

Theory of analytical energy derivatives for the variational quantum eigensolver

Kosuke Mitarai,^{1,2,*} Yuya O. Nakagawa,² and Wataru Mizukami^{1,3}

¹Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan

²QunaSys Inc., High-tech Hongo Building 1F, 5-25-18 Hongo, Bunkyo, Tokyo 113-0033, Japan

³Quantum Information and Quantum Biology Division, Institute for Open and Transdisciplinary Research Initiatives, Osaka University, Osaka 560-8531, Japan



(Received 1 October 2019; accepted 20 November 2019; published 5 February 2020)

The variational quantum eigensolver (VQE) and its variants, which is a method for finding eigenstates and eigenenergies of a given Hamiltonian, are appealing applications of near-term quantum computers. Although the eigenenergies are certainly important quantities which determine properties of a given system, their derivatives with respect to parameters of the system, such as positions of nuclei if we target a quantum chemistry problem, are also crucial to analyze the system. Here, we describe methods to evaluate analytical derivatives of the eigenenergy of a given Hamiltonian, including the excited state energy as well as the ground-state energy, with respect to the system parameters in the framework of the VQE. We give explicit, low-depth quantum circuits which can measure essential quantities to evaluate energy derivatives, incorporating with proof-of-principle numerical simulations. This work extends the theory of the variational quantum eigensolver, by enabling it to measure more physical properties of a quantum system than before and to explore chemical reactions.

DOI: [10.1103/PhysRevResearch.2.013129](https://doi.org/10.1103/PhysRevResearch.2.013129)

I. INTRODUCTION

The variational quantum eigensolver (VQE) has attracted much attention as a potential application of near-term quantum computers [1]. The VQE is an iterative algorithm to construct a quantum circuit that outputs eigenstates and eigenenergy of a Hamiltonian which describes the system under consideration. Originally, the method was devised for finding the ground state of a system [1]. It has subsequently been extended for excited states by several proposals [2–5]. From the generated eigenstates, one can measure its associated physical quantities, such as the particle densities and transition amplitudes between the different eigenstates.

Though eigenenergy and associated particle density are certainly important quantities, the wave function of a quantum system has valuable information besides those. Quantum chemistry calculations, which would be one of the most promising applications of the VQE, often utilize such information. Among such, we focus on energy derivatives in this work. Many time-independent physical/chemical properties can be defined as derivatives of the energy [6–9]. For example, the first derivatives of the energy with respect to nuclear coordinates give us the forces acting on atoms, which can be utilized for the task of locating energy extrema on the potential energy surface (i.e., geometry optimization

[10]. The second-order derivatives give the force-constant matrix that not only helps to locate and verify transition states but also allows us to compute vibrational frequencies and partition functions within the harmonic approximation [10,11]. The energy derivatives with respect to external fields have to do with various spectroscopy: intensities of infrared and Raman spectroscopy are proportional to the cross derivatives with respect to vibrational normal modes and external electric fields [11]; NMR chemical shifts can be obtained using the cross derivatives with respect to nuclear spin and magnetic fields [12,13]. Computing such derivatives is a core part of simulations or analysis of molecular spectra.

A simple way to compute energy derivatives is to use the finite difference method and calculate them numerically. This approach, however, suffers from high computational costs as well as numerical errors and instabilities [7,9]. Say, the number of energy points needed to evaluate the forces increases linearly to the number of atoms. This high computational cost makes the numerical approach impractical in many cases. Moreover, with near-term quantum devices, where noise is inevitable, the numerical difference approach would give us poor results. The other way—analytical approach—is therefore vital. The theory and program codes of analytical energy derivatives indeed support the high practicality (and popularity) of today’s molecular electronic structure theory [9]. Methods to calculate the derivatives of excited state energy on classical computers has also been widely developed [14–34], but still suffers from its high computational cost and relatively low accuracy. The task of computing such derivatives on quantum computers has been addressed in the traditional methods which utilize the quantum phase estimation [35], but not for the VQE.

*mitarai@qc.ee.es.osaka-u.ac.jp

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.

In this work, we derive the analytical formulas and explicit quantum circuits to address the task of measuring the energy derivatives. More specifically, we describe the methods to obtain the derivatives of the energy with respect to the system parameters up to the third order, from which one can extract the physical properties. We also present a method to extract the derivatives of excited state energy based on the technique presented in Refs. [3,4]. The analysis on the computational cost shows that the analytical differentiation provided in this work can be more practical than a finite difference approach, in a sense that one does not need to consider the best step size to extract the energy derivative information with an optimal precision. Proof-of-principle numerical simulations are also performed to verify the correctness of the derived equation and circuits. The presented methods extend the applicability of the VQE by enabling it to evaluate more physical properties than before.

II. VARIATIONAL QUANTUM EIGENSOLVER

Here, we briefly review the algorithm of the VQE. In the VQE, we construct a parameterized quantum circuit $U(\theta)$ and the corresponding ansatz state $|\psi(\theta)\rangle = U(\theta)|0\rangle^{\otimes n}$, where $|0\rangle^{\otimes n}$ is an initialized n -qubit state and $\theta = (\theta_1, \dots, \theta_{N_\theta}) \in \mathbb{R}^{N_\theta}$ is a vector of parameters implemented on the circuit with N_θ being the number of them. A set of parameters θ are variationally optimized so that the expectation value $E(\theta) = \langle \psi(\theta) | H | \psi(\theta) \rangle$ of a given Hamiltonian H is minimized. At the optimal point $\theta = \theta^*$, one naturally expects

$$\frac{\partial E(\theta^*)}{\partial \theta_a} = 0, \quad (1)$$

for all a . Let

$$|\partial_a \psi(\theta)\rangle = \frac{\partial}{\partial \theta_a} |\psi(\theta)\rangle. \quad (2)$$

The condition of Eq. (1) reduces to

$$\text{Re}\langle \psi(\theta^*) | H | \partial_a \psi(\theta^*) \rangle = 0. \quad (3)$$

Higher derivatives of the wave function will be denoted by

$$|\partial_a \partial_b \dots \partial_c \psi(\theta)\rangle = \frac{\partial}{\partial \theta_a} \frac{\partial}{\partial \theta_b} \dots \frac{\partial}{\partial \theta_c} |\psi(\theta)\rangle, \quad (4)$$

to shorten the notation.

As stated in the introduction, many physical properties of a quantum system are calculated from the derivative of the energy with respect to the system parameter in a given Hamiltonian. The parameter can be, for example, the coordinates of atoms or the electric field applied to the system. We denote such parameters by an N_x -dimensional vector $x \in \mathbb{R}^{N_x}$. In this case, both of the Hamiltonian H and the optimal parameter θ^* of the wave function at the specific value of x is also a function of x , which will be denoted by $H(x)$ and $\theta^*(x)$, respectively. Corresponding to the change of this problem setting, we redefine the energy as,

$$E(\theta, x) = \langle \psi(\theta) | H(x) | \psi(\theta) \rangle. \quad (5)$$

Let the optimal ground-state energy be $E^*(x)$, that is, $E^*(x) = E(\theta^*(x), x)$. In the following sections, we show the analytical forms of the derivatives such as $\frac{\partial E^*}{\partial x_i}$ and $\frac{\partial^2 E^*}{\partial x_i \partial x_j}$, which are the essential quantities for extracting physical properties of the target system.

III. ANALYTICAL EXPRESSION OF DERIVATIVES

The derivation of the formulas presented in this section is in Appendix for completeness, or you can also refer to Ref. [36]. In the following, we use the following notation,

$$\frac{\partial}{\partial \theta_a} \frac{\partial E(\theta^*(x), x)}{\partial x_j} := \frac{\partial}{\partial \theta_a} \frac{\partial E(\theta, x)}{\partial x_j} \Big|_{\theta=\theta^*(x)}, \quad (6)$$

and likewise for terms similar to this. In this manuscript, the partial derivative symbols are meant to act only on the term immediately on its right, rather than on all the symbols appearing to its right.

A. Derivatives of ground-state energy

The analytical expressions for the derivatives of ground-state energy are the following:

$$\frac{\partial E^*(x)}{\partial x_i} = \langle \psi(\theta^*(x)) | \frac{\partial H(x)}{\partial x_i} | \psi(\theta^*(x)) \rangle, \quad (7)$$

$$\frac{\partial}{\partial x_i} \frac{\partial E^*(x)}{\partial x_j} = \sum_a \frac{\partial \theta_a^*(x)}{\partial x_i} \frac{\partial}{\partial \theta_a} \frac{\partial E(\theta^*(x), x)}{\partial x_j} + \langle \psi(\theta^*(x)) | \frac{\partial}{\partial x_i} \frac{\partial H(x)}{\partial x_j} | \psi(\theta^*(x)) \rangle, \quad (8)$$

$$\begin{aligned} \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} \frac{\partial E^*(x)}{\partial x_k} &= \sum_{a,b,c} \frac{\partial \theta_a^*(x)}{\partial x_i} \frac{\partial \theta_b^*(x)}{\partial x_j} \frac{\partial \theta_c^*(x)}{\partial x_k} \frac{\partial}{\partial \theta_a} \frac{\partial}{\partial \theta_b} \frac{\partial E(\theta^*(x), x)}{\partial \theta_c} + \sum_{a,b} \left[\frac{\partial \theta_a^*(x)}{\partial x_i} \frac{\partial \theta_b^*(x)}{\partial x_j} \frac{\partial}{\partial \theta_b} \frac{\partial}{\partial \theta_a} \frac{\partial E(\theta^*(x), x)}{\partial x_k} \right. \\ &\quad \left. + \frac{\partial \theta_a^*(x)}{\partial x_k} \frac{\partial \theta_b^*(x)}{\partial x_i} \frac{\partial}{\partial \theta_b} \frac{\partial}{\partial \theta_a} \frac{\partial E(\theta^*(x), x)}{\partial x_j} + \frac{\partial \theta_a^*(x)}{\partial x_j} \frac{\partial \theta_b^*(x)}{\partial x_k} \frac{\partial}{\partial \theta_b} \frac{\partial}{\partial \theta_a} \frac{\partial E(\theta^*(x), x)}{\partial x_i} \right] \\ &\quad + \sum_a \left[\frac{\partial \theta_a^*(x)}{\partial x_i} \frac{\partial}{\partial \theta_a} \frac{\partial}{\partial x_j} \frac{\partial E(\theta^*(x), x)}{\partial x_k} + \frac{\partial \theta_a^*(x)}{\partial x_k} \frac{\partial}{\partial \theta_a} \frac{\partial}{\partial x_i} \frac{\partial E(\theta^*(x), x)}{\partial x_j} + \frac{\partial \theta_a^*(x)}{\partial x_j} \frac{\partial}{\partial \theta_a} \frac{\partial}{\partial x_k} \frac{\partial E(\theta^*(x), x)}{\partial x_i} \right] \\ &\quad + \langle \psi(\theta^*(x)) | \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} \frac{\partial H(x)}{\partial x_k} | \psi(\theta^*(x)) \rangle, \end{aligned} \quad (9)$$

where we assumed $\frac{\partial E(\theta^*(x), x)}{\partial \theta} = 0$. Note that, in general, the formulas for the d th derivative of $E^*(x)$ contains θ derivatives up to the d th in the form of $\frac{\partial^q}{\partial \theta^q} \frac{\partial^{d-q} E}{\partial x^{d-q}}$ for $q = 1, \dots, d$. Also, Wigner's $(2n + 1)$ rule [36] ensures that x derivatives of the optimal parameter, $\theta^*(x)$, up to the n th are sufficient for calculating $(2n + 1)$ th derivative of $E^*(x)$. In other words, for d th derivative of $E^*(x)$, we only need $\lfloor d/2 \rfloor$ th derivative of $\theta^*(x)$, where $\lfloor y \rfloor$ is the floor function denoting the greatest integer less than or equal to y . The term $\frac{\partial \theta_a^*(x)}{\partial x_i}$ in the above equation can be calculated by solving the response equation, which we write down in the next section.

B. Derivatives of optimal parameters

The first and second derivatives of the optimal parameter, $\theta^*(x)$, can be obtained from the following response equation:

$$\sum_b \frac{\partial}{\partial \theta_a} \frac{\partial E(\theta^*(x), x)}{\partial \theta_b} \frac{\partial \theta_b^*(x)}{\partial x_i} = - \frac{\partial}{\partial \theta_a} \frac{\partial E(\theta^*(x), x)}{\partial x_i}, \quad (10)$$

$$\sum_b \frac{\partial}{\partial \theta_a} \frac{\partial E(\theta^*(x), x)}{\partial \theta_b} \frac{\partial}{\partial x_i} \frac{\partial \theta_b^*(x)}{\partial x_j} = -\gamma_a^{(ij)}(\theta^*(x), x), \quad (11)$$

where

$$\begin{aligned} \gamma_c^{(ij)} &= \sum_{a,b} \frac{\partial}{\partial \theta_c} \frac{\partial}{\partial \theta_a} \frac{\partial E}{\partial \theta_b} \frac{\partial \theta_a^*}{\partial x_i} \frac{\partial \theta_b^*}{\partial x_j} \\ &+ 2 \sum_a \frac{\partial}{\partial \theta_c} \frac{\partial}{\partial \theta_a} \frac{\partial E}{\partial x_j} \frac{\partial \theta_a^*}{\partial x_i} + \frac{\partial}{\partial \theta_c} \frac{\partial}{\partial x_i} \frac{\partial E}{\partial x_j}. \end{aligned} \quad (12)$$

Equations (10) and (11) is obtained by taking the x derivative of Eq. (1), and then exchanging the order of the partial derivatives. By Wigner's $(2n + 1)$ rule [36], the above equations are enough to obtain derivatives of the energy up to the fifth order.

IV. MEASUREMENT AND CALCULATION OF DERIVATIVES OF GROUND STATE

In this section, we describe the methodology for calculating the derivatives of the ground-state energy, whose analytical expressions are shown in the previous section, when given an optimal circuit parameter $\theta^*(x)$ at some x that gives the local minimum of $E(\theta, x)$.

A. Notations and assumptions

In the VQE, we target a Hamiltonian which acts on an n -qubit system and is decomposed into a sum of Pauli strings, $\mathcal{P} = \{I, X, Y, Z\}^{\otimes n}$, as follows

$$H(x) = \sum_{P \in \mathcal{P}} h_P(x) P, \quad (13)$$

where $h_P(x) \in \mathbb{R}$. $h_P(x)$ is assumed to be nonzero only for poly(n) terms. We also assume that $h_P(x)$ is differentiable to any order. For a quantum chemistry problem, the original Hamiltonian has the following form:

$$H(x) = \sum_{i,j} h_{ij}(x) c_i^\dagger c_j + \sum_{i,j,k,l} h_{ijkl}(x) c_i^\dagger c_j^\dagger c_k c_l, \quad (14)$$

where c_i^\dagger and c_i are the fermion creation and annihilation operators acting on the i th orbital, respectively. Equation (14)

is converted to the form of Eq. (13) by, for example, Jordan-Wigner or Bravyi-Kitaev transformation [37,38]. Note that because we always work in the second quantization formalism, the effect of the change of the molecular orbital corresponding to the change of molecular geometry is totally absorbed in the coefficients $h(x)$. Therefore, the change of the molecular orbital does not explicitly appear in the following discussion.

To calculate the energy derivatives of such Hamiltonian, first of all, we assume that the derivatives of Hamiltonian, $\frac{\partial H}{\partial x_i}$, $\frac{\partial}{\partial x_i} \frac{\partial H}{\partial x_j}$, and so on, can be calculated by the classical computer, e.g., using the conventional quantum chemistry. In other words, we are able to calculate the quantities such as $\frac{\partial h_P(x)}{\partial x_i}$ and $\frac{\partial}{\partial x_i} \frac{\partial h_P(x)}{\partial x_j}$. For quantum chemistry problems given in terms of Hartree-Fock orbitals, these calculations correspond to solving the coupled perturbed Hartree-Fock equation [30,36].

Notice that under this assumption, we only need to consider how to evaluate quantities which involve the differentiation with respect to θ such as $\frac{\partial}{\partial \theta_a} \frac{\partial}{\partial \theta_b} \frac{\partial E}{\partial x_i}$, because the expectation values such as $\frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} \frac{\partial E}{\partial x_k} = \langle \psi(\theta) | \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} \frac{\partial H(x)}{\partial x_k} | \psi(\theta) \rangle$ at $\theta = \theta^*(x)$, can be evaluated with the exactly same procedure as the usual VQE, i.e., we can measure the expectation value of each Pauli term which appears in $\frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} \frac{\partial H(x)}{\partial x_k}$. After the measurement of all quantities which appear in Eqs. (7)–(9), one can compute the energy derivative using a classical computer by summing up the terms.

The parameterized quantum state, $|\psi(\theta)\rangle$, is constructed by applying a parameterized unitary matrix, that is, a parameterized quantum circuit, $U(\theta)$ to an initialized state; $|\psi(\theta)\rangle = U(\theta)|0\rangle^{\otimes n}$. Following [39], we assume $U(\theta)$ to be a product of unitary matrices each with one parameter,

$$U(\theta) = U_{N_\theta}(\theta_{N_\theta}) \cdots U_2(\theta_2) U_1(\theta_1). \quad (15)$$

We define each unitary $U_a(\theta_a)$ to be generated by a generator G_a as $U_a = e^{i\theta_a G_a}$, which can be decomposed into a sum of Pauli strings;

$$G_a = \sum_{\mu} g_{a,\mu} P_{a,\mu}, \quad (16)$$

where $g_{a,\mu} \in \mathbb{R}$ and $P_{a,\mu} \in \mathcal{P}$. We often use this form of parameterized quantum circuits, for example see Refs. [40–43].

B. Overview of the algorithm

The presented algorithm in this work for evaluating the d th derivative is the following: (1) perform the VQE and obtain the optimal parameter $\theta^*(x)$; (2) compute derivatives of the Hamiltonian, $\frac{\partial^q H(x)}{\partial x^q}$, for $q = 1, \dots, d$ on a classical computer; (3) evaluate x derivatives of E , $\frac{\partial^q E}{\partial x^q}$, for $q = 1, \dots, d$ at $\theta = \theta^*(x)$ by the method described above (Sec. IV A); (4) evaluate terms involving differentiations with respect to θ , $\frac{\partial^q}{\partial \theta^q} \frac{\partial^{d-q} E}{\partial x^{d-q}}$ and $\frac{\partial^q E}{\partial \theta^q}$, for $q = 1, \dots, d$ at $\theta = \theta^*(x)$ on a quantum device by the method described in the following sections (Secs. IV C–IV E); (5) solve the response equations to obtain $\frac{\partial^q \theta^*}{\partial x^q}$ for $q = 1, \dots, \lfloor d/2 \rfloor$ (for $d \leq 3$, solving Eq. (10) suffices); and (6) substitute all terms into Eqs. (7)–(9) to obtain the energy derivatives.

Let us recall N_H and N_θ be the number of terms in the target Hamiltonian and the number of the VQE parameters. The cost

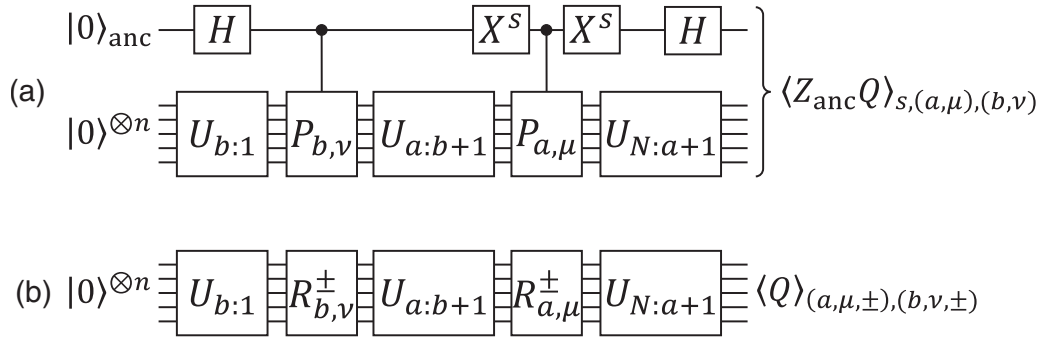


FIG. 1. Quantum circuit to evaluate $\frac{\partial}{\partial \theta_a} \frac{\partial}{\partial \theta_b} E(\theta^*(x), x)$. $U_{a:b} = U_a \cdots U_{b+1} U_b$. (a) Ancilla-based approach. This circuit is a variant of the one presented in Ref. [39]. In the figure, Z_{anc} is the Pauli Z operator acting only on the ancilla qubit, and $s \in \{0, 1\}$. The expectation value of the observable $Z_{\text{anc}} \otimes Q := Z_{\text{anc}} Q$ is obtained by sampling it repeatedly. See Eq. (22) for detailed procedure to obtain the derivative. (b) Low-depth version of (a), derived with the strategy presented in Ref. [45]. In the figure, $R_{a,\mu}^{\pm} = \exp(\pm i\pi P_{a,\mu}/4)$. See Eq. (23) for detailed procedure.

to evaluate d th derivatives by the above algorithm is roughly $O(N_H N_\theta^d)$ when we ignore the cost of the task performed on a classical computer. This is because the most time-consuming part of the algorithm is steps 3 and 4 on a quantum computer, which take $O(N_H N_\theta^d)$ as shall be clear in Secs. IV C–IV E and V.

C. Measurement of $\frac{\partial}{\partial \theta_a} \frac{\partial E}{\partial \theta_b}$

The key quantities for evaluating $\frac{\partial}{\partial x_i} \frac{\partial E^*}{\partial x_j}$ and higher-order derivatives are $\frac{\partial}{\partial \theta_a} \frac{\partial}{\partial \theta_b} \cdots \frac{\partial E}{\partial \theta_c}$ and $\frac{\partial}{\partial \theta_a} \frac{\partial}{\partial \theta_b} \cdots \frac{\partial}{\partial \theta_c} \frac{\partial E}{\partial x_i}$. Note that the first-order derivative, $\frac{\partial E}{\partial \theta_a}$ can be evaluated by the method presented in Ref. [44]. First, we show how to measure $\frac{\partial}{\partial \theta_a} \frac{\partial E}{\partial \theta_b}$. A detailed expression of $\frac{\partial}{\partial \theta_a} \frac{\partial E}{\partial \theta_b}$ is

$$\frac{\partial}{\partial \theta_a} \frac{\partial E(\theta^*(x), x)}{\partial \theta_b} = 2\text{Re}[\langle \partial_a \partial_b \psi(\theta^*(x)) | H(x) | \psi(\theta^*(x)) \rangle + \langle \partial_a \psi(\theta^*(x)) | H(x) | \partial_b \psi(\theta^*(x)) \rangle]. \quad (17)$$

$|\partial_a \psi(\theta)\rangle$ can be expressed as

$$|\partial_a \psi(\theta)\rangle = i \sum_{\mu} g_{a,\mu} U_N(\theta_N) \cdots P_{a,\mu} U_a(\theta_a) \cdots U_2(\theta_2) U_1(\theta_1) |0\rangle^{\otimes n}, \quad (18)$$

and $|\partial_a \partial_b \psi(\theta)\rangle$ is

$$|\partial_a \partial_b \psi(\theta)\rangle = - \sum_{\mu, \nu} g_{a,\mu} g_{b,\nu} U_N(\theta_N) \cdots P_{a,\mu} U_a(\theta_a) \cdots P_{b,\nu} U_b(\theta_b) \cdots U_2(\theta_2) U_1(\theta_1) |0\rangle^{\otimes n}. \quad (19)$$

For convenience, we define

$$|\phi_{(a,\mu),(b,\nu),\dots(c,\rho)}(\theta)\rangle := U_N(\theta_N) \cdots (iP_{a,\mu}) U_a(\theta_a) \cdots (iP_{b,\nu}) U_b(\theta_b) \cdots (iP_{c,\rho}) U_c(\theta_c) \cdots U_1(\theta_1) |0\rangle^{\otimes n}. \quad (20)$$

Then,

$$\frac{\partial}{\partial \theta_a} \frac{\partial E(\theta^*(x), x)}{\partial \theta_b} = 2 \sum_{\mu, \nu} \sum_{Q \in \mathcal{P}} h_Q(x) g_{a,\mu} g_{b,\nu} \text{Re}[\langle \phi_{(a,\mu),(b,\nu)}(\theta^*(x)) | Q | \psi(\theta^*(x)) \rangle + \langle \phi_{(a,\mu)}(\theta^*(x)) | Q | \phi_{(b,\nu)}(\theta^*(x)) \rangle]. \quad (21)$$

In Fig. 1(a), we show quantum circuits for evaluation of each of the terms in the above equation, which is a variant of the circuit presented in Ref. [39]. By sampling the observable $Z_{\text{anc}} \otimes Q$ of the circuit in Fig. 1(a) and evaluating the expectation value $\langle Z_{\text{anc}} \otimes Q \rangle$, each term of Eq. (21) can be calculated by

$$\begin{aligned} \text{Re}[\langle \phi_{(a,\mu),(b,\nu)}(\theta^*(x)) | Q | \psi(\theta^*(x)) \rangle] &= \langle Z_{\text{anc}} Q \rangle_{0,(a,\mu),(b,\nu)}, \\ \text{Re}[\langle \phi_{(a,\mu)}(\theta^*(x)) | Q | \phi_{(b,\nu)}(\theta^*(x)) \rangle] &= \langle Z_{\text{anc}} Q \rangle_{1,(a,\mu),(b,\nu)}. \end{aligned} \quad (22)$$

The strategy proposed in Ref. [45] gives low-depth versions of the circuits to measure the same quantities. We show the low-depth quantum circuit in Fig. 1(b). From the measurement of $\langle Q \rangle_{(a,\mu,\pm),(b,\nu,\pm)}$ with the circuit in Fig. 1(b), we can evaluate each term of Eq. (21) by the following formula:

$$\begin{aligned} &2\text{Re}[\langle \phi_{(a,\mu),(b,\nu)}(\theta^*(x)) | Q | \psi(\theta^*(x)) \rangle + \langle \phi_{(a,\mu)}(\theta^*(x)) | Q | \phi_{(b,\nu)}(\theta^*(x)) \rangle] \\ &= \langle Q \rangle_{(a,\mu,+),(b,\nu,+)} + \langle Q \rangle_{(a,\mu,-),(b,\nu,-)} - \langle Q \rangle_{(a,\mu,-),(b,\nu,+)} - \langle Q \rangle_{(a,\mu,+),(b,\nu,-)}. \end{aligned} \quad (23)$$

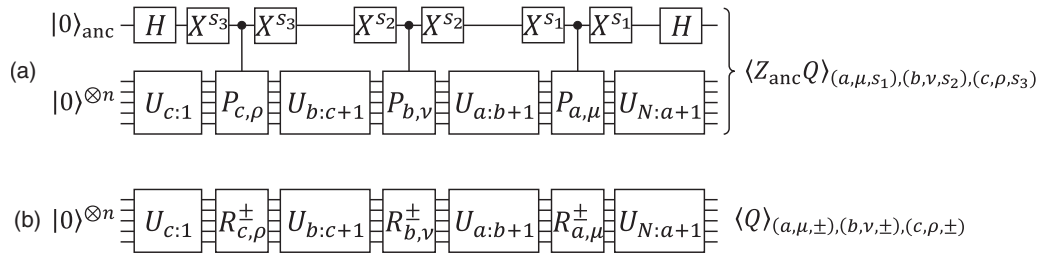


FIG. 2. Quantum circuit to evaluate $\frac{\partial}{\partial \theta_a} \frac{\partial}{\partial \theta_b} \frac{\partial}{\partial \theta_c} E(\theta^*(x), x)$. For the definition of the notations, refer to Fig. 1. (a) Ancilla-based approach. See Eq. (25) for detailed procedure. (b) Low-depth version of (a). See Eq. (26) for detailed procedure.

Notice that the low-depth version doubles the number of the circuit runs, and therefore, if the device can execute the circuit in Fig. 1(a) while maintaining sufficient overall fidelity, it is advantageous to utilize Fig. 1(a) and Eq. (22). Otherwise the low-depth version should be used to obtain meaningful result.

D. Evaluation of $\frac{\partial}{\partial \theta_a} \frac{\partial}{\partial \theta_b} \frac{\partial E}{\partial \theta_c}$

Here, we show how to evaluate $\frac{\partial}{\partial \theta_a} \frac{\partial}{\partial \theta_b} \frac{\partial E}{\partial \theta_c}$. It should be clear how one can extend the method to the order higher than the third. $\frac{\partial}{\partial \theta_a} \frac{\partial}{\partial \theta_b} \frac{\partial E}{\partial \theta_c}$ can be written down as

$$\begin{aligned} & \frac{\partial}{\partial \theta_a} \frac{\partial}{\partial \theta_b} \frac{\partial E(\theta^*(x), x)}{\partial \theta_c} \\ &= 2 \sum_{\mu, \nu, \rho} \sum_{Q \in \mathcal{P}} h_Q(x) g_{a,\mu} g_{b,\nu} g_{c,\rho} \text{Re}[\langle \phi_{(a,\mu),(b,\nu),(c,\rho)}(\theta^*(x)) | Q | \psi(\theta^*(x)) \rangle + \langle \phi_{(a,\mu),(b,\nu)}(\theta^*(x)) | Q | \phi_{(c,\rho)}(\theta^*(x)) \rangle \\ & \quad + \langle \phi_{(a,\mu),(c,\rho)}(\theta^*(x)) | Q | \phi_{(b,\nu)}(\theta^*(x)) \rangle + \langle \phi_{(b,\nu),(c,\rho)}(\theta^*(x)) | Q | \phi_{(a,\mu)}(\theta^*(x)) \rangle]. \end{aligned} \quad (24)$$

Each term can be evaluated with the circuit in Fig. 2(a):

$$\begin{aligned} \text{Re}[\langle \phi_{(a,\mu),(b,\nu),(c,\rho)}(\theta^*(x)) | Q | \psi(\theta^*(x)) \rangle] &= \langle Z_{\text{anc}} Q \rangle_{(a,\mu,0),(b,\nu,0),(c,\rho,0)}, \\ \text{Re}[\langle \phi_{(a,\mu),(b,\nu)}(\theta^*(x)) | Q | \phi_{(c,\rho)}(\theta^*(x)) \rangle] &= \langle Z_{\text{anc}} Q \rangle_{(a,\mu,0),(b,\nu,0),(c,\rho,1)}, \\ \text{Re}[\langle \phi_{(a,\mu),(c,\rho)}(\theta^*(x)) | Q | \phi_{(b,\nu)}(\theta^*(x)) \rangle] &= \langle Z_{\text{anc}} Q \rangle_{(a,\mu,0),(b,\nu,1),(c,\rho,0)}, \\ \text{Re}[\langle \phi_{(b,\nu),(c,\rho)}(\theta^*(x)) | Q | \phi_{(a,\mu)}(\theta^*(x)) \rangle] &= \langle Z_{\text{anc}} Q \rangle_{(a,\mu,1),(b,\nu,0),(c,\rho,0)}. \end{aligned} \quad (25)$$

The circuits can also be reduced to the low-depth version using the same strategy [45]. The low-depth circuit for $\frac{\partial}{\partial \theta_a} \frac{\partial}{\partial \theta_b} \frac{\partial E}{\partial \theta_c}$ is shown in Fig. 2(b). The circuit can evaluate each term of the summation by the following formula:

$$\begin{aligned} & -2\text{Re}[\langle \phi_{(a,\mu),(b,\nu),(c,\rho)}(\theta^*(x)) | Q | \psi(\theta^*(x)) \rangle + \langle \phi_{(a,\mu),(b,\nu)}(\theta^*(x)) | Q | \phi_{(c,\rho)}(\theta^*(x)) \rangle \\ & \quad + \langle \phi_{(a,\mu),(c,\rho)}(\theta^*(x)) | Q | \phi_{(b,\nu)}(\theta^*(x)) \rangle + \langle \phi_{(b,\nu),(c,\rho)}(\theta^*(x)) | Q | \phi_{(a,\mu)}(\theta^*(x)) \rangle] \\ &= \langle Q \rangle_{(a,\mu,+),(b,\nu,+),(c,\rho,+)} - \langle Q \rangle_{(a,\mu,-),(b,\nu,-),(c,\rho,-)} + \langle Q \rangle_{(a,\mu,-),(b,\nu,-),(c,\rho,+)} + \langle Q \rangle_{(a,\mu,-),(b,\nu,+),(c,\rho,-)} \\ & \quad + \langle Q \rangle_{(a,\mu,+),(b,\nu,-),(c,\rho,-)} - \langle Q \rangle_{(a,\mu,-),(b,\nu,+),(c,\rho,+)} - \langle Q \rangle_{(a,\mu,+),(b,\nu,-),(c,\rho,+)} - \langle Q \rangle_{(a,\mu,+),(b,\nu,+),(c,\rho,-)}. \end{aligned} \quad (26)$$

Notice the pattern in the signs associating $\langle Q \rangle_{(a,\mu,\pm),(b,\nu,\pm),(c,\rho,\pm)}$, that is, the signs are determined in the parity of \pm appearing in the subscript. Higher-order derivatives of E with respect to θ can be evaluated by following the strategy described in this and the previous section.

Similar to the previous case, the low-depth version doubles the number of the circuit runs, and therefore, the same argument about the tradeoff between the noise and the number of the circuit runs also applies in this case.

E. Evaluation of $\frac{\partial}{\partial \theta_a} \frac{\partial}{\partial \theta_b} \dots \frac{\partial}{\partial \theta_c} \frac{\partial E}{\partial x_i}$

$\frac{\partial}{\partial \theta_a} \frac{\partial}{\partial \theta_b} \dots \frac{\partial}{\partial \theta_c} \frac{\partial E}{\partial x_i}$ can be measured with the same protocol as described in the previous sections; one can substitute h_Q

with $\frac{\partial h_Q}{\partial x_i}$. With this substitution, Eqs. (21) and (24) give us the analytical formula for $\frac{\partial}{\partial \theta_a} \frac{\partial}{\partial \theta_b} \dots \frac{\partial}{\partial \theta_c} \frac{\partial E}{\partial x_i}$, where each term in the summation can be evaluated with the same procedure. Also, $\frac{\partial}{\partial \theta_a} \frac{\partial}{\partial x_i} \frac{\partial E}{\partial x_i}$, which appears in Eq. (9), can be measured using the same strategy.

V. COMPUTATIONAL COST

Here, we give a comparison between the presented algorithm and the numerical differentiation to take d th energy derivatives. We specifically discuss the cost for each step in the overall algorithm presented in Sec. IV B for quantum chemistry problem, where a problem Hamiltonian and its

derivatives always has $O(n^4)$ terms as expressed by Eq. (14), as it is the main target of the proposed algorithm. This analysis should be easy to be extended to more general cases. We assume that classical computational cost involved in algorithms is much smaller than the quantum ones and thus can be ignored. The classical computation required to perform the proposed protocol is to calculate the derivative of the Hamiltonian (and the Hamiltonian itself) which formally takes $O(n^5)$ computational time [9,36] for each $\frac{\partial H}{\partial x_i}, \frac{\partial}{\partial x_i} \frac{\partial H}{\partial x_j}, \dots$. Note that, if we are only interested in evaluating analytical derivative of the Hartree-Fock energy, which is not the case considering our algorithm, it can be reduced to $O(n^2)$ [46]. We believe that, although the time required to estimate the expectation value of the Hamiltonian by repeatedly running the NISQ devices takes the order of $O(n^4)$, the classical computation performed at the level of some tens or hundreds of qubits is much faster than the time used by the NISQ device by a large constant prefactor ignored in the order notation.

First, we ignore the cost of performing the VQE (step 1), since both of the numerical differentiation and the presented algorithm need this step. Steps 2, 5, and 6 is merely classical, and therefore, we also ignore the cost for this part. Step 3 of the presented algorithm requires us to run a quantum computer $O(n^4/\epsilon^2)$ times to estimate each term up to an additional error of ϵ . For step 4, it takes $O(n^4 \sum_{k=0}^{d-1} (N_\theta^{d-k})/\epsilon^2) = O(n^4 N_\theta^d/\epsilon^2)$ runs. Note that even for the calculation of the terms involving x - and θ -differentiation at the same time, such as $\frac{\partial}{\partial \theta_a} \frac{\partial E}{\partial x_i}$, we only need to perform measurement of the θ -derivatives of each term, such as $\frac{\partial}{\partial \theta_a} \langle \psi(\theta) | c_i^\dagger c_j^\dagger c_k c_l | \psi(\theta) \rangle$, in the Hamiltonian on the quantum device and then combine them by $\sum_{ijkl} \frac{\partial h_{ijkl}(x)}{\partial x} \frac{\partial}{\partial \theta_a} \langle \psi(\theta) | c_i^\dagger c_j^\dagger c_k c_l | \psi(\theta) \rangle$ where x -differentiated $h(x)$ is computed classically. Therefore N_x does not appear in the number of NISQ runs. Overall, the cost for the quantum part of the computation scales as $O(n^4 N_\theta^d/\epsilon^2)$. In this simplified analysis, we ignored the cost caused by the condition number of the Hessian matrix, $\{\frac{\partial^2 E}{\partial \theta_a \partial \theta_b}\}$, which depends highly on the ansatz circuit. See Appendix C for a detailed analysis of the error.

Let us compare this cost against the numerical differentiation. More specifically, we consider the case of evaluating the second derivative by the formula,

$$\frac{\partial^2 E^*(x)}{\partial x_i^2} \approx \frac{E^*(x+h) + E^*(x-h) - 2E^*(x)}{h^2}, \quad (27)$$

where $h > 0$. Let $\tilde{E}^*(x)$ be the estimate of $E^*(x)$ obtained by measuring the state $|\psi(\theta^*(x))\rangle$. If $|\tilde{E}^*(x) - E^*(x)| \leq \epsilon_E$, the precision of Eq. (27) is

$$\left| \frac{\partial^2 E^*(x)}{\partial x_i^2} - \frac{\tilde{E}^*(x+h) + \tilde{E}^*(x-h) - 2\tilde{E}^*(x)}{h^2} \right| \leq O\left(h^2 \left| \frac{\partial^4 E}{\partial x_i^4} \right| \right) + O(\epsilon_E/h^2). \quad (28)$$

See Appendix C 2 for a detailed analysis. For classical computation where the source of ϵ_E is mainly the round-off error, the second term of the right-hand side is usually negligibly

small for a decent h . On the other hand, for the VQE, this term is the leading factor for the precision, since $\tilde{E}^*(x)$ has to be calculated by sampling. To achieve $|\tilde{E}^*(x) - E^*(x)| \leq \epsilon_E$ with high probability, we need to run the quantum computer for $O[(\sum_P |h_P|)^2/\epsilon_E^2]$ times (Appendix C). Therefore, if we want to achieve the precision of ϵ at least for $O(\epsilon_E/h^2)$ term, we need $O[(\sum_P |h_P|)^2/(\epsilon^2 h^4)]$ samples from $|\psi(\theta^*(x))\rangle$, and overall cost is $O[n^4 N_x^2 (\sum_P |h_P|)^2/(\epsilon^2 h^4)]$. h^4 factor, which is not present in the analytical differentiation, makes the scaling of the finite different approach worse. Furthermore, $O(h^2 |\frac{\partial^4 E}{\partial x_i^4}|)$ term makes the computation somewhat unreliable, since we do not know the value of $\frac{\partial^4 E}{\partial x_i^4}$ a priori. In practice, we would be forced to try some different h to certify that the computation is correct. On the other hand, in the analytical differentiation approach, we can estimate the error bounds from the measured terms, for example, by estimating the condition number of the hessian matrix, $\{\frac{\partial^2 E}{\partial \theta_a \partial \theta_b}\}$.

VI. DERIVATIVES OF EXCITED STATE ENERGY

The generation of excited states can be a powerful application of the VQE, because the classical computation, despite the recent significant improvement in the theory and the computational power, still suffers in the calculation of them [47]. Among the several proposals [2–5] to generate excited states with the VQE, we adopt the one proposed in Refs. [3,4] to compute the derivatives of the excited state energy.

The algorithm [3,4] works as follows. First, we find the ground state of the given Hamiltonian $H_0(x)$, which we denote by $|\psi^{(0)}(\theta^{(0)}(x))\rangle$, where $\theta^{(0)}(x)$ is the optimal parameter for the ground state ($\theta^*(x)$ in the previous sections). Then, we iteratively define a Hamiltonian $H_r(x)$ for $r = 1, 2, \dots$ as

$$H_r(x) := H_0(x) + \sum_{s=0}^{r-1} \beta_s |\psi^{(s)}(\theta^{(s)}(x))\rangle \langle \psi^{(s)}(\theta^{(s)}(x))|, \quad (29)$$

where $|\psi^{(r)}(\theta^{(r)}(x))\rangle$ is the ground state of $H_r(x)$. If β_s is sufficiently large, each $H_r(x)$ has r th excited state of the original Hamiltonian, $H_0(x)$, as its ground state. Therefore, by finding the ground state of each $H_r(x)$, one can generate the series of excited states of $H_0(x)$. We assume $|\psi^{(r)}(\theta)\rangle = U^{(r)}(\theta)|0\rangle$, where $U^{(r)}(\theta)$ has the same structure as $U(\theta)$ in previous sections. In this algorithm, it is also assumed that we have a device which can measure the magnitude of overlap between $|\psi^{(r)}(\theta)\rangle$ and $|\psi^{(s)}(\varphi)\rangle$, that is, we assume we can measure $|\langle \psi^{(r)}(\theta) | \psi^{(s)}(\varphi) \rangle|^2$. Let the expectation value of $H_r(x)$ with respect to the state $|\psi^{(r)}(\theta)\rangle$ be $E_r(\theta, x)$; $E_r(\theta, x) = \langle \psi^{(r)}(\theta) | H_r(x) | \psi^{(r)}(\theta) \rangle$. We define the optimal energy by $E_r^*(x) = E_r(\theta^{(r)}(x), x)$.

The task is to compute the derivatives such as $\frac{\partial^2 E_r^*}{\partial x_i \partial x_j}$. Since E_r^* is the ground-state energy for H_r , Eqs. (7)–(9) can be adapted for this task. The only difference from that of the actual ground state is that the derivative of the Hamiltonian, such as $\frac{\partial^2 H_r}{\partial x_i \partial x_j}$, cannot be computed classically. For example,

the expression of the first derivative of the Hamiltonian is

$$\begin{aligned}\frac{\partial H_r}{\partial x_i}(x) &= \frac{\partial H_0}{\partial x_i}(x) + \sum_{s=0}^{r-1} \beta_s \left(\frac{\partial |\psi^{(s)}(\theta^{(s)}(x))\rangle}{\partial x_i} \langle \psi^{(s)}(\theta^{(s)}(x)) | + \text{H.c.} \right) \\ &= \frac{\partial H_0}{\partial x_i}(x) + \sum_{s=0}^{r-1} \sum_a \beta_s \frac{\partial \theta_a^{(s)}}{\partial x_i}(x) (|\partial_a \psi^{(s)}(\theta^{(s)}(x))\rangle \langle \psi^{(s)}(\theta^{(s)}(x)) | + \text{H.c.}).\end{aligned}\quad (30)$$

In the analytical expression for $\frac{\partial E_r^*}{\partial x_i}$ [Eq. (7)], $\frac{\partial H_r}{\partial x_i}(x)$ appears as the expectation value with respect to $|\psi^{(r)}(\theta^{(r)}(x))\rangle$; $\langle \psi^{(r)}(\theta^{(r)}(x)) | \frac{\partial H_r}{\partial x_i}(x) | \psi^{(r)}(\theta^{(r)}(x)) \rangle$. On the other hand, in the expression of $\frac{\partial^2 E_r^*}{\partial x_i \partial x_j}$ [Eq. (8)], it appears as $\text{Re}[\langle \psi^{(r)}(\theta^{(r)}(x)) | \frac{\partial H_r}{\partial x_i}(x) | \partial_a \psi^{(r)}(\theta^{(r)}(x)) \rangle]$. The quantities that cannot be computed classically in Eqs. (7) to (9) are the terms which involves the inner product between the states, such as $\text{Re}[\langle \psi^{(r)}(\theta^{(r)}(x)) | \partial_a \psi^{(s)}(\theta^{(s)}(x)) \rangle \langle \psi^{(s)}(\theta^{(s)}(x)) | \psi^{(r)}(\theta^{(r)}(x)) \rangle]$. However, notice that if the condition $\langle \psi^{(r)}(\theta^{(r)}(x)) | \psi^{(s)}(\theta^{(s)}(x)) \rangle = 0$ holds for all x , which we naturally expect at the optimal parameter, we obtain by differentiating with respect to x the both hand side of the equation,

$$\begin{aligned}\sum_a \frac{\partial \theta_a^{(s)}}{\partial x_i}(x) \text{Re}[\langle \psi^{(r)}(\theta^{(r)}(x)) | \partial_a \psi^{(s)}(\theta^{(s)}(x)) \rangle \\ \times \langle \psi^{(s)}(\theta^{(s)}(x)) | \psi^{(r)}(\theta^{(r)}(x)) \rangle] = 0.\end{aligned}\quad (31)$$

We can therefore ignore the inner-product terms for the evaluation of derivatives of excited state energy and utilize the same procedure as the ground-state energy in this case. Likewise, the inner-product terms that appear in the higher-order derivatives can also be ignored when the orthogonality condition $\langle \psi^{(r)}(\theta^{(r)}(x)) | \psi^{(s)}(\theta^{(s)}(x)) \rangle = 0$ is satisfied. The so-called subspace search VQE method [2], which guarantees the orthogonality condition of the resultant states, can be advantageous to fulfill such condition.

VII. NUMERICAL SIMULATION

We provide proof-of-principle numerical simulations using the electronic Hamiltonian of the hydrogen molecule. The Hamiltonians are calculated with the open source library PYSCF [48] and OPENFERMION [49]. The simulation of quantum circuits are performed with QULACS [50].

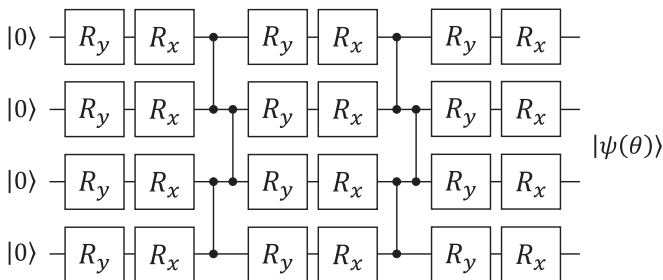


FIG. 3. Ansatz used in simulations. R_x and R_y is single-qubit x and y rotation gate, respectively. The parameters θ are implemented as rotation angles of R_x and R_y .

A. Approximation of the potential energy surface

First, we perform a simulation of the VQE and the methods described in Sec. IV with the Hamiltonian of a hydrogen molecule, calculated with the STO-3G basis set, at the bonding distance $r = 0.735 \text{ \AA}$. We use the ansatz circuit shown in Fig. 3, which is a variant of so-called hardware efficient circuit [41]. The result of the simulation is shown as Fig. 4. The ansatz used in this simulation could achieve the exact ground state, which is called full configuration interaction (Full-CI) state in the context of chemistry, and therefore, we could draw the harmonic and the third-order approximation of the Full-CI energy as shown in Fig. 4. The harmonic approximation can be used to calculate the vibrational spectra of a molecule, and the third-order approximation can be utilized for more accurate description of the vibration.

B. Continuous determination of the optimal parameter

The response equation, Eq. (10), can be used to determine the optimal parameter $\theta^*(x)$ from the one at the slightly different system parameter, $\theta^*(x + \delta x)$, that is, to the first order,

$$\theta^*(x + \delta x) \approx \theta^*(x) + \frac{\partial \theta^*}{\partial x}(x) \delta x, \quad (32)$$

holds up to the additive error of $O(\delta x^2)$. One can iteratively use the equation above, which resembles the Euler method, to obtain the optimal parameter $\theta_{\text{Euler}}(x) \approx \theta^*(x)$ from some $\theta^*(x_0)$ in a range of x around x_0 where the error term is sufficiently small.

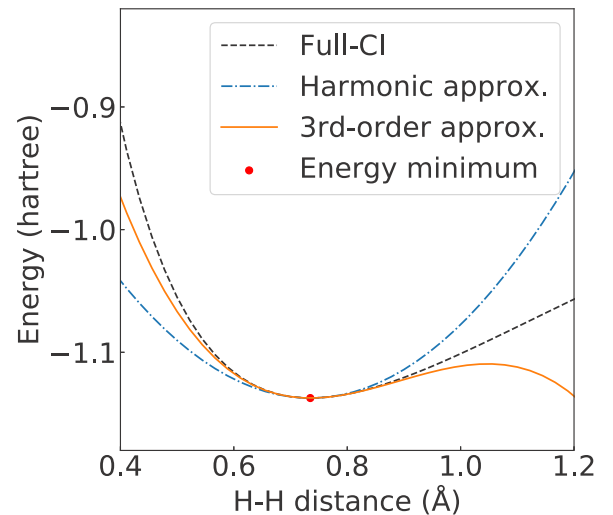


FIG. 4. The harmonic and third-order approximation of the energy curve of the hydrogen molecule at the bonding distance, determined by the simulation of the VQE and the proposed method.

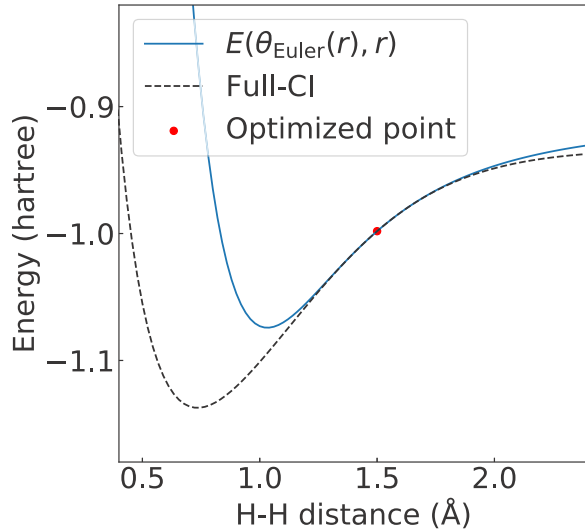


FIG. 5. The evolution of the energy expectation value when the circuit parameters are evolved according to Eq. (32), starting from the optimized parameter at interatomic distance of 1.5 Å.

We demonstrate this parameter update in Fig. 5, by plotting the energy expectation value of $|\psi(\theta_{\text{Euler}}(r))\rangle$ where $\theta_{\text{Euler}}(r)$ is determined with the iterative use of Eq. (32), starting from the optimal parameter at $r = 1.5 \text{ \AA}$ with the step size $\delta r = 0.02 \text{ \AA}$. Equation (32) resembles the Euler method, therefore we refer to the update of θ according to Eq. (32) as so in Fig. 5. It can be seen that in the range where r is sufficiently close to the optimized point, $r = 1.5 \text{ \AA}$, the parameter determined from Eq. (32) is almost optimal, but as r goes far from the optimized point, they deviate fast from the optima. This is because, as the error of the each update accumulates, the state becomes a nonvariational state, that is, $\frac{\partial E}{\partial \theta} \neq 0$, which results in the breakdown of the response equation, Eq. (10), where we assumed $\frac{\partial E}{\partial \theta} = 0$. This method should be useful for drawing potential energy surfaces using the VQE, because it can reduce the time for the optimization of the circuit parameter by predicting the optimal parameter from the one determined with the slightly different system. We believe it can also be combined with the parameter interpolation approach proposed in Ref. [51].

VIII. CONCLUSION

We have described a methodology for computing the derivatives of the ground state and the excited state energy. We have shown the straight-forwardly constructed quantum circuit to measure the quantities necessary for the calculation of energy derivatives, and also the low-depth version of the circuit. The low-depth protocol for the measurement makes them suitable for the use in NISQ devices. We also performed numerical simulations of the proposed method, which validates the correctness. This work enables to extract the physical properties of the quantum system under investigation, thus widening the application range of the NISQ devices.

Note added. During the final revision of this paper, we have become aware of recent article [52] which also describes a methodology for evaluating energy derivatives. Their work is based on perturbative sum-over-state approach and differs from our methods which provides a way to evaluate analytic derivatives.

ACKNOWLEDGMENTS

K.M. thanks for METI and IPA for its support through MITOU Target program. K.M. is supported by JSPS KAKENHI Grant No. 19J10978. W.M. thanks for KAKENHI Grant No. 18K14181. This work is supported by MEXT Quantum Leap Flagship Program (MEXT Q-LEAP) Grant No. JPMXS0118067394.

APPENDIX A: DERIVATIVE OF THE OPTIMAL PARAMETER

We derive the expression for the derivatives of the optimal parameter, such as $\frac{\partial \theta_a^*(x)}{\partial x_i}$, by Taylor expansion. Assume that at $x = \alpha$, we know the optimal parameter $\theta^*(\alpha)$. We perform Taylor expansion of $|\psi(\theta^*(x))\rangle$ and $H(x)$ around α . For $H(x)$, we obtain

$$H(\alpha + x) = H(\alpha) + \sum_i \frac{\partial H(\alpha)}{\partial x_i} x_i + \frac{1}{2} \sum_i \frac{\partial}{\partial x_i} \frac{\partial H(\alpha)}{\partial x_j} x_i x_j + \dots \quad (\text{A1})$$

For $|\psi(\theta^*(x))\rangle$, we can expand it as follows:

$$\begin{aligned} |\psi(\theta^*(\alpha + x))\rangle &= \left| \psi \left(\theta^*(\alpha) + \sum_i \frac{\partial \theta^*(\alpha)}{\partial x_i} x_i + \frac{1}{2} \sum_{i,j} \frac{\partial}{\partial x_i} \frac{\partial \theta^*(\alpha)}{\partial x_j} x_i x_j + \dots \right) \right\rangle \\ &= |\psi(\theta^*(\alpha))\rangle + \sum_a |\partial_a \psi(\theta^*(\alpha))\rangle \left[\sum_i \frac{\partial \theta_a^*(\alpha)}{\partial x_i} x_i + \frac{1}{2} \sum_{i,j} \frac{\partial}{\partial x_i} \frac{\partial \theta_a^*(\alpha)}{\partial x_j} x_i x_j + \dots \right] \\ &\quad + \frac{1}{2} \sum_{a,b} |\partial_a \partial_b \psi(\theta^*(\alpha))\rangle \left[\sum_i \frac{\partial \theta_a^*(\alpha)}{\partial x_i} x_i + \frac{1}{2} \sum_{i,j} \frac{\partial}{\partial x_i} \frac{\partial \theta_a^*(\alpha)}{\partial x_j} x_i x_j + \dots \right] \\ &\quad \times \left[\sum_i \frac{\partial \theta_b^*(\alpha)}{\partial x_i} x_i + \frac{1}{2} \sum_{i,j} \frac{\partial}{\partial x_i} \frac{\partial \theta_b^*(\alpha)}{\partial x_j} x_i x_j + \dots \right] + \dots \end{aligned}$$

When grouped by the order of x ,

$$\begin{aligned}
 |\psi(\theta^*(\alpha + x))\rangle = & |\psi(\theta^*(\alpha))\rangle + \sum_{a,i} \frac{\partial \theta_a^*(\alpha)}{\partial x_i} |\partial_a \psi(\theta^*(\alpha))\rangle x_i + \frac{1}{2} \sum_{i,j} \left[\sum_a \frac{\partial}{\partial x_i} \frac{\partial \theta_a^*(\alpha)}{\partial x_j} |\partial_a \psi(\theta^*(\alpha))\rangle \right. \\
 & + \sum_{a,b} \frac{\partial \theta_a^*(\alpha)}{\partial x_i} \frac{\partial \theta_b^*(\alpha)}{\partial x_j} |\partial_a \partial_b \psi(\theta^*(\alpha))\rangle \left. \right] x_i x_j + \frac{1}{6} \sum_{i,j,k} \left[\sum_a \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} \frac{\partial \theta_a^*(\alpha)}{\partial x_k} |\partial_a \psi(\theta^*(\alpha))\rangle \right. \\
 & + 3 \sum_{a,b} \frac{\partial}{\partial x_i} \frac{\partial \theta_a^*(\alpha)}{\partial x_j} \frac{\partial \theta_b^*(\alpha)}{\partial x_j} |\partial_a \partial_b \psi(\theta^*(\alpha))\rangle + \sum_{a,b,c} \frac{\partial \theta_a^*(\alpha)}{\partial x_j} \frac{\partial \theta_b^*(\alpha)}{\partial x_j} \frac{\partial \theta_c^*(\alpha)}{\partial x_k} |\partial_a \partial_b \partial_c \psi(\theta^*(\alpha))\rangle \left. \right] x_i x_j x_k + \dots .
 \end{aligned}
 \tag{A2}$$

We can derive a similar expression for $|\partial_a \psi(\theta^*(\alpha + x))\rangle$. Now, we use the condition of Eq. (1) to derive the expression of the derivatives of $\theta^*(\alpha)$. Plugging Eqs. (A1) and (A2) into Eq. (1) and imposing the coefficient of each order in x to be zero, we get the analytical expression for the derivatives of $\theta^*(\alpha)$. In the following two sections, we derive the expression for the first and second derivatives of $\theta^*(\alpha)$.

1. First derivative

For the first order in x , we get

$$\text{Re} \left[\sum_a \left(\langle \partial_b \psi(\theta^*(\alpha)) | H(\alpha) | \partial_a \psi(\theta^*(\alpha)) \rangle + \langle \psi(\theta^*(\alpha)) | H(\alpha) | \partial_b \partial_a \psi(\theta^*(\alpha)) \rangle \right) \frac{\partial \theta_a^*(\alpha)}{\partial x_i} + \langle \partial_b \psi(\theta^*(\alpha)) | \frac{\partial H(\alpha)}{\partial x_i} | \psi(\theta^*(\alpha)) \rangle \right] = 0.
 \tag{A3}$$

Note that the first term of Eq. (A3) includes the Hessian of $E(\theta, x)$, that is,

$$\frac{\partial}{\partial \theta_b} \frac{\partial E(\theta^*(\alpha), \alpha)}{\partial \theta_a} = 2\text{Re}[\langle \partial_b \psi(\theta^*(\alpha)) | H(\alpha) | \partial_a \psi(\theta^*(\alpha)) \rangle + \langle \psi(\theta^*(\alpha)) | H(\alpha) | \partial_b \partial_a \psi(\theta^*(\alpha)) \rangle].
 \tag{A4}$$

Also, notice that

$$\frac{\partial}{\partial \theta_b} \frac{\partial E(\theta, x)}{\partial x_i} = 2\text{Re} \left[\langle \partial_b \psi(\theta) | \frac{\partial H}{\partial x_i}(x) | \psi(\theta) \rangle \right].
 \tag{A5}$$

We define $\nabla_{\theta}^d \frac{\partial E}{\partial x_i}$ likewise. Finally, we obtain

$$\sum_a \frac{\partial}{\partial \theta_b} \frac{\partial E(\theta^*(\alpha), \alpha)}{\partial \theta_a} \frac{\partial \theta_a^*(\alpha)}{\partial x_i} = - \frac{\partial}{\partial \theta_b} \frac{\partial E(\theta^*(x), x)}{\partial x_i},
 \tag{A6}$$

which is Eq. (10) of the main text. Note that we expect the matrix $\frac{\partial}{\partial \theta_b} \frac{\partial E(\theta^*(\alpha), \alpha)}{\partial \theta_a}$ to be positive definite because $\theta^*(\alpha)$ is a local minimum, and thus Eq. (A6) is solvable to obtain $\frac{\partial \theta_a^*(\alpha)}{\partial x_i}$.

2. Second derivative

The second derivative, $\frac{\partial}{\partial x_i} \frac{\partial \theta_a}{\partial x_j}$, is derived from the second order in x of Eq. (1). We have, from Eq. (1), for all i, j , and c ,

$$\begin{aligned}
 \text{Re} \left[\frac{1}{2} \sum_{a,b} \frac{\partial \theta_a^*(\alpha)}{\partial x_i} \frac{\partial \theta_b^*(\alpha)}{\partial x_j} \left(\langle \partial_c \psi(\theta^*(\alpha)) | H(\alpha) | \partial_a \partial_b \psi(\theta^*(\alpha)) \rangle + \langle \partial_a \partial_b \partial_c \psi(\theta^*(\alpha)) | H(\alpha) | \psi(\theta^*(\alpha)) \rangle \right) \right. \\
 + \langle \partial_a \partial_c \psi(\theta^*(\alpha)) | H(\alpha) | \partial_b \psi(\theta^*(\alpha)) \rangle + \langle \partial_b \partial_c \psi(\theta^*(\alpha)) | H(\alpha) | \partial_a \psi(\theta^*(\alpha)) \rangle \\
 + \frac{1}{2} \sum_a \frac{\partial}{\partial x_i} \frac{\partial \theta_a^*(\alpha)}{\partial x_j} \left(\langle \partial_c \psi(\theta^*(\alpha)) | H(\alpha) | \partial_a \psi(\theta^*(\alpha)) \rangle + \langle \partial_a \partial_c \psi(\theta^*(\alpha)) | H(\alpha) | \psi(\theta^*(\alpha)) \rangle \right) \\
 + \sum_a \frac{\partial \theta_a^*(\alpha)}{\partial x_j} \left(\langle \partial_c \psi(\theta^*(\alpha)) | \frac{\partial H(\alpha)}{\partial x_j} | \partial_a \psi(\theta^*(\alpha)) \rangle + \langle \partial_a \partial_c \psi(\theta^*(\alpha)) | \frac{\partial H(\alpha)}{\partial x_j} | \psi(\theta^*(\alpha)) \rangle \right) \\
 \left. + \frac{1}{2} \langle \partial_c \psi(\theta^*(\alpha)) | \frac{\partial}{\partial x_i} \frac{\partial H(\alpha)}{\partial x_j} | \psi(\theta^*(\alpha)) \rangle \right] = 0.
 \end{aligned}
 \tag{A7}$$

$\frac{\partial}{\partial \theta_a} \frac{\partial}{\partial \theta_b} \frac{\partial E}{\partial \theta_c}, \frac{\partial}{\partial \theta_a} \frac{\partial E}{\partial \theta_b}$ can be used to greatly simplify the above equation, which gives us

$$\begin{aligned} & \frac{1}{4} \sum_{a,b} \frac{\partial \theta_a^*(\alpha)}{\partial x_i} \frac{\partial \theta_b^*(\alpha)}{\partial x_j} \frac{\partial}{\partial \theta_c} \frac{\partial}{\partial \theta_a} \frac{\partial E(\theta^*(\alpha), \alpha)}{\partial \theta_b} + \frac{1}{4} \sum_a \frac{\partial}{\partial x_i} \frac{\partial \theta_a^*(\alpha)}{\partial x_j} \frac{\partial}{\partial \theta_c} \frac{\partial E(\theta^*(\alpha), \alpha)}{\partial \theta_a} \\ & + \frac{1}{2} \sum_a \frac{\partial \theta_a^*(\alpha)}{\partial x_i} \frac{\partial}{\partial \theta_c} \frac{\partial}{\partial \theta_a} \frac{\partial E(\theta^*(\alpha), \alpha)}{\partial x_j} + \frac{1}{2} \text{Re} \left[\langle \partial_c \psi(\theta^*(\alpha)) | \frac{\partial}{\partial x_i} \frac{\partial H(\alpha)}{\partial x_j} | \psi(\theta^*(\alpha)) \rangle \right] = 0. \end{aligned} \quad (\text{A8})$$

This is equivalent to Eq. (11).

APPENDIX B: DERIVATIVES OF THE GROUND-STATE ENERGY

1. First derivative

The first derivative of the energy is calculated as

$$\begin{aligned} \frac{\partial E^*}{\partial x_i}(x) &= \frac{\partial}{\partial x_i} \langle \psi(\theta^*(x)) | H(x) | \psi(\theta^*(x)) \rangle, \\ &= 2 \text{Re} \left[\langle \psi(\theta^*(x)) | H(x) \frac{\partial | \psi(\theta^*(x)) \rangle}{\partial x_i} \right] + \langle \psi(\theta^*(x)) | \frac{\partial H}{\partial x_i}(x) | \psi(\theta^*(x)) \rangle, \\ &= 2 \sum_a \frac{\partial \theta_a^*(x)}{\partial x_i} \text{Re} [\langle \psi(\theta^*(x)) | H(x) | \partial_a \psi(\theta^*(x)) \rangle] + \langle \psi(\theta^*(x)) | \frac{\partial H}{\partial x_i}(x) | \psi(\theta^*(x)) \rangle, \end{aligned}$$

and first term of the above equation vanishes by Eq. (1) of the main text. Thus we get

$$\frac{\partial E^*}{\partial x_i}(x) = \langle \psi(\theta^*(x), x) | \frac{\partial H}{\partial x_i}(x) | \psi(\theta^*(x), x) \rangle, \quad (\text{B1})$$

which is Eq. (7) of the main text.

2. Second derivative

Here, we derive the expression for the second derivative:

$$\begin{aligned} \frac{\partial}{\partial x_i} \frac{\partial E^*}{\partial x_j}(x) &= \frac{\partial}{\partial x_i} \langle \psi(\theta^*(x)) | \frac{\partial H}{\partial x_j}(x) | \psi(\theta^*(x)) \rangle \\ &= 2 \text{Re} \left[\langle \psi(\theta^*(x)) | \frac{\partial H}{\partial x_j}(x) \frac{\partial | \psi(\theta^*(x)) \rangle}{\partial x_i} \right] + \langle \psi(\theta^*(x)) | \frac{\partial}{\partial x_i} \frac{\partial H}{\partial x_j}(x) | \psi(\theta^*(x)) \rangle \\ &= 2 \sum_a \text{Re} \left[\langle \psi(\theta^*(x)) | \frac{\partial H}{\partial x_j}(x) | \partial_a \psi(\theta^*(x)) \rangle \right] \frac{\partial \theta_a^*(x)}{\partial x_i} + \langle \psi(\theta^*(x)) | \frac{\partial}{\partial x_i} \frac{\partial H}{\partial x_j}(x) | \psi(\theta^*(x)) \rangle \\ &= \sum_a \frac{\partial \theta_a^*(x)}{\partial x_i} \frac{\partial}{\partial \theta_a} \frac{\partial E}{\partial x_j}(\theta^*(x), x) + \langle \psi(\theta^*(x)) | \frac{\partial}{\partial x_i} \frac{\partial H}{\partial x_j}(x) | \psi(\theta^*(x)) \rangle. \end{aligned} \quad (\text{B2})$$

This is Eq. (8) of the main text.

3. Third derivative

Third derivative can be calculated as follows:

$$\begin{aligned} \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} \frac{\partial E^*}{\partial x_k}(x) &= \frac{\partial}{\partial x_i} \left(\sum_a \frac{\partial \theta_a^*(x)}{\partial x_j} \frac{\partial}{\partial \theta_a} \frac{\partial E}{\partial x_k}(\theta^*(x), x) + \langle \psi(\theta^*(x)) | \frac{\partial}{\partial x_j} \frac{\partial H}{\partial x_k}(x) | \psi(\theta^*(x)) \rangle \right) \\ &= \sum_a \frac{\partial}{\partial x_i} \frac{\partial \theta_a^*(x)}{\partial x_j} \frac{\partial}{\partial \theta_a} \frac{\partial E}{\partial x_k}(\theta^*(x), x) + \sum_a \frac{\partial \theta_a^*(x)}{\partial x_j} \frac{\partial}{\partial x_i} \left(\frac{\partial}{\partial \theta_a} \frac{\partial E}{\partial x_k}(\theta^*(x), x) \right) \\ &\quad + 2 \text{Re} \left[\frac{\partial \langle \psi(\theta^*(x)) |}{\partial x_i} \frac{\partial}{\partial x_j} \frac{\partial H}{\partial x_k}(x) | \psi(\theta^*(x)) \right] + \langle \psi(\theta^*(x)) | \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} \frac{\partial H}{\partial x_k}(x) | \psi(\theta^*(x)) \rangle \\ &= \sum_a \frac{\partial}{\partial x_i} \frac{\partial \theta_a^*(x)}{\partial x_j} \frac{\partial}{\partial \theta_a} \frac{\partial E}{\partial x_k}(\theta^*(x), x) + \sum_{a,b} \frac{\partial \theta_a^*(x)}{\partial x_j} \frac{\partial \theta_b^*(x)}{\partial x_i} \left(\frac{\partial}{\partial \theta_b} \frac{\partial}{\partial \theta_a} \frac{\partial E}{\partial x_k}(\theta^*(x), x) \right) \end{aligned}$$

$$\begin{aligned}
& + 2 \sum_a \frac{\partial \theta_a^*}{\partial x_j}(x) \text{Re} \left[\langle \partial_a \psi(\theta^*(x)) | \frac{\partial}{\partial x_i} \frac{\partial H}{\partial x_k} | \psi(\theta^*(x)) \rangle \right] \\
& + 2 \sum_a \frac{\partial \theta_a^*}{\partial x_i}(x) \text{Re} \left[\langle \partial_a \psi(\theta^*(x)) | \frac{\partial}{\partial x_j} \frac{\partial H}{\partial x_k}(x) | \psi(\theta^*(x)) \rangle \right] + \langle \psi(\theta^*(x)) | \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} \frac{\partial H}{\partial x_k}(x) | \psi(\theta^*(x)) \rangle. \quad (\text{B3})
\end{aligned}$$

This expression [Eq. (B3)] can be simplified so as to avoid the explicit calculation of $\frac{\partial}{\partial x_i} \frac{\partial \theta_a^*}{\partial x_j}$. By multiplying $\frac{\partial \theta_a}{\partial x_k}$ to Eq. (10) and $\frac{\partial}{\partial x_i} \frac{\partial \theta_a}{\partial x_j}$ to Eq. (11), and combining them, we obtain

$$\frac{\partial}{\partial x_i} \frac{\partial \theta_a}{\partial x_j} \frac{\partial}{\partial \theta_a} \frac{\partial E}{\partial x_k} = \sum_{b,c} \frac{\partial}{\partial \theta_a} \frac{\partial}{\partial \theta_b} \frac{\partial E}{\partial \theta_c} \frac{\partial \theta_a}{\partial x_k} \frac{\partial \theta_b}{\partial x_i} \frac{\partial \theta_c}{\partial x_j} + 2 \sum_b \frac{\partial}{\partial \theta_a} \frac{\partial}{\partial \theta_b} \frac{\partial E}{\partial x_j} \frac{\partial \theta_b}{\partial x_i} \frac{\partial \theta_a}{\partial x_k} + 2 \text{Re} \left[\langle \partial_a \psi | \frac{\partial}{\partial x_i} \frac{\partial H}{\partial x_j} | \psi \rangle \right] \frac{\partial \theta_a}{\partial x_k}. \quad (\text{B4})$$

This yields

$$\begin{aligned}
\frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} \frac{\partial E^*}{\partial x_k}(x) & = \sum_{a,b,c} \frac{\partial}{\partial \theta_a} \frac{\partial}{\partial \theta_b} \frac{\partial E}{\partial \theta_c}(\theta^*(x), x) \frac{\partial \theta_a}{\partial x_i}(x) \frac{\partial \theta_b}{\partial x_j}(x) \frac{\partial \theta_c}{\partial x_k}(x) + \langle \psi(\theta^*(x)) | \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} \frac{\partial H}{\partial x_k}(x) | \psi(\theta^*(x)) \rangle \\
& + \sum_{a,b} \left[\frac{\partial \theta_a^*}{\partial x_i}(x) \frac{\partial \theta_b^*}{\partial x_j}(x) \frac{\partial}{\partial \theta_b} \frac{\partial}{\partial \theta_a} \frac{\partial E}{\partial x_k}(\theta^*(x), x) + \frac{\partial \theta_a^*}{\partial x_k}(x) \frac{\partial \theta_b^*}{\partial x_i}(x) \frac{\partial}{\partial \theta_b} \frac{\partial}{\partial \theta_a} \frac{\partial E}{\partial x_j}(\theta^*(x), x) \right. \\
& + \left. \frac{\partial \theta_a^*}{\partial x_j}(x) \frac{\partial \theta_b^*}{\partial x_k}(x) \frac{\partial}{\partial \theta_b} \frac{\partial}{\partial \theta_a} \frac{\partial E}{\partial x_i}(\theta^*(x), x) \right] + 2 \sum_a \left[\frac{\partial \theta_a^*}{\partial x_i}(x) \text{Re} \left[\langle \partial_a \psi(\theta^*(x)) | \frac{\partial}{\partial x_j} \frac{\partial H}{\partial x_k}(x) | \psi(\theta^*(x)) \rangle \right] \right. \\
& + \left. \frac{\partial \theta_a^*}{\partial x_k}(x) \text{Re} \left[\langle \partial_a \psi(\theta^*(x)) | \frac{\partial}{\partial x_i} \frac{\partial H}{\partial x_j}(x) | \psi(\theta^*(x)) \rangle \right] + \frac{\partial \theta_a^*}{\partial x_j}(x) \text{Re} \left[\langle \partial_a \psi(\theta^*(x)) | \frac{\partial}{\partial x_k} \frac{\partial H}{\partial x_i}(x) | \psi(\theta^*(x)) \rangle \right] \right], \quad (\text{B5})
\end{aligned}$$

which is Eq. (9) of the main text.

APPENDIX C: COMPARISON OF THE COST WITH THE FINITE-DIFFERENCE METHOD

1. Cost in estimating the second derivative with the proposed method for quantum chemistry problem

First, we give a simple analysis of the cost of the proposed method in the case of computing the second derivative. We assume that the derivatives of the given Hamiltonian, $\frac{\partial^2 H}{\partial x^d}$, are decomposed into a sum of $N_{\text{Pauli}} = O(n^4)$ Pauli operators like Eq. (13) by the use of the Jordan-Wigner transformation. The expectation values are taken by evaluating every Pauli operator with N_{sample} samples independently for simplicity.

a. Errors in reduced density matrices

The evaluation of $O(n^4)$ Pauli operators provides us single-particle and two-particle reduced density matrix (1,2-RDM), that is, $\{c_i^\dagger c_j\}_{i,j=1}^n$ and $\{c_i^\dagger c_j^\dagger c_k c_l\}_{i,j,k,l=1}^n$. By the Hoeffding's bound, the estimated expectation values of each Pauli operators, $\langle \widehat{P} \rangle$, satisfies

$$|\langle \widehat{P} \rangle - \langle P \rangle| \leq \epsilon_{\text{Pauli}} \quad (\text{C1})$$

with probability at least $1 - \exp(-N_{\text{sample}} \epsilon_{\text{Pauli}}^2 / 2)$. Under the Jordan-Wigner transformation, $c_i^\dagger c_j^\dagger c_k c_l$ is decomposed into sum of 16 Pauli operators, each with coefficient of 1/16. The probability of all such 16 Pauli operators being estimated within the error of ϵ_{Pauli} is at least $1 - 16 \exp(-N_{\text{sample}} \epsilon_{\text{Pauli}}^2 / 2)$ by the union bound, which implies that each element of 2-RDM are estimated within an error of $\epsilon_{2\text{RDM}} = \epsilon_{\text{Pauli}}$ with at least this probability. Since the 2-RDM has n^4 elements, the probability that all of the estimated value of elements of 2-RDM is accurate up to $\epsilon_{2\text{RDM}}$ is at least $1 - 16n^4 \exp(-N_{\text{sample}} \epsilon_{\text{Pauli}}^2 / 2)$. Therefore, to achieve the error of $\epsilon_{2\text{RDM}}$ for each of the term with probability $1 - \delta$, we can take $N_{\text{sample}} = O(\log(n/\delta) / \epsilon_{2\text{RDM}}^2)$.

By evaluating parameter-shifted expectation values of the Pauli operators, such as $\langle Q \rangle_{(a,\mu,\pm),(b,\nu,\pm)}$, we obtain the derivative of the RDMs, such as $\frac{\partial}{\partial \theta_a} \frac{\partial}{\partial \theta_b} \langle \psi(\theta) | c_i^\dagger c_j^\dagger c_k c_l | \psi(\theta) \rangle$. Here, we assume for simplicity that each unitary $U_a(\theta_a)$ in the ansatz has the form of $U_a(\theta_a) = \exp(i\theta_a P_a / 2)$. To obtain the second derivative, this requires us to evaluate $4N_\theta^2$ expectation values for each Pauli operator appearing in the Hamiltonian. We combine them according to Eq. (23) and then Eq. (21). By denoting the estimated value of θ -derivatives of each Pauli operators as $\frac{\partial^2 \langle P \rangle}{\partial \theta_a \partial \theta_b}$, we can again state by the Hoeffdings bound and the union bound,

$$\left| \frac{\partial^2 \langle \widehat{P} \rangle}{\partial \theta_a \partial \theta_b} - \frac{\partial^2 \langle P \rangle}{\partial \theta_a \partial \theta_b} \right| \leq \epsilon_{\text{Pauli},2} \quad (\text{C2})$$

with probability at least $1 - 4 \exp(-N_{\text{sample}} \epsilon_{\text{Pauli},2}^2/2)$. Likewise to the above discussion, 16 θ derivatives of Pauli operators contribute to one element of the 2-RDM derivative. This implies that, by the union bound, the estimation of each element is within the error of $\epsilon_{2\text{RDM},2} = \epsilon_{\text{Pauli},2}$ with probability at least $1 - 64 \exp(-N_{\text{sample}} \epsilon_{\text{Pauli},2}^2/2)$. Hence the probability that all of the element of the 2-RDM derivative is within the error of $\epsilon_{2\text{RDM},2}$ is at least $1 - 64n^4 N_{\theta}^2 \exp(-N_{\text{sample}} \epsilon_{\text{Pauli},2}^2/2)$. Therefore, also in this case, $N_{\text{sample}} = O(\log(nN_{\theta}/\delta)/\epsilon_{2\text{RDM},2}^2)$ suffices to obtain the accuracy bounded by $\epsilon_{2\text{RDM},2}$ with probability $1 - \delta$. The same bound also holds for the estimated value of 1-RDM and its derivatives.

b. Error in the solution of response equation

For simplicity, we denote each element of 1,2-RDM by C_{α} , that is, $C_{\alpha} \in \{c_i^{\dagger} c_j, c_i^{\dagger} c_j^{\dagger} c_k c_l\}_{i,j,k,l=1}^n$, and corresponding coefficients as $h_{\alpha} \in \{h_{ij}, h_{ijkl}\}_{i,j,k,l=1}^n$. Here, we analyze the error in $\frac{\partial}{\partial \theta_a} \frac{\partial E(\theta^*(x), x)}{\partial x_j}$. This term is evaluated by solving the response equation

$$\frac{\partial}{\partial \theta_a} \frac{\partial E(\theta^*(x), x)}{\partial x_j} = \sum_{\alpha} \frac{\partial h_{\alpha}(x)}{\partial x_j} \frac{\partial \langle C_{\alpha} \rangle}{\partial \theta_a}. \quad (\text{C3})$$

Taking $N_{\text{sample}} = O(\log(nN_{\theta}/\delta)/\epsilon'^2)$,

$$\left| \frac{\partial}{\partial \theta_a} \frac{\partial \widehat{E}(\theta^*(x), x)}{\partial x_j} - \frac{\partial}{\partial \theta_a} \frac{\partial E(\theta^*(x), x)}{\partial x_j} \right| \leq \epsilon' \sum_{\alpha} \left| \frac{\partial h_{\alpha}(x)}{\partial x_j} \right|, \quad (\text{C4})$$

holds for all $a \in \{1, \dots, N_{\theta}\}$ with the high probability $1 - \delta$. It also follows that

$$\sum_{a=1}^{N_{\theta}} \left| \frac{\partial}{\partial \theta_a} \frac{\partial \widehat{E}(\theta^*(x), x)}{\partial x_j} - \frac{\partial}{\partial \theta_a} \frac{\partial E(\theta^*(x), x)}{\partial x_j} \right|^2 \leq N_{\theta} \epsilon'^2 \left(\sum_{\alpha} \left| \frac{\partial h_{\alpha}(x)}{\partial x_j} \right| \right)^2. \quad (\text{C5})$$

Similarly,

$$\sum_{a=1}^{N_{\theta}} \sum_{b=1}^{N_{\theta}} \left| \frac{\partial^2 \widehat{E}(\theta^*(x), x)}{\partial \theta_a \partial \theta_b} - \frac{\partial^2 E(\theta^*(x), x)}{\partial \theta_a \partial \theta_b} \right|^2 \leq N_{\theta}^2 \epsilon'^2 \left(\sum_{\alpha} |h_{\alpha}(x)| \right)^2. \quad (\text{C6})$$

$\frac{\partial \theta_a^*(x)}{\partial x_j}$ is computed on a classical computer by solving Eq. (10). To bound the error in $\frac{\partial \theta_a^*(x)}{\partial x_j}$ which originates in the error in $\frac{\partial^2 E}{\partial \theta_a \partial \theta_b}$ and $\frac{\partial}{\partial \theta_a} \frac{\partial E}{\partial x_j}$, we use the following lemma [53].

Lemma 1. Let A and \mathbf{b} be a matrix and a vector, respectively. Let ΔA and $\Delta \mathbf{b}$ be their errors. Suppose \mathbf{y} and $\widehat{\mathbf{y}}$ are vectors satisfying $A\mathbf{y} = \mathbf{b}$ and $(A + \Delta A)\widehat{\mathbf{y}} = \mathbf{b} + \Delta \mathbf{b}$. Then,

$$\frac{\|\widehat{\mathbf{y}} - \mathbf{y}\|_2}{\|\mathbf{y}\|_2} \leq \frac{\kappa(A)}{1 - \kappa(A) \frac{\|\Delta A\|_F}{\|A\|_F}} \left(\frac{\|\Delta \mathbf{b}\|_2}{\|\mathbf{b}\|_2} + \frac{\|\Delta A\|_F}{\|A\|_F} \right), \quad (\text{C7})$$

holds under the condition that $\|A^{-1}\|_F \|\Delta A\|_F < 1$, where $\|\cdot\|_2$ is the vector 2-norm, $\|\cdot\|_F$ is the Frobenius norm, and $\kappa(A) = \|A^{-1}\|_F \|A\|_F$ is the condition number of A .

This can be utilized to state,

$$\frac{\left\| \frac{\partial \widehat{\theta}^*(x)}{\partial x_j} - \frac{\partial \theta^*(x)}{\partial x_j} \right\|}{\left\| \frac{\partial \theta^*(x)}{\partial x_j} \right\|} \leq \frac{\kappa}{1 - \kappa \frac{N_{\theta} \epsilon' \sum_{\alpha} |h_{\alpha}(x)|}{\|\{\frac{\partial^2 E}{\partial \theta_a \partial \theta_b}\}\|_F}} \left(\frac{N_{\theta}^{1/2} \epsilon' \sum_{\alpha} \left| \frac{\partial h_{\alpha}}{\partial x_j} \right|}{\|\{\frac{\partial}{\partial \theta_a} \frac{\partial E}{\partial x_j}\}\|_F} + \frac{N_{\theta} \epsilon' \sum_{\alpha} |h_{\alpha}(x)|}{\|\{\frac{\partial^2 E}{\partial \theta_a \partial \theta_b}\}\|_F} \right), \quad (\text{C8})$$

where κ is the condition number of $\frac{\partial^2 E(\theta^*(x), x)}{\partial \theta_a \partial \theta_b}$.

c. Overall error

The last term in Eq. (8), $\frac{\partial^2 E(\theta^*(x), x)}{\partial x_i \partial x_j}$, is evaluated by

$$\frac{\partial^2 E(\theta^*(x), x)}{\partial x_i \partial x_j} = \sum_{\alpha} \frac{\partial^2 h_{\alpha}(x)}{\partial x_i \partial x_j} \langle C_{\alpha} \rangle. \quad (\text{C9})$$

By taking $N_{\text{sample}} = O(\log(n/\delta)/\epsilon'^2)$ for some $\epsilon' > 0$, the error in each expectation value is at most ϵ' with a high probability $1 - \delta$. This implies that the error in the estimation is

$$\left| \frac{\partial^2 \widehat{E}(\theta^*(x), x)}{\partial x_i \partial x_j} - \frac{\partial^2 E(\theta^*(x), x)}{\partial x_i \partial x_j} \right| \leq \epsilon' \sum_{\alpha} \left| \frac{\partial^2 h_{\alpha}}{\partial x_i \partial x_j} \right|. \quad (\text{C10})$$

The error in the first term in Eq. (8), $\sum_a \frac{\partial \theta_a^*}{\partial x_i} \frac{\partial}{\partial \theta_a} \frac{\partial E}{\partial x_j}$, can be evaluated as follows:

$$\begin{aligned} \left| \sum_a \frac{\partial \theta_a^*}{\partial x_i} \frac{\partial}{\partial \theta_a} \frac{\partial E}{\partial x_j} - \sum_a \widehat{\frac{\partial \theta_a^*}{\partial x_i}} \widehat{\frac{\partial}{\partial \theta_a}} \widehat{\frac{\partial E}{\partial x_j}} \right| &= \left| \sum_a \left(\frac{\partial \theta_a^*}{\partial x_i} - \widehat{\frac{\partial \theta_a^*}{\partial x_i}} \right) \widehat{\frac{\partial}{\partial \theta_a}} \widehat{\frac{\partial E}{\partial x_j}} + \sum_a \frac{\partial \theta_a^*}{\partial x_i} \left(\frac{\partial}{\partial \theta_a} \frac{\partial E}{\partial x_j} - \widehat{\frac{\partial}{\partial \theta_a}} \widehat{\frac{\partial E}{\partial x_j}} \right) \right| \\ &\leq \left\| \frac{\partial \theta_a^*}{\partial x_i} - \widehat{\frac{\partial \theta_a^*}{\partial x_i}} \right\|_2 \left\| \widehat{\frac{\partial}{\partial \theta_a}} \widehat{\frac{\partial E}{\partial x_j}} \right\|_2 + \left\| \frac{\partial \theta_a^*}{\partial x_i} \right\|_2 \left\| \frac{\partial}{\partial \theta_a} \frac{\partial E}{\partial x_j} - \widehat{\frac{\partial}{\partial \theta_a}} \widehat{\frac{\partial E}{\partial x_j}} \right\|_2. \end{aligned}$$

Therefore overall error is

$$\begin{aligned} \left| \frac{\partial^2 E^*}{\partial x_i \partial x_j} - \widehat{\frac{\partial^2 E^*}{\partial x_i \partial x_j}} \right| &\leq \left\| \frac{\partial \theta_a^*}{\partial x_i} - \widehat{\frac{\partial \theta_a^*}{\partial x_i}} \right\|_2 \left\| \widehat{\frac{\partial}{\partial \theta_a}} \widehat{\frac{\partial E}{\partial x_j}} \right\|_2 + \left\| \frac{\partial \theta_a^*}{\partial x_i} \right\|_2 \left\| \frac{\partial}{\partial \theta_a} \frac{\partial E}{\partial x_j} - \widehat{\frac{\partial}{\partial \theta_a}} \widehat{\frac{\partial E}{\partial x_j}} \right\|_2 + \epsilon' \sum_\alpha \left| \frac{\partial^2 h_\alpha}{\partial x_i \partial x_j} \right| \\ &\leq \epsilon' \left\{ \left\| \frac{\partial \theta_a^*}{\partial x_i} \right\|_2 \left[\frac{\kappa N_\theta^{1/2} \sum_\alpha \left| \frac{\partial h_\alpha}{\partial x_j} \right|}{1 - \kappa \cdot \frac{N_\theta \sum_\alpha |h_\alpha(x)|}{\|\{\frac{\partial^2 E}{\partial \theta_a \partial \theta_b}\}\|_F}} \left(\frac{N_\theta^{1/2} \sum_\alpha \left| \frac{\partial h_\alpha}{\partial x_i} \right|}{\|\{\frac{\partial}{\partial \theta_a} \frac{\partial E}{\partial x_i}\}\|} + \frac{N_\theta \sum_\alpha |h_\alpha|}{\|\{\frac{\partial^2 E}{\partial \theta_a \partial \theta_b}\}\|_F} \right) + N_\theta^{1/2} \sum_\alpha \left| \frac{\partial h_\alpha}{\partial x_j} \right| \right] + \sum_\alpha \left| \frac{\partial^2 h_\alpha}{\partial x_i \partial x_j} \right| \right\}. \end{aligned}$$

To upper-bound this error by ϵ , we want ϵ' to satisfy

$$\epsilon = \epsilon' \left\{ \left\| \frac{\partial \theta_a^*}{\partial x_i} \right\|_2 \left[\frac{\kappa N_\theta^{1/2} \sum_\alpha \left| \frac{\partial h_\alpha}{\partial x_j} \right|}{1 - \kappa \cdot \frac{N_\theta \sum_\alpha |h_\alpha(x)|}{\|\{\frac{\partial^2 E}{\partial \theta_a \partial \theta_b}\}\|_F}} \left(\frac{N_\theta^{1/2} \sum_\alpha \left| \frac{\partial h_\alpha}{\partial x_i} \right|}{\|\{\frac{\partial}{\partial \theta_a} \frac{\partial E}{\partial x_i}\}\|} + \frac{N_\theta \sum_\alpha |h_\alpha|}{\|\{\frac{\partial^2 E}{\partial \theta_a \partial \theta_b}\}\|_F} \right) + N_\theta^{1/2} \sum_\alpha \left| \frac{\partial h_\alpha}{\partial x_j} \right| \right] + \sum_\alpha \left| \frac{\partial^2 h_\alpha}{\partial x_i \partial x_j} \right| \right\}. \tag{C11}$$

The number of samples per measurement we need to achieve this bound with probability $1 - \delta$ is $N_{\text{sample}} = O(\log(n/\delta)/\epsilon^2)$. Therefore, in total, the number of measurement required for our protocol is

$$N_{\text{sample}}^{\text{tot}} = O(N_\theta^2 n^4 / \epsilon'^2 \log(n/\delta)). \tag{C12}$$

Note that κ should not be too big for a good ansatz, because it implies the VQE would suffer in the optimization of the parameter.

2. Cost in estimating the second derivative with the finite difference

Here, we analyze the cost of the finite difference approach where we use the formula,

$$\frac{\partial^2 E^*(x)}{\partial x_i^2} = (E^*(x + h e_i) + E^*(x - h e_i) - 2E^*(x)) / h^2, \tag{C13}$$

with some step size h . e_i is a unit vector with an entry 1 in the i th element. The following lemma, which helps to bound the error, can easily be shown with simple arithmetics.

Lemma 2. For a function $f : \mathbb{R} \rightarrow \mathbb{R}$, a step size h , and an approximated value of the function $\widehat{f}(y)$ at a point $y \in \mathbb{R}$, such that $|\widehat{f}(y) - f(y)| \leq \epsilon'$, the following inequality holds:

$$\left| f''(y) - \frac{\widehat{f}(y+h) + \widehat{f}(y-h) - 2\widehat{f}(y)}{h^2} \right| \leq \frac{h^2}{12} \max_{z \in [y-h, y+h]} |f^{(4)}(z)| + \frac{4\epsilon'}{h^2}. \tag{C14}$$

Taking $N_{\text{sample}} = O(\log(n/\delta)/\epsilon'^2)$ bounds the error in the estimation of $E^*(x)$ to $|\widehat{E^*(x)} - E^*(x)| \leq \epsilon'' \sum_\alpha |h_\alpha|$ with probability $1 - \delta$. Assume that we want to suppress the error of the second derivative to ϵ , that is, $|\widehat{\frac{\partial^2 E^*(x)}{\partial x_i^2}} - \frac{\partial^2 E^*(x)}{\partial x_i^2}| \leq \epsilon$. To achieve this, we need

$$\frac{h^2}{12} \max_{[x_i-h, x_i+h]} \left| \frac{\partial^4 E^*}{\partial x_i^4} \right| + \frac{4\epsilon''}{h^2} \sum_\alpha |h_\alpha| = \epsilon, \tag{C15}$$

from which we find

$$\epsilon'' = \frac{h^2 \epsilon}{4 \sum_\alpha |h_\alpha|} - \frac{h^4}{48 \sum_\alpha |h_\alpha|} \max_{[x_i-h, x_i+h]} \left| \frac{\partial^4 E^*}{\partial x_i^4} \right|. \tag{C16}$$

Therefore, the total number of NISQ runs to achieve the error of ϵ can be written as

$$N_{\text{sample}}^{\text{tot}} = O \left[N_x^2 n^4 \frac{\log(n/\delta)}{\left(\frac{h^2 \epsilon}{\sum_\alpha |h_\alpha|} - \frac{h^4}{12 \sum_\alpha |h_\alpha|} \max_{[x_i-h, x_i+h]} \left| \frac{\partial^4 E^*}{\partial x_i^4} \right| \right)^2} \right]. \tag{C17}$$

- [1] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, *Nat. Comm.* **5**, 4213 (2014).
- [2] K. M. Nakanishi, K. Mitarai, and K. Fujii, *Phys. Rev. Res.* **1**, 033062 (2019).
- [3] T. Jones, S. Endo, S. McArdle, X. Yuan, and S. C. Benjamin, *Phys. Rev. A* **99**, 062304 (2019).
- [4] O. Higgott, D. Wang, and S. Brierley, *Quantum* **3**, 156 (2019).
- [5] J. R. McClean, M. E. Kimchi-Schwartz, J. Carter, and W. A. de Jong, *Phys. Rev. A* **95**, 042308 (2017).
- [6] F. Jensen, *Introduction to Computational Chemistry* (John Wiley & Sons, Hoboken, NJ, 2017).
- [7] J. Gauss, in *Modern Methods and Algorithms of Quantum Chemistry* (NIC-John von Neumann Institute for Computing, Juelich, 2000), Vol. 3, pp. 541–592.
- [8] T. Helgaker, S. Coriani, P. Jørgensen, K. Kristensen, J. Olsen, and K. Ruud, *Chem. Rev.* **112**, 543 (2012).
- [9] P. Pulay, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **4**, 169 (2014).
- [10] H. B. Schlegel, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **1**, 790 (2011).
- [11] E. B. Wilson, J. C. Decius, and P. C. Cross, *Molecular Vibrations: the Theory of Infrared and Raman Vibrational Spectra* (Dover Publications, Mineola, NY, 1980).
- [12] N. F. Ramsey, *Phys. Rev.* **78**, 699 (1950).
- [13] T. Helgaker, M. Jaszuński, and K. Ruud, *Chem. Rev.* **99**, 293 (1999).
- [14] J. B. Foresman, M. Head-Gordon, J. A. Pople, and M. J. Frisch, *J. Phys. Chem.* **96**, 135 (1992).
- [15] D. Maurice and M. Head-Gordon, *Mol. Phys.* **96**, 1533 (1999).
- [16] C. Van Caillie and R. D. Amos, *Chem. Phys. Lett.* **308**, 249 (1999).
- [17] C. Van Caillie and R. D. Amos, *Chem. Phys. Lett.* **317**, 159 (2000).
- [18] F. Furche and R. Ahlrichs, *J. Chem. Phys.* **117**, 7433 (2002).
- [19] Y. A. Bernard, Y. Shao, and A. I. Krylov, *J. Chem. Phys.* **136**, 204103 (2012).
- [20] A. Köhn and C. Hättig, *J. Chem. Phys.* **119**, 5021 (2003).
- [21] C. Hättig, *Adv. Quantum Chem.* **50**, 37 (2005).
- [22] D. Tew, C. Hättig, and N. Graf, *Phys. Chem. Chem. Phys.* **21**, 14063 (2019).
- [23] J. F. Stanton, *J. Chem. Phys.* **99**, 8840 (1993).
- [24] J. F. Stanton and J. Gauss, *J. Chem. Phys.* **101**, 8938 (1994).
- [25] J. F. Stanton and J. Gauss, *J. Chem. Phys.* **103**, 8931 (1995).
- [26] T. Nakajima and H. Nakatsuji, *Chem. Phys.* **242**, 177 (1999).
- [27] M. Kállay and J. Gauss, *J. Chem. Phys.* **121**, 9257 (2004).
- [28] H. Nakano, K. Hirao, and M. S. Gordon, *J. Chem. Phys.* **108**, 5660 (1998).
- [29] P. Celani and H.-J. Werner, *J. Chem. Phys.* **119**, 5044 (2003).
- [30] Y. Yamaguchi, *A New Dimension to Quantum Chemistry: Analytic Derivative Methods in ab initio Molecular Electronic Structure Theory* (Oxford University Press, New York, 1994).
- [31] M. K. MacLeod and T. Shiozaki, *J. Chem. Phys.* **142**, 051103 (2015).
- [32] A. M. Sand, C. E. Hoyer, K. Sharkas, K. M. Kidder, R. Lindh, D. G. Truhlar, and L. Gagliardi, *J. Chem. Theory Comp.* **14**, 126 (2017).
- [33] E. G. Hohenstein, M. E. Bouduban, C. Song, N. Luehr, I. S. Ufimtsev, and T. J. Martínez, *J. Chem. Phys.* **143**, 014111 (2015).
- [34] B. S. Fales, Y. Shu, B. G. Levine, and E. G. Hohenstein, *J. Chem. Phys.* **147**, 094104 (2017).
- [35] I. Kassal and A. Aspuru-Guzik, *J. Chem. Phys.* **131**, 224102 (2009).
- [36] P. Pulay, *Adv. Chem. Phys.* **69**, 241 (1987).
- [37] J. T. Seeley, M. J. Richard, and P. J. Love, *J. Chem. Phys.* **137**, 224109 (2012).
- [38] S. McArdle, S. Endo, A. Aspuru-Guzik, S. Benjamin, and X. Yuan, [arXiv:1808.10402](https://arxiv.org/abs/1808.10402).
- [39] Y. Li and S. C. Benjamin, *Phys. Rev. X* **7**, 021050 (2017).
- [40] P.-L. Dallaire-Demers, J. Romero, L. Veis, S. Sim, and A. Aspuru-Guzik, *Quantum Sci. Technol.* **4**, 045005 (2019).
- [41] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, *Nature (London)* **549**, 242 (2017).
- [42] Y. Nam, J.-S. Chen, N. C. Pienti, K. Wright, C. Delaney, D. Maslov, K. R. Brown, S. Allen, J. M. Amini, J. Apisdorf, K. M. Beck, A. Blinov, V. Chaplin, M. Chmielewski, C. Collins, S. Debnath, A. M. Ducore, K. M. Hudek, M. Keenan, S. M. Kreikemeier, J. Mizrahi, P. Solomon, M. Williams, J. D. Wong-Campos, C. Monroe, and J. Kim, [arXiv:1902.10171](https://arxiv.org/abs/1902.10171).
- [43] H. R. Grimsley, S. E. Economou, E. Barnes, and N. J. Mayhall, *Nat. Commun.* **10**, 3007 (2019).
- [44] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, *Phys. Rev. A* **98**, 032309 (2018).
- [45] K. Mitarai and K. Fujii, *Phys. Rev. Res.* **1**, 013006 (2019).
- [46] J. Kussmann, A. Luenser, M. Beer, and C. Ochsenfeld, *J. Chem. Phys.* **142**, 094101 (2015).
- [47] H. Lischka, D. Nachtigallova, A. J. A. Aquino, P. G. Szalay, F. Plasser, F. B. C. Machado, and M. Barbatti, *Chem. Rev.* **118**, 7293 (2018).
- [48] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters, and G. K. Chan, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **8**, e1340 (2017).
- [49] J. R. McClean, K. J. Sung, I. D. Kivlichan, Y. Cao, C. Dai, E. S. Fried, C. Gidney, B. Gimby, P. Gokhale, T. Häner, T. Hardikar, V. Havlíček, O. Higgott, C. Huang, J. Izaac, Z. Jiang, X. Liu, S. McArdle, M. Neeley, T. O'Brien, B. O'Gorman, I. Ozfidan, M. D. Radin, J. Romero, N. Rubin, N. P. D. Sawaya, K. Setia, S. Sim, D. S. Steiger, M. Steudtner, Q. Sun, W. Sun, D. Wang, F. Zhang, and R. Babbush, [arXiv:1710.07629](https://arxiv.org/abs/1710.07629).
- [50] Qulacs github.com/qulacs/qulacs.
- [51] K. Mitarai, T. Yan, and K. Fujii, *Phys. Rev. Appl.* **11**, 044087 (2019).
- [52] T. E. O'Brien, B. Senjean, R. Sagastizabal, X. Bonet-Monroig, A. Dutkiewicz, F. Buda, L. DiCarlo, and L. Visscher, *npj Quantum Inf.* **5**, 113 (2019).
- [53] R. A. Horn, *Matrix Analysis*, 2nd ed. (Cambridge University Press, New York, NY, 2012).