as the statistical ensemble  $\rho(\boldsymbol{\phi}, \{\boldsymbol{\theta}_i\}) = \sum_i p_i(\boldsymbol{\phi}) |\psi_i(\boldsymbol{\theta}_i)\rangle\langle\psi_i(\boldsymbol{\theta}_i)|$ . Ref. [57] proposed to use a simple product distribution based on physical insights, while a more general proposals for energy based models was introduced in [59]. More recently, there has been a proposal to generate mixed states which uses the autoregressive model [60].

*Ansatz expressibility.* Given the wide range of ansätze one can employ, a relevant question is whether a given architecture can prepare a target state by optimizing its parameters. In this sense, there are different ways to judge the quality of an ansatz [61] by considering two different notions: the expressibility and the entangling capability of an ansatz. An ansatz is expressible if the circuit can be used to uniformly explore the entire space of quantum states. Thus one way to quantify the expressibility of an ansatz  $U(\boldsymbol{\theta})$  is to compare the distribution of states obtained from  $U(\boldsymbol{\theta})$  to the maximally expressive uniform (Haar) distribution of states  $U_{\text{Haar}}$ . Motivated by this line of thought, the expressibility of a circuit is measured by [61]  $\|A^{(t)}\|_2$ , where

$$A^{(t)}(U) := \int dU_{\text{Haar}} U_{\text{Haar}}^{\otimes t} |0\rangle\langle 0| (U_{\text{Haar}}^\dagger)^{\otimes t} - \int dU U^{\otimes t} |0\rangle\langle 0| (U^\dagger)^{\otimes t}. \quad (5)$$

Other expressibility measures can be considered as well [61], and the expressibility of different ansätze is investigated further in [62]. We remark that Ref. [61] also introduced a measure of entangling capability for ansätze, which quantifies the average entanglement of states produced from randomly sampling the circuit parameters  $\boldsymbol{\theta}$ .

### C. Gradients

Once the cost function and ansatz have been defined, the next step is to train the parameters  $\boldsymbol{\theta}$  and solve the optimization problem of Eq. (1). It is known that for many optimization tasks using information in the cost function gradient (or in higher-order derivatives) can help in speeding up and guaranteeing the convergence of the

optimizer. One of the main advantages of many VQAs is that, as discussed below, one can analytically evaluate the cost function gradient.

*Parameter-shift rule.* Let us consider for simplicity a cost function of the form (2) with  $f_k(x) = x$ , and let  $\theta_l$  be the  $l$ -th element in  $\boldsymbol{\theta}$  which parametrize a unitary  $e^{i\theta_l \sigma_l}$  in the ansatz. Here,  $\sigma_l$  is a Pauli operator. Surprisingly, there is a hardware-friendly protocol to evaluate the partial derivative of  $C(\boldsymbol{\theta})$  with respect to  $\theta_l$  often referred to as the parameter-shift rule [63–66]. Explicitly, the parameter-shift rule states that the equality

$$\frac{\partial C}{\partial \theta_l} = \sum_k \frac{1}{2 \sin \alpha} \left( \text{Tr} [O_k U^\dagger(\boldsymbol{\theta}_+) \rho_k U(\boldsymbol{\theta}_+)] - \text{Tr} [O_k U^\dagger(\boldsymbol{\theta}_-) \rho_k U(\boldsymbol{\theta}_-)] \right), \quad (6)$$

with  $\boldsymbol{\theta}_\pm = \boldsymbol{\theta} \pm \alpha \mathbf{e}_l$ , holds for any real number  $\alpha$ . Here  $\mathbf{e}_l$  is a vector having 1 as its  $l$ -th element and 0 otherwise. Equation (6) shows that one can evaluate the gradient by shifting the  $l$ -th parameter by some amount  $\alpha$ . Note that the accuracy of the evaluation depends on the coefficient  $1/(2 \sin \alpha)$  since each of the  $\pm\alpha$ -term is evaluated by sampling  $O_k$ . This accuracy is maximized at  $\alpha = \pi/4$ , since  $1/\sin \alpha$  is minimized at this point. While the parameter-shift rule might resemble a naive finite difference, it evaluates the analytic gradient of the parameter by virtue of the coefficient  $1/\sin \alpha$ . A detailed comparison between the parameter-shift rule and the finite difference can be found in Ref. [67]. Finally, we remark that the gradient for more general  $f_k(x)$  can be obtained from (6) by using the chain rule.

*Other derivatives.* Higher-order derivatives of the cost function can be evaluated by straight-forward extensions of the parameter-shift rule. For example, the second derivative for the previous example can be written as

$$\begin{aligned} \frac{\partial^2 C}{\partial \theta_l^2} = \sum_k \frac{1}{4 \sin^2 \alpha} & \left( \text{Tr} [O_k U^\dagger(\boldsymbol{\theta} + 2\alpha \mathbf{e}_l) \rho_k U(\boldsymbol{\theta} + 2\alpha \mathbf{e}_l)] \right. \\ & + \text{Tr} [O_k U^\dagger(\boldsymbol{\theta} - 2\alpha \mathbf{e}_l) \rho_k U(\boldsymbol{\theta} - 2\alpha \mathbf{e}_l)] \\ & \left. - 2 \text{Tr} [O_k U^\dagger(\boldsymbol{\theta}) \rho_k U(\boldsymbol{\theta})] \right), \end{aligned}$$

by applying the parameter-shift rule twice. Other higher-order ones such as  $\frac{\partial^2 C}{\partial \theta_l \partial \theta_{l'}}$  or  $\frac{\partial^3 C}{\partial \theta_l^3}$  can be obtained similarly. Explicit formulas can be found in [67, 68]. These observations relate to the fact that the cost function can be expanded into a trigonometric series that admits a classically efficient, analytical approximation around any reference point. One can thus infer a classical model of the cost function, and minimise it, to offload more work from the quantum processor to the classical supervising system [69, 70].

Other types of derivatives of the parametrized quantum state not directly related to the cost function, such as a metric tensor of a state  $\frac{\partial \langle \psi(\boldsymbol{\theta}) |}{\partial \theta_l} \frac{\partial | \psi(\boldsymbol{\theta}) \rangle}{\partial \theta_{l'}}$  (with  $|\psi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})|\psi_0\rangle$  for some initial state  $|\psi_0\rangle$ ), are sometimes used in sophisticated optimization algorithms [71–73] (see Section IID) and variational quantum simulation [74–76]

(see Section III B). As quantities like  $\frac{\partial \langle \psi(\boldsymbol{\theta}) |}{\partial \theta_l} \frac{\partial | \psi(\boldsymbol{\theta}) \rangle}{\partial \theta_l'}$  are essentially overlaps of different states, this can be evaluated via Hadamard-test like protocols [75]. However, as shown in [77], those can also be reduced to the parameter-shift technique.

#### D. Optimizers

As for any variational approach, the success of a VQA depends on the efficiency and reliability of the optimization method employed. In addition to the typical difficulties encountered in classical optimization, it has been shown that when training a VQA one can encounter novel challenges. These include issues such as the inherently stochastic environment due to the finite budget for measurements, hardware noise, and the presence of barren plateaus (see Section IV). This has led to the development of many quantum-aware optimizers, with the optimal choice still being an active topic of debate. Here we discuss a selection of optimizers that have been designed or promoted for use with VQAs. For convenience, these will be grouped into two categories based on whether or not they implement some version of gradient descent.

*Gradient descent methods.* One of the most common approaches to optimization is to make iterative steps in directions indicated by the gradient. Given that only statistical estimates are available for these gradients, these strategies fall under the umbrella of Stochastic Gradient Descent (SGD). One SGD method which has been imported from the machine learning community is Adam, which adapts the size of the steps taken during the optimization to allow for more efficient and precise solutions than those obtained through basic SGD [78]. An alternative method inspired by the machine learning literature adapts the precision (i.e. the number of shots taken for each estimate), rather than the step size, at each iteration in an attempt to be frugal with the quantum resources used [79].

A different gradient-based approach is based on simulating an imaginary time evolution [71], or equivalently by employing the quantum natural gradient descent method, which is based on notions of information geometry [72, 73]. While standard gradient descent takes steps in the steepest descent direction in the  $l_2$  (Euclidean) geometry of the parameter space, natural gradient descent works instead on a space with a metric tensor that encodes the sensitivity of the quantum state to variations in the parameters. Using this metric tensor, typically accelerates the convergence of the gradient update steps, allowing a given level of precision to be attained with fewer iterations. This method has also been extended to incorporate the effects of noise [73].

*Other methods.* A different method which uses gradients but has a more complicated update step than SGD is meta-learning [80]. In this context, the optimizer “learns to learn” by training a neural network to make a good update step based on the optimization history and current

gradient with similar optimization problems. Because the update steps taken are based on rules learned from similar cost functions, this meta-learning approach has significant potential to be highly efficient when used on a new instance of a common class of optimizations.

Of the optimization methods proposed for use with VQAs which do not directly utilize gradients, the one that is perhaps the most closely related to SGD is the simultaneous perturbation stochastic approximation (SPSA) method [81]. SPSA can be considered as an approximation to gradient descent where the gradient is approximated by a single partial derivative computed by a finite difference along a randomly chosen direction. SPSA has thus been put forward as an efficient method for VQAs as it avoids the expense of computing many gradient components at each iteration. Moreover, it has been shown that for a restricted set of problems, SPSA has a faster theoretical convergence rate (in terms of the number of function calls) than SGD performed with finite differences [81].

Finally, another noteworthy gradient-free approach has been developed specifically for the context of VQAs for problems where the objective function is a linear function of an operator expectation value, so that  $C(\boldsymbol{\theta})$  can be expressed as a sum of trigonometric functions. Using this insight, one can fit the functional dependence on a few parameters (with the rest held fixed) allowing one to make local parameter updates [69, 82]. Performing such local updates sequentially over all parameters, or subsets of parameters, and iterating over all parameters one then has an optimization method that is gradient-free and which doesn’t depend on hyper-parameters. Additionally, a variation of this method using Anderson acceleration to speed up convergence has been proposed [82].

*Convergence analysis.* The cost landscapes of VQAs are generally non-convex and can be complicated [83], making it difficult to make general guarantees about the computational expense of the optimizations. However, for simplified landscapes, SGD convergence guarantees have been derived which are similar to those provided in the machine learning literature [84, 85]. Furthermore, within a convex region about a minimum, SGD methods using gradients calculated via the parameter-shift rule have been shown to have smaller upper bounds on the optimization complexity than methods using only objective values (including finite difference methods) [84].

### III. APPLICATIONS

One of the main advantages of the VQA paradigm is that it allows for task-oriented programming. That is, VQAs provide a framework that can be used to tackle a wide array of tasks. This has led to VQAs being proposed for essentially all applications envisioned for quantum computers, and in fact, it has been shown that VQAs allow for universal quantum computing [86]. Their versatility makes the field of VQAs a thriving and fast-

moving one, where new VQAs architectures are regularly being proposed and where existing algorithms are constantly being improved upon. In this section we provide an overview of some of the main applications of VQAs and their state-of-the implementation.

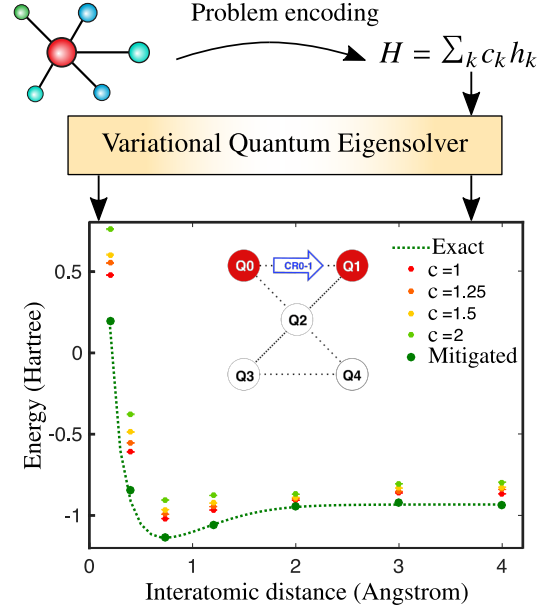
### A. Finding ground and excited states

Perhaps the most paradigmatic application of VQAs is estimating low-lying eigenstates and corresponding eigenvalues of a given Hamiltonian. Previous quantum algorithms to find the ground state of a given Hamiltonian  $H$  were based on adiabatic state preparation and quantum phase estimation subroutines [87, 88], both of which have circuit depth requirements beyond those available in the NISQ era. Hence, the first proposed VQA, the Variational Quantum Eigensolver (VQE), was developed to provide a near-term solution to this task. Here we review both the original VQE architecture as well as some more advanced methods for finding ground and excited states.

*Variational quantum eigensolver.* As shown in Fig. 3, VQE is aimed at finding the ground state energy  $E_G$  of a Hamiltonian  $H$  [16]. Here the cost function is defined as  $C(\theta) = \langle \psi(\theta) | H | \psi(\theta) \rangle$ . That is, one seeks to minimize the expectation value of  $H$  over a trial state  $|\psi(\theta)\rangle = U(\theta)|\psi_0\rangle$  for some ansatz  $U(\theta)$  and initial state  $|\psi_0\rangle$ . According to the Rayleigh-Ritz variational principle, the cost is meaningful and faithful as  $C(\theta) \geq E_G$ , with equality holding if  $|\psi(\theta)\rangle$  is the ground state  $|\psi_G\rangle$  of  $H$ . In practice, the Hamiltonian  $H$  is usually represented as a linear combination of products of Pauli operators  $\sigma_k$  as  $H = \sum_k c_k \sigma_k$  ( $c_k \in \mathbb{R}$ ), so that the cost function  $C(\theta)$  is obtained from a linear combination of expectation values of  $\sigma_k$ . Since practical physical systems are generally described by sparse Hamiltonians, the cost function can be efficiently estimated on quantum computers with a computational cost that usually grows at most polynomially with the system size.

*Orthogonality constrained VQE.* Once an approximated ground state  $|\tilde{\psi}_G\rangle = U(\theta^*)|\psi_0\rangle$  has been obtained, one can use it to find excited states of  $H$ . Let  $a$  be a positive constant that is much larger than the energy gap between the ground state and the first excited states. Then,  $H' = H + a|\tilde{\psi}_G\rangle\langle\tilde{\psi}_G|$  is a Hamiltonian whose ground state is the first excited state of  $H$  [90]. Thus, by employing the VQE for  $H'$  with an updated cost  $C(\theta) = \langle \psi(\theta) | H | \psi(\theta) \rangle + a\langle \psi(\theta) | \tilde{\psi}_G \rangle \langle \tilde{\psi}_G | \psi(\theta) \rangle$ , one may find the first excited state of  $H$ . The first term here is evaluated as in VQE, while the second term can be obtained by computing the state overlap between  $|\tilde{\psi}_G\rangle$  and  $|\psi(\theta)\rangle$  [91, 92]. This procedure can be further generalized to approximate higher excited states. Moreover, it has been shown that incorporating imaginary time evolution can help to improve the calculation robustness [93].

*Subspace expansion method.* Another way to discover low energy excited states using information of the es-



**FIG. 3. Variational Quantum Eigensolver (VQE) implementation.** The VQE algorithm can be used to estimate the ground state energy  $E_G$  of a molecule. The interactions of the system are encoded in a Hamiltonian  $H$ , usually expressed as a linear combination of simple operators  $h_k$ . Taking  $H$  as input, VQE outputs an estimate  $\tilde{E}_G$  of the ground-state energy. In the figure, we show results of a VQE implementation to the electronic structure problem of an  $H_2$  molecule, whose exact energy is shown as a dashed line. The experimental results were obtained using two qubits in one of IBM's superconducting quantum processors (the inset shows qubit connectivity). Due to the presence of hardware noise the estimated energy  $\tilde{E}_G$  has a gap with the true energy. In fact, amplifying the noise strength (i.e., increasing the quantity  $c$ ), deteriorates the solution quality. However, as discussed in Section IV C 3, one can employ error mitigation techniques to improve the solution quality. Data and plot adapted from [89].

timated ground state  $|\tilde{\psi}_G\rangle$  is via the subspace expansion method [44]. Here one runs an additional optimization in a subspace of states  $\{|\psi_k\rangle\}$  generated from  $|\tilde{\psi}_G\rangle$ . For instance, one creates states  $|\psi_k\rangle = \sigma_k|\tilde{\psi}_G\rangle$  for low-weight Pauli operators  $\sigma_k$ , and expands the candidates for the eigenstates as  $|E\rangle = \sum_k \alpha_k |\psi_k\rangle$ . Then, one obtains approximations to the lowest eigenstates by training the coefficients  $\alpha = (\alpha_0, \alpha_1, \alpha_2, \dots)$  while solving the generalised eigenvalue problem  $H\alpha = ES\alpha$ , with  $H_{k,j} = \langle \psi_k | H | \psi_j \rangle$  and  $S_{k,j} = \langle \psi_k | \psi_j \rangle$ .

*Subspace VQE.* The main idea behind subspace VQE is to train a unitary for preparing states in the lowest energy subspace of  $H$  [94]. There are two variants of subspace VQE called weighted and non-weighted subspace VQE. For the weighted subspace VQE, one considers a cost function  $C(\theta) = \sum_{i=0}^m w_i \langle \varphi_i | U(\theta) H U(\theta) | \varphi_i \rangle$  with ordered weights  $w_0 > w_1 > \dots > w_m$  and easily prepared mutually-orthogonal states  $\{|\varphi_i\rangle\}$ . By minimizing the cost function, one approximates the subspace of the lowest eigenstates as  $\{U(\theta^*)|\varphi_i\rangle\}_{i=0}^m$ . Since the

weights are in decreasing order, each state  $U(\boldsymbol{\theta}^*)|\varphi_i\rangle$  corresponds to an eigenstate of the (non-degenerate) Hamiltonian with increasing energies. On the other hand, the non-weighted subspace VQE makes use of the cost function  $C_1(\boldsymbol{\theta}) = \sum_{i=0}^m \langle \varphi_i | U^\dagger(\boldsymbol{\theta}) H U(\boldsymbol{\theta}) | \varphi_i \rangle$ . Minimizing  $C_1$  again gives the subspace of lowest eigenstates. While each state  $U(\boldsymbol{\theta}^*)|\varphi_i\rangle$  is in a superposition of the eigenstates, one needs to further optimize a second cost  $C_2(\boldsymbol{\theta}^*, \boldsymbol{\phi}) = \langle \varphi_i | V^\dagger(\boldsymbol{\phi}) U^\dagger(\boldsymbol{\theta}^*) H U(\boldsymbol{\theta}^*) V(\boldsymbol{\phi}) | \varphi_i \rangle$  over parameters  $\boldsymbol{\phi}$  to rotate each state  $U(\boldsymbol{\theta}^*)V(\boldsymbol{\phi})|\varphi_i\rangle$  to an eigenstate.

*Multistate contracted VQE.* The multistate contracted VQE [45] can be regarded as a midway point between subspace expansion and subspace VQE. It first obtains the lowest energy subspace  $\{U(\boldsymbol{\theta}^*)|\varphi_i\rangle\}_{i=0}^m$  by optimizing  $C_1(\boldsymbol{\theta})$  as in the non-weighted subspace VQE. While instead of optimizing an additional unitary, the multistate contracted VQE approximates each eigenstate as  $|E\rangle = \sum_i \alpha_i U(\boldsymbol{\theta}^*)|\varphi_i\rangle$  with coefficients  $\alpha_i$  which are obtained by solving a generalised eigenvalue problem similar to subspace expansion with  $S = I$ .

*Adiabatically assisted VQE.* Quantum adiabatic optimization seeks to find a solution to an optimization problem by slowly transforming the ground state of a simple problem to that of a complex problem. These methods have a close connection with classical homotopy schemes that are used to find the solutions of classical problems in optimization [95]. In light of this connection, the adiabatically assisted VQE [96] uses a cost function  $C(\boldsymbol{\theta}) = \langle \psi(\boldsymbol{\theta}) | H(s) | \psi(\boldsymbol{\theta}) \rangle$  where  $H(s) = (1-s)H_0 + sH_P$ , and where we recall that  $|\psi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})|\psi_0\rangle$ . Here  $H_P$  is the problem Hamiltonian of interest and  $H_0$  is a simple Hamiltonian whose known ground state is taken as the initial state  $|\psi_0\rangle$ . During the parameter optimization, one slowly changes  $s$  from 0 to 1. The idea of Hamiltonian transformation has been used as a type of ansatz to obtain solutions near the more challenging endpoint [97].

*Accelerated VQE.* As previously mentioned, while Quantum Phase Estimation (QPE) provides a means to estimate eigenenergies in the fault-tolerant era, it is not implementable in the near-term. However, one of the positive features of this algorithm is that a precision  $\epsilon$  can be obtained with a number of measurements which scale as  $\mathcal{O}(\log(\frac{1}{\epsilon}))$ . This is in contrast with VQE, which requires  $\mathcal{O}(\frac{1}{\epsilon^2})$  measurement for the same precision. This motivated the Accelerated VQE algorithm, which interpolates between the VQE and QPE algorithms [98]. This involves taking the VQE algorithm and replacing the measurement process with a tunable version of QPE called  $\alpha$ -QPE. This allows the measurement cost to interpolate between that of VQE and QPE.

## B. Dynamical quantum simulation

Apart from static eigenstate problems, VQAs can also be applied to simulate the dynamical evolution of a quantum system. Conventional quantum Hamiltonian sim-

ulation algorithms, such as the Trotter-Suzuki product formula, generally discretize time into small time steps and simulate each time evolution with a quantum circuit. Therefore, the circuit depth generally increases polynomially with the system size and simulated time. Given the noise inherent in NISQ devices, the accumulated hardware errors for such deep quantum circuits can prove prohibitive. On the other hand, VQAs for dynamical quantum simulation only use a shallow depth circuit, significantly reducing the impact of hardware noise.

*Iterative approach.* Instead of directly implementing the unitary evolution described by the Schrödinger equation  $\frac{d|\psi(t)\rangle}{dt} = -iH|\psi(t)\rangle$ , iterative variational algorithms [74, 75] consider trial states  $|\psi(\boldsymbol{\theta})\rangle$  and map the evolution of the state to the evolution of the parameters  $\boldsymbol{\theta}$ . By iteratively updating the parameters, the quantum state is effectively updated and hence evolved. Specifically, by using variational principles, e.g., McLachlan's principle to solve the minimization  $\min_{\boldsymbol{\theta}} \delta \|\left(\frac{d}{dt} + iH\right)|\psi(\boldsymbol{\theta})\rangle\|$ , one obtains a linear equation for the parameters as  $M \cdot \dot{\boldsymbol{\theta}} = V$ . Here  $\|\psi\rangle\| = \sqrt{\langle \psi | \psi \rangle}$ ,  $\dot{\boldsymbol{\theta}} = \frac{d\boldsymbol{\theta}}{dt}$ ,  $M_{i,j} = \text{Re}(\partial_i \langle \psi(\boldsymbol{\theta}) | \partial_j |\psi(\boldsymbol{\theta}) \rangle)$ ,  $V_i = \text{Im}(\langle \psi(\boldsymbol{\theta}) | H \partial_i |\psi(\boldsymbol{\theta}) \rangle)$ , and  $\partial_i |\psi(\boldsymbol{\theta})\rangle = \frac{\partial |\psi(\boldsymbol{\theta})\rangle}{\partial \theta_i}$ . Each element of  $M$  and  $V$  can be efficiently measured with a modified Hadamard test circuit. By solving the linear equation, one can iteratively update the parameters from  $\boldsymbol{\theta}$  to  $\boldsymbol{\theta} + \dot{\boldsymbol{\theta}}\Delta t$  with a small time step  $\Delta t$ . Similar variational algorithms could be applied for simulating the Wick-rotated Schrödinger equation of imaginary time evolution [71] and general first-order derivative equations with non-Hermitian Hamiltonians [76]. A systematic comparison between different variational principles for different problems can be found in Ref. [75]. Recent works also extend the algorithms to use adaptive ansatz to reduce the circuit depth [99, 100].

*Subspace approach.* The weighted subspace VQE [94] provides an alternative way for simulating dynamics in the subspace of the low energy eigenstates [101]. Here one uses the weighted subspace VQE unitary operator  $U(\boldsymbol{\theta}^*)$  that maps computational basis states  $\{|\varphi_j\rangle\}$  to the low energy eigenstates  $\{|E_j\rangle\}$  as  $U(\boldsymbol{\theta}^*)|\varphi_j\rangle \approx e^{i\delta_j}|E_j\rangle$ , with  $\delta_j$  an unknown phase. Considering the low energy subspace, the time evolution operator can be approximated as  $\exp(-iHt) \approx U(\boldsymbol{\theta}^*)\mathcal{T}(t)U^\dagger(\boldsymbol{\theta}^*)$  with  $\mathcal{T}(t) = \sum_j \exp(-iE_j t)|\psi_j\rangle\langle\psi_j|$ . The procedure could intuitively be understood as (1) rotating the state to the computational basis with  $U^\dagger(\boldsymbol{\theta}^*)$ , (2) evolving the state with  $\mathcal{T}(t)$ , and (3) rotating the basis back with  $U(\boldsymbol{\theta}^*)$ . Therefore, for any state  $|\psi(0)\rangle = \sum_j \alpha_j |E_j\rangle$  that is a superposition of the low energy eigenstates, its time evolution can be simulated as  $|\psi(t)\rangle = U(\boldsymbol{\theta}^*)\mathcal{T}(t)U^\dagger(\boldsymbol{\theta}^*)|\psi(0)\rangle$ . Since the time evolution is directly implemented via  $\mathcal{T}(t)$ , it does not involve iterative parameter update and the circuit depth is independent of the simulation time.

*Variational fast forwarding.* Similar to the subspace approach, variational fast forwarding [102] simulates the time evolution operation  $\exp(-iHt)$  as  $U(\boldsymbol{\theta}^*)\mathcal{T}(\mathbf{E}, t)U^\dagger(\boldsymbol{\theta}^*)$  with



$\mathcal{T}(\mathbf{E}, t) = \sum_j \exp(-iE_j t) |\psi_j\rangle \langle \psi_j|$  a trainable diagonal matrix and  $U(\boldsymbol{\theta}^*)$  a trainable unitary that maps between the eigenstates of  $H$  and the computational basis. While the subspace approach obtains  $\mathcal{T}(\mathbf{E}, t)$  and  $U(\boldsymbol{\theta}^*)$  via weighted subspace VQE, variational fast forwarding optimises a cost given by the fidelity between  $e^{-iH\delta t}$  and  $U(\boldsymbol{\theta}^*)\mathcal{T}(\mathbf{E}, \delta t)U^\dagger(\boldsymbol{\theta}^*)$  for a small time step  $\delta t$  via the so-called local Hilbert-Schmidt test [103]. Then, according to the Trotter-Suzuki product formula, one has  $e^{-iHT} = (e^{-iH\Delta t})^M \approx U(\boldsymbol{\theta}^*)(\mathcal{T}(\mathbf{E}, t))^M U^\dagger(\boldsymbol{\theta}^*)$ . Again, since the time evolution is implemented in  $\mathcal{T}(\mathbf{E}, t)$ , one can simulate the evolution for arbitrary time  $t$  with the same circuit structure. As shown in [104], the ensuing Trotter error of this approach can be removed by diagonalizing instead the Hamiltonian  $H$  that generates the evolution.

*Simulating open systems.* The VQA framework can also be extended to simulate dynamical evolution of open quantum systems. Suppose that the dynamics of the system is described by  $\frac{d\rho}{dt} = \mathcal{L}(\rho)$ , where  $\mathcal{L}$  denotes a super-operator for a dissipative process. Similarly to the iterative approach for pure states [74], one maps the evolution of the mixed state to one of the variational parameters via McLachlan's principle, which solves the minimization  $\min_{\boldsymbol{\theta}} \|(\frac{d}{dt} - \mathcal{L})\rho(\boldsymbol{\theta})\|$ . The solution determines the evolution of the parameters  $M \cdot \dot{\boldsymbol{\theta}} = V$  with  $M_{i,j} = \text{Tr}[\partial_i \rho(\boldsymbol{\theta})^\dagger \partial_j \rho(\boldsymbol{\theta})]$ ,  $V_i = \text{Tr}[\partial_i \rho(\boldsymbol{\theta})^\dagger \mathcal{L}(\rho)]$  and  $\partial_i \rho(\boldsymbol{\theta}) = \frac{\partial \rho(\boldsymbol{\theta})}{\partial \theta_i}$ . Each term of  $M$  and  $V$  can be computed by applying the SWAP test circuit on two copies of the purified states [75]. Here, to simulate an open system of  $n$  qubits, one needs to apply operations on  $4n + 1$  qubits. An alternative approach [76] which reduces this overhead is to simulate the stochastic Schrödinger equation, which unravels the evolution of the density matrix into trajectories of pure states. Each pure state trajectory experiences continuous damping effect and jump processes due to the noise operators, both of which can be efficiently simulated. Since this method one only controls a single copy of the pure state, it only requires  $n + 1$  qubits.

### C. Optimization

Thus far we have discussed using VQAs for tasks which are inherently quantum in nature, i.e., finding ground states and simulating the evolution of quantum states. Here we discuss a different possibility where one employs a VQA to solve a classical optimization problem [105].

The most famous VQA for quantum-enhanced optimization is the Quantum Approximate Optimization Algorithm (QAOA) [22], originally introduced to approximately solve combinatorial problems such as Constraint-Satisfaction (SAT) [106] and Max-Cut problems [107]. Note that we henceforth use the acronym QAOA when referring to the algorithm, while in Section II we used the same term for the quantum alternating operator ansatz.

Combinatorial optimization problems are defined on binary strings  $\mathbf{s} = (s_1, \dots, s_N)$  with the task of maxi-

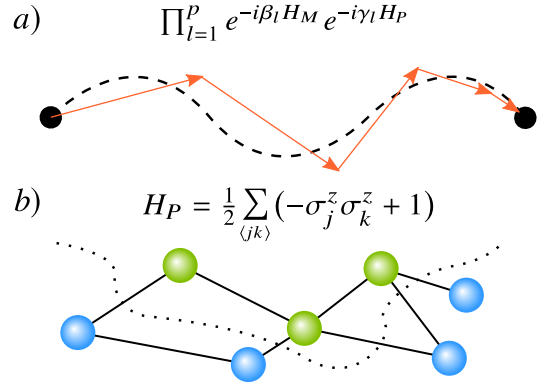


FIG. 4. **Quantum Approximate Optimization Algorithm (QAOA).** a) Schematic representation of the Trotterized adiabatic transformation in the ansatz. The algorithm only loosely follows the evolution of the ground state of  $H(t) = (1 - t)H_M + tH_P$  for every  $t \in [0, 1]$ , as one is interested in making the final state close to the ground state of  $H_P$ . b) Problem Hamiltonian  $H_P$  and graph  $\langle jk \rangle$  for a Max-Cut task. **Each node in the graph (circle) represents a spin.** Vertices connecting two nodes indicate an interaction  $\sigma_j^z \sigma_k^z$  in  $H_P$ , with  $\sigma_k^z$  the Pauli  $z$  operator on spin  $k$ . The solution is encoded in the ground state of  $H_P$  where some spins are pointing up (green) while others point down (blue).

mizing a given classical objective function  $L(\mathbf{s})$ . QAOA encodes  $L(\mathbf{s})$  in a quantum Hamiltonian  $H_P$  by promoting each classical variable  $s_j$  to a Pauli spin-1/2 operator  $\sigma_j^z$ , so that the goal is to prepare the ground state of  $H_P$ . **Motivated by the quantum adiabatic algorithm, QAOA replaces adiabatic evolution with  $p$  rounds of alternating time propagation between the problem Hamiltonian  $H_P$  and appropriately chosen mixer Hamiltonian  $H_M$ , see Fig. 4.** As discussed in the quantum alternating operator ansatz of Section II B, **the evolution time intervals are treated as variational parameters and are optimized classically.** Hence, defining  $\boldsymbol{\theta} = \{\boldsymbol{\gamma}, \boldsymbol{\beta}\}$ , the cost function is  $C(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \langle \psi_p(\boldsymbol{\gamma}, \boldsymbol{\beta}) | H_P | \psi_p(\boldsymbol{\gamma}, \boldsymbol{\beta}) \rangle$  with

$$|\psi_p(\boldsymbol{\gamma}, \boldsymbol{\beta})\rangle = e^{-i\beta_p H_M} e^{-i\gamma_p H_P} \dots e^{-i\beta_1 H_M} e^{-i\gamma_1 H_P} |\psi_0\rangle, \quad (7)$$

and where  $|\psi_0\rangle$  is the ground state of  $H_M$ .

**Finding optimal values  $\boldsymbol{\gamma}$  and  $\boldsymbol{\beta}$  is a hard problem since the optimization landscape in QAOA is non-convex with many local optima [108].** Hence, great efforts have been devoted to finding a good classical optimizer that would require as few calls to the quantum computer as possible. Gradient-based [109, 110], derivative-free [32, 111] as well as reinforcement learning [112] methods were investigated, and this still remains an active field to guarantee a good performance for the QAOA.

### D. Mathematical applications

Several VQAs have been proposed to tackle relevant mathematical problems such as solving linear systems of equations or integer factorization. Since in many cases

there exist quantum algorithms for the fault-tolerant era aimed for these tasks, the goal of VQAs is to have heuristic scalings comparable to the provable scaling of these non-near-term algorithms while keeping the algorithm requirements compatible with the NISQ era.

*Linear systems.* Solving systems of linear equations has wide-ranging applications in science and engineering. Quantum computers offer the possibility of exponential speedup for this task. Specifically, for an  $N \times N$  linear system  $A\mathbf{x} = \mathbf{b}$ , one considers the Quantum Linear Systems Problem (QLSP) where the task is to prepare a normalized state  $|x\rangle$  such that  $A|x\rangle \propto |b\rangle$ , where  $|b\rangle = \mathbf{b}/\|\mathbf{b}\|$  is also a normalized state. The classical algorithmic complexity for this task scales polynomially in the dimension  $N$ , whereas the now-famous HHL quantum algorithm [4] has a complexity that scales logarithmically in  $N$ , with some scaling improvements having been proposed [113–116]. These pioneering quantum algorithms, however, will be difficult to implement in the near term due to the enormous circuit depth requirements [117].

This has motivated researchers to propose VQAs for the QLSP [118–120]. A common feature in these algorithms is the assumption that  $A = \sum_k c_k A_k$  is given as a linear combination of unitaries  $A_k$  that can be efficiently implemented. One can then construct a Hamiltonian whose ground state is the solution to the QLSP and apply a variational approach to minimize the cost  $C(\boldsymbol{\theta}) = \langle \psi(\boldsymbol{\theta}) | H_G | \psi(\boldsymbol{\theta}) \rangle$ . Refs. [118–120] considered the Hamiltonian  $H_G = A(\mathbb{1} - |b\rangle\langle b|)A^\dagger$  (which was also considered outside of the variational setting [121]). It is worth noting that the aforementioned cost can have gradients that vanish exponentially in  $n$  (i.e., a so-called barren plateau in the cost landscape, see Section IV). This problem can be mitigated by considering a local Hamiltonian with the same ground state [118] or by using a hybrid ansatz strategy [120] where  $|\psi(\boldsymbol{\theta})\rangle = \sum_i \alpha_i |\psi_i(\boldsymbol{\theta}_1)\rangle$  with  $\alpha_i$  being variational parameters. Numerical heuristics for random (sparse) linear systems showed efficient scaling in  $N$ ,  $\kappa$ , and  $\epsilon$  [118]. In particular, the heuristic logarithmic scaling in  $N$  suggests that VQAs could potentially give an exponential speedup, analogous to HHL, for the QLSP.

*Matrix-vector multiplication.* Another related problem is to realize matrix-vector multiplication, i.e., to prepare a normalized state  $|x\rangle$  such that  $|x\rangle \propto A|b\rangle$  with normalized vector  $|b\rangle$ . When  $A = 1 - iH\delta t$ , then the problem becomes the task of Hamiltonian simulation. Similar to solving the QLSP, one constructs the Hamiltonian  $H_M = \mathbb{1} - A|b\rangle\langle b|A^\dagger/\|A|b\rangle\|^2$ , whose ground state is  $|x\rangle$  with zero energy [119]. Here  $\|A|b\rangle\| = \sqrt{\langle b|A^\dagger A|b\rangle}$  is the Euclidean norm. Given an approximate solution  $|\psi(\boldsymbol{\theta}^*)\rangle$ , one can lower bound the fidelity to the exact solution as  $|\langle \psi(\boldsymbol{\theta}^*) | x \rangle|^2 \geq 1 - \langle \psi(\boldsymbol{\theta}^*) | H_M | \psi(\boldsymbol{\theta}^*) \rangle$ , thus verify the solution’s correctness whenever the cost function is small.

*Non-linear equations.* Non-linear equations are important to various fields, especially in the form of non-linear partial differential equations. However, mapping such equations onto quantum computers requires care-

ful thought since the underlying mathematics of quantum mechanics is linear. To address this, a VQA for such non-linear problems was proposed in [122]. The approach was illustrated for the time-independent non-linear Schrödinger equation, where the cost function is the total energy (sum of potential, kinetic, and interaction energies), and where the space was discretized into a finite grid. By employing multiple copies of variational quantum states in the cost-evaluation circuit, this VQA can compute non-linear functions.

*Factoring.* While Shor’s algorithm for factoring is very well known, the time horizon for its large-scale implementation is long. Hence, a VQA for factoring as a potential near-term alternative was introduced in [123]. This proposal relies on the fact that factoring can be formulated as an optimization problem, and in particular, as a ground state problem for a classical Ising model. The authors employed the QAOA to variationally search for the ground state. Their numerical heuristics suggest that a linear number of layers in the ansatz ( $p \in \mathcal{O}(n)$ ) leads to a large overlap with the ground state.

*Principal Component Analysis.* An important primitive in data science is reducing the dimensionality of one’s data with Principal Component Analysis (PCA). This involves diagonalizing the covariance matrix for a data set and selecting the eigenvectors with the largest eigenvalues as the key features of the data. Because the covariance matrix is positive semi-definite, one can store it in a density matrix, i.e., in a quantum state, and then any diagonalization method for quantum states can be used for PCA. This idea was exploited in [124] to propose a quantum algorithm for PCA. However, quantum phase estimation and density matrix exponentiation were sub-routines in this algorithm, making it non-implementable in the NISQ era. To potentially make this application more near term, Ref. [125] proposed a variational quantum state diagonalization algorithm, where the cost function  $C(\boldsymbol{\theta})$  quantifies the Hilbert-Schmidt distance between the state  $\tilde{\rho}(\boldsymbol{\theta}) = U(\boldsymbol{\theta})\rho U(\boldsymbol{\theta})^\dagger$  and  $\mathcal{Z}(\tilde{\rho}(\boldsymbol{\theta}))$ , and where  $\mathcal{Z}$  is the dephasing channel. While this VQA outputs estimates of all the eigenvalues and eigenvectors of  $\rho$ , it comes at the cost of requiring  $2n$  qubits for an  $n$  qubit state. This qubit requirement can be reduced with the VQA of [97], which requires only  $n$  qubits. Here one exploits the connection between diagonalization and majorization to define a cost function of the form  $C(\boldsymbol{\theta}) = \text{Tr}[\tilde{\rho}(\boldsymbol{\theta})H]$  where  $H$  is a non-degenerate Hamiltonian. Due to Schur concavity, this cost function is minimized when  $\tilde{\rho}(\boldsymbol{\theta})$  is diagonalized.

## E. Compilation and Unsampling

A very natural task that NISQ devices can potentially accelerate is the compiling of quantum programs. In quantum compiling, the goal is to transform a given unitary  $V$  into native gate sequence  $U(\boldsymbol{\theta})$  with an optimally short circuit depth. Quantum compiling plays

a major role in error mitigation, as errors increase with circuit depth. Quantum compiling is a challenging problem for classical computers to perform optimally, due to the exponential complexity of classically simulating quantum dynamics. Hence, several VQAs have been introduced that can potentially be used to accelerate this task [103, 126–129]. These algorithms can be categorized as either Full Unitary Matrix Compiling (FUMC) or Fixed Input State Compiling (FISC), which respectively aim to compile the target unitary  $V$  over all input states or for a particular input state. In [103] a VQA for FUMC was presented, which employs cost functions closely related to entanglement fidelities to quantify the distance between  $V$  and  $U(\theta)$ . The proposal in [126] also treats the FUMC case, but with an alternative approach to quantifying the cost using the average gate fidelity, averaged over many input and output states. The FISC case was treated in [127], where the problem was reformulated as a ground state energy task, hence making the connection with VQE. The connection with VQE was also generalized to FUMC [128], showing that variational quantum compiling, in general, is a special kind of VQE problem. Ref. [129] introduced and experimentally implemented a compiling scheme which can be thought of as FISC, although the architecture here is focused on the application of unpreparing a quantum state. Finally, it is worth noting that both FUMC and FISC exhibit resilience to hardware noise, in that the global minimum of the cost landscape is unaffected by various types of noise [128]. This noise resilience feature is crucial for the utility of variational quantum compiling for error mitigation, and we discuss this in more detail in Sec. IV C.

## F. Error correction

Quantum Error Correction (QEC) protects qubits from hardware noise. Due to the large qubit requirements of QEC schemes, their implementation is beyond NISQ device capabilities. Nevertheless, QEC could still benefit NISQ hardware by suppressing the error to a certain extent and by combining it with other error mitigation methods. Specifically, conventional universal approaches for implementing QEC codes generally involve an unnecessarily long circuit that does not take into account the hardware structure or the type of noise. Hence, two VQAs have been introduced to solve these problems to automatically discover or compile a small quantum error-correcting code for any quantum hardware and any noise.

The Variational Quantum Error Corrector (QVECTOR) was first proposed to discover a device-tailored quantum error-correcting code for a quantum memory [130]. For any  $k$ -qubit input state  $|\psi\rangle = U_S|\mathbf{0}\rangle$ , prepared by a unitary  $U_S$  acting on a reference state  $|\mathbf{0}\rangle$ , QVECTOR considered two parametrized circuits  $V(\theta_1)$  (on  $n \geq k$  qubits) and  $W(\theta_2)$  (on  $n + r$  qubits), which respectively encode the input logical state into  $n$  qubits with  $n - k$  ancillary qubits and realize recovery

operations with  $r$  ancillary qubits. By sequentially applying encoding, recovery, and decoding on the input state, one obtains an output  $\rho_{out} = W(\theta_1)V(\theta_1)(\psi \otimes |0\rangle\langle 0|^{\otimes n-k+r})V(\theta_1)^\dagger W(\theta_1)^\dagger$ . Projecting the  $n - k$  ancillary qubits back to  $|0\rangle\langle 0|$  and discarding the last  $r$  ancillary qubits, one finds a quantum channel  $\rho = \mathcal{E}(\theta_1, \theta_2)(\psi)$  on the input state  $\psi$ . The target of QVECTOR is to maximize the fidelity  $\int_\psi d\psi F(\psi, \mathcal{E}(\theta_1, \theta_2)(\psi))$  between the output  $\rho$  and the input  $\psi$  averaged overall all  $\psi$  or any  $U_S$  that forms a unitary 2-design. The solution will give the quantum circuit that maximally protects the input state. Numerical simulations showed that QVECTOR can find quantum codes that outperform existing ones [130].

Instead of discovering new device-tailored QEC codes, Ref. [131] considered how to compile conventional QEC codes into a given quantum hardware with specific noise. Suppose one aims to implement the logical state  $|\psi\rangle_L = \alpha|0\rangle_L + \beta|1\rangle_L$  with logical state basis  $\{|0\rangle_L, |1\rangle_L\}$ . Note that  $|\psi\rangle_L$  is the ground state of the stabilizers  $G_k$  as well as the logical operator  $P = |\psi\rangle_L\langle\psi|_L - |\psi^\perp\rangle_L\langle\psi^\perp|_L$  with orthogonal state  $|\psi^\perp\rangle_L$ . Then one can construct a frustration-free Hamiltonian  $H = -a_0P - \sum_{k \geq 1} a_k G_k$  with positive coefficients  $a_0, a_k$ , and with  $|\psi\rangle_L$  the ground state with energy  $E_G = -(a_0 + \sum_{k \geq 1} a_k)$ . One then employs a VQA to discover the circuit that implements  $|\psi\rangle_L$  with a given hardware structure. Since the eigenstate energies are known, the fidelity of the discovered state can be bounded by  $F \geq 1 - (E - E_G)/a$  with the discovered energy  $E$  and  $a = \min\{a_0, a_k\}$ . Numerical studies showed the encoding circuits for the five- and seven-qubit codes with different noisy hardware [131].

## G. Machine learning and data science

Quantum machine learning (QML) generally refers to the tasks of learning patterns with the goal of making accurate predictions on unknown, and unseen data. While providing an in-depth discussion on the field of QML is beyond the scope of this review, we here present several QML applications for which the VQA framework can be readily implemented. Specifically, here one learns a parametrized quantum circuit to solve a given task [64, 132]. This connection between VQAs and (typical) QML applications shows that the lessons learned in one field can be of great use in the other, hence providing a close connection between these two fields.

*Classifiers.* The classification of data is a ubiquitous task in machine learning. Given training data of the form  $\{x^{(i)}, y^{(i)}\}$ , where  $x^{(i)}$  are inputs, and  $y^{(i)}$  labels, the goal is to train a classifier to accurately predict the label of each input. Since a key aspect for the success of classical neural networks is their non-linearity, one can expect this property to also arise in a quantum classifier. As shown in [133], parametrized quantum circuits can support linear transformations and non-linearity can be exploited from the tensor product structure of a quantum

system. More precisely, defining an input data dependent unitary  $V(x)$ , then the tensor product  $V(x) \otimes V'(x)$  or the multiplication  $V(x)V'(x)$  results in a non-linear function of the input data  $x$ . In this sense, the unitary  $V(x)$  can be used as a quantum non-linear feature map, where the Hilbert space can be exploited for a feature space [134, 135]. Interestingly, the tensor network structure of quantum mechanics has even inspired classical machine learning methods [136].

Here, after embedding the input data  $x$  into the quantum state, a linear transformation is performed using a parametrized quantum circuit,  $U(\theta)V(x)|\psi_0\rangle$ . The cost function is then defined as the error between the true label and the expectation value of an easily measurable observable  $A$ , i.e.,  $C(\theta) = \sum_i [y^{(i)} - \langle\psi_0|V^\dagger(x^{(i)})U^\dagger(\theta)AU(\theta)V(x^{(i)})|\psi_0\rangle]^2$ . This approach has been used in generalization and in classification tasks [64, 133], as well as in an experimental demonstration of variational classification [135]. Moreover, as shown in Refs. [134, 135], instead of using a parametrized unitary  $U(\theta)$  one can employ estimate products of quantum feature vectors  $\langle\psi_0|V^\dagger(x')V(x)|\psi_0\rangle$  to perform a kernel method. Finally, the quantum kernel trick, which means that the dimensions of the quantum-enhanced feature space are larger than the number of data sets, has been demonstrated experimentally by using an ensemble nuclear spin system [137].

*Autoencoders.* The autoencoder for data compression is an important primitive in machine learning. Here the idea is to force information through a bottleneck while still maintaining the recoverability of the data. As a quantum analog, Ref. [138] introduced a VQA for quantum autoencoding, with the goal of compressing quantum data. (See [139, 140] for alternative approaches to quantum autoencoders.) The input to the algorithm is an ensemble of pure quantum states  $\{p_\mu, |\psi_\mu\rangle\}$  on a bipartite system  $AB$ . The goal is then to train an ansatz  $U(\theta)$  to compress this ensemble into the  $A$  subsystem, such that one can recover each state  $|\psi_\mu\rangle$  with high fidelity from subsystem  $A$ . The  $B$  subsystem is discarded and hence can be thought of as the “trash”. Given the close connection between data compression and decoupling [138], the cost function is based on the overlap between the output state on  $B$  and a fixed pure state. Recently, a local version of this cost function was also proposed and was shown to train well for large-scale problems [141]. Quantum autoencoders have seen experimental implementation on quantum hardware [142], and will likely be an important primitive in quantum machine learning.

*Generative models.* The idea of training a parametrized quantum circuit for a QML implementation can also be applied for a generative model [143, 144], which is an unsupervised statistical learning task with the goal of learning a probability distribution that generates a given data set. Let  $\{x^{(i)}\}_{i=1}^D$  be a data set sampled from a probability distribution  $q(x)$ . Here one learns  $q(x)$  as the parametrized probability distribution  $p_\theta(x) = |\langle x|U(\theta)|\psi_0\rangle|^2$  obtained by applying  $U(\theta)$  to an

input state and measuring in the computational basis, i.e., this corresponds then to a quantum circuit Born machine [145]. In principle one wishes to minimize the difference between the two distributions. However, since  $q(x)$  is not available, the cost function is defined by the negative log-likelihood  $C(\theta) = -\frac{1}{D} \sum_i \log(p_\theta(x^{(i)}))$ . In Ref. [146], cat and coherent thermal states were generated experimentally in this way from specific classical data sets. In [145], an implicit generative model has been constructed by comparing the distance in the Gaussian kernel feature space. The representation power of the generative model has been investigated in [143]. Finally, it has been shown that quantum circuit Born machines can simulate the restricted Boltzmann machine and perform a sampling task that is hard for a classical computer [147].

*Variational Quantum Generators.* Generative Adversarial Networks (GANs) play an important role in classical machine learning for applications such as image synthesis and molecular discovery. Ref. [148] proposed a VQA for learning continuous distributions which is meant to be a quantum version of GANs. Here one still considers classical data, but encoded into a quantum circuit. This encoding is followed by a variational quantum circuit that generates quantum states, which are then measured to produce a fake sample. This fake sample then enters either a classical discriminator or a quantum discriminator, and the cost function is optimized to minimize the discrimination probability with respect to real samples. The target application is to accelerate classical GANs using quantum computers.

*Quantum Neural Network architectures.* Several Quantum Neural Network (QNNs) architectures have been proposed, and which can be used for some of the aforementioned tasks. For instance, Refs. [132, 149, 150] proposed perceptron-based QNNs. In these architectures each node in the neural network represents a qubit, and their connections are given by parametrized unitaries of the form (3) acting on the input states. On the other hand, Ref. [151] introduced a Quantum Convolutional Neural Networks (QCNN). QCNNs have been employed for error correction, image recognition [152], and to discriminate quantum state belonging to different topological phases [151]. Moreover, it has been shown that QCNN and QNNs with tree tensor network architectures do not exhibit barren plateaus [153, 154] (see Section IV A 1 for more details on barren plateaus), potentially making them a generically trainable architecture for large-scale implementations.

## H. New frontiers

In this section we discuss an exciting application of the VQA framework where one draws on the fact that VQAs deal with systems that are quantum mechanical. That is, many VQAs have been proposed to understand and exploit the mathematical and physical structure of



quantum states, and quantum theory in general.

*Quantum foundations.* NISQ computers will likely play an important role in understanding the foundations of quantum mechanics. In a sense, these devices offer experimental platforms to test foundational ideas ranging from quantum gravity to quantum Darwinism. For example, the emergence of classicality in quantum systems will be soon be a computationally tractable field of study due to the increasing size of NISQ computers. Along these lines, Ref. [155] proposed the Variational Consistent Histories (VCH) algorithm. Consistent Histories is a formal approach to quantum mechanics that has proven to be useful in studying the quantum-to-classical transition as well as quantum cosmology; however, it involves computing the decoherence functional between all pairs of histories. Since the number of histories grows exponentially in both the system size and the number of dynamical times considered, classical numerics are intractable for this formalism. The VCH algorithm aims to remedy this by storing the decoherence functional in a quantum density matrix and then using a quantum device to efficiently compute a cost function that quantifies how close this matrix is to being diagonal. Fully diagonal corresponds to the consistency condition, and hence VCH actively searches for consistent families of histories. VCH has the potential to elucidate the quantum-to-classical transition in quantum dynamics that otherwise could not be efficiently studied.

*Quantum information theory.* Another field that will likely see renewed interest due to NISQ computers is quantum information theory or quantum Shannon theory [156]. For example, in [138] it was remarked that the quantum autoencoder algorithm could potentially be used to learn encodings and achievable rates for quantum channel transmission. Another area of research is using NISQ computers to compute key quantities in quantum information theory, such as the von Neumann entropy or distinguishability measures like the trace distance. While it is known that these problems are hard for general quantum states [157], Ref. [158] introduced a VQA to estimate the quantum fidelity between an arbitrary state  $\sigma$  and a low-rank state  $\rho$ . Moreover, in [59] a VQA was introduced to learn modular Hamiltonians, which provides an upper bound on the von Neumann entropy of a quantum state. Here one attempts to variationally decorrelate a quantum state by minimizing the relative entropy to a product distribution, and hence this method is suited for states that can be easily decorrelated.

*Entanglement Spectroscopy.* Characterizing entanglement is crucial for understanding condensed matter systems, and the entanglement spectrum has proven to be useful in studying topological order. Several VQAs have been introduced to extract the entanglement spectrum of a quantum state [97, 125, 159]. Since the entanglement spectrum can be viewed as the principal components of a reduced density matrix, algorithms for PCA can be used for this purpose. This includes the VQAs discussed in Sec. III G. In addition, one can also employ

the variational algorithm for quantum singular value decomposition introduced in [159]. One can envision then using these algorithms to characterize the entanglement (and for example, topological order) in a ground state that was prepared by VQE, and hence different VQAs can be used together in a complementary manner.

*Quantum metrology.* Quantum metrology is a research field where one seeks the optimal setup for probing a parameter of interest with minimal shot noise. The probed parameter is typically a magnetic field. In the absence of noise during the probing process, the analytical solution about the optimal probe state has been given; however, when general physical noises are present, an analytical argument is hard. Variational-state quantum metrology variationally searches for the optimal probe state [160–163]. For state preparation, variational quantum circuits are used in Refs. [160, 162, 163] while optical tweezer arrays are considered in [161]. More concretely, one prepares a probe state with variational parameters, probe the magnetic field with physical noises, measure quantum Fisher information (QFI) as a cost function, and update the parameters to maximize it. Note that since QFI cannot be efficiently computed, an approximation of QFI can be heuristically found by optimizing measurement basis, or by computing upper and lower bounds on the QFI [163].

#### IV. CHALLENGES AND POTENTIAL SOLUTIONS

Despite the tremendous developments in the field of VQAs, there are still many challenges that need to be addressed to maintain the hope of achieving quantum speedups when scaling up these near-term architectures. Understanding the limitations of VQAs is crucial to developing strategies that can be used to construct better algorithms, prove certain guarantees on their performance, and even to build better quantum hardware. In this section we review some of the known challenges of VQAs and how they can be potentially addressed. The results are presented in three main areas pertaining the trainability of the cost function, the efficiency of estimating the cost, and the effect of noise in VQAs.

##### A. Trainability

###### 1. Barren plateaus

The so-called Barren Plateau (BP) phenomenon in the cost function landscape has recently received considerable attention as one of the main bottlenecks for VQAs. When a given cost function  $C(\theta)$  exhibits a BP the magnitude of its partial derivatives will be, on average, exponentially vanishing with the system size [164]. As shown in Fig. 5 this has the effect of the landscape being essentially flat. Hence, in a BP one needs an exponentially

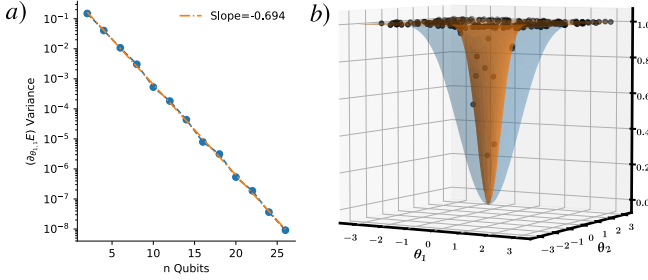


FIG. 5. **Barren Plateau (BP) phenomenon.** a) Variance of the cost function partial derivative versus number of qubits. Results were obtained from a VQE implementation with a deep unstructured ansatz. The  $y$ -axis is on a log scale. As the number of qubits increases the variance vanishes exponentially with the system size. Plot from [164]. b) Visualization of the landscape of a global cost function which exhibits a BP for the quantum compilation implementation, taken from [68]. The orange (blue) landscape was obtained for  $n = 24$  ( $n = 4$ ) qubits. As the number of qubits increases, the landscape becomes flatter. Moreover, this cost also exhibits the narrow gorge phenomena, where the volume of parameters leading to small cost values shrinks exponentially with  $n$  [68].

large precision to resolve against finite sampling noise and determine a cost-minimizing direction, with this being valid independently of using a gradient-based [68] or gradient-free optimization method [165]. Note that the exponential scaling in the precision due to BPs destroys any hope of achieving a quantum advantage with a VQA, as its computational complexity will be no better than that of a classical algorithm. Hence, analyzing the existence of BP in a given VQA is fundamental to preserve the hope of using it to achieve a quantum advantage.

The phenomenon of BPs was originally discovered in [164] where it was shown that deep unstructured parametrized quantum circuits exhibit BPs when randomly initialized. The proof of this result relies on the fact that these unstructured ansatzes become 2-designs when their depth grows polynomially with the number of qubits  $n$  [166, 167]. One can view this phenomenon as spanning from the fact that the ansatz is problem-agnostic and hence needs to explore an exponentially large space to find the solution. Therefore, the probability of finding the solution when randomly initializing the ansatz is exponentially small.

The analysis of BPs was extended to shallow random layered ansatz [141]. Here it was shown that the BP phenomenon is cost-function dependent: Global cost functions (i.e., when one compares operators or states living in exponentially large Hilbert spaces) exhibit BPs, while local cost functions (i.e., when one compares objects at the single-qubit level), exhibit gradients which vanish polynomially in  $n$  so long as the circuit depth is at most logarithmic in  $n$ . This implies a connection between the locality and trainability and informs our intuition as to what types of cost functions might have to be avoided. These results have been numerically verified

in [168] and further extended in [169]. Here one can understand the presence of BPs for global costs as spanning from the fact that the Hilbert space grows exponentially with  $n$  and hence the probability of two objects being close is exponentially small.

In addition, it has been shown that BPs can arise in more general problems such as learning a scrambler [170] where for any choice of variational ansatz will lead to, with high probability, a BP. This again is due to the intrinsic randomness in a scrambler. In addition, the BP phenomenon has been studied when training randomly initialized perceptron-based quantum neural networks [171, 172]. Here, BPs arise from the significant amount of entanglement created by the perceptrons connecting large number of qubits in visible and hidden layers. Specifically, when tracing out the qubits in the hidden layers the state of the visible qubits becomes exponentially close to being maximally mixed (due to concentration of measure). Evidently, extracting information from such a state is hard.

While the previous results rely on the randomness of the ansatz there is a conceptually different phenomena that can lead to BPs. Recently, it was shown in [173] that noise can induce barren plateaus, regardless of the ansatz employed. Here, the presence of noise acting throughout the circuit progressively corrupts the state towards the fixed point of the noise model, usually the maximally mixed state [174]. Such phenomenon was shown to arise when the circuit depth needs to be linear (or larger) with the system size, meaning that it will affect many widely used ansatzes.

## 2. Ansatz and initialization strategies

Hand-in-hand with the theoretical progress on the analysis on the BP phenomenon, there has been great research efforts focused on avoiding or mitigating the effect of BP. The main strategy here being to breaking the assumptions leading to the BPs. In what follows we present two main approaches: parameters initialization and choice of ansatz.

*Parameter initialization.* Randomly initializing an ansatz can lead to the algorithm starting far from the solution, near a local minima, or even in a region with barren plateaus. Hence, optimally choosing the seed for  $\theta$  at the beginning of the optimization is an important task. The importance of parameter initialization was made clear in [175] where it was noted that the optimal parameters in QAOA exhibit persistent patterns. Based on these observations initialization strategies were proposed and which were heuristically shown to outperform randomly initialized optimizations. Additionally, in Refs. [27, 176] an initialization strategy was developed specifically to address BPs in deep circuits. Here, one selects at random a subset of parameters in  $\theta$ , and chooses the value of the remaining ones so that the circuit is a sequence of shallow unitaries that evaluates to the iden-

tity. The main idea behind this method is to reduce the randomness and depth of the circuit to break the assumption that the circuit approximates a 2-design, a condition necessary for BPs to arise in deep ansatz. Similar to the previous method, other schemes have been introduced to prevent BP by restricting the randomization of the ansatz. For instance, the proposal of [177] showed that correlating the parameters in the ansatz effectively reduces the dimension of the hyperparameter space and can lead to large cost function gradients. In addition, Ref. [178] introduced a method where one employs layer-by-layer training. Here, one initially trains shallow circuits and progressively adds components to the circuit. While the latter guarantees that the number of parameters and randomness remains small for the first steps of the training, it has been shown [179] this method can lead to an abrupt transition in the ability of quantum circuits to be trained. Finally, a method was introduced in [180] where one pre-trains the parameters in the quantum circuits by employing classical neural networks.

*Ansatz strategies.* Another strategy for preventing BPs is using structured ansatzes which are problem-inspired. The goal here is to restrict the space explored by the ansatz during the optimization. As discussed in Section II B, the UCC ansatz for VQE of the quantum alternating operator ansatz [22, 23] for optimization are problem-inspired ansatzes which are usually trainable even when randomly initialized. Other ansatz strategies include the proposals in [180] to learning a mixed state, where one leverages knowledge of the target Hamiltonian to create a Hamiltonian variational ansatz. In addition, Refs. [49, 50] presented an approach where the ansatz for the solution is  $|\psi(\{c_\mu\})\rangle = \sum_\mu c_\mu |\psi_\mu\rangle$ , for a fixed set of states  $\{|\psi_\mu\rangle\}$  determined by the problem at hand. Here the optimization over the coefficients  $\{c_\mu\}$  can be solved using a quadratically constrained quadratic program.

Finally, we remark that along with ansatz strategies there are other ways of potentially addressing BPs. These include optimizers tailored to mitigate the effect of BPs [181], using local cost functions [97], or employing an architecture such as the Quantum Convolutional Neural Network, which has been shown to avoid BPs [153].

## B. Efficiency

Another requirement that must be met for VQAs to provide a quantum advantage is having an efficient way to estimate expectation values (and thus more general cost functions). The existence of BPs can exponentially increase the precision requirements needed for the optimization portion of VQAs, as discussed above in IV A 1, but even in the absence of such BPs these expectation value estimations are not guaranteed to be efficient. Indeed, early estimations of resource requirements suggested that the number of measurements that would be required for interesting quantum chemistry VQE problems would be astronomical, making addressing this issue

essential for quantum advantage [26]. More reasonable resource estimates can be reached for restricted problems such as the Hubbard model [182, 183], how this need to efficiently estimate the expected values of operators from digital quantum circuits has inspired a variety of approaches [184]. While in principle one could always take projective measurements onto the eigenbasis of the operator in question, in general both the computational complexity of finding the required unitary, as well as the depth required to implement that transformation, may be intractable. However, given that arbitrary Pauli operators are diagonalizable with one layer of single qubit rotations, it is common for the operators of interest (such as quantum chemistry Hamiltonians) to be expressed by their decomposition into such Pauli operators. That is,  $H = \sum_i c_i \sigma_i$ , where  $\{c_i\}$  are real coefficients and  $\sigma_i$  Pauli operators. The drawback of this approach is that for many interesting Hamiltonians this decomposition contains many terms. For example, for chemical Hamiltonians the number of distinct Pauli strings scales as  $n^4$  where  $n$  is the number of orbitals (and thus qubits) for large molecules. In what follows we discuss several methods whose goal is to obtain measurement frugality in estimating the cost function.

*Commuting sets of operators.* In the interests of reducing the number of measurements required to estimate an operator expectation value, a number of methods have been proposed for partitioning sets of Pauli strings into commuting (i.e. simultaneously measurable) subsets. The choice of the subsets is also of course non-unique and has been mapped onto the combinatorial problems of graph coloring [185, 186], finding the minimum clique cover [187–190], or finding the maximal flow in network flow graphs [191], which makes it possible to import the heuristics and formal results from those problems.

Perhaps the simplest approach to such a partitioning is to look for subsets that are qubit-wise commuting (QWC), which is to say that the Pauli operators on each qubit commute. Indeed, this was the first method introduced [192]. However, while the QWC methods help reduce the number of operators, they do not change the asymptotic scaling for quantum chemistry applications, motivating more general commutative groupings to be considered. To this end, it has been shown that by considering general commutations (and increasing the number of gates of the circuit quadratically with  $n$ ) the scaling of the number of measurements can be reduced to  $n^3$  [185, 186, 188–191].

For using VQE on fermionic systems, this scaling can actually be brought down to either quadratic or, for simpler cases even linear in  $n$  [193]. This significant improvement is found by considering factorizations of the two-electron integral tensors, rather than working at the operator level. The success of this approach suggests that using background information on the problem may significantly improve the measurement efficiency of estimating an expectation value.

*Optimized sampling.* In addition to reducing the num-

ber of individual operators that need to be measured, measurement efficiency can also be improved by carefully allocating the number of shots among the Pauli operators. Since operators with smaller coefficients will tend to contribute less to the overall variance, assigning the same number of shots to each operator is usually inefficient. Instead, the optimal approach is to give each Pauli operator a number of shots proportional to  $|c_i|\sqrt{\text{Var}(\sigma_i)}$ , where  $\text{Var}(\sigma_i)$  is the variance of  $\langle\sigma_i\rangle$  [194]. During an optimization where low precision steps may be allowed early on, this allocation can instead be performed randomly with probabilities proportional to  $|c_i|\sqrt{\text{Var}(\sigma_i)}$ . Making the allocation randomly in this way allows for unbiased estimates with as little as one shot, potentially significantly increasing the efficiency of the optimization [195]. Optimizing the sampling of the metric tensor (Section II C) has also been explored, with the conclusion that these costs need not be dominant in metric-aware VQAs [196].

*Classical shadows.* Another promising approach to efficient measurements is the construction of classical shadows [197], also known as shadow tomography. In this approach, an approximate classical representation of the state (the classical shadow) is constructed by summing over the collection of states that a sequence of different measurements projects onto. These measurements are taken in the basis of randomly chosen strings so that a partial tomography of the state is completed. Combining the measurements in this way, each shot contributes to the estimation of each Pauli operator expectation value, resulting in a number of measurements that scales logarithmically with  $n$ . As with direct measurement approaches discussed above, this approach can also be further optimized by tuning the probability distribution for the Pauli operators that define the measurements to match the properties of the operator and state [198].

*Neural network tomography.* A different approach using partial tomography is to train an approximate restricted Boltzmann machine (RBM) representation of the desired quantum state [199]. This RBM is fitted using measurements of the Pauli operators that are needed to directly estimate a given operator’s expectation value, and so does not inherently reduce the number of operators to measure. However, by computing the expectation value on an approximate RBM instead of directly from measurements the sampling variance for a given number of shots is substantially reduced at the cost of introducing a small, positive bias [199].

### C. Accuracy

One of the main goals for VQAs is to provide a practical use for near-term noisy devices. In this sense, VQAs provide a strategy to deal with hardware noise as they can potentially minimize quantum circuit depth. However, one can still ask what the impact of hardware noise will be on the accuracy of a VQA. There are multiple

aspects of the impact: hardware noise could potentially slow down the training process, it could bias the landscape so that the noisy global optimum no longer corresponds to the noise-free global optimum, and it could affect the final value of the optimal cost.

#### 1. Impact of hardware noise

*Effect of noise on training.* The question of whether noise can help with the training process was posed in [200]. In practice, it is typical to observe that noise slows down the training. For example, it was heuristically observed that the noise-free cost achieves lower values with noise-free training than with noisy training [79, 195, 201]. As discussed in Section IV A 1, the intuition behind this slowing down is that the cost landscape is flattened, and hence gradient magnitudes are reduced, by the presence of incoherent noise [173, 202, 203]. Moreover, gradients decay exponentially with the algorithm’s depth, meaning that the deeper the circuit, the more it will be affected. This can be further understood from the fact that cost functions are typically extremized by pure states, and since incoherent noise reduces state purity, one expects this noise to erode the extremal points of the landscape [174]. The presence of the noise-induced barren plateaus and their effect on the trainability is one of the leading challenges for VQAs, with potential solutions being developing better quantum hardware or shorter-depth algorithms. It is worth remarking that the results discussed here do not account for the use of error mitigation techniques (see below), and the scope to which these could help is still an open question.

*Effect of noise on cost evaluation.* In Refs. [173, 174] it was also shown that in the presence of local Pauli noise, the cost landscape concentrated exponentially with the depth of the ansatz around the value of the cost associated with the maximally mixed state. While the proof of this exponential concentration of the cost was for general VQAs, some previous works had also observed this effect for the special case of the QAOA [202, 203]. The exponential concentration of the cost is of course important beyond the issue of trainability. Even if one is able to train, the final cost value will be corrupted by noise. There are certain VQAs where this is not an important issue (e.g., in QAOA where one can classically compute the cost after sampling). However, for VQE problems, this is important, since one is ultimately interested in an accurate estimation of the energy. This emphasizes the importance of understanding to what degree error mitigation methods (discussed below) can correct for this issue.

#### 2. Noise resilience

*Inherent resilience to coherent noise.* One reason for interest in variational algorithms is their ability to nat-



urally suppress certain types of noise in hardware, especially in near-term implementations. This is due to the fact that by construction, variational algorithms are insensitive to the specific parameter values, ultimately only sampling physical observables from the resulting state. More specifically, if the physical implementation of a unitary results in a coherent error within the parameter space, or  $U(\theta)$  actually results in  $U(\theta + \delta)$ , then under mild assumptions the optimizer can calibrate this block unitary on the fly to improve the physical state produced. This effect was first conjectured theoretically [192] and later seen experimentally in superconducting qubits [37], where errors after the variational procedure were reduced in some cases by over an order of magnitude. Success in this endeavor depends upon the ability to optimize faster than the drift of calibration in the device, and sufficient variational flexibility in the ansatz, but may continue to be effective even into the early fault-tolerant regime where coherent errors can be especially insidious.

*Inherent resilience to incoherent noise.* It is an interesting question as to what degree incoherent noise, such as decoherence, random gate errors, and measurement errors, will impact VQAs. For example, it was shown that optimization in the presence of some noise channels can automatically move the state into subspaces that are resilient to those channels as an energetic trade-off [44]. On the other hand, one could operate under the assumption of perfect training (which may still be possible for either weak noise or shallow ansatzes), and ask whether the global optimum in the cost landscape is robust to such noise. This was the approach of Ref. [128], where it was shown that VQAs for quantum compiling (see Section III E), exhibit a special type of noise resilience known as Optimal Parameter Resilience (OPR). OPR is the notion that global minima of the noisy cost function correspond to global minima of the noise-free cost function. In this sense, if an algorithm exhibits OPR, then minimizing the cost in the presence of noise will still obtain the correct optimal parameters, and hence the optimal parameters are resilient. Since quantum compilation is a special case of VQE the question still remains open to whether other VQAs exhibit this type of noise resilience for certain noise models.

### 3. Error mitigation

Quantum Error Mitigation (QEM) generally suppress physical errors to expectation values of observables via classical post-processing of measurement outcomes [204]. An intuitive but powerful example is the extrapolation method [74, 205]. Even if the error rate cannot be reduced, in many cases it can be deliberately boosted, for example, as shown in Fig. 6, by inserting additional noisy pulses or making gate operations longer, the quantum device undergoes more physical errors. Then, by obtaining measurement outcomes at several noise levels and extrapolating them, one can estimate the error-

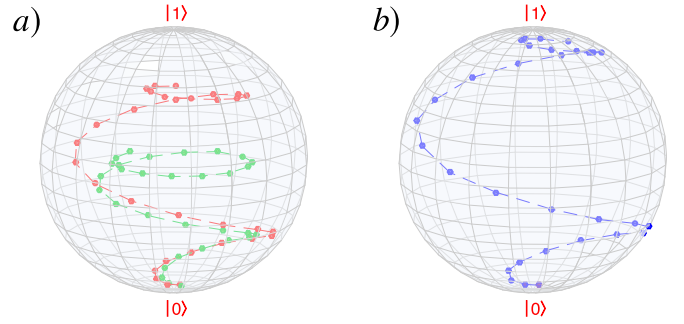


FIG. 6. **Qubit trajectories in the Bloch sphere with the Zero-Noise Extrapolation (ZNE) technique.** The accuracy of a noisy quantum computer can be improved with the ZNE error mitigation method. a) Here, one repeats a given calculation with different levels of noise. The green curve corresponds to a rotation in the Bloch sphere with a higher noise level than that leading to the red curve. b) Taking data from the red and green curves, ZNE can be used to estimate what the trajectory (blue) would be like in the absence of noise. Figure adapted from [89].

free result using the so-called zero-noise extrapolation method. Due to propagation of uncertainty, the variance of the error-mitigated result is amplified and hence one needs to have a larger sampling cost, which is the overhead of QEM. First, Richardson extrapolation was proposed [74, 89, 205], and it was shown that single- and multi-exponential extrapolation work well for Markovian gate errors, with the latter subsequently shown to have very broad efficacy [206, 207]. In addition, extrapolation using least square fitting for several noise parameters has also been proposed [208]. Furthermore, it has been observed that the extrapolation method can mitigate algorithmic errors that arise due to insufficiency in the number of time steps [209].

Although extrapolation methods by construction cannot fully mitigate physical errors [205, 206], probabilistic error cancellation in theory can obtain unbiased expectation values by inverting the noise process with additional probabilistic gate operations (if a complete characterization of noise is provided). Note that since an inverse map of physical errors is generally unphysical, it is necessary to post-process measurement outcomes according to applied recovery operations. In Ref. [205] this method was first introduced and in [206] it was found that gate set tomography is a suitable noise characterization strategy, and a set of operations was proposed which can compensate for general Markovian errors. Furthermore, based on probabilistic error cancellation, stochastic error mitigation which works for general continuous systems such as analog quantum simulators as well as digital quantum computers was introduced [210].

A different approach to QEM relies on the classical simulability of near-Clifford circuits. The basic idea behind this approach is to compare the classically computed exact expectation values for near-Clifford circuits with their noisy counterparts evaluated on actual hard-

ware [211–213]. Taking this approach can allow one to implement a probabilistic error mitigation protocol without needing to construct a full error model for an experiment [211]. Alternatively, one can perform a simple regression with this Clifford data to estimate how the observables have been affected and invert this regression to estimate desired noise-free expectation values [212]. Finally, zero-noise extrapolation can be merged with this regression to have an extrapolation to zero-noise whose form is tuned via the Clifford data, reducing the risk of blind extrapolations [213].

Several additional QEM methods have been proposed. Symmetry verification is especially useful for ansatzes that preserve symmetries such as particle and spin number [12, 214, 215]. Since physical errors break the symmetry, by measuring and ignoring the undesired case (similarly to error detection), one can mitigate physical noise. Unlike other QEM methods, symmetry verification can recover the quantum state itself. One can also take a post-processing approach using the information of the symmetry with a larger sample number [215]. The use of symmetry verification to augment error extrapolation and probabilistic error cancellation was taken still further in [207].

In an alternative and complementary approach, the subspace expansion method was also shown to be useful for QEM in [216]. Here, using subspace expansion one can mitigate physical noise for eigenstates of the Hamiltonian as well as evaluating excited states because the state is expanded in a larger subspace. Note that this method works better for coherent noise than for stochastic noise. A distinct approach was introduced in [217, 218] which comes at the cost of increasing the number of qubits. Here, by entangling and measuring  $M$  copies of a noisy state  $\rho$ , one can compute expectation values with respect to the state  $\frac{\rho^M}{\text{Tr}[\rho^M]}$ . Under the assumption that the principal eigenvector of  $\rho$  is the desired state, this method can exponentially suppress errors with  $M$ . Finally we remark that Ref. [219] introduced a method to mitigate expectation values against correlated measurement errors, while [220] implemented an error mitigation technique to suppress the effects of photon loss for a Gaussian Boson sampling device.

## V. OPPORTUNITIES FOR NEAR TERM QUANTUM ADVANTAGE

VQAs are largely regarded as the best candidate for providing quantum advantage for practical applications. That is, it is expected that one will be able to employ a VQA to solve a problem more efficiently than any classical state-of-the-art method. As discussed in the previous sections, tremendous effort has been recently put forward towards this goal with the development of efficient ansatz strategies, quantum-aware optimization methods, novel VQAs, and error mitigation techniques. While many challenges still remain to be addressed, like the need for

larger and better quantum devices, one can nevertheless pose the question as to what specific applications will provide the first quantum advantage for a practical scenario. In this section we discuss some of the most exciting possibilities where quantum advantage could arise.

### A. Chemistry and material sciences

The ability to simulate and understand the static and dynamical properties of molecules and strongly correlated electronic systems is a fundamental task in many areas of science. For instance, this task is relevant in biology to understand protein folding dynamics, while in pharmaceutical sciences one could analyze drug–receptor interactions to improve drug discovery capabilities [221–223]. Similarly, analyzing the electronic structure of complex correlated materials is highly important for studying high-temperature superconductivity or to analyze transition metal materials near a Mott transition.

*Molecular structure.* In the past few decades there have been great developments in the classical treatment of the structure of molecular systems. These include approximate methods such as Hartree-Fock or density functional theory, or methods closely connected to quantum information, like the density matrix renormalization group approach that utilizes matrix product states as an ansatz [224, 225]. However, even for these sophisticated approaches, systems of interest like FeMoCo are beyond the reach of an accurate description due to the entanglement structure of the electrons and orbitals. The relevant electronic space that one needs to treat correlations accurately in for these systems is relatively modest, and for that reason, these may be good targets for near-term quantum computers to play a role. As discussed in Section III A the VQE algorithm [16] (and associated architectures) have shown promising advances towards the goal of performing molecular quantum chemistry on quantum computers [226], with large scale implementation already being executed [227].

*Molecular dynamics.* As for the dynamics of chemical and other quantum systems, there have been a number of strides in evaluating or compressing these evolutions using variational approaches [74, 75]. Much like variational principles connected to the ground state, there are a number of time-dependent variational principles that can be used to approximate time-dynamics. Here there are two timescales of interest. The first is the electronic timescales over which electrons rearrange upon excitation. The second, much slower than the first, is the rearrangement of nuclei that is induced by forces derived from the electrons in their respective configurations, excited or not. Generally speaking, treating the detailed dynamics of the electrons accurately has been extremely challenging for classical approaches despite its relevance in phenomena related to photovoltaics and light-emitting diodes [228, 229]. The scale between the two timescales has motivated the development of meth-

ods that treat them separately, often using a classical or semi-classical representation for the nuclei and quantum representation for the electrons [230]. Variational methods can be applied incrementally in these cases, by stepping the electronic wavefunction forward with time-dependent variational principles [74, 75] and sampling the forces [231] to move the nuclei classically, resulting in a Born-Oppenheimer type molecular dynamics. Early test systems for quantum molecular dynamics often include photo-dissociation reactions and conical intersections of small molecular systems [232]. Ultimately, these methods may help unlock proton-coupled electron transfer mechanisms [233] in proteins and help with the design of novel organic photovoltaics [228] and related systems.

*Materials science.* Classical methods for materials simulations usually employ density-functional theory coupled with approximation methods, such as the local density approximation [234] to tackle weakly correlated materials. However, many effects arising from strongly correlated systems are beyond the reach of such classical methods. Since long-term algorithms for material simulation require phase estimation [235–237], these lie beyond the scope of near-term devices. In contrast, near-term VQAs for analyzing strong correlation problem are aimed at reducing the circuit depth by using smart initializations [238], or by optimizing the circuit structure itself [29, 30].

## B. Nuclear and particle physics

*Nuclear physics.* Similar to the chemistry applications discussed above, VQAs have the potential to convey a quantum advantage in studying nuclear structure and dynamics. The most studied potential contribution is the utility of the VQE method to find nuclear ground states. This was first demonstrated for computing the deuteron ( $^2H$ ) binding energy [239], and has been extended to other light nuclei such as the triton ( $^3H$ ),  $^3He$ , and an alpha particle ( $^4He$ ) [240]. Additionally, using VQE to prepare the ground state of a triton has been an initial step as a demonstration of simulating neutrino-nucleon scattering [241]. Considering these low-energy applications along with the general progress towards studying higher energy nuclear interactions (i.e., quantum chromodynamics) via VQA lattice gauge theory approaches (discussed below) shows that VQAs have the potential to provide a significant advantage over classical methods for nuclear physics.

*Particle physics.* In particle physics many analytical tools have been developed to describe and study theories, but there are many areas that remain intractable. In particular, the study of important gauge theories like quantum chromodynamics is often handled by mapping the problem onto a lattice to allow for numerical studies. One of the major drawbacks of such Lattice Gauge Theories (LGTs) for classical computation is that they exhibit the sign problem and as a result are usually not

classically simulable. While large scale, fault-tolerant quantum computers will eventually be able to handle this difficulty, there is also the potential for achieving a significant quantum advantage in this area with VQAs in the NISQ era [242]. Recent advances in this direction include work on VQAs for LGT simulation [14] as well as variational determinations of mass gaps, Green’s functions, and running coupling constants [243–245]. In addition, an approach using a VQA to determine interpolation operators to accelerate classical LGT computations has been proposed [246]. Finally, the impacts of decoherence by hardware noise on LGT calculations have been studied, finding that gauge violations caused by decoherence only grow linearly at short times, suggesting that short depth approaches may be possible [247]. Taken together, these results show that studying LGTs is a viable candidate for NISQ quantum advantage.

## C. Optimization and machine learning

While it is natural to consider that VQAs can bring an advantage on task which are inherently quantum in nature, the prospect of employing quantum algorithms to solve classical problems is also an exciting one. Generally, one here aims to employ the large dimension of the Hilbert space to encode large problems or large amounts of data, with the premise that the quantum nature of the algorithm (such as coherence or entanglement between qubits [248]) helps in speeding up a given task.

*Optimization.* Many optimization problems can be encoded in relatively simple mathematical models such as the Max-Cut or the Max-Sat problems. These include tasks like electronic circuitry layout design, state problems in statistical physics [249], and even automotive configuration [250]. Applying QAOA to classical optimization problems is widely considered to be one of the leading candidates for achieving quantum advantage on NISQ devices [110]. There are several reasons for this optimism. QAOA has provable performance guarantees [22, 251] for  $p = 1$ . In general, even  $p = 1$  QAOA ansatz cannot be efficiently simulated on any classical device [252]. At the same time, QAOA performance can only improve by increasing  $p$ . It was also shown that “bang-bang” evolution that motivates QAOA ansatz is the optimal approach given fixed quantum computation time [32]. On the other hand, there are problems for which a shallow QAOA ansatz does not perform well [253, 254] suggesting that  $p$  may have to grow with the problem size. Larger  $p$  requires improvements in the parametrization and optimization [175]. Similarly to quantum chemistry, large scale experiments of QAOA have already been implemented [255].

*Machine Learning.* In the past few decades, the use of machine learning has become common in most, if not all, areas of science. While the problem of loading classical data on quantum computers is still an active topic of research, there has been significant efforts put forward

to use quantum algorithms for machine learning applications [134, 148, 256, 257]. For instance, it has been shown that quantum neural networks can achieve a significantly better effective dimension than comparable classical neural networks [258]. Moreover, it has also been pointed out that quantum algorithms can outperform classical ones in deep learning problems [259], and more recently a VQA has been proposed for deep reinforcement learning [260]. An exciting prospect for using quantum neural networks is that certain architectures are immune to barren plateaus, and hence are trainable even for large problems [153, 154].

## VI. OUTLOOK

In the quest for quantum advantage, analytical and heuristic scaling analysis of VQAs will be increasingly important. Better methods to analyze VQA scalability are anticipated in the future. This will likely include both gradient scaling as well as other scaling aspects, such as the density of local minima and the shape of the cost landscape. These fundamental results will help to guide the search for quantum advantage.

At the same time, the future will also see an improved toolbox for VQAs. Quantum-aware optimizers will exploit knowledge gained about the cost landscape. These improved optimizers will mitigate the impacts of small gradients and avoid local minima to facilitate rapid training of the parameters in VQAs.

Application-specific ansatzes will continue to be developed. Better ansatzes will enhance gradient magnitudes to improve trainability and they may also reduce the impact of noise on VQAs. This will likely include adaptive ansatz strategies, which appear promising.

New error mitigation strategies are anticipated in the future. These will be crucial for obtaining accurate results from VQAs and will improve accuracy by orders of magnitude. Error mitigation will be hard-coded into cloud-based quantum computing platforms, to allow users to obtain accurate results with ease.

The future will also see better quantum hardware become available, both in terms of qubit count and noise levels. VQAs will certainly benefit from such improved hardware. Moreover, VQAs will play a central role in benchmarking the capabilities of these new platforms.

In the near future, VQAs will likely see a shift from the proposal and development phase to the implementation phase. Researchers will aim to implement larger, more realistic problems with VQAs instead of toy problems. These implementations will incorporate multiple state-of-the-art strategies for enhancing VQA performance. Combining strategies for improving the accuracy, trainability, and efficiency of VQAs will test their ultimate capabilities and will push the boundaries of NISQ devices, with the grand vision of obtaining quantum advantage.

In the more distant future, VQAs will even find use

even when the fault-tolerant era arrives. Transitioning from estimating expectation values from Hamiltonian averaging to phase estimation may be an important component here [98]. QAOA may be a good candidate VQA to find usage in the fault-tolerant era, albeit with caveats about the overhead [261]. Strategies that address challenges in the NISQ era, such as keeping circuit depth shallow and avoiding barren plateaus, could still play a role in the fault-tolerant era. Therefore, current research on VQAs will likely remain useful even when fault-tolerant quantum devices arrive.

## VII. REFERENCES

- [1] Richard P Feynman, “Simulating physics with computers,” *Int. J. Theor. Phys* **21** (1982).
- [2] Peter W Shor, “Algorithms for quantum computation: discrete logarithms and factoring,” in *Proceedings 35th annual symposium on foundations of computer science* (Ieee, 1994) pp. 124–134.
- [3] Seth Lloyd, “Universal quantum simulators,” *Science* **273**, 1073–1078 (1996).
- [4] Aram W Harrow, Avinandan Hassidim, and Seth Lloyd, “Quantum algorithm for linear systems of equations,” *Physical Review Letters* **103**, 150502 (2009).
- [5] “IBM Makes Quantum Computing Available on IBM Cloud to Accelerate Innovation,” (2016), press release at <https://www-03.ibm.com/press/us/en/pressrelease/49661.wss>.
- [6] Adetokunbo Adedoyin, John Ambrosiano, Petr Anisimov, Andreas Bärtzsch, William Casper, Gopinath Chennupati, Carleton Coffrin, Hristo Djidjev, David Gunter, Satish Karra, *et al.*, “Quantum algorithm implementations for beginners,” *arXiv preprint arXiv:1804.03719* (2018).
- [7] J. Preskill, “Quantum computing in the NISQ era and beyond,” *Quantum* **2**, 79 (2018).
- [8] Frank Arute *et al.*, “Quantum supremacy using a programmable superconducting processor,” *Nature* **574**, 505–510 (2019).
- [9] Han-Sen Zhong, Hui Wang, Yu-Hao Deng, Ming-Cheng Chen, Li-Chao Peng, Yi-Han Luo, Jian Qin, Dian Wu, Xing Ding, Yi Hu, *et al.*, “Quantum computational advantage using photons,” *Science* (2020).
- [10] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, “Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets,” *Nature* **549**, 242 (2017).
- [11] Bryan T Gard, Linghua Zhu, George S Barron, Nicholas J Mayhall, Sophia E Economou, and Edwin Barnes, “Efficient symmetry-preserving state preparation circuits for the variational quantum eigensolver algorithm,” *npj Quantum Information* **6**, 1–9 (2020).
- [12] Matthew Otten, Cristian L Cortes, and Stephen K Gray, “Noise-resilient quantum dynamics using symmetry-preserving ansatzes,” *arXiv preprint arXiv:1910.06284* (2019).
- [13] Nikolay V Tkachenko, James Sud, Yu Zhang, Sergei Tretiak, Petr M Anisimov, Andrew T Arrasmith, Patrick J Coles, Lukasz Cincio, and Pavel A



- Dub, “Correlation-informed permutation of qubits for reducing ansatz depth in vqe,” [arXiv preprint arXiv:2009.04996](#) (2020).
- [14] Christian Kokail, Christine Maier, Rick van Bijnen, Tiff Brydges, Manoj K Joshi, Petar Jurcevic, Christine A Muschik, Pietro Silvi, Rainer Blatt, Christian F Roos, *et al.*, “Self-verifying variational quantum simulation of lattice models,” *Nature* **569**, 355–360 (2019).
- [15] Andrew G. Taube and Rodney J. Bartlett, “New perspectives on unitary coupled-cluster theory,” *International Journal of Quantum Chemistry* **106**, 3393–3401 (2006).
- [16] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O’Brien, “A variational eigenvalue solver on a photonic quantum processor,” *Nature Communications* **5**, 4213 (2014).
- [17] Joonho Lee, William J. Huggins, Martin Head-Gordon, and K. Birgitta Whaley, “Generalized unitary coupled cluster wave functions for quantum computation,” *Journal of Chemical Theory and Computation* **15**, 311–324 (2019).
- [18] Mario Motta, Erika Ye, Jarrod R McClean, Zhen-dong Li, Austin J Minnich, Ryan Babbush, and Garnet Kin Chan, “Low rank representations for quantum simulation of electronic structure,” [arXiv preprint arXiv:1808.02625](#) (2018).
- [19] Yuta Matsuzawa and Yuki Kurashige, “Jastrow-type decomposition in quantum chemistry for low-depth quantum circuits,” *Journal of Chemical Theory and Computation* **16**, 944–952 (2020).
- [20] Ian D Kivlichan, Jarrod McClean, Nathan Wiebe, Craig Gidney, Alán Aspuru-Guzik, Garnet Kin-Lic Chan, and Ryan Babbush, “Quantum simulation of electronic structure with linear depth and connectivity,” *Physical Review Letters* **120**, 110501 (2018).
- [21] Kanav Setia, Sergey Bravyi, Antonio Mezzacapo, and James D Whitfield, “Superfast encodings for fermionic quantum simulation,” *Physical Review Research* **1**, 033033 (2019).
- [22] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann, “A quantum approximate optimization algorithm,” [arXiv preprint arXiv:1411.4028](#) (2014).
- [23] S. Hadfield, Z. Wang, B. O’Gorman, E. G. Rieffel, D. Venturelli, and R. Biswas, “From the quantum approximate optimization algorithm to a quantum alternating operator ansatz,” *Algorithms* **12**, 34 (2019).
- [24] Seth Lloyd, “Quantum approximate optimization is computationally universal,” [arXiv preprint arXiv:1812.11075](#) (2018).
- [25] Zhihui Wang, Nicholas C Rubin, Jason M Dominy, and Eleanor G Rieffel, “X y mixers: Analytical and numerical results for the quantum alternating operator ansatz,” *Physical Review A* **101**, 012320 (2020).
- [26] Dave Wecker, Matthew B Hastings, and Matthias Troyer, “Progress towards practical quantum variational algorithms,” *Physical Review A* **92**, 042303 (2015).
- [27] Roeland Wiersema, Cunlu Zhou, Yvette de Sereville, Juan Felipe Carrasquilla, Yong Baek Kim, and Henry Yuen, “Exploring entanglement and optimization within the hamiltonian variational ansatz,” *P R X Quantum* **1**, 020319 (2020).
- [28] Wen Wei Ho and Timothy H Hsieh, “Efficient variational simulation of non-trivial quantum states,” *SciPost Phys* **6**, 029 (2019).
- [29] Harper R Grimsley, Sophia E Economou, Edwin Barnes, and Nicholas J Mayhall, “An adaptive variational algorithm for exact molecular simulations on a quantum computer,” *Nature Communications* **10**, 1–9 (2019).
- [30] Ho Lun Tang, VO Shkolnikov, George S Barron, Harper R Grimsley, Nicholas J Mayhall, Edwin Barnes, and Sophia E Economou, “qubit-adapt-vqe: An adaptive algorithm for constructing hardware-efficient ansatzes on a quantum processor,” [arXiv preprint arXiv:1911.10205](#) (2019).
- [31] Arthur G. Rattew, Shaohan Hu, Marco Pistoia, Richard Chen, and Steve Wood, “A domain-agnostic, noise-resistant, hardware-efficient evolutionary variational quantum eigensolver,” [arXiv preprint arXiv:1910.09694](#) (2019).
- [32] Zhi-Cheng Yang, Armin Rahmani, Alireza Shabani, Hartmut Neven, and Claudio Chamón, “Optimizing variational quantum algorithms using pontryagin’s minimum principle,” *Phys. Rev. X* **7**, 021027 (2017).
- [33] Alicia B Magann, Christian Arenz, Matthew D Grace, Tak-San Ho, Robert L Kosut, Jarrod R McClean, Herschel A Rabitz, and Mohan Sarovar, “From pulses to circuits and back again: A quantum optimal control perspective on variational quantum algorithms,” [arXiv preprint arXiv:2009.06702](#) (2020).
- [34] Alexandre Choquette, Agustin Di Paolo, Panagiotis Kl Barkoutsos, David Sénéchal, Ivano Tavernelli, and Alexandre Blais, “Quantum-optimal-control-inspired ansatz for variational quantum algorithms,” [arXiv preprint arXiv:2008.01098](#) (2020).
- [35] Jun Li, Xiaodong Yang, Xinhua Peng, and Chang-Pu Sun, “Hybrid quantum-classical approach to quantum optimal control,” *Phys. Rev. Lett.* **118**, 150503 (2017).
- [36] Dawei Lu, Keren Li, Jun Li, Hemant Katiyar, Annie Jihyun Park, Guanru Feng, Tao Xin, Hang Li, Guilu Long, Aharon Brodutch, Jonathan Baugh, Bei Zeng, and Raymond Laflamme, “Enhancing quantum control by bootstrapping a quantum processor of 12 qubits,” *npj Quantum Information* **3**, 45 (2017).
- [37] Peter JJ O’Malley, Ryan Babbush, Ian D Kivlichan, Jonathan Romero, Jarrod R McClean, Rami Barends, Julian Kelly, Pedram Roushan, Andrew Tranter, Nan Ding, *et al.*, “Scalable quantum simulation of molecular energies,” *Physical Review X* **6**, 031007 (2016).
- [38] Tyler Takeshita, Nicholas C. Rubin, Zhang Jiang, Eun-seok Lee, Ryan Babbush, and Jarrod R. McClean, “Increasing the representation accuracy of quantum simulations of chemistry without extra quantum resources,” *Phys. Rev. X* **10**, 011004 (2020).
- [39] Leslie G. Valiant, “Quantum circuits that can be simulated classically in polynomial time,” *SIAM Journal on Computing* **31**, 1229–1254 (2002).
- [40] Barbara M. Terhal and David P. DiVincenzo, “Classical simulation of noninteracting-fermion quantum circuits,” *Phys. Rev. A* **65**, 032325 (2002).
- [41] Richard Jozsa and Akimasa Miyake, “Matchgates and classical simulation of quantum circuits,” *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **464**, 3089–3106 (2008).
- [42] Wataru Mizukami, Kosuke Mitarai, Yuya O. Nakagawa, Takahiro Yamamoto, Tennin Yan, and Yu-ya Ohnishi, “Orbital optimized unitary coupled cluster theory for quantum computer,” *Phys. Rev. Research* **2**, 033421

- (2020).
- [43] Igor O. Sokolov, Panagiotis Kl. Barkoutsos, Pauline J. Ollitrault, Donny Greenberg, Julia Rice, Marco Pistoia, and Ivano Tavernelli, “Quantum orbital-optimized unitary coupled cluster methods in the strongly correlated regime: Can quantum algorithms outperform their classical equivalents?” *The Journal of Chemical Physics* **152**, 124107 (2020).
  - [44] Jarrod R McClean, Mollie E Kimchi-Schwartz, Jonathan Carter, and Wibe A de Jong, “Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states,” *Physical Review A* **95**, 042308 (2017).
  - [45] Robert M Parrish, Edward G Hohenstein, Peter L McMahon, and Todd J Martínez, “Quantum computation of electronic transitions using a variational quantum eigensolver,” *Physical Review Letters* **122**, 230401 (2019).
  - [46] Robert M Parrish and Peter L McMahon, “Quantum filter diagonalization: Quantum eigendecomposition without full quantum phase estimation,” *arXiv preprint arXiv:1909.08925* (2019).
  - [47] William J Huggins, Joonho Lee, Unpil Baek, Bryan O’Gorman, and K Birgitta Whaley, “A non-orthogonal variational quantum eigensolver,” *New Journal of Physics* **22**, 073009 (2020).
  - [48] Nicholas H Stair, Renke Huang, and Francesco A Evangelista, “A multireference quantum krylov algorithm for strongly correlated electrons,” *Journal of Chemical Theory and Computation* **16**, 2236–2245 (2020).
  - [49] Kishor Bharti and Tobias Haug, “Iterative quantum assisted eigensolver,” *arXiv preprint arXiv:2010.05638* (2020).
  - [50] Kishor Bharti and Tobias Haug, “Quantum assisted simulator,” *arXiv preprint arXiv:2011.06911* (2020).
  - [51] Igor L. Markov and Yaoyun Shi, “Simulating quantum computation by contracting tensor networks,” *SIAM Journal on Computing* **38**, 963–981 (2008).
  - [52] Jin-Guo Liu, Yi-Hong Zhang, Yuan Wan, and Lei Wang, “Variational quantum eigensolver with fewer qubits,” *Phys. Rev. Research* **1**, 023025 (2019).
  - [53] Fergus Barratt, James Dborin, Matthias Bal, Vid Stojevic, Frank Pollmann, and Andrew G Green, “Parallel quantum simulation of large systems on small quantum computers,” *arXiv preprint arXiv:2003.12087* (2020).
  - [54] Xiao Yuan, Jinzhao Sun, Junyu Liu, Qi Zhao, and You Zhou, “Quantum simulation with hybrid tensor networks,” *arXiv preprint arXiv:2007.00958* (2020).
  - [55] Keisuke Fujii, Kosuke Mitarai, Wataru Mizukami, and Yuya O Nakagawa, “Deep variational quantum eigensolver: a divide-and-conquer method for solving a larger problem with smaller size quantum computers,” *arXiv preprint arXiv:2007.10917* (2020).
  - [56] Guglielmo Mazzola, Pauline J. Ollitrault, Panagiotis Kl. Barkoutsos, and Ivano Tavernelli, “Nonunitary operations for ground-state calculations in near-term quantum computers,” *Phys. Rev. Lett.* **123**, 130501 (2019).
  - [57] John Martyn and Brian Swingle, “Product spectrum ansatz and the simplicity of thermal states,” *Phys. Rev. A* **100**, 032107 (2019).
  - [58] Nobuyuki Yoshioka, Yuya O Nakagawa, Kosuke Mitarai, and Keisuke Fujii, “Variational quantum algorithm for nonequilibrium steady states,” *Physical Review Research* **2**, 043289 (2020).
  - [59] Guillaume Verdon, Jacob Marks, Sasha Nanda, Stefan Leichenauer, and Jack Hidary, “Quantum hamiltonian-based models and the variational quantum thermalizer algorithm,” *arXiv preprint arXiv:1910.02071* (2019).
  - [60] JinGuo Liu, Liang Mao, Pan Zhang, and Lei Wang, “Solving quantum statistical mechanics with variational autoregressive networks and quantum circuits,” *Machine Learning: Science and Technology* (2020).
  - [61] Sukin Sim, Peter D Johnson, and Alán Aspuru-Guzik, “Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms,” *Advanced Quantum Technologies* **2**, 1900070 (2019).
  - [62] Kouhei Nakaji and Naoki Yamamoto, “Expressibility of the alternating layered ansatz for quantum computation,” *arXiv preprint arXiv:2005.12537* (2020).
  - [63] Gian Giacomo Guerreschi and Mikhail Smelyanskiy, “Practical optimization for hybrid quantum-classical algorithms,” *arXiv preprint arXiv:1701.01450* (2017).
  - [64] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, “Quantum circuit learning,” *Phys. Rev. A* **98**, 032309 (2018).
  - [65] Maria Schuld, Ville Bergholm, Christian Gogolin, Josh Izaac, and Nathan Killoran, “Evaluating analytic gradients on quantum hardware,” *Phys. Rev. A* **99**, 032331 (2019).
  - [66] Ville Bergholm, Josh Izaac, Maria Schuld, Christian Gogolin, M Sohaib Alam, Shahnawaz Ahmed, Juan Miguel Arrazola, Carsten Blank, Alain Delgado, Soran Jahangiri, *et al.*, “Pennylane: Automatic differentiation of hybrid quantum-classical computations,” *arXiv preprint arXiv:1811.04968* (2018).
  - [67] Andrea Mari, Thomas R Bromley, and Nathan Killoran, “Estimating the gradient and higher-order derivatives on quantum hardware,” *arXiv preprint arXiv:2008.06517* (2020).
  - [68] M. Cerezo and Patrick J Coles, “Impact of barren plateaus on the hessian and higher order derivatives,” *arXiv preprint arXiv:2008.07454* (2020).
  - [69] Ken M Nakanishi, Keisuke Fujii, and Synge Todo, “Sequential minimal optimization for quantum-classical hybrid algorithms,” *Physical Review Research* **2**, 043158 (2020).
  - [70] Bálint Koczor and Simon C Benjamin, “Quantum analytic descent,” *arXiv preprint arXiv:2008.13774* (2020).
  - [71] Sam McArdle, Tyson Jones, Suguru Endo, Ying Li, Simon C Benjamin, and Xiao Yuan, “Variational ansatz-based quantum simulation of imaginary time evolution,” *npj Quantum Information* **5**, 1–6 (2019).
  - [72] James Stokes, Josh Izaac, Nathan Killoran, and Giuseppe Carleo, “Quantum natural gradient,” *Quantum* **4**, 269 (2020).
  - [73] Bálint Koczor and Simon C Benjamin, “Quantum natural gradient generalised to non-unitary circuits,” *arXiv preprint arXiv:1912.08660* (2019).
  - [74] Ying Li and Simon C Benjamin, “Efficient variational quantum simulator incorporating active error minimization,” *Physical Review X* **7**, 021050 (2017).
  - [75] Xiao Yuan, Suguru Endo, Qi Zhao, Ying Li, and Simon C Benjamin, “Theory of variational quantum simulation,” *Quantum* **3**, 191 (2019).
  - [76] Suguru Endo, Jinzhao Sun, Ying Li, Simon C Benjamin, and Xiao Yuan, “Variational quantum simulation of general processes,” *Physical Review Letters* **125**, 010501 (2020).

- (2020).
- [77] Kosuke Mitarai and Keisuke Fujii, “Methodology for replacing indirect measurements with direct measurements,” *Phys. Rev. Research* **1**, 013006 (2019).
  - [78] Diederik P Kingma and Jimmy Ba, “Adam: A method for stochastic optimization,” in *Proceedings of the 3rd International Conference on Learning Representations (ICLR)* (2015).
  - [79] Jonas M Kübler, Andrew Arrasmith, Lukasz Cincio, and Patrick J Coles, “An adaptive optimizer for measurement-frugal variational algorithms,” *Quantum* **4**, 263 (2020).
  - [80] Max Wilson, Sam Stromswold, Filip Wudarski, Stuart Hadfield, Norm M Tubman, and Eleanor Rieffel, “Optimizing quantum heuristics with meta-learning,” *arXiv preprint arXiv:1908.03185* (2019).
  - [81] James C Spall, “Multivariate stochastic approximation using a simultaneous perturbation gradient approximation,” *IEEE transactions on automatic control* **37**, 332–341 (1992).
  - [82] Robert M Parrish, Joseph T Iosue, Asier Ozaeta, and Peter L McMahon, “A jacobi diagonalization and anderson acceleration algorithm for variational quantum algorithm parameter optimization,” *arXiv preprint arXiv:1904.03206* (2019).
  - [83] Patrick Huembeli and Alexandre Dauphin, “Characterizing the loss landscape of variational quantum circuits,” *arXiv preprint arXiv:2008.02785* (2020).
  - [84] Aram Harrow and John Napp, “Low-depth gradient measurements can improve convergence in variational hybrid quantum-classical algorithms,” *arXiv preprint arXiv:1901.05374* (2019).
  - [85] Ryan Sweke, Frederik Wilde, Johannes Jakob Meyer, Maria Schuld, Paul K Fährmann, Barthélemy Meynard-Piganeau, and Jens Eisert, “Stochastic gradient descent for hybrid quantum-classical optimization,” *Quantum* **4**, 314 (2020).
  - [86] Jacob Biamonte, “Universal variational quantum computation,” *arXiv preprint arXiv:1903.04500* (2019).
  - [87] Daniel S. Abrams and Seth Lloyd, “Quantum algorithm providing exponential speed increase for finding eigenvalues and eigenvectors,” *Phys. Rev. Lett.* **83**, 5162–5165 (1999).
  - [88] Alán Aspuru-Guzik, Anthony D Dutoi, Peter J Love, and Martin Head-Gordon, “Simulated quantum computation of molecular energies,” *Science* **309**, 1704–1707 (2005).
  - [89] Abhinav Kandala, Kristan Temme, Antonio D Córcoles, Antonio Mezzacapo, Jerry M Chow, and Jay M Gambetta, “Error mitigation extends the computational reach of a noisy quantum processor,” *Nature* **567**, 491–495 (2019).
  - [90] Oscar Higgott, Daochen Wang, and Stephen Brierley, “Variational quantum computation of excited states,” *Quantum* **3**, 156 (2019).
  - [91] Harry Buhrman, Richard Cleve, John Watrous, and Ronald De Wolf, “Quantum fingerprinting,” *Physical Review Letters* **87**, 167902 (2001).
  - [92] L. Cincio, Y. Subaşı, A. T. Sornborger, and P. J. Coles, “Learning the quantum algorithm for state overlap,” *New Journal of Physics* **20**, 113022 (2018).
  - [93] Tyson Jones, Suguru Endo, Sam McArdle, Xiao Yuan, and Simon C Benjamin, “Variational quantum algorithms for discovering hamiltonian spectra,” *Physical Review A* **99**, 062304 (2019).
  - [94] Ken M Nakanishi, Kosuke Mitarai, and Keisuke Fujii, “Subspace-search variational quantum eigensolver for excited states,” *Physical Review Research* **1**, 033062 (2019).
  - [95] Jarrod R McClean, Matthew P Harrigan, Masoud Mohseni, Nicholas C Rubin, Zhang Jiang, Sergio Boixo, Vadim N Smelyanskiy, Ryan Babbush, and Hartmut Neven, “Low depth mechanisms for quantum optimization,” *arXiv preprint arXiv:2008.08615* (2020).
  - [96] A Garcia-Saez and JI Latorre, “Addressing hard classical problems with adiabatically assisted variational quantum eigensolvers,” *arXiv preprint arXiv:1806.02287* (2018).
  - [97] M. Cerezo, Kunal Sharma, Andrew Arrasmith, and Patrick J Coles, “Variational quantum state eigensolver,” *arXiv preprint arXiv:2004.01372* (2020).
  - [98] Daochen Wang, Oscar Higgott, and Stephen Brierley, “Accelerated variational quantum eigensolver,” *Physical Review Letters* **122**, 140504 (2019).
  - [99] Yong-Xin Yao, Niladri Gomes, Feng Zhang, Thomas Iadecola, Cai-Zhuang Wang, Kai-Ming Ho, and Peter P. Orth, “Adaptive variational quantum dynamics simulations,” *arXiv preprint arXiv:2011.00622* (2020).
  - [100] Zi-Jian Zhang, Jinzhao Sun, Xiao Yuan, and Man-Hong Yung, “Low-depth hamiltonian simulation by adaptive product formula,” *arXiv preprint arXiv:2011.05283* (2020).
  - [101] Kentaro Heya, Ken M Nakanishi, Kosuke Mitarai, and Keisuke Fujii, “Subspace variational quantum simulator,” *arXiv preprint arXiv:1904.08566* (2019).
  - [102] Cristina Cirstoiu, Zoe Holmes, Joseph Iosue, Lukasz Cincio, Patrick J Coles, and Andrew Sornborger, “Variational fast forwarding for quantum simulation beyond the coherence time,” *npj Quantum Information* **6**, 1–10 (2020).
  - [103] S. Khatiri, R. LaRose, A. Poremba, L. Cincio, A. T. Sornborger, and P. J. Coles, “Quantum-assisted quantum compiling,” *Quantum* **3**, 140 (2019).
  - [104] Benjamin Commeau, M. Cerezo, Zoë Holmes, Lukasz Cincio, Patrick J Coles, and Andrew Sornborger, “Variational hamiltonian diagonalization for dynamical quantum simulation,” *arXiv preprint arXiv:2009.02559* (2020).
  - [105] Nikolaj Moll, Panagiotis Barkoutsos, Lev S Bishop, Jerry M Chow, Andrew Cross, Daniel J Egger, Stefan Filipp, Andreas Fuhrer, Jay M Gambetta, Marc Ganzhorn, *et al.*, “Quantum optimization using variational algorithms on near-term quantum devices,” *Quantum Science and Technology* **3**, 030503 (2018).
  - [106] Cedric Yen-Yu Lin and Yechao Zhu, “Performance of qaoa on typical instances of constraint satisfaction problems with bounded degree,” *arXiv preprint arXiv:1601.01744* (2016).
  - [107] Z. Wang, S. Hadfield, Z. Jiang, and E. G. Rieffel, “Quantum approximate optimization algorithm for MaxCut: A fermionic view,” *Phys. Rev. A* **97**, 022304 (2018).
  - [108] Ruslan Shaydulin, Ilya Safro, and Jeffrey Larson, “Multistart methods for quantum approximate optimization,” in *2019 IEEE High Performance Extreme Computing Conference (HPEC)* (IEEE, 2019) pp. 1–8.
  - [109] Jonathan Romero, Ryan Babbush, Jarrod R McClean, Cornelius Hempel, Peter J Love, and Alán Aspuru-



- Guzik, “Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz,” *Quantum Science and Technology* **4**, 014008 (2018).
- [110] Gavin E Crooks, “Performance of the quantum approximate optimization algorithm on the maximum cut problem,” *arXiv preprint arXiv:1811.08419* (2018).
- [111] Dave Wecker, Matthew B Hastings, and Matthias Troyer, “Training a quantum optimizer,” *Physical Review A* **94**, 022309 (2016).
- [112] Sami Khairy, Ruslan Shaydulin, Lukasz Cincio, Yuri Alexeev, and Prasanna Balaprakash, “Learning to optimize variational quantum circuits to solve combinatorial problems,” *Proceedings of the AAAI Conference on Artificial Intelligence* **34**, 2367–2375 (2020).
- [113] A Ambainis, “Variable time amplitude amplification and a faster quantum algorithm for solving systems of linear equations 29th int,” in *Symp. Theoretical Aspects of Computer Science (STACS 2012)*, Vol. 14 (2012) pp. 636–47.
- [114] Y. Subaşı, R. D. Somma, and D. Orsucci, “Quantum algorithms for systems of linear equations inspired by adiabatic quantum computing,” *Phys. Rev. Lett.* **122**, 060504 (2019).
- [115] A. Childs, R. Kothari, and R. Somma, “Quantum algorithm for systems of linear equations with exponentially improved dependence on precision,” *SIAM J. Computing* **46**, 1920–1950 (2017).
- [116] Shantanav Chakraborty, András Gilyén, and Stacey Jeffery, “The Power of Block-Encoded Matrix Powers: Improved Regression Techniques via Faster Hamiltonian Simulation,” in *46th International Colloquium on Automata, Languages, and Programming (ICALP 2019)*, Leibniz International Proceedings in Informatics (LIPIcs), Vol. 132 (2019) pp. 33:1–33:14.
- [117] A. Scherer, B. Valiron, S.-C. Mau, S. Alexander, E. Van den Berg, and T. E. Chapuran, “Concrete resource analysis of the quantum linear-system algorithm used to compute the electromagnetic scattering cross section of a 2D target,” *Quantum Information Processing* **16**, 60 (2017).
- [118] Carlos Bravo-Prieto, Ryan LaRose, M. Cerezo, Yigit Subasi, Lukasz Cincio, and Patrick J Coles, “Variational quantum linear solver: A hybrid algorithm for linear systems,” *arXiv preprint arXiv:1909.05820* (2019).
- [119] Xiaosi Xu, Jinzhao Sun, Suguru Endo, Ying Li, Simon C Benjamin, and Xiao Yuan, “Variational algorithms for linear algebra,” *arXiv preprint arXiv:1909.03898* (2019).
- [120] Hsin-Yuan Huang, Kishor Bharti, and Patrick Rebentrost, “Near-term quantum algorithms for linear systems of equations,” *arXiv preprint arXiv:1909.07344* (2019).
- [121] Yiğit Subaşı, Rolando D Somma, and Davide Orsucci, “Quantum algorithms for systems of linear equations inspired by adiabatic quantum computing,” *Physical Review Letters* **122**, 060504 (2019).
- [122] Michael Lubasch, Jaewoo Joo, Pierre Moinier, Martin Kiffner, and Dieter Jaksch, “Variational quantum algorithms for nonlinear problems,” *Physical Review A* **101**, 010301 (2020).
- [123] Eric Anschuetz, Jonathan Olson, Alán Aspuru-Guzik, and Yudong Cao, “Variational quantum factoring,” in *International Workshop on Quantum Technology and Optimization Problems* (Springer, 2019) pp. 74–85.
- [124] Seth Lloyd, Masoud Mohseni, and Patrick Rebentrost, “Quantum principal component analysis,” *Nature Physics* **10**, 631–633 (2014).
- [125] Ryan LaRose, Arkin Tikku, Étude O’Neel-Judy, Lukasz Cincio, and Patrick J Coles, “Variational quantum state diagonalization,” *npj Quantum Information* **5**, 1–10 (2019).
- [126] Kentaro Heya, Yasunari Suzuki, Yasunobu Nakamura, and Keisuke Fujii, “Variational quantum gate optimization,” *arXiv preprint arXiv:1810.12745* (2018).
- [127] Tyson Jones and Simon C Benjamin, “Quantum compilation and circuit optimisation via energy dissipation,” *arXiv preprint arXiv:1811.03147* (2018).
- [128] Kunal Sharma, Sumeet Khatri, M. Cerezo, and Patrick J Coles, “Noise resilience of variational quantum compiling,” *New Journal of Physics* **22**, 043006 (2020).
- [129] Jacques Carolan, Masoud Mohseni, Jonathan P Olson, Mihika Prabhu, Changchen Chen, Darius Bunandar, Murphy Yuezhen Niu, Nicholas C Harris, Franco NC Wong, Michael Hochberg, *et al.*, “Variational quantum unsampling on a quantum photonic processor,” *Nature Physics*, 1–6 (2020).
- [130] Peter D Johnson, Jonathan Romero, Jonathan Olson, Yudong Cao, and Alán Aspuru-Guzik, “Qvector: an algorithm for device-tailored quantum error correction,” *arXiv preprint arXiv:1711.02249* (2017).
- [131] Xiaosi Xu, Simon C Benjamin, and Xiao Yuan, “Variational circuit compiler for quantum error correction,” *arXiv preprint arXiv:1911.05759* (2019).
- [132] Edward Farhi and Hartmut Neven, “Classification with quantum neural networks on near term processors,” *arXiv preprint arXiv:1802.06002* (2018).
- [133] Maria Schuld, Alex Bocharov, Krysta M Svore, and Nathan Wiebe, “Circuit-centric quantum classifiers,” *Physical Review A* **101**, 032308 (2020).
- [134] Maria Schuld and Nathan Killoran, “Quantum machine learning in feature hilbert spaces,” *Physical Review Letters* **122**, 040504 (2019).
- [135] Vojtěch Havlíček, Antonio D Córcoles, Kristan Temme, Aram W Harrow, Abhinav Kandala, Jerry M Chow, and Jay M Gambetta, “Supervised learning with quantum-enhanced feature spaces,” *Nature* **567**, 209–212 (2019).
- [136] Edwin Stoudenmire and David J Schwab, “Supervised learning with tensor networks,” in *Advances in Neural Information Processing Systems* (2016) pp. 4799–4807.
- [137] Takeru Kusumoto, Kosuke Mitarai, Keisuke Fujii, Masahiro Kitagawa, and Makoto Negoro, “Experimental quantum kernel machine learning with nuclear spins in a solid,” *arXiv preprint arXiv:1911.12021* (2019).
- [138] J. Romero, J. P. Olson, and A. Aspuru-Guzik, “Quantum autoencoders for efficient compression of quantum data,” *Quantum Science and Technology* **2**, 045001 (2017).
- [139] Kwok Ho Wan, Oscar Dahlsten, Hlér Kristjánsson, Robert Gardner, and MS Kim, “Quantum generalisation of feedforward neural networks,” *npj Quantum Information* **3**, 36 (2017).
- [140] Guillaume Verdon, Jason Pye, and Michael Broughton, “A universal training algorithm for quantum deep learning,” *arXiv preprint arXiv:1806.09729* (2018).
- [141] M. Cerezo, Akira Sone, Tyler Volkoff, Lukasz Cincio, and Patrick J Coles, “Cost-function-dependent barren plateaus in shallow quantum neural networks,” *arXiv preprint arXiv:2001.00550* (2020).



- [142] Alex Pepper, Nora Tischler, and Geoff J Pryde, “Experimental realization of a quantum autoencoder: The compression of qutrits via machine learning,” *Physical Review Letters* **122**, 060501 (2019).
- [143] Yuxuan Du, Min-Hsiu Hsieh, Tongliang Liu, and Dacheng Tao, “Expressive power of parametrized quantum circuits,” *Physical Review Research* **2**, 033125 (2020).
- [144] Guillaume Verdon, Michael Broughton, and Jacob Biamonte, “A quantum algorithm to train neural networks using low-depth circuits,” *arXiv preprint arXiv:1712.05304* (2017).
- [145] Jin-Guo Liu and Lei Wang, “Differentiable learning of quantum circuit born machines,” *Physical Review A* **98**, 062324 (2018).
- [146] Marcello Benedetti, Delfina Garcia-Pintos, Oscar Perdomo, Vicente Leyton-Ortega, Yunseong Nam, and Alejandro Perdomo-Ortiz, “A generative modeling approach for benchmarking and training shallow quantum circuits,” *npj Quantum Information* **5**, 1–9 (2019).
- [147] Brian Coyle, Daniel Mills, Vincent Danos, and Elham Kashefi, “The born supremacy: Quantum advantage and training of an ising born machine,” *npj Quantum Information* **6**, 1–11 (2020).
- [148] Jonathan Romero and Alan Aspuru-Guzik, “Variational quantum generators: Generative adversarial quantum machine learning for continuous distributions,” *arXiv preprint arXiv:1901.00848* (2019).
- [149] MV Altaisky, “Quantum neural network,” *arXiv preprint quant-ph/0107012* (2001).
- [150] Kerstin Beer, Dmytro Bondarenko, Terry Farrelly, Tobias J Osborne, Robert Salzmann, Daniel Scheiermann, and Ramona Wolf, “Training deep quantum neural networks,” *Nature communications* **11**, 1–6 (2020).
- [151] Iris Cong, Soonwon Choi, and Mikhail D Lukin, “Quantum convolutional neural networks,” *Nature Physics* **15**, 1273–1278 (2019).
- [152] Lukas Franken and Bogdan Georgiev, “Explorations in quantum neural networks with intermediate measurements,” in *Proceedings of ESANN* (2020).
- [153] Arthur Pesah, M. Cerezo, Samson Wang, Tyler Volkoff, Andrew T Sornborger, and Patrick J Coles, “Absence of barren plateaus in quantum convolutional neural networks,” *arXiv preprint arXiv:2011.02966* (2020).
- [154] Kaining Zhang, Min-Hsiu Hsieh, Liu Liu, and Dacheng Tao, “Toward trainability of quantum neural networks,” *arXiv preprint arXiv:2011.06258* (2020).
- [155] A. Arrasmith, L. Cincio, A. T. Sornborger, W. H. Zurek, and P. J. Coles, “Variational consistent histories as a hybrid algorithm for quantum foundations,” *Nature communications* **10**, 3438 (2019).
- [156] Mark M Wilde, *Quantum information theory* (Cambridge University Press, 2013).
- [157] B. Rosgen and J. Watrous, “On the hardness of distinguishing mixed-state quantum computations,” in *20th Annual IEEE Conference on Computational Complexity (CCC’05)* (2005) pp. 344–354.
- [158] M. Cerezo, Alexander Poremba, Lukasz Cincio, and Patrick J Coles, “Variational quantum fidelity estimation,” *Quantum* **4**, 248 (2020).
- [159] Carlos Bravo-Prieto, Diego García-Martín, and José I Latorre, “Quantum singular value decomposer,” *Physical Review A* **101**, 062310 (2020).
- [160] Bálint Koczor, Suguru Endo, Tyson Jones, Yuichiro Matsuzaki, and Simon C Benjamin, “Variational-state quantum metrology,” *New Journal of Physics* **22**, 083038 (2020).
- [161] Raphael Kaubruegger, Pietro Silvi, Christian Kokail, Rick van Bijnen, Ana Maria Rey, Jun Ye, Adam M Kaufman, and Peter Zoller, “Variational spin-squeezing algorithms on programmable quantum sensors,” *Physical Review Letters* **123**, 260505 (2019).
- [162] Ziqi Ma, Pranav Gokhale, Tian-Xing Zheng, Sisi Zhou, Xiaofei Yu, Liang Jiang, Peter Maurer, and Frederic T Chong, “Adaptive circuit learning for quantum metrology,” *arXiv preprint arXiv:2010.08702* (2020).
- [163] Jacob L Beckey, M. Cerezo, Akira Sone, and Patrick J Coles, “Variational quantum algorithm for estimating the quantum fisher information,” *arXiv preprint arXiv:2010.10488* (2020).
- [164] Jarrod R McClean, Sergio Boixo, Vadim N Smelyanskiy, Ryan Babbush, and Hartmut Neven, “Barren plateaus in quantum neural network training landscapes,” *Nature communications* **9**, 4812 (2018).
- [165] Andrew Arrasmith, M. Cerezo, Piotr Czarnik, Lukasz Cincio, and Patrick J Coles, “Effect of barren plateaus on gradient-free optimization,” *arXiv preprint arXiv:2011.12245* (2020).
- [166] Aram W Harrow and Richard A Low, “Random quantum circuits are approximate 2-designs,” *Communications in Mathematical Physics* **291**, 257–302 (2009).
- [167] Fernando GSL Brandao, Aram W Harrow, and Michał Horodecki, “Local random quantum circuits are approximate polynomial-designs,” *Communications in Mathematical Physics* **346**, 397–434 (2016).
- [168] Alexey Uvarov, Jacob D. Biamonte, and Dmitry Yudin, “Variational quantum eigensolver for frustrated quantum systems,” *Phys. Rev. B* **102**, 075104 (2020).
- [169] Alexey Uvarov and Jacob Biamonte, “On barren plateaus and cost function locality in variational quantum algorithms,” *arXiv preprint arXiv:2011.10530* (2020).
- [170] Zoë Holmes, Andrew Arrasmith, Bin Yan, Patrick J Coles, Andreas Albrecht, and Andrew T Sornborger, “Barren plateaus preclude learning scramblers,” *arXiv preprint arXiv:2009.14808* (2020).
- [171] Kunal Sharma, M. Cerezo, Lukasz Cincio, and Patrick J Coles, “Trainability of dissipative perceptron-based quantum neural networks,” *arXiv preprint arXiv:2005.12458* (2020).
- [172] Carlos Ortiz Marrero, Mária Kieferová, and Nathan Wiebe, “Entanglement induced barren plateaus,” *arXiv preprint arXiv:2010.15968* (2020).
- [173] Samson Wang, Enrico Fontana, M. Cerezo, Kunal Sharma, Akira Sone, Lukasz Cincio, and Patrick J Coles, “Noise-induced barren plateaus in variational quantum algorithms,” *arXiv preprint arXiv:2007.14384* (2020).
- [174] Daniel Stilck Franca and Raul Garcia-Patron, “Limitations of optimization algorithms on noisy quantum devices,” *arXiv preprint arXiv:2009.05532* (2020).
- [175] Leo Zhou, Sheng-Tao Wang, Soonwon Choi, Hannes Pichler, and Mikhail D Lukin, “Quantum approximate optimization algorithm: Performance, mechanism, and implementation on near-term devices,” *Physical Review X* **10**, 021067 (2020).
- [176] Edward Grant, Leonard Wossnig, Mateusz Ostaszewski, and Marcello Benedetti, “An initialization strategy for

- addressing barren plateaus in parametrized quantum circuits,” *Quantum* **3**, 214 (2019).
- [177] Tyler Volkoff and Patrick J Coles, “Large gradients via correlation in random parameterized quantum circuits,” *arXiv preprint arXiv:2005.12200* (2020).
  - [178] Andrea Skolik, Jarrod R McClean, Masoud Mohseni, Patrick van der Smagt, and Martin Leib, “Layerwise learning for quantum neural networks,” *arXiv preprint arXiv:2006.14904* (2020).
  - [179] Ernesto Campos, Aly Nasrallah, and Jacob Biamonte, “Abrupt transitions in variational quantum circuit training,” *arXiv preprint arXiv:2010.09720* (2020).
  - [180] Guillaume Verdon, Michael Broughton, Jarrod R McClean, Kevin J Sung, Ryan Babbush, Zhang Jiang, Hartmut Neven, and Masoud Mohseni, “Learning to learn with quantum neural networks via classical neural networks,” *arXiv preprint arXiv:1907.05415* (2019).
  - [181] Abhinav Anand, Matthias Degroote, and Alán Aspuru-Guzik, “Natural evolutionary strategies for variational quantum computation,” *arXiv preprint arXiv:2012.00101* (2020).
  - [182] Zhenyu Cai, “Resource estimation for quantum variational simulations of the hubbard model,” *Phys. Rev. Applied* **14**, 014059 (2020).
  - [183] Chris Cade, Lana Mineh, Ashley Montanaro, and Stasja Stanisic, “Strategies for solving the fermi-hubbard model on near-term quantum computers,” *Physical Review B* **102**, 235122 (2020).
  - [184] Guoming Wang, Dax Enshan Koh, Peter D Johnson, and Yudong Cao, “Bayesian inference with engineered likelihood functions for robust amplitude estimation,” *arXiv preprint arXiv:2006.09350* (2020).
  - [185] Andrew Jena, Scott Genin, and Michele Mosca, “Pauli partitioning with respect to gate sets,” *arXiv preprint arXiv:1907.07859* (2019).
  - [186] Ophelia Crawford, Barnaby van Straaten, Daochen Wang, Thomas Parks, Earl Campbell, and Stephen Brierley, “Efficient quantum measurement of pauli operators,” *arXiv preprint arXiv:1908.06942* (2019).
  - [187] Vladyslav Verteletskyi, Tzu-Ching Yen, and Artur F Izmaylov, “Measurement optimization in the variational quantum eigensolver using a minimum clique cover,” *The Journal of Chemical Physics* **152**, 124114 (2020).
  - [188] Artur F Izmaylov, Tzu-Ching Yen, Robert A Lang, and Vladyslav Verteletskyi, “Unitary partitioning approach to the measurement problem in the variational quantum eigensolver method,” *Journal of Chemical Theory and Computation* **16**, 190–195 (2019).
  - [189] Andrew Zhao, Andrew Tranter, William M Kirby, Shu Fay Ung, Akimasa Miyake, and Peter J Love, “Measurement reduction in variational quantum algorithms,” *Physical Review A* **101**, 062322 (2020).
  - [190] Tzu-Ching Yen, Vladyslav Verteletskyi, and Artur F Izmaylov, “Measuring all compatible operators in one series of single-qubit measurements using unitary transformations,” *Journal of Chemical Theory and Computation* **16**, 2400–2409 (2020).
  - [191] Pranav Gokhale and Frederic T Chong, “ $\mathcal{O}(n^3)$  measurement cost for variational quantum eigensolver on molecular hamiltonians,” *arXiv preprint arXiv:1908.11857* (2019).
  - [192] Jarrod R McClean, Jonathan Romero, Ryan Babbush, and Alán Aspuru-Guzik, “The theory of variational hybrid quantum-classical algorithms,” *New Journal of Physics* **18**, 023023 (2016).
  - [193] William J Huggins, Jarrod McClean, Nicholas Rubin, Zhang Jiang, Nathan Wiebe, K Birgitta Whaley, and Ryan Babbush, “Efficient and noise resilient measurements for quantum chemistry on near-term quantum computers,” *arXiv preprint arXiv:1907.13117* (2019).
  - [194] Nicholas C Rubin, Ryan Babbush, and Jarrod McClean, “Application of fermionic marginal constraints to hybrid quantum algorithms,” *New Journal of Physics* **20**, 053020 (2018).
  - [195] Andrew Arrasmith, Lukasz Cincio, Rolando D Somma, and Patrick J Coles, “Operator sampling for shot-frugal optimization in variational algorithms,” *arXiv preprint arXiv:2004.06252* (2020).
  - [196] Barnaby van Straaten and Bálint Koczor, “Measurement cost of metric-aware variational quantum algorithms,” *arXiv preprint arXiv:2005.05172* (2020).
  - [197] Hsin-Yuan Huang, Richard Kueng, and John Preskill, “Predicting many properties of a quantum system from very few measurements,” *Nature Physics* **16**, 1050–1057 (2020).
  - [198] Charles Hadfield, Sergey Bravyi, Rudy Raymond, and Antonio Mezzacapo, “Measurements of quantum hamiltonians with locally-biased classical shadows,” *arXiv preprint arXiv:2006.15788* (2020).
  - [199] Giacomo Torlai, Guglielmo Mazzola, Giuseppe Carleo, and Antonio Mezzacapo, “Precise measurement of quantum observables with neural-network estimators,” *Physical Review Research* **2**, 022060 (2020).
  - [200] Laura Gentini, Alessandro Cuccoli, Stefano Pirandola, Paola Verrucchi, and Leonardo Banchi, “Noise-resilient variational hybrid quantum-classical optimization,” *Physical Review A* **102**, 052414 (2020).
  - [201] Enrico Fontana, M. Cerezo, Andrew Arrasmith, Ivan Rungger, and Patrick J Coles, “Optimizing parametrized quantum circuits via noise-induced breaking of symmetries,” *arXiv preprint arXiv:2011.08763* (2020).
  - [202] Cheng Xue, Zhao-Yun Chen, Yu-Chun Wu, and Guo-Ping Guo, “Effects of quantum noise on quantum approximate optimization algorithm,” *arXiv preprint arXiv:1909.02196* (2019).
  - [203] Jeffrey Marshall, Filip Wudarski, Stuart Hadfield, and Tad Hogg, “Characterizing local noise in QAOA circuits,” *IOP SciNotes* **1**, 025208 (2020).
  - [204] Suguru Endo, Zhenyu Cai, Simon C Benjamin, and Xiao Yuan, “Hybrid quantum-classical algorithms and quantum error mitigation,” *arXiv preprint arXiv:2011.01382* (2020).
  - [205] Kristan Temme, Sergey Bravyi, and Jay M Gambetta, “Error mitigation for short-depth quantum circuits,” *Physical Review Letters* **119**, 180509 (2017).
  - [206] Suguru Endo, Simon C Benjamin, and Ying Li, “Practical quantum error mitigation for near-future applications,” *Physical Review X* **8**, 031027 (2018).
  - [207] Zhenyu Cai, “Multi-exponential error extrapolation and combining error mitigation techniques for nisq applications,” *arXiv preprint arXiv:2007.01265* (2020).
  - [208] Matthew Otten and Stephen K Gray, “Recovering noise-free quantum observables,” *Physical Review A* **99**, 012338 (2019).
  - [209] Suguru Endo, Qi Zhao, Ying Li, Simon Benjamin, and Xiao Yuan, “Mitigating algorithmic errors in a hamiltonian simulation,” *Physical Review A* **99**, 012334 (2019).

- [210] Jinzhao Sun, Xiao Yuan, Takahiro Tsunoda, Vlatko Vedral, Simon C Benjamin, and Suguru Endo, "Mitigating realistic noise in practical noisy intermediate-scale quantum devices," *arXiv preprint arXiv:2001.04891* (2020).
- [211] Armands Strikis, Dayue Qin, Yanzhu Chen, Simon C Benjamin, and Ying Li, "Learning-based quantum error mitigation," *arXiv preprint arXiv:2005.07601* (2020).
- [212] Piotr Czarnik, Andrew Arrasmith, Patrick J Coles, and Lukasz Cincio, "Error mitigation with clifford quantum-circuit data," *arXiv preprint arXiv:2005.10189* (2020).
- [213] Angus Lowe, Max Hunter Gordon, Piotr Czarnik, Andrew Arrasmith, Patrick J Coles, and Lukasz Cincio, "Unified approach to data-driven quantum error mitigation," *arXiv arXiv:2011.01157* (2020).
- [214] Sam McArdle, Xiao Yuan, and Simon Benjamin, "Error-mitigated digital quantum simulation," *Physical Review Letters* **122**, 180501 (2019).
- [215] Xavi Bonet-Monroig, Ramiro Sagastizabal, M Singh, and TE O'Brien, "Low-cost error mitigation by symmetry verification," *Physical Review A* **98**, 062339 (2018).
- [216] Jarrod R McClean, Zhang Jiang, Nicholas C Rubin, Ryan Babbush, and Hartmut Neven, "Decoding quantum errors with subspace expansions," *Nature Communications* **11**, 1–9 (2020).
- [217] Bálint Koczor, "Exponential error suppression for near-term quantum devices," *arXiv preprint arXiv:2011.05942* (2020).
- [218] William J Huggins, Sam McArdle, Thomas E O'Brien, Joonho Lee, Nicholas C Rubin, Sergio Boixo, K Birgitta Whaley, Ryan Babbush, and Jarrod R McClean, "Virtual distillation for quantum error mitigation," *arXiv preprint arXiv:2011.07064* (2020).
- [219] Sergey Bravyi, Sarah Sheldon, Abhinav Kandala, David C McKay, and Jay M Gambetta, "Mitigating measurement errors in multi-qubit experiments," *arXiv preprint arXiv:2006.14044* (2020).
- [220] Daiqin Su, Robert Israel, Kunal Sharma, Haoyu Qi, Ish Dhand, and Kamil Brádler, "Error mitigation on a near-term quantum photonic device," *arXiv preprint arXiv:2008.06670* (2020).
- [221] Yudong Cao, Jhonathan Romero, and Alán Aspuru-Guzik, "Potential of quantum computing for drug discovery," *IBM Journal of Research and Development* **62**, 6–1 (2018).
- [222] Yudong Cao, Jonathan Romero, Jonathan P Olson, Matthias Degroote, Peter D Johnson, Mária Kieferová, Ian D Kivlichan, Tim Menke, Borja Peropadre, Nicolas PD Sawaya, *et al.*, "Quantum chemistry in the age of quantum computing," *Chemical reviews* **119**, 10856–10915 (2019).
- [223] Carlos Outeiral, Martin Strahm, Jiye Shi, Garrett M Morris, Simon C Benjamin, and Charlotte M Deane, "The prospects of quantum computing in computational molecular biology," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, e1481 (2020).
- [224] Steven R White, "Density matrix formulation for quantum renormalization groups," *Physical Review Letters* **69**, 2863 (1992).
- [225] Garnet Kin-Lic Chan and Sandeep Sharma, "The density matrix renormalization group in quantum chemistry," *Annual Review of Physical Chemistry* **62**, 465–481 (2011).
- [226] Sam McArdle, Suguru Endo, Alan Aspuru-Guzik, Simon C Benjamin, and Xiao Yuan, "Quantum computational chemistry," *Reviews of Modern Physics* **92**, 015003 (2020).
- [227] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C Bardin, Rami Barends, Sergio Boixo, Michael Broughton, Bob B Buckley, David A Buell, *et al.*, "Hartree-fock on a superconducting qubit quantum computer," *arXiv preprint arXiv:2004.04174* (2020).
- [228] Artem A Bakulin, Stoichko D Dimitrov, Akshay Rao, Philip CY Chow, Christian B Nielsen, Bob C Schroeder, Iain McCulloch, Huib J Bakker, James R Durrant, and Richard H Friend, "Charge-transfer state dynamics following hole and electron transfer in organic photovoltaic devices," *The Journal of Physical Chemistry Letters* **4**, 209–215 (2013).
- [229] Markus Gross, David C Müller, Heinz-Georg Nothofer, Ulrich Scherf, Dieter Neher, Christoph Bräuchle, and Klaus Meerholz, "Improving the performance of doped  $\pi$ -conjugated polymers for use in organic light-emitting diodes," *Nature* **405**, 661–665 (2000).
- [230] JR Schmidt, Priya V Parandekar, and John C Tully, "Mixed quantum-classical equilibrium: Surface hopping," *The Journal of Chemical Physics* **129**, 044104 (2008).
- [231] Thomas E O'Brien, Bruno Senjean, Ramiro Sagastizabal, Xavier Bonet-Monroig, Alicja Dutkiewicz, Francesco Buda, Leonardo DiCarlo, and Lucas Visscher, "Calculating energy derivatives for quantum chemistry on a quantum computer," *npj Quantum Information* **5**, 1–12 (2019).
- [232] John C Tully and Richard K Preston, "Trajectory surface hopping approach to nonadiabatic molecular collisions: the reaction of  $\text{H}^+$  with  $\text{d}_2$ ," *The Journal of Chemical Physics* **55**, 562–572 (1971).
- [233] David R Weinberg, Christopher J Gagliardi, Jonathan F Hull, Christine Fecenko Murphy, Caleb A Kent, Brittany C Westlake, Amit Paul, Daniel H Ess, Dewey Granville McCafferty, and Thomas J Meyer, "Proton-coupled electron transfer," *Chemical Reviews* **112**, 4016–4093 (2012).
- [234] Walter Kohn and Lu Jeu Sham, "Self-consistent equations including exchange and correlation effects," *Physical review* **140**, A1133 (1965).
- [235] Bela Bauer, Dave Wecker, Andrew J Millis, Matthew B Hastings, and Matthias Troyer, "Hybrid quantum-classical approach to correlated materials," *Physical Review X* **6**, 031045 (2016).
- [236] Ryan Babbush, Craig Gidney, Dominic W Berry, Nathan Wiebe, Jarrod McClean, Alexandru Paler, Austin Fowler, and Hartmut Neven, "Encoding electronic spectra in quantum circuits with linear  $t$  complexity," *Physical Review X* **8**, 041015 (2018).
- [237] Dominic W Berry, Mária Kieferová, Artur Scherer, Yuval R Sanders, Guang Hao Low, Nathan Wiebe, Craig Gidney, and Ryan Babbush, "Improved techniques for preparing eigenstates of fermionic hamiltonians," *npj Quantum Information* **4**, 1–7 (2018).
- [238] Pierre-Luc Dallaire-Demers, Jonathan Romero, Libor Veis, Sukin Sim, and Alán Aspuru-Guzik, "Low-depth circuit ansatz for preparing correlated fermionic states on a quantum computer," *arXiv preprint arXiv:1801.01053* (2018).
- [239] Eugene F Dumitrescu, Alex J McCaskey, Gaute Hagen, Gustav R Jansen, Titus D Morris, T Papenbrock,



- Raphael C Pooser, David Jarvis Dean, and Pavel Lougovski, “Cloud quantum computing of an atomic nucleus,” *Physical Review Letters* **120**, 210501 (2018).
- [240] Hsuan-Hao Lu, Natalie Klco, Joseph M Lukens, Titus D Morris, Aaina Bansal, Andreas Ekström, Gaute Hagen, Thomas Papenbrock, Andrew M Weiner, Martin J Savage, *et al.*, “Simulations of subatomic many-body physics on a quantum frequency processor,” *Physical Review A* **100**, 012320 (2019).
- [241] Alessandro Roggero, Andy CY Li, Joseph Carlson, Rajan Gupta, and Gabriel N Perdue, “Quantum computing for neutrino-nucleus scattering,” *Physical Review D* **101**, 074038 (2020).
- [242] John Preskill, “Simulating quantum field theory with a quantum computer,” *arXiv preprint arXiv:1811.10085* (2018).
- [243] Suguru Endo, Iori Kurata, and Yuya O Nakagawa, “Calculation of the green’s function on near-term quantum computers,” *Physical Review Research* **2**, 033281 (2020).
- [244] Chinmay Mishra, Shane Thompson, Raphael Pooser, and George Siopsis, “Quantum computation of an interacting fermionic model,” *Quantum Science and Technology* **5**, 035010 (2020).
- [245] Danny Paulson, Luca Dellantonio, Jan F Haase, Alessio Celi, Angus Kan, Andrew Jena, Christian Kokail, Rick van Bijnen, Karl Jansen, Peter Zoller, *et al.*, “Towards simulating 2d effects in lattice gauge theories on a quantum computer,” *arXiv preprint arXiv:2008.09252* (2020).
- [246] A Avkhadiiev, PE Shanahan, and RD Young, “Accelerating lattice quantum field theory calculations via interpolator optimization using noisy intermediate-scale quantum computing,” *Physical Review Letters* **124**, 080501 (2020).
- [247] Jad C Halimeh, Valentin Kasper, and Philipp Hauke, “Fate of lattice gauge theories under decoherence,” *arXiv preprint arXiv:2009.07848* (2020).
- [248] Kunal Sharma, M. Cerezo, Zoë Holmes, Lukasz Cincio, Andrew Sornborger, and Patrick J Coles, “Reformulation of the no-free-lunch theorem for entangled data sets,” *arXiv preprint arXiv:2007.04900* (2020).
- [249] Francisco Barahona, Martin Grötschel, Michael Jünger, and Gerhard Reinelt, “An application of combinatorial optimization to statistical physics and circuit layout design,” *Operations Research* **36**, 493–513 (1988).
- [250] Wolfgang Küchlin and Carsten Sinz, “Proving consistency assertions for automotive product data management,” *Journal of Automated Reasoning* **24**, 145–163 (2000).
- [251] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann, “A Quantum Approximate Optimization Algorithm Applied to a Bounded Occurrence Constraint Problem,” *arXiv preprint arXiv:1412.6062* (2014).
- [252] Edward Farhi and Aram W Harrow, “Quantum supremacy through the quantum approximate optimization algorithm,” *arXiv preprint arXiv:1602.07674* (2016).
- [253] Matthew B Hastings, “Classical and quantum bounded depth approximation algorithms,” *arXiv preprint arXiv:1905.07047* (2019).
- [254] Sergey Bravyi, Alexander Kliesch, Robert Koenig, and Eugene Tang, “Obstacles to state preparation and variational optimization from symmetry protection,” *arXiv preprint arXiv:1910.08980* (2019).
- [255] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C Bardin, Rami Barends, Sergio Boixo, Michael Broughton, Bob B Buckley, David A Buell, *et al.*, “Quantum approximate optimization of non-planar graph problems on a planar superconducting processor,” *arXiv preprint arXiv:2004.04197* (2020).
- [256] Maria Schuld, Ilya Sinayskiy, and Francesco Petruccione, “An introduction to quantum machine learning,” *Contemporary Physics* **56**, 172–185 (2015).
- [257] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd, “Quantum machine learning,” *Nature* **549**, 195–202 (2017).
- [258] Amira Abbas, David Sutter, Christa Zoufal, Aurélien Lucchi, Alessio Figalli, and Stefan Woerner, “The power of quantum neural networks,” *arXiv preprint arXiv:2011.00027* (2020).
- [259] Nathan Wiebe, Ashish Kapoor, and Krysta M Svore, “Quantum deep learning,” *arXiv preprint arXiv:1412.3489* (2014).
- [260] Samuel Yen-Chi Chen, Chao-Han Huck Yang, Jun Qi, Pin-Yu Chen, Xiaoli Ma, and Hsi-Sheng Goan, “Variational quantum circuits for deep reinforcement learning,” *IEEE Access* **8**, 141007–141024 (2020).
- [261] Yuval R Sanders, Dominic W Berry, Pedro CS Costa, Louis W Tessler, Nathan Wiebe, Craig Gidney, Hartmut Neven, and Ryan Babbush, “Compilation of fault-tolerant quantum heuristics for combinatorial optimization,” *Physical Review X Quantum* **1**, 020312 (2020).

## ACKNOWLEDGEMENTS

MC is thankful to Kunal Sharma for helpful discussions. MC was initially supported by the Laboratory Directed Research and Development (LDRD) program of Los Alamos National Laboratory (LANL) under project number 20180628ECR, and later supported by the Center for Nonlinear Studies at LANL. AA was initially supported by the LDRD program of LANL under project number 20200056DR, and later supported by the by the U.S. Department of Energy (DOE), Office of Science, Office of High Energy Physics QuantISED program under under Contract Nos. DE-AC52-06NA25396 and KA2401032. SCB acknowledges financial support from EPSRC Hub grants under the agreement numbers EP/M013243/1 and EP/T001062/1, and from EU H2020-FETFLAG-03-2018 under the grant agreement No 820495 (AQTION). SE was supported by MEXT Quantum Leap Flagship Program (MEXT QLEAP) Grant Number JPMXS0120319794, JPMXS0118068682 and JST ERATO Grant Number JPMJER1601. KF was supported by JSPS KAKENHI Grant No. 16H02211, JST ERATO JPMJER1601, and JST CREST JPMJCR1673. KM was supported by JST PRESTO Grant No. JPMJPR2019 and JSPS KAKENHI Grant No. 20K22330. KM and KF were also supported by MEXT Quantum Leap Flagship Program (MEXT QLEAP) Grant Number JPMXS0118067394 and JPMXS0120319794. XY acknowledges support from the



Simons Foundation. LC was initially supported by the LDRD program of LANL under project number 20190065DR, and later supported by the U.S. DOE, Office of Science, Office of Advanced Scientific Computing Research, under the Quantum Computing Application Teams (QCAT) program. PJC was initially supported by the LANL ASC Beyond Moore's Law project, and later supported by the U.S. DOE, Office of Science, Office of Advanced Scientific Computing Research, under the Accelerated Research in Quantum Computing (ARQC) program. Most recently, MC, LC, and PJC were supported by the Quantum Science Center (QSC), a National Quantum Information Science Research Center of

the U.S. Department of Energy (DOE).

## **AUTHOR CONTRIBUTIONS**

All authors have read, discussed and contributed to the writing of the manuscript.

## **COMPETING INTERESTS**

The authors declare no competing interests.