Quadratic Clifford expansion for efficient benchmarking and initialization of variational quantum algorithms

Kosuke Mitarai,[1,2,3,*] Yasunari Suzuki,[4,3] Wataru Mizukami,[2,3] Yuya O. Nakagawa,[5] and Keisuke Fujii[1,2,6]

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

[2]*Center for Quantum Information and Quantum Biology, Osaka University, 1-2 Machikaneyama, Toyonaka, Osaka, 560-0043, Japan*

[3]*JST, PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan*

[4]*NTT Secure Platform Laboratories, NTT Corporation, Musashino 180-8585, Japan*

[5]*QunaSys Inc., Aqua Hakusan Building 9F, 1-13-7 Hakusan, Bunkyo, Tokyo 113-0001, Japan*

[6]*Center for Emergent Matter Science, RIKEN, Wako Saitama 351-0198, Japan*

(Received 14 December 2020; accepted 26 May 2022; published 6 July 2022)

Variational quantum algorithms are considered to be appealing applications of near-term quantum computers. However, it has been unclear whether they can outperform classical algorithms or not. To reveal their limitations, we must seek a technique to benchmark them on large-scale problems. Here we propose a perturbative approach for efficient benchmarking of variational quantum algorithms. The proposed technique performs perturbative expansion of a circuit consisting of Clifford and Pauli rotation gates, which is enabled by exploiting the classical simulatability of Clifford circuits. Our method can be applied to a wide family of parameterized quantum circuits consisting of Clifford gates and single-qubit rotation gates. The approximate optimal parameter obtained by the method can also serve as an initial guess for further optimizations on a quantum device. As the first application of the method, we perform a benchmark of so-called hardware-efficient-type ansatzes when they are applied to the variational quantum eigensolver (VQE) of one-dimensional hydrogen chains up to H_{24}, which corresponds to a 48-qubit system using a standard workstation. This is the largest scale benchmark of the VQE to the best of our knowledge and reveals the limitation of hardware-efficient-type ansatzes.

DOI: 10.1103/PhysRevResearch.4.033012

I. INTRODUCTION

As promising candidates for possible applications of early days quantum devices, variational quantum algorithms (VQAs) [1] have been developed rapidly. VQAs utilize parameterized quantum circuits $U(\boldsymbol{\theta})$ with parameters $\boldsymbol{\theta}$ that are optimized with respect to some suitably defined cost function $\mathcal{L}(\boldsymbol{\theta})$ depending on specific tasks. Target applications of VQAs range among quantum chemistry calculations [2–4], combinational optimization [5], and machine learning [6–10].

Despite the vast amount of theoretical proposals, demonstrations, and benchmarks of algorithms [1,8,11–13], they are limited to relatively small-scale problems, where classical simulations are still feasible. Efficient techniques for their benchmark in large-scale problems are strongly demanded to understand the limitations of VQAs and to develop more sophisticated algorithms.

Here we aim to resolve the above problem with a perturbative expansion of the cost function. We assume that a parametrized quantum circuit utilized in an algorithm is made of Clifford gates and single-qubit rotation gates whose angles are the circuit parameters and initialized to zero. The above form of the circuit includes a wide family of parameterized circuits, which guarantees the wide applicability of the proposed method. With this assumption, we can efficiently compute the first and second derivatives of a cost function by exploiting the classical simulatability [14,15] of Clifford circuits. This allows us to perform a simple minimization of a quadratic function to obtain an approximately optimal value of parameters and a cost function. In particular, for the variational quantum eigensolver (VQE) [2], which is an algorithm to obtain an approximate ground state of a quantum system, the perturbative treatment can be justified because classically tractable variational solutions such as Hartree-Fock or mean-field states are expected to be close to true ones. In such cases, we can obtain an approximately optimal energy that can be achieved with a certain ansatz, thereby our method serves as a benchmark of the ansatz performance measured by the energy it can produce. The method can also be seen as an efficient initializer of the circuit parameters as the above procedure corresponds to the first step of Newton-Raphson optimization. The obtained Hessian together with the obtained parameter can be passed as an initial guess to quasi-Newton optimizers such as the BFGS method for further optimization on quantum devices.

In the following, we first give the concrete algorithm of the proposed method. Then, as an application of the method, we benchmark the so-called hardware-efficient-type ansatzes applied to the VQE up to 48 qubits using the hydrogen chain

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

as a testbed. Benchmarks are performed by approximately evaluating the energy that the hardware-efficient ansatz can achieve. This is the largest scale benchmark of the VQE to date and reveals limitations of the ansatzes. This benchmark can be performed because the proposed technique provides us an approximately optimal value of the cost function, which is the energy expectation value in this case. The benchmark itself can be seen as a demonstration of a "quantum-inspired" quantum chemistry calculation, where the ansatz wave function is constructed with the language of the quantum circuit. Finally, we also show the effectiveness of our perturbative initialization approach by another numerical experiment.

II. ALGORITHMS

A. VQA

VQA generally refers to a family of algorithms that involve the use of parametrized quantum circuits $U(\boldsymbol{\theta})$ whose parameters $\boldsymbol{\theta}$ are optimized with respect to a suitable cost function $\mathcal{L}(\boldsymbol{\theta})$; $U(\boldsymbol{\theta})$ is used to generate a n-qubit parametrized state $|\psi(\boldsymbol{\theta})\rangle := U(\boldsymbol{\theta})|0\rangle^{\otimes n}$. Hereafter we abbreviate $|0\rangle^{\otimes n}$ by $|0\rangle$ when it is clear from the context. A famous example of VQAs is the VQE [2], where the cost function is defined as an energy expectation value $E(\boldsymbol{\theta})$ with respect to a Hamiltonian H, i.e., $\mathcal{L}(\boldsymbol{\theta}) = E(\boldsymbol{\theta}) = \langle\psi(\boldsymbol{\theta})| H |\psi(\boldsymbol{\theta})\rangle$. The cost function is usually computed from expectation values of observables also in other examples such as machine learning [6–10] and combinational optimization [5]. A general form of $\mathcal{L}(\boldsymbol{\theta})$ can be written as $\mathcal{L}(\boldsymbol{\theta}) = L(\langle O(\boldsymbol{\theta})\rangle)$ where O denotes the measured observable and $\langle O(\boldsymbol{\theta})\rangle := \langle\psi(\boldsymbol{\theta})| O |\psi(\boldsymbol{\theta})\rangle$. O is typically expressed as a sum of N_o n-qubit Pauli operators $\{P_i\} \subset \{I, X, Y, Z\}^{\otimes n}$ as $O = \sum_{i=1}^{N_o} c_i P_i$ with coefficients $\{c_i\}$.

B. Main result: Quadratic Clifford expansion

We consider an ansatz in the form of

$$U(\boldsymbol{\theta}) = R_K(\theta_K)C_K \cdots R_2(\theta_2)C_2 R_1(\theta_1)C_1, \tag{1}$$

where $\boldsymbol{\theta} = \{\theta_k\}_{k=1}^K$, R_k is a single-qubit rotation gate generated by a Pauli operator P_k, i.e., $R_k(\theta_k) = e^{i\theta_k P_k}$, and C_k is a circuit consisting of Clifford gates. Note that this form of the ansatz is quite general. When we wish to build a hardware-efficient ansatz [4], it is frequently in the form of Eq. (1) because the two-qubit gates which are tuned to give a high fidelity on the hardware are usually Clifford gates such as controlled-NOT or controlled-Z gates. More sophisticated ansatz such as unitary coupled cluster [2] can also be written in this form.

This form of the ansatz allows us to efficiently compute the perturbative form of the cost function $\mathcal{L}(\boldsymbol{\theta})$. More concretely, a Taylor expansion of $\langle O(\boldsymbol{\theta})\rangle$ around $\boldsymbol{\theta} = 0$ can be written as

$$\langle O(\boldsymbol{\theta})\rangle = \langle O(0)\rangle + \sum_k g_k\theta_k + \frac{1}{2}\sum_{k,m} A_{km}\theta_k\theta_m + \mathcal{O}(\|\boldsymbol{\theta}\|^3), \tag{2}$$

where

$$g_k = 2\mathrm{Re}\left[\langle 0| U^\dagger(0) O \frac{\partial U(0)}{\partial\theta_k}|0\rangle\right], \tag{3}$$

$$A_{km} = 2\mathrm{Re}\left[\langle 0|\frac{\partial U^\dagger(0)}{\partial\theta_k} O \frac{\partial U(0)}{\partial\theta_m}|0\rangle\right] + 2\mathrm{Re}\left[\langle 0| U^\dagger(0) O \frac{\partial}{\partial\theta_k}\frac{\partial U(0)}{\partial\theta_m}|0\rangle\right], \tag{4}$$

and $\frac{\partial U(0)}{\partial\theta_k} := \frac{\partial U(\boldsymbol{\theta})}{\partial\theta_k}|_{\boldsymbol{\theta}=0}$, which can then be used to expand $\mathcal{L}(\boldsymbol{\theta}) = L(\langle O(\boldsymbol{\theta})\rangle)$ itself. Since we assumed U to be in the form of Eq. (1), $\frac{\partial U(0)}{\partial\theta_i}$ can be efficiently computed. To see this, observe that

$$\frac{\partial U(0)}{\partial\theta_k} = iC_K \cdots P_k C_k \cdots C_2 C_1. \tag{5}$$

P_k can be efficiently passed through $C_K \cdots C_{k+1}$. Let

$$C_K \cdots C_{k+1} P_k = P_k' C_K \cdots C_{k+1}, \tag{6}$$

for some Pauli operator P_k'. P_k' can be found in time $\mathcal{O}(nK)$ on a classical computer if $\{C_k\}$ are local, that is, $\{C_k\}$ act only on $\mathcal{O}(1)$ qubits. Using P_k', the coefficients appearing in the second-order Taylor expansion [Eq. (2)] can be written in terms of the expectation values $\langle\psi(0)|OP_k'|\psi(0)\rangle$, $\langle\psi(0)|P_k'OP_m'|\psi(0)\rangle$, and $\langle\psi(0)|OP_k'P_m'|\psi(0)\rangle$. The decomposition of the operators OP_k', $P_k'OP_m'$, and $OP_k'P_m'$ into a sum of Pauli operators can be computed in time $\mathcal{O}(nN_o)$ on a classical computer. This can be performed simply by multiplying P_k' and P_m' to each Pauli operator in O. The expectation values of these operators can be evaluated efficiently on a classical computer because $|\psi(0)\rangle$ is a stabilizer state under the assumption that $U(0)$ is Clifford. More concretely, we evaluate expectation values of each Pauli operator constituting OP_k', $P_k'OP_m'$, and $OP_k'P_m'$ and then take the summation. This process can be performed in time $\mathcal{O}(n^2N_oK^2)$ using a standard simulation technique [15], which gives the leading-order complexity of the perturbative expansion. We call this technique quadratic Clifford expansion. The technique itself might be useful for classical simulations of near-Clifford circuits, i.e., we can approximately simulate the output of such circuits as long as the non-Clifford rotation angles are small and the perturbative treatment is valid.

The perturbative expansion given in Eq. (2) is justified especially for the VQE [2], where we can obtain an approximate ground state classically by using techniques such as Hartree-Fock methods. If we construct $U(\boldsymbol{\theta})$ in such a way that $|\psi(0)\rangle$ becomes the Hartree-Fock ground state, $|\psi(0)\rangle$ is considered to be close to the true ground state. This type of strategy in which we start the optimization from a state close to a solution has been shown to be effective to alleviate the barren plateau problem [16]. Therefore in this case, we can presume the optimal value of $\boldsymbol{\theta}$ to be small, which justifies the perturbative treatment of the cost function $E(\boldsymbol{\theta})$. As long as the perturbation is accurate enough, we can obtain the optimal value of $\langle O(\boldsymbol{\theta})\rangle$ by simply minimizing the quadratic function obtained with the second-order expansion, which can be done in time $\mathcal{O}(K^3)$ and provides us an optimal parameter $\boldsymbol{\theta}^* = -A^+\boldsymbol{g}$, where A^+ is the Moore-Penrose pseudoinverse of the Hessian A. The approximate, perturbative optimal value of $\langle O(\boldsymbol{\theta})\rangle$ can be calculated by substituting $\boldsymbol{\theta}^*$ into Eq. (2) and neglecting cubic error term, which we denote by $\langle O\rangle^*$.

One might think that the condition the ground state must be approximated by a stabilizer state is rather restrictive.

While it is true for general quantum systems, there are certain interesting cases such as the transverse-field toric code model [17] and the transverse-field cluster model [18] where this condition holds. Moreover, we can readily improve the approximation by using an arbitrary superposition of the polynomial number of stabilizer states instead of $|0\rangle$ as an input to the ansatz $U(\boldsymbol{\theta})$ while maintaining the polynomial runtime. This strategy, for instance, allows us to use wave functions obtained by the configuration interaction singles and doubles method, which is a classical post-Hartree-Fock method, as $|\psi(0)\rangle$.

Whether the above perturbation is accurate or not cannot be efficiently determined classically in general, since we cannot obtain the expectation value $\langle O(\boldsymbol{\theta})\rangle$ for general $\boldsymbol{\theta}$ on a classical computer. This indicates the need for a quantum device for validating the result. Therefore a possible strategy of using the proposed technique is to evaluate $\langle O(\boldsymbol{\theta}^*)\rangle$ on a quantum computer, and if it returns a value close to the perturbative one, then we just assume that we have found an optimal approximate ground state; otherwise, we further optimize the parameters starting from $\boldsymbol{\theta}^*$.

We remark that $\boldsymbol{\theta}^*$ can also be seen as the parameter obtained by the first step of the Newton-Raphson method. It provides a good starting point for further optimization using sophisticated techniques like stochastic gradient descent [19]. Moreover, the Hessian A_{km} can be passed to quasi-Newton optimizers such as the BFGS method which are frequently utilized in the VQE. This method can also be applied to calculate the quantum natural gradient [20–22], which can be calculated from overlaps between derivatives of $|\psi(\boldsymbol{\theta})\rangle$. We also note that a recently proposed optimization method called the quantum analytic descent [23] can be easily combined with our method, as it uses the gradient and the Hessian of the cost function to approximate its landscape with trigonometric functions.

We can also apply the proposed method to the machine-learning algorithms [6–10]. In this direction, techniques for computing a "good" initial guess like the mean-field solution are not yet developed. We leave such an extension for the future.

III. NUMERICAL EXPERIMENT

To demonstrate the effectiveness of our idea, we apply the method described in the previous section to the VQE to benchmark the performance of the so-called hardware-efficient ansatz [4]. For this purpose, we use electronic Hamiltonians of evenly spaced one-dimensional chains of hydrogen atoms H_m, which are frequently used as a benchmark system for quantum chemistry calculations [24]. All benchmarks are performed on a workstation with two Intel Xeon Silver 4108 processors. For quantum circuit simulations, we utilized an NVIDIA Tesla-V100 GPU.

A. Experimental details

Electronic Hamiltonians of hydrogen chains are generated by OpenFermion [25] and PySCF [26,27] using the STO-3G minimal basis set. The generated fermionic Hamiltonians are mapped to qubit ones by the Jordan-Wigner transformation implemented in OpenFermion, which results in a $2m$-qubit Hamiltonian for an m-hydrogen chain H_m. A thorough review of these procedures can be found in, e.g., Ref. [28]. All conventional quantum chemistry calculations are also performed with PySCF.

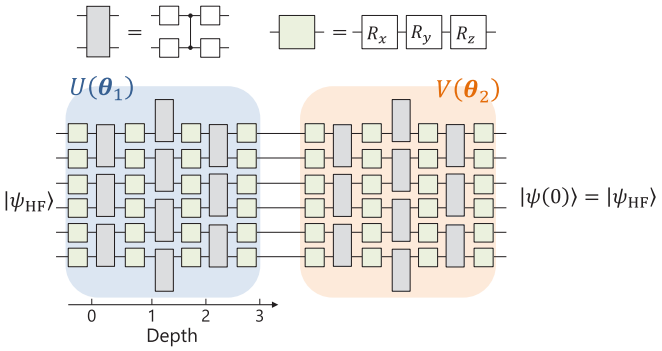

FIG. 1. Ansatz used in the numerical experiment. Gray boxes represent (fixed) 2-qubit Clifford gates in the form of the upper left, where white boxes are randomly chosen from 24 single-qubit Clifford gates [31]. Green boxes represent parametrized single-qubit rotations consisting of x, y and z-axis rotations, respectively.

As for the ansatz, we use the one shown in Fig. 1, which can be regarded as a "hardware-efficient" ansatz constructed on a one-dimensional qubit array. It consists of alternating layers of two-qubit Clifford gates and single-qubit rotation gates. This form of the ansatz can generate sufficiently nonlocal evolutions that give nonzero gradients. In Fig. 1, the two-qubit Clifford gates in the region shaded by blue are randomly chosen as shown in the upper left of the figure. We define $U(\boldsymbol{\theta}_1)$ to be the circuit in the blue region. The circuit in the orange region, $V(\boldsymbol{\theta}_2)$, is chosen to satisfy $V(0) = U^\dagger(0)$. This allows us to easily guarantee $|\psi(0)\rangle$ to be the Hartree-Fock state $|\psi_{\mathrm{HF}}\rangle$; we can just inject $|\psi_{\mathrm{HF}}\rangle$ to the input of the circuit. Note that Hartree-Fock states are computational basis states under fermion-to-qubit mappings such as the Jordan-Wigner transformation, and its evolution under Clifford gates can efficiently be simulated. The single-qubit rotations hold the parameter $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\}$ to be optimized. Each has three parameters as x-, y-, and z-rotation angles. Note that, although the two-qubit Clifford gates in $V(\boldsymbol{\theta}_2)$ are chosen to satisfy $V(0) = U^\dagger(0)$, the parameters implemented in single-qubit rotations in $U(\boldsymbol{\theta}_1)$ and $V(\boldsymbol{\theta}_2)$ are independent.

Using this ansatz, we calculate the gradient and Hessian based on the method described in the previous section. Then, we perform the minimization of the second-order perturbative energy to obtain approximately optimal energies and parameters. Optimization is performed simply by calculating $\boldsymbol{\theta}^* = -A^+\boldsymbol{g}$. This provides us $\boldsymbol{\theta}^*$ and perturbative energies $E^* = \langle H\rangle^*$. Finally, when possible, we simulate an ansatz whose parameters are set to the perturbatively optimal ones to check if the perturbative treatment can be justified. This simulation is performed with Qulacs [29].

To make the benchmark systematic, we set the depth of the ansatzes equal to the number of hydrogen atoms, m. This scaling of the depth can be considered as the largest possible value for today's most advanced quantum computer [30]. Note that this choice corresponds to $K = \mathcal{O}(m^2)$. Combining with the fact that $N_o = \mathcal{O}(m^4)$ and $n = 2m$ in this case, the

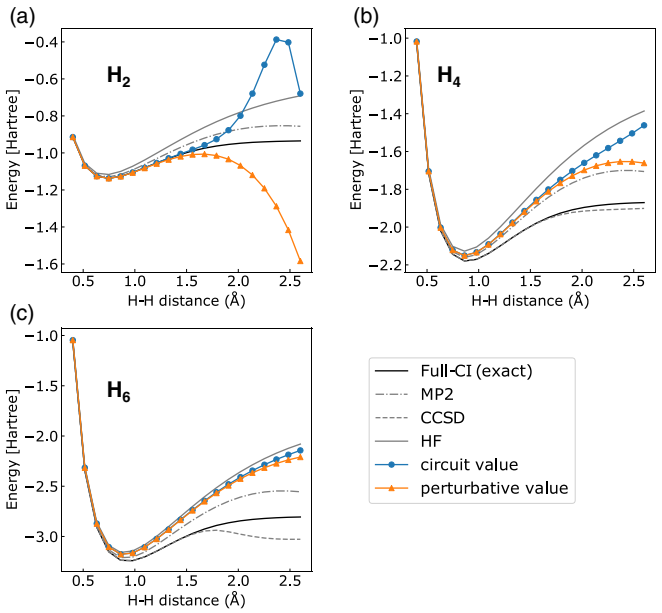

FIG. 2. Results of the numerical experiments at (a) $m = 2$, (b) $m = 4$, and (c) $m = 6$. Full-CI is the exact ground-state energy, and MP2, CCSD, and HF, respectively, are the mean approximate ground-state energy obtained with second-order Møller-Plesset perturbation, coupled cluster with single and double excitations, and Hartree-Fock. Circuit value and perturbative value represent $E(\boldsymbol{\theta}^*)$ and its perturbative approximation computed from Eq. (2), respectively.

overall time complexity of computing Hessian for this system is $\mathcal{O}(m^8)$. Since the minimization of the quadratic function can be done in time $\mathcal{O}(K^3) = \mathcal{O}(m^6)$, the Hessian part contributes the most to the total time.

Finally, to somewhat relax the randomness of the ansatz, we first randomly generate 200 ansatzes in the form of Fig. 1 for each m used in the experiment. For each generated ansatz, we calculate g_l using the Hamiltonian with the spacing of 1.0 Å. Then, the circuit with the largest $\sum_l |g_l|$ is chosen as the ansatz to be used for each hydrogen chain with different spacings. This is based on our expectation that an ansatz with large gradients would provide the highest performance.

B. Results and discussion

1. Preliminary results of small-scale systems

Figure 2 shows the result of the numerical experiment at $m = 2$, 4, and 6 along with the energy obtained from standard quantum chemistry calculations [32] as references. For $m = 2$ and 4, we can observe that the energies obtained from the circuit simulation and the one from the perturbative optimization match well at small spacings. Here the Hartree-Fock method gives a relatively accurate description of the ground state, and the perturbative treatment works fine as expected. The effectiveness of the perturbation also means that we can achieve the optimal parameter with this technique. This implies that the hardware-efficient ansatzes considered in this work can only achieve the accuracy of second-order Møller-Plesset perturbation (MP2) [32], which is a technique used widely in current quantum chemistry calculations as one of the easiest post-Hartree-Fock methods, for H_4 as we can observe from Fig. 2(b). There is a possibility of improving the accuracy by optimizing from a randomly initialized $\boldsymbol{\theta}$ as the above discussion only considers the case where we take $\boldsymbol{\theta} = 0$ as the initial parameter. However, such a strategy would not be generally scalable because of the barren plateau problem [33]. On the other hand, the perturbative treatment breaks down at the larger spacings where the electronic correlation becomes larger. As mentioned in the previous section, one has to perform further optimization in such a case. Note that for the $m = 2$ case, we can generate the exact ground state by a Clifford circuit at a large spacing in principle since the ground state approaches $(|0011\rangle + |1100\rangle)/\sqrt{2}$. However, it is not possible to generate this state using the ansatz construction strategy utilized here because we restrict the ansatz to become identity at $\boldsymbol{\theta} = 0$. This is a limitation of this strategy in the ansatz construction but not of the quadratic clifford expansion. We believe this situation can be improved by using more sophisticated strategy for choosing the Clifford part of the circuit, e.g., it would be interesting to combine the strategy presented in Ref. [34] with the proposed method.

In the case of $m = 6$, we cannot observe the clear breakdown of the perturbative treatment, i.e., the energies obtained from the perturbation match well with those from the circuit simulation. Again, it means that the optimal parameters and corresponding energies can be obtained with the proposed technique. We can see that the hardware-efficient circuit cannot even achieve the MP2 energy for H_6. Note that MP2 considers up to double electron excitations and involves $\mathcal{O}(n^4)$ parameters in its construction. This scaling is considerably greater than the number of parameters implemented in the ansatz of Fig. 1 with depth n. In this sense, the performance worse than MP2 is expected behavior. This trend of decreasing accuracy will also be certified with the result in the next subsection.

The important message of Fig. 2 is that the perturbative treatment becomes more accurate as we increase the system size in this setup. Assuming this trend persists in larger systems, our method can provide a performance benchmark in terms of the energy achievable by the ansatz at a scale where classical simulation of VQE is infeasible. In the next subsection, we perform such a benchmark up to a system size $m = 24$ which corresponds to 48 qubits.

2. Benchmark results of large-scale systems

Although the complexity of $\mathcal{O}(m^8)$ is polynomial in m, the large exponent prohibits us from extending the analysis to larger scales. Due to this complexity, we need some modifications to the experimental settings. First, we modify the ansatz to only involve real numbers by generating an ansatz that has the same form as the one shown in Fig. 1, but the random single-qubit Clifford gates are randomly chosen from identity and Hadamard gates only, and the single-qubit rotations just contain a y rotation. This modification reduces the number of parameters by a third. We find that this modification does not significantly alter the result as shown in the Appendix, which can be explained by the fact that the eigenstates of nonrelativistic quantum chemistry Hamiltonians can be described with states that are real in the computational basis. Second,

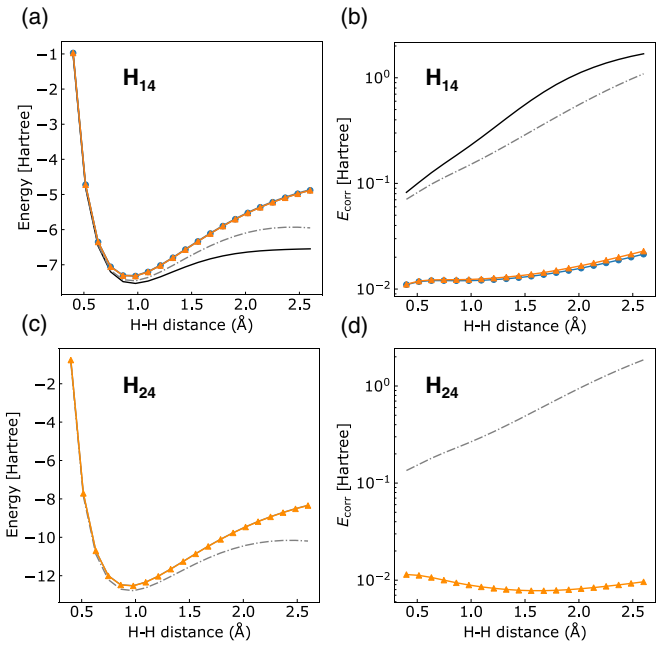

FIG. 3. Results of the numerical experiments at (a), (b) $m = 14$ and (c), (d) $m = 24$. Graph legends follow those of Fig. 2. Panels (a) and (c) show the total energy, while (b) and (d) show the correlation energy.

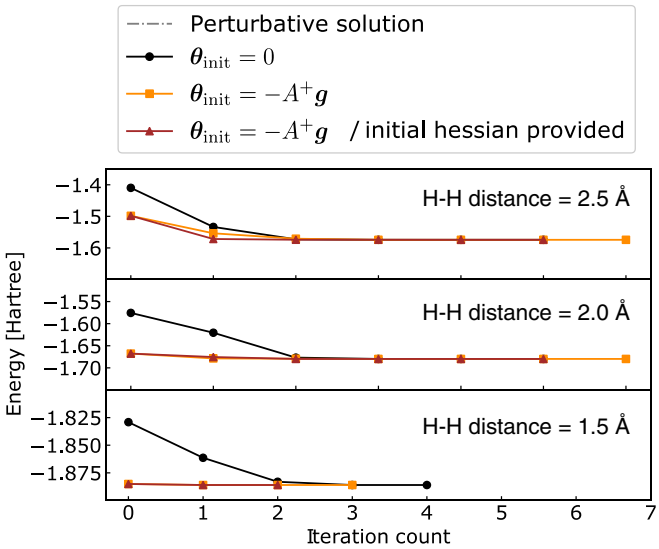

FIG. 4. Comparison of the convergence of optimization procedures with different $\boldsymbol{\theta}_{\text{init}}$ using $m = 4$ Hamiltonian at different spacings of hydrogen atoms.

to further reduce the number of parameters involved in the Hessian calculation, we "drop out" the parameters that give zero gradients to the energy, i.e., y rotations that do not give the first-order contribution to the energy are removed from the ansatz after calculating the gradient. This is motivated by our observation in the preliminary experiment where we have found that the gradients with respect to most of the parameters are exactly zero. In the following experiment, the results are obtained by removing the rotation gates with $|g_l| < 10^{-6}$ Hartree. In the Appendix, we show that this modification does not significantly alter the results either.

Figure 3 shows the benchmark results for $m = 14$ and $m = 24$ cases. Note that $m = 14$ corresponds to 28 qubits, which is the largest number of qubits that can be handled with the Qulacs-GPU simulator [29] using an NVIDIA Tesla-V100 processor. At $m = 16$, the required memory far exceeds its capacity. For this reason, we do not show the circuit value in the case of $m = 24$ which corresponds to 48 qubits [Figs. 3(c) and 3(d)]. Also, as the exact diagonalization at $m = 24$ could not be performed under our environment, it is not shown in the figure. The energy of coupled-cluster with single and double excitations (CCSD) is not depicted in both cases because we experienced its numerical instability. Note that coupled-cluster methods generally encounter instability when applied to a system with strong correlation [35,36].

In Figs. 3(b) and 3(d), we show the correlation energy E_{corr} defined as the difference between the obtained energy and the Hartree-Fock reference energy to illustrate the performance of the proposed method and the hardware-efficient ansatz itself. From Fig. 3(b), we can observe that the perturbative energy and the circuit value match almost exactly, indicating that we can analyze the performance of the ansatz itself with this perturbative treatment as we have expected from the results at smaller m. Therefore Figs. 3(b) and 3(d) show that the correlation energy that can be achieved by the hardware-efficient ansatz used in this work is a tenth or a hundredth smaller than that of MP2. This is, to the best of our knowledge, the first benchmark of the hardware efficient ansatz at this scale. Note that the perturbative minimum obtained by this method is a local one that locates around $\boldsymbol{\theta} = 0$, and the global optimal solution for this ansatz can perform better. Nevertheless, this result indicates that we should use types of ansatz other than this work since the one-dimensional hardware-efficient ansatz utilized in this work does not seem to be a promising choice.

3. Initialization performance

Finally, we show that the approximate optimal parameter $\boldsymbol{\theta}^* = -A^+\boldsymbol{g}$ together with the initial Hessian A can indeed serve as a good initial guess of the VQE. To this end, we take Hamiltonians of H_4 at different atom spacings as examples and compare the convergence of the optimization procedure of the VQE when using different initial parameters $\boldsymbol{\theta}_{\text{init}}$. The BFGS method implemented in SciPy [37], which is a popular quasi-Newton technique, is employed as the optimizer. We compare three cases: $\boldsymbol{\theta}_{\text{init}} = 0$, $\boldsymbol{\theta}_{\text{init}} = -A^+\boldsymbol{g}$, and $\boldsymbol{\theta}_{\text{init}} = -A^+\boldsymbol{g}$ with the initial Hessian provided to the optimizer.

Figure 4 shows the result of the numerical experiment. We can observe that the $\boldsymbol{\theta}_{\text{init}} = -A^+\boldsymbol{g}$ cases exhibit the faster convergence than $\boldsymbol{\theta}_{\text{init}} = 0$ in all cases, saving about 2 iterations. Also, the optimizer with the initial Hessian performs equally well or better than the case $\boldsymbol{\theta}_{\text{init}} = -A^+\boldsymbol{g}$ without providing Hessian. Although the improvement is not substantial as the $\boldsymbol{\theta}_{\text{init}} = 0$ case also converge in few iterations, we can conclude that out method can reduce the optimization steps to be performed on quantum devices.

IV. CONCLUSION

We proposed a technique to efficiently compute an approximate optimal parameter and the corresponding value of the cost function in the VQAs. It is based on the observation that we can efficiently compute the gradient and Hessian of the cost function if an ansatz is in the form of Eq. (1) which includes a wide range of circuits. Since the method is based on a perturbative expansion, we can obtain an accurate solution when the initial guess of the parameter from which we perform the Taylor expansion of Eq. (2) is close to an optimal one. Even if we do not have such an initial guess, the gradient and Hessian can be used to perform the first-step optimization, and those quantities can be passed to optimizers. The generality of the ansatz allows us to apply the proposed method to various VQAs such as VQE [2–4], quantum approximate optimization [5], and variational machine-learning algorithms [6–10].

We applied the method to the VQE of hydrogen chains with a one-dimensional hardware-efficient ansatz shown in Fig. 1 for its benchmark. The simulation involving 48 qubits is the largest scale to date. The numerical experiments showed that the performance of such a hardware-efficient ansatz in the VQE cannot even achieve that of classical MP2 calculation. To the best of our knowledge, the proposed method is the only one that enables us the benchmark of the VQAs beyond the scale that is classically simulatable. Although the benchmark results are pessimistic, it also motivates us to use other types of ansatzes and to make other initialization strategies such as the one presented in Ref. [38]. For example, one might be able to use genetic optimization to improve the ansatz in Fig. 1 from the random choice of Clifford gates. One might also be able to improve the performance of this ansatz by using localized orbitals instead of the naive Hartree-Fock orbitals utilized in this work to express the Hamiltonian, which would make it easier for the ansatz to capture the electronic correlation. We believe that the proposed technique will be of use to a wide range of the VQAs.

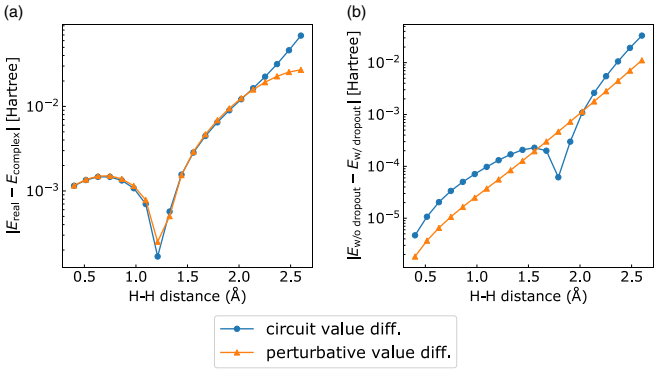

FIG. 5. (a) Difference of the results obtained with the original ansatz in Fig. 1 and the one modified to generate only real-valued state vectors at $m = 6$. (b) Difference of the result with and without dropout at $m = 6$ using the modified ansatz similar to Fig. 1, which only generates real wave functions.

ACKNOWLEDGMENTS

K.M. is supported by JST PRESTO Grant No. JPMJPR2019 and JSPS KAKENHI Grant No. 20K22330. Y.S. is supported by JST PRESTO Grant No. JPMJPR1916. W.M. is supported by JST PRESTO Grant No. JPMJPR191A. K.F. is supported by JSPS KAKENHI Grant No. 16H02211, JST ERATO JPMJER1601, and JST CREST JPMJCR1673. This work is supported by MEXT Quantum Leap Flagship Program (MEXT QLEAP) Grants No. JPMXS0118067394 and No. JPMXS0120319794.

APPENDIX: REDUCING THE NUMBER OF PARAMETERS IN THE ANSATZ

First, we demonstrate that the modification of the ansatz to be real in the computational basis (see Sec. III B 2) does not alter the results significantly. It is illustrated in Fig. 5(a) where we plot the difference of the energy obtained by the modified and original ansatz, respectively, is denoted as E_{real} and E_{complex}. The energy difference does not exceed 10^{-2} Hartree in the figure when the spacing is less than 2.0 Å, which is negligible compared with the correlation energy.

The effect of the "drop-out" utilized in Sec. III B 2 does not alter results either. Figure 5(b) shows the comparison of the results with and without dropout at $m = 6$ using the modified real ansatz. We can observe that the "dropout" only slightly alters the result by about the same magnitude as Fig. 5(a).

[1] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Variational quantum algorithms, Nat. Rev. Phys. **3**, 625 (2021).

[2] 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, Nat. Commun. **5**, 4213(2014).

[3] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, The theory of variational hybrid quantum-classical algorithms, New J. Phys. **18**, 023023 (2016).

[4] 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 (London) **549**, 242 (2017).

[5] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm, arXiv:1411.4028 [quant-ph].

[6] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, Quantum circuit learning, Phys. Rev. A **98**, 032309 (2018).

[7] M. Schuld and N. Killoran, Quantum Machine Learning in Feature Hilbert Spaces, Phys. Rev. Lett. **122**, 040504 (2019).

[8] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, Supervised learning with quantum-enhanced feature spaces, Nature (London) **567**, 209 (2019).

[9] E. Farhi and H. Neven, Classification with quantum neural networks on near term processors, arXiv:1802.06002 [quant-ph].

[10] M. Benedetti, D. Garcia-Pintos, O. Perdomo, V. Leyton-Ortega, Y. Nam, and A. Perdomo-Ortiz, A generative modeling approach for benchmarking and training shallow quantum circuits, npj Quantum Inf. **5**, 45 (2019).

[11] M. P. Harrigan, K. J. Sung, M. Neeley, K. J. Satzinger, F. Arute, K. Arya, J. Atalaya, J. C. Bardin, R. Barends, S. Boixo, M. Broughton, B. B. Buckley, D. A. Buell, B. Burkett, N. Bushnell, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney *et al.*, Quantum approximate optimization of non-planar graph problems on a planar superconducting processor, Nat. Phys. **17**, 332 (2021).

[12] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, S. Boixo, M. Broughton, B. B. Buckley, D. A. Buell, B. Burkett, N. Bushnell, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, S. Demura, A. Dunsworth, E. Farhi *et al.*, Hartree-fock on a superconducting qubit quantum computer, Science **369**, 1084 (2020).

[13] Y. Nam, J.-S. Chen, N. C. Pisenti, 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, K. M. Hudek, A. M. Ducore, M. Keesan, S. M. Kreikemeier *et al.*, Ground-state energy estimation of the water molecule on a trapped-ion quantum computer, npj Quantum Inf. **6**, 33 (2020).

[14] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, 10th ed. (Cambridge University Press, USA, 2011).

[15] S. Aaronson and D. Gottesman, Improved simulation of stabilizer circuits, Phys. Rev. A **70**, 052328 (2004).

[16] Z. Holmes, K. Sharma, M. Cerezo, and P. J. Coles, Connecting ansatz expressibility to gradient magnitudes and barren plateaus, PRX Quantum **3**, 010313 (2022).

[17] X. G. WEN, Topological orders in rigid states, Int. J. Mod. Phys. B **04**, 239 (1990).

[18] A. C. Doherty and S. D. Bartlett, Identifying Phases of Quantum Many-Body Systems That Are Universal for Quantum Computation, Phys. Rev. Lett. **103**, 020506 (2009).

[19] R. Sweke, F. Wilde, J. Meyer, M. Schuld, P. K. Faehrmann, B. Meynard-Piganeau, and J. Eisert, Stochastic gradient descent for hybrid quantum-classical optimization, Quantum **4**, 314 (2020).

[20] J. Stokes, J. Izaac, N. Killoran, and G. Carleo, Quantum Natural Gradient, Quantum **4**, 269 (2020).

[21] A. Mari, T. R. Bromley, and N. Killoran, Estimating the gradient and higher-order derivatives on quantum hardware, Phys. Rev. A **103**, 012405 (2021).

[22] N. Yamamoto, On the natural gradient for variational quantum eigensolver, arXiv:1909.05074 [quant-ph].

[23] B. Koczor and S. C. Benjamin, Quantum analytic descent, Phys. Rev. Research **4**, 023017 (2022).

[24] M. Motta, D. M. Ceperley, G. K.-L. Chan, J. A. Gomez, E. Gull, S. Guo, C. A. Jiménez-Hoyos, T. N. Lan, J. Li, F. Ma, A. J. Millis, N. V. Prokof'ev, U. Ray, G. E. Scuseria, S. Sorella, E. M. Stoudenmire, Q. Sun, I. S. Tupitsyn, S. R. White, D. Zgid, and S. Zhang (Simons Collaboration on the Many-Electron Problem), Towards the Solution of the Many-Electron Problem in Real Materials: Equation of State of the Hydrogen Chain with State-of-the-Art Many-Body Methods, Phys. Rev. X **7**, 031059 (2017).

[25] J. R. McClean, N. C. Rubin, K. J. Sung, I. D. Kivlichan, X. Bonet-Monroig, 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 *et al.*, OpenFermion: the electronic structure package for quantum computers, Quantum Sci. Technol. **5**, 034014 (2020).

[26] 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 *et al.*, Pyscf: the python-based simulations of chemistry framework, WIREs Comput. Mol. Sci. **8**, e1340 (2018).

[27] Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt, N. A. Bogdanov, G. H. Booth, J. Chen, Z.-H. Cui *et al.*, Recent developments in the pyscf program package, J. Chem. Phys. **153**, 024109 (2020).

[28] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, Quantum computational chemistry, Rev. Mod. Phys. **92**, 015003 (2020).

[29] Y. Suzuki, Y. Kawase, Y. Masumura, Y. Hiraga, M. Nakadai, J. Chen, K. M. Nakanishi, K. Mitarai, R. Imai, S. Tamiya, T. Yamamoto, T. Yan, T. Kawakubo, Y. O. Nakagawa, Y. Ibe, Y. Zhang, H. Yamashita, H. Yoshimura, A. Hayashi, and K. Fujii, Qulacs: a fast and versatile quantum circuit simulator for research purpose, Quantum **5**, 559 (2021).

[30] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler *et al.*, Quantum supremacy using a programmable superconducting processor, Nature (London) **574**, 505 (2019).

[31] P. Selinger, Generators and relations for n-qubit clifford operators, Logical Methods Comp. Sci. **11** (2015).

[32] T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic-Structure Theory* (John Wiley & Sons, Ltd, New York, 2000).

[33] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, Nat. Commun. **9**, 4812 (2018).

[34] G. S. Ravi, P. Gokhale, Y. Ding, W. M. Kirby, K. N. Smith, J. M. Baker, P. J. Love, H. Hoffmann, K. R. Brown, and F. T. Chong, Cafqa: Clifford ansatz for quantum accuracy, arXiv:2202.12924 [quant-ph].

[35] P. R. Surján and Á. Szabados, On the coupled-cluster equations. stability analysis and nonstandard correction schemes, in *Recent Progress in Coupled Cluster Methods* (Springer, Berlin, 2010), pp. 513–534.

[36] T. J. Lee and P. R. Taylor, A diagnostic for determining the quality of single-reference electron correlation methods, Int. J. Quantum Chem. **36**, 199 (1989).

[37] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright *et al.*, Scipy 1.0: fundamental algorithms for scientific computing in python, Nat. Methods **17**, 261 (2020).

[38] E. Grant, L. Wossnig, M. Ostaszewski, and M. Benedetti, An initialization strategy for addressing barren plateaus in parametrized quantum circuits, Quantum **3**, 214 (2019).