Adaptive construction of shallower quantum circuits with quantum spin projection for fermionic systems

Takashi Tsuchimochi,[1,2,*] Masaki Taii,[3] Taisei Nishimaki,[1] and Seiichiro L. Ten-no[3,1]

[1]*Graduate School of System Informatics, Kobe University, 1-1 Rokkodai-cho, Nada-ku, Kobe, Hyogo 657-8501, Japan*

[2]*Japan Science and Technology Agency (JST), Precursory Research for Embryonic Science and Technology (PRESTO), 4-1-8 Honcho Kawaguchi, Saitama 332-0012, Japan*

[3]*Graduate School of Science, Technology, and Innovation, Kobe University, 1-1 Rokkodai-cho, Nada-ku, Kobe, Hyogo 657-8501, Japan*

(Received 1 March 2022; revised 14 June 2022; accepted 15 July 2022; published 4 August 2022)

Quantum computing is a promising approach to harnessing strong correlation in molecular systems; however, current devices only allow for hybrid quantum-classical algorithms with a shallow circuit depth, such as the variational quantum eigensolver (VQE). In this paper, we report the importance of the Hamiltonian symmetry in constructing VQE circuits adaptively (ADAPT-VQE). This treatment often violates symmetry, thereby deteriorating the convergence of fidelity to the exact solution and ultimately resulting in deeper circuits. We demonstrate that spin-symmetry projection can provide a simple yet effective solution to this problem, by keeping the quantum state in the correct symmetry space, to reduce the overall gate operations. To further investigate the role of spin-symmetry in computing molecular properties with ADAPT-VQE, we have derived the analytical derivative of symmetry-projected VQE energy. Our illustrative calculations reveal the significance of preserving symmetry in providing accurate dipole moments and geometries with variational approximations.

DOI: 10.1103/PhysRevResearch.4.033100

I. INTRODUCTION

Recent advances in quantum technology have attracted widespread attention in various disciplines. Electronic structure theory is one of the fields expected to potentially benefit from the development of quantum computers because strong correlation can be handled by mapping the wave function directly onto entangled qubits without an exponential increase in computational cost. However, current noisy intermediate-scale quantum (NISQ) devices cannot implement error correction owing to the limited quantum resources; hence, they suffer from errors triggered by noise and short coherent times [1–3], which limit the application of general quantum algorithms such as quantum phase estimation. Therefore, several hybrid quantum-classical algorithms have been proposed to address the technical challenges emerging in NISQ hardware. The variational quantum eigensolver (VQE) is one such algorithm [4,5], in which a quantum state is propagated by a parameterized short quantum circuit that can be optimized in classical postprocessing, to variationally lower the energy expectation value and obtain the Hamiltonian ground state.

In quantum chemistry, the most conventional *Ansatz* for VQE circuit is the unitary coupled cluster (UCC) [6–10], where fermionic excitation operators with respect to the Fermi vacuum [e.g., Hartree-Fock (HF)] are anti-hermitized and exponentiated to form unitary gates. Given the remarkable success of the standard coupled cluster (CC) in a wide range of classical applications, the UCC with single and double substitutions (UCCSD) is expected to be a feasible starting point for chemical Hamiltonians. Nevertheless, the UCCSD is manifestly incapable of describing strong correlations [11], which is the primary target of quantum computing. Extending the excitation manifold to include general orbitals, the generalized UCCSD captures higher excitation effects in a compact manner and is thus known to often provide accurate results for strongly correlated systems [11,12]. However, these UCC-based methods are inhibited by a limited representability because of the fixed structure of unitary gates; in addition, they also require several parameters and deep circuits, which makes them difficult for NISQ devices to handle.

Ansatz-free algorithms are more flexible and can control the accuracy and circuit depth in a highly black-box manner [13–17]. A prominent algorithm in this context is ADAPT-VQE proposed by Grimsley *et al.* [13], which adaptively adds a set of short unitary gates to the quantum circuit successively. Each unitary is chosen from an operator pool comprising generalized-UCCSD excitation operators, based on the VQE energy gradient. Later, Tang *et al.* [15] proposed qubit-ADAPT, which employs Pauli rotations instead of a UCC-like operator to drastically reduce the gate depth. To minimize the size of the pool in qubit-ADAPT, Shkolnikov *et al.* [18] chose Pauli operators that conserve symmetries against the initial HF state. Recently, Yordanov *et al.* [19] demonstrated that qubit excitations, each of which is a linear combination of Pauli strings in qubit-ADAPT, can be implemented with only

*tsuchimochi@gmail.com

a handful of CNOT gates. Based on this result, they employed qubit excitations to form an operator pool in what they called qubit-excitation-based (QEB) ADAPT [17]. However, these ADAPT-based methods are prone to break important symmetries of chemical Hamiltonians.

On many occasions, breaking symmetry triggers several undesired consequences. For instance, the result of the $\hat{S}^2$ (spin) symmetry breaking in molecular applications is well documented in the literature, which includes inaccurate energies and properties, a slow convergence to the exact state, and difficulty in interpreting obtained results [20–22]. ADAPT-based methods can result in symmetry breakages because of the use of fermionic or qubit excitations, for which spin-symmetry adaptation is especially difficult when exponentiated. This is more the rule than the exception. In fact, in strongly correlated systems, spin symmetry is completely broken because ADAPT typically requires large rotation parameters to represent the multiconfiguration character. As will be demonstrated, this causes the slow convergence of ADAPT algorithms and inaccurate results with truncated operators.

We note that there have been several studies that consider the importance of symmetries for the performance of VQE. For hardware-efficient circuits, Barkoutsos *et al.* [23] introduced particle-conserving gates to keep the number ($\hat{N}$) symmetry. Gate operations that preserve the $\hat{S}^2$ symmetry were proposed and used to singlet-pair states for a Heisenberg model [24,25]. Gard *et al.* [26] also developed general circuits that conserve all $\hat{N}$, $\hat{S}_z$, and $\hat{S}^2$ symmetries. From a different perspective, recently, symmetry projection has been also combined with VQE, to restore broken symmetries of prepared quantum states [27–31], showing promise for even more accurate VQE simulations. A desirable feature of symmetry projection is that it can be applied straightforwardly to general circuits including Pauli rotations used in ADAPT-VQE. Therefore, to address the above symmetry issue in ADAPT-VQE, in this paper, we focus on the effect of spin projection. It should be mentioned that, related to this idea, Romero *et al.* [31] have very recently combined the number-symmetry projection with ADAPT-VQE to tackle the spontaneous symmetry-breaking of the particle number in the Lipkin model. Here, we particularly consider the $\hat{S}^2$ symmetry, noting that the $\hat{N}$ and $\hat{S}_z$ symmetries are rather straightforward to conserve in chemical systems. We employ a spin-projection circuit [27], which restores the total spin quantum number to restrict ADAPT and explore the operator candidate in the correct symmetry space. We demonstrate that our algorithm can generally reduce the circuit depth, although it increases the measurement cost. In particular, our numerical tests show that the spin-projected (SP) ADAPT outperforms the UCCSD as a better tradeoff between accuracy and circuit depth, even in the case where ADAPT (without projection) turns out not to be particularly advantageous over the UCCSD in terms of these criteria. To investigate the importance of spin-symmetry in molecular property calculations, we further derive the first-order energy derivative in the presence of the spin-projection operator.

The remainder of this paper is organized as follows. In Secs, II A and II B, we review the original algorithms for ADAPT-VQE, and we describe how they break symmetries in Sec. II C. Section II D presents detailed discussions on the application of spin projection to each ADAPT scheme, including an explanation on how the simulated qubits can be tapered off in the presence of spin projection. Subsequently, we derive the energy derivative for ADAPT in Sec. III, especially focusing on the importance of orbital response. In Sec. IV B, we demonstrate that the convergence in ADAPT becomes increasingly slower for strongly correlated systems, and in Sec. IV C, we present our benchmark results on the SP-ADAPT, comparing the accuracy and gate efficiency between different methods. Section IV D demonstrates how the number of operators in a pool can be reduced in QEB-ADAPT. The calculations for dipole moment and geometry are explained in Sec. IV E. Finally, we conclude this paper in Sec. V.

II. ADAPT-VQE ALGORITHM

A. Fermionic ADAPT-VQE

First, we review the ADAPT-VQE algorithm proposed by Grimsley *et al.* [13]. This algorithm is based on the observation that the state obtained with exact diagonalization, also called full configuration interaction (FCI), can be reached from the HF state using the exponential form [13,32]:

$$|\Psi^{\mathrm{FCI}}\rangle = \prod_k \exp[t_k \hat{\tau}(k)]|\Phi^{\mathrm{HF}}\rangle, \tag{1}$$

where $\hat{\tau}(k)$ is the kth instance of arbitrarily ordered operators comprising the antisymmetrized single and double excitations:

$$\hat{\tau}_q^p = a_p^\dagger a_q - a_q^\dagger a_p, \tag{2a}$$

$$\hat{\tau}_{rs}^{pq} = a_p^\dagger a_q^\dagger a_r a_s - a_s^\dagger a_r^\dagger a_q a_p. \tag{2b}$$

Here, we have adopted $p > q, r > s$ to denote general spin orbitals. Note that the same antisymmetrized excitation operators can appear repeatedly in Eq. (1) but with different variational parameters t_k. Although Eq. (1) includes only singles and doubles, the effect of higher excitations such as triples and quadruples can be conveniently considered by their exponential products [32].

Regardless of its exactness, Eq. (1) is extremely cumbersome and long because it requires a large number of k; hence, its realization is prohibitively difficult on quantum computers. On the one hand, from a circuit complexity perspective, such as the number of CNOT gates, one can employ only a limited number of unitaries $\exp[t_k \hat{\tau}(k)]$ on NISQ computers. On the other hand, it is expected that such a truncated *Ansatz* with a reasonable number n of unitaries can still outperform fixed *Ansätze* such as the UCCSD, especially for strongly correlated systems, with appropriately chosen $\hat{\tau}(k)$ $(k = 1, \cdots, n)$ and variationally optimized t_k. Therefore, the objective of ADAPT-VQE is to dynamically create an *Ansatz* that approaches FCI, using a maximally compact sequence of n unitary operators, which are successively determined based on the energy gradient. To do so, an operator pool $\mathcal{P}$ is first defined, which comprises M operators that are used to construct the trial state. There exist several different definitions of such operator pools. As considered in Ref. [13], fermionic ADAPT-VQE prepares the following pool:

$$\mathcal{P} = \left\{\hat{\tau}_{q_\alpha}^{p_\alpha} + \hat{\tau}_{q_\beta}^{p_\beta}, \hat{\tau}_{r_\alpha s_\alpha}^{p_\alpha q_\alpha} + \hat{\tau}_{r_\beta s_\beta}^{p_\beta q_\beta}, \hat{\tau}_{r_\alpha s_\beta}^{p_\alpha q_\beta} + \hat{\tau}_{r_\beta s_\alpha}^{p_\beta q_\alpha}\right\}, \tag{3}$$

where we explicitly consider the spins α and β, and each linear combination of operators is labeled by $m = 1, \cdots, M$. Note that each combination of doubles does not comprise the spin-adapted form, which we address later in this paper.

A trial state of ADAPT-VQE at the nth cycle is given by

$$|\psi_n\rangle = \exp(\theta_n \hat{\mathcal{A}}_n) \exp(\theta_{n-1} \hat{\mathcal{A}}_{n-1}) \times \cdots \exp(\theta_1 \hat{\mathcal{A}}_1) |\psi^{\mathrm{HF}}\rangle, \tag{4}$$

where $\hat{\mathcal{A}}_k \in \mathcal{P}$, and the parameters $\{\theta_k\}$ are variationally optimized to minimize the energy expectation value. To select the most contributing operator from $\mathcal{P}$ for the $(n+1)$th cycle, we create trial states with each $\hat{A}_m \in \mathcal{P}$ $(m = 1, \cdots, M)$ as

$$|\psi_{n+1,m}\rangle = \exp(\theta_{n+1} \hat{A}_m) |\psi_n\rangle, \tag{5}$$

and calculate the energy gradient for each $|\psi_{n+1,m}\rangle$ around $\theta_{n+1} = 0$:

$$R_m^{(n)} = \left. \frac{\partial \langle \psi_n | \exp(-\theta \hat{A}_m) \hat{H} \exp(\theta \hat{A}_m) | \psi_n \rangle}{\partial \theta} \right|_{\theta=0} = \langle \psi_n | [\hat{H}, \hat{A}_m] | \psi_n \rangle, \tag{6}$$

which can be evaluated with up to a three-body reduced density matrix because of the commutator property. The operator with the largest gradient $R_m^{(n)}$ is selected as $\hat{\mathcal{A}}_{n+1}$, with which the parameters $\{\theta_k : k = 1, \cdots, n+1\}$ in $|\psi_{n+1}\rangle$ are all optimized by the standard VQE.

The ADAPT algorithm presented above is converged when the norm of Eq. (6) is smaller than the threshold ϵ. It is worth noting that such a condition is closely related to the k-particle generalized Brillouin theorem [33]. Kutzelnigg [33] demonstrated that an exact FCI wave function satisfies $\langle \Psi_{\mathrm{FCI}} | [\hat{H}, \hat{\tau}_\mu] | \Psi_{\mathrm{FCI}} \rangle = 0$ for all excitation ranks. The convergence in fermionic ADAPT-VQE with $\epsilon = 0$ only corresponds to the 1- and 2-particle generalized Brillouin conditions, and therefore, fermionic ADAPT-VQE may not necessarily be equivalent to FCI at convergence. However, in practice, it approaches the FCI accuracy as $||\mathbf{R}||$ decreases [34], although there are cases where ADAPT-VQE with only single and double excitations in the pool fails to converge to the FCI state. Appendix A showcases such an example using N_2 with a π active space.

B. Qubit-based ADAPT-VQE

The aforementioned fermionic ADAPT is based on the Jordan-Wigner transformation for mapping fermionic excitation operators Eq. (2) to Pauli operators. This introduces the chain of Z operations because of the antisymmetric character of fermions, e.g.,

$$\hat{\tau}_{rs}^{pq} = \frac{i}{8} (Y_p X_q X_r X_s + X_p Y_q X_r X_s - X_p X_q Y_r X_s - X_p X_q X_r Y_s - X_p Y_q Y_r Y_s - Y_p X_q Y_r Y_s + Y_p Y_q X_r Y_s + Y_p Y_q Y_r X_s) \bigotimes_{t=q+1}^{p-1} Z_t \bigotimes_{u=s+1}^{r-1} Z_u, \tag{7}$$

where we assume $p > q > r > s$, but similar relations can be obtained for other cases. The exponential of $\hat{\tau}_{rs}^{pq}$ is usually implemented by decomposing the eight Pauli strings presented in Eq. (7), for which each requires the ladders of CNOT gates to consider the parities for $t = q+1, \cdots, p-1$ and $u = s+1, \cdots, r-1$; refer to Fig. 1 as an example. Therefore, such a naïve implementation of $\exp(\theta \hat{\tau}_{rs}^{pq})$ necessitates $8(p-q+r-s+2) \geqslant 48$ CNOTs. However, several studies [35–37] have demonstrated that, even if all the Z operators from the Pauli string in Eq. (7) are discarded to define the qubit excitation:

$$\hat{\tilde{\tau}}_{rs}^{pq} = \frac{i}{8} (Y_p X_q X_r X_s + X_p Y_q X_r X_s - X_p X_q Y_r X_s - X_p X_q X_r Y_s - X_p Y_q Y_r Y_s - Y_p X_q Y_r Y_s + Y_p Y_q X_r Y_s + Y_p Y_q Y_r X_s), \tag{8}$$

the accuracy obtained by VQE will be similar to that with the full $\hat{\tau}_{rs}^{pq}$, although it eliminates a considerable number of CNOT gates. Tang *et al.* [15] attempted to further decompose $\hat{\tilde{\tau}}_{rs}^{pq}$ and construct a reduced operator pool for what they called qubit-ADAPT-VQE, noting that each Pauli string has a similar action. Using circuits depicted in Fig. 1(a), each operator $\exp(\theta_m \hat{A}_m)$ requires only six CNOT gates (similarly, two CNOT gates for single excitation variants, $Y_p X_q$). However, as will be discussed, this treatment significantly breaks several symmetries in the Hamiltonian and slows down the convergence of the algorithm for chemical applications. Moreover, because it violates the number symmetry, it cannot be applied to ionized states or half-filled Hubbard models.

Recently, Yordanov *et al.* [19] demonstrated that the role of $\exp(\theta \hat{\tilde{\tau}}_{rs}^{pq})$ is to rotate the bit strings between $|0_p 0_q 1_r 1_s\rangle$ and $|1_p 1_q 0_r 0_s\rangle$ by θ, while everything else is unaffected, and proposed the application of a controlled-Ry gate, as illustrated in Fig. 1(b), up to the phase. This implementation requires only 13 CNOT gates, although the connectivity between $pqrs$ is assumed. They have also established that a standard double excitation $\exp(\theta \tau_{rs}^{pq})$ can be similarly implemented [Fig. 1(c)], where the number of CNOT gates is reduced to $2(p-q+r-s)+9$. Using Eq. (8) and Fig. 1(b), they proposed QEB ADAPT, which was shown to perform significantly better than qubit-ADAPT of Ref. [15] in terms of CNOT gate counts and number of parameters [17].

C. Symmetry-breaking in ADAPT-VQE

In the proposal by Grimsley *et al.* [13], it appears to be convenient to choose $\hat{\tau}_{r_\alpha s_\alpha}^{p_\alpha q_\alpha} + \hat{\tau}_{r_\beta s_\beta}^{p_\beta q_\beta}$ and $\hat{\tau}_{r_\alpha s_\beta}^{p_\alpha q_\beta} + \hat{\tau}_{r_\beta s_\alpha}^{p_\beta q_\alpha}$ as the elements of the operator pool for fermionic ADAPT because operators in each linear combination are symmetric in terms of spin. However, they are neither generally commutative ($\hat{\tau}_{r_\alpha s_\beta}^{p_\alpha q_\beta}$ and $\hat{\tau}_{r_\beta s_\alpha}^{p_\beta q_\alpha}$ do not commute if $p = q$ or $r = s$) nor spin adapted. This results in two limitations. First, because the operator selection is based on Eq. (6) while the VQE part (implicitly) introduces no Trotter decomposition, i.e., a single Trotterization step, the gradients evaluated in these two parts can be inconsistent, which may cause an occasional numerical instability. Specifically, for $\hat{P}_1 = \hat{\tau}_{r_\alpha s_\beta}^{p_\alpha p_\beta}$ and $\hat{P}_2 = \hat{\tau}_{r_\beta s_\alpha}^{p_\beta p_\alpha}$, where $[\hat{P}_1, \hat{P}_2] \neq 0$,

$$\exp[\theta(\hat{P}_1 + \hat{P}_2)] \neq \exp(\theta \hat{P}_1) \exp(\theta \hat{P}_2), \tag{9}$$

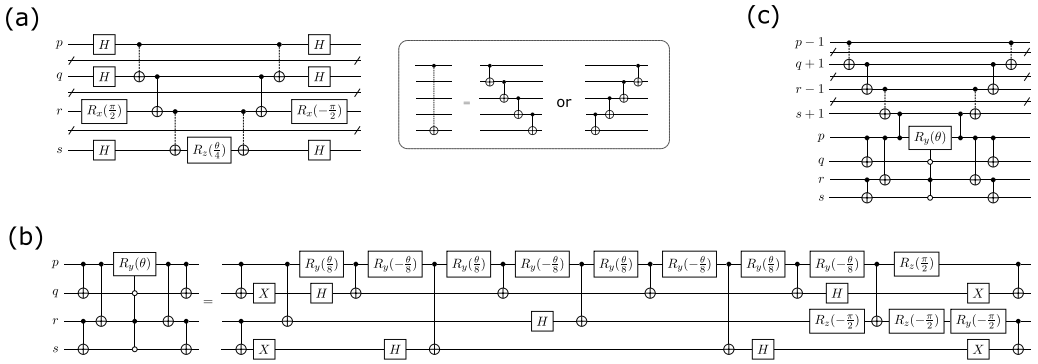

FIG. 1. Quantum circuits for (a) $\exp(-\frac{i\theta}{8}X_pX_qY_rX_s)$, where the CNOT with a dashed line is shorthand for the CNOT ladder to convey the parities (refer to the inset), (b) $\exp(\theta\hat{\tau}^{pq}_{rs})$, and (c) $\exp(\theta\hat{\tau}^{pq}_{rs})$ using the controlled-Ry gate.

and therefore,

$$\frac{\partial}{\partial\theta}\exp[\theta(\hat{P}_1+\hat{P}_2)] = (\hat{P}_1+\hat{P}_2)\exp[\theta(\hat{P}_1+\hat{P}_2)], \tag{10}$$

is different from

$$\frac{\partial}{\partial\theta}[\exp(\theta\hat{P}_1)\exp(\theta\hat{P}_2)] = \hat{P}_1\exp(\theta\hat{P}_1)\exp(\theta P_2) + \exp(\theta\hat{P}_1)\hat{P}_2\exp(\theta\hat{P}_2). \tag{11}$$

The operator selection step Eq. (6) relies on Eq. (10), whereas the VQE step is based on Eq. (11). This indicates that the same operator could potentially be chosen between consecutive ADAPT cycles, although in practice we do not see such an ill behavior. Second, the application of a spin-dependent operator basis sacrifices the conservation of spin s. Although a more recent study has employed the spin-adapted doubles generator, that is, the sum of these two combinations [15], it necessitates an undesired Trotter decomposition, as demonstrated in our previous study [27]. Using such spin-incomplete operators provokes a serious issue widely known in electronic structure as spin contamination. This is especially problematic in strongly correlated systems where the initial HF state significantly deviates from the FCI state, and therefore, large amplitudes θ_k are essential. Certainly, a lost spin symmetry would be eventually recovered by the ADAPT algorithm, provided the target spin state is the ground state and the aforementioned instability is circumvented. However, large spin-symmetry breaking is inevitable in the early steps of the ADAPT procedure, which often leads to unphysical gradients and a poor convergence, as demonstrated below. The qubit-ADAPT algorithm [15] is also expected to suffer from the same problem. Furthermore, the candidate operators in qubit-ADAPT break two other important symmetries, $\hat{S}_z$ and $\hat{N}$, and the problem becomes even more severe. This fact can be easily confirmed by transforming each of the strings in Eq. (8) back to their corresponding fermionic representations; for example, for the case where $p=6, q=3, r=1$, and $s=0$, we have

$$\tfrac{i}{8}Y_6X_3X_1X_0 = \tfrac{1}{8}(a_6^\dagger a_3^\dagger a_1 a_0 + a_6^\dagger a_3^\dagger a_1^\dagger a_0^\dagger - 2a_6^\dagger a_4^\dagger a_1^\dagger a_0^\dagger a_4 a_3 + \cdots - \text{H.c.}). \tag{12}$$

Hence, it is important to maintain the correct symmetry throughout the ADAPT algorithm, as much as possible, to prevent any unnecessary diversion toward FCI. Accordingly, in this paper, we investigate the importance of a symmetry-conserving sequence of operators, with a special focus on spin symmetry. One may employ the constraint approach where $\lambda(\hat{O}-o)^2$ is added to the energy functional as a penalty to enforce the target state to be an (approximate) eigenfunction of the symmetry operator $\hat{O}$ with the eigenvalue o. Here, we take a different approach and restore the broken symmetry in each step of ADAPT by adopting appropriate posttreatments, that is, spin projection.

D. SP-ADAPT

In electronic structure theory and nuclear physics, symmetry projection is a vital tool [38–43] and has recently exhibited a significant potential in quantum computing in the context of VQE [27,28,31]. Specifically spin projection, which corresponds to the angular momentum symmetry projection in the field of nuclear physics [39], restores the spin of the broken-symmetry quantum state $|\psi\rangle$, where $\hat{S}^2|\psi\rangle \neq s(s+1)|\psi\rangle$, by applying a spin-projection operator $\hat{P}$, such that

$$\hat{S}^2\hat{P}|\psi\rangle = s(s+1)\hat{P}|\psi\rangle, \tag{13}$$

$$\hat{S}_z\hat{P}|\psi\rangle = m_s\hat{P}|\psi\rangle, \tag{14}$$

where $m_s = (N_\alpha - N_\beta)/2$. Following our previous work, we express $\hat{P}$ as a linear combination of unitary operators:

$$\hat{P} = |s;m_s\rangle\langle s;m_s| = \frac{2s+1}{8\pi^2}\int_\Omega d\Omega D^{s*}_{m_sm_s}\exp(-i\alpha\hat{S}_z) \times \exp(-i\beta\hat{S}_y)\exp(-i\gamma\hat{S}_z) \approx \sum_g^{N_g} w_g\hat{U}_g, \tag{15}$$

where $\Omega=(\alpha,\beta,\gamma)$ are Euler angles [39]. We have used the numerical integration with $\hat{U}_g = \exp(-i\alpha_g\hat{S}_z)\exp(-i\beta_g\hat{S}_y)\exp(-i\gamma_g\hat{S}_z)$ and w_g are weights defined by the Wigner D-matrix $D^s_{m_sm_s}$. In VQE, we are interested in the expectation value of energy:

$$E = \frac{\langle\psi|\hat{P}^\dagger\hat{H}\hat{P}|\psi\rangle}{\langle\psi|\hat{P}^\dagger\hat{P}|\psi\rangle} \approx \frac{\sum_g w_g\langle\psi|\hat{H}\hat{U}_g|\psi\rangle}{\sum_g w_g\langle\psi|\hat{U}_g|\psi\rangle}, \tag{16}$$

where we have used $\hat{P}^2 = \hat{P} = \hat{P}^\dagger$ and $[\hat{H}, \hat{P}] = 0$. Note that, using transfer operators $|s;m\rangle\langle s;k|$, which generate a spin s state with $m_s = m$ from $m_s = k$, a more general projector $\sum_{m'_s} c_{m'_s}|s;m_s\rangle\langle s;m'_s|$ introduces more variational freedom, provided $|\psi\rangle$ is not an eigenstate of $\hat{S}_z$, by diagonalizing the Hamiltonian in the subspace of $\{|s;m_s\rangle\langle s;m'_s|\psi\rangle\}$ to determine the coefficients c_k [41]; however, for simplicity, we only consider $|s;m_s\rangle\langle s;m_s|$, which results in Eqs. (15) and (16). When $|\psi\rangle$ is an eigenstate of $\hat{S}_z$, Eq. (16) can be further simplified using the commutativity between $\hat{H}$ and $\hat{S}_z$, and one only needs to evaluate the Hamiltonian coupling $\langle\psi|\hat{H}\exp(-i\beta_g\hat{S}_y)|\psi\rangle$ and the overlap $\langle\psi|\exp(-i\beta_g\hat{S}_y)|\psi\rangle$. These expectation values are estimated, for instance, by the Hadamard test with one ancilla qubit, with a constant number of CNOTs ($4n_{\text{qubit}}$). With the Gauss-Legendre quadrature (see Appendix B), we take $N_g = 2$ in the applications below; hence, the overall measurements will be doubled, while the circuit length is almost unchanged with or without $\hat{P}$.

The basic idea here is to introduce $\hat{P}$ to ADAPT-VQE and constrain the operator search in the correct spin space for a better convergence. Accordingly, we simply define the SP-ADAPT by applying $\hat{P}$ to Eq. (4):

$$|\psi_n^{\text{SP}}\rangle = \hat{P}\ \exp(\theta_n\hat{\mathcal{A}}_n)\exp(\theta_{n-1}\hat{\mathcal{A}}_{n-1}) \times\cdots\exp(\theta_1\hat{\mathcal{A}}_1)|\psi^{\text{HF}}\rangle. \tag{17}$$

The energy gradient evaluated to find the $(n+1)$th operator is then like Eq. (6), as follows:

$$R_m^{(n)} = \frac{\partial}{\partial\theta}\frac{\langle\psi_n|\exp(-\theta\hat{A}_m)\hat{H}\hat{P}\ \exp(\theta\hat{A}_m)|\psi_n\rangle}{\langle\psi_n|\exp(-\theta\hat{A}_m)\hat{P}\ \exp(\theta\hat{A}_m)|\psi_n\rangle}\bigg|_{\theta=0} = \frac{\langle\psi_n|[(\hat{H}-E)\hat{P},\hat{A}_m]|\psi_n\rangle}{\langle\psi_n|\hat{P}|\psi_n\rangle}. \tag{18}$$

The commutator involves higher rank operators owing to the presence of $\hat{P}$, and $R_m^{(n)}$ requires a four-body transition reduced density matrix between $|\psi_n\rangle$ and $\hat{U}_g|\psi_n\rangle$. However, in practice, they can be handled by the (fermionic-)shift rule without explicitly computing reduced density matrices [44–46] consistently, both in the operator selection and VQE steps.

The operator candidates used in the original fermionic ADAPT algorithm are pairwise in terms of spin because, if only either of the spin-dependent excitations is employed in ADAPT-VQE, the other would become almost indispensable to restore the spin symmetry. However, when $\hat{P}$ is introduced, such a pairwise treatment is no longer necessary. For example, consider the single excitation operators $\hat{\tau}_{q_\alpha}^{p_\alpha}$ and $\hat{\tau}_{q_\beta}^{p_\beta}$. Although each operator excites (and de-excites) an electron from the qth orbital to the pth orbital, the role of $\hat{\tau}_{q_\beta}^{p_\beta}$ is simply to complement the lost spin in $\hat{\tau}_{q_\alpha}^{p_\alpha}$ and vice versa, which can be handled by $\hat{P}$. Based on this consideration, it is advisable to employ completely spin-dependent excitation operators as a pool for the fermionic ADAPT *Ansatz*:

$$\mathcal{P}^{\text{spin}} = \left\{\hat{\tau}_{q_\alpha}^{p_\alpha}, \hat{\tau}_{q_\beta}^{p_\beta}, \hat{\tau}_{r_\alpha s_\alpha}^{p_\alpha q_\alpha}, \hat{\tau}_{r_\beta s_\beta}^{p_\beta q_\beta}, \hat{\tau}_{r_\alpha s_\beta}^{p_\alpha q_\beta}, \hat{\tau}_{r_\beta s_\alpha}^{p_\beta q_\alpha}\right\}. \tag{19}$$

In the presence of $\hat{P}$, the role of $\hat{\tau}_{r_\alpha s_\beta}^{p_\alpha q_\beta}$ becomes less significant if its spin-flipped operator $\hat{\tau}_{r_\beta s_\alpha}^{p_\beta q_\alpha}$ is already treated in the *Ansatz*. Therefore, some spin-flipped operators that primarily produce effects similar to $\hat{P}$ would not be chosen in the algorithm, which will potentially minimize the overall circuit depth. Furthermore, the exponent of each operator in the pool $\mathcal{P}^{\text{spin}}$ can be treated without the Trotter decomposition; therefore, the energy derivatives estimated in the VQE part and Eq. (18) are equivalent for the last chosen operator. This eliminates the algorithmic problem discussed in Sec. II C.

Spin projection is also applicable to qubit-based ADAPT. However, as we have discussed, qubit-ADAPT, which adopts the decomposed Pauli strings as candidate operators, violates the $\hat{S}_z$ symmetry and therefore prevents us from using the simplified $\exp(-i\beta_g\hat{S}_y)$ rotation for spin projection. The full rotation operator $\exp(-i\alpha_g\hat{S}_z)\exp(-i\beta_g\hat{S}_y)\exp(-i\gamma_g\hat{S}_z)$ allows one to project the correct m_s state but entails both α and γ rotation grids, which results in a drastic increase in the number of measurements, by two orders of magnitude. In contrast, in QEB-ADAPT, $\hat{\tilde{\tau}}_{rs}^{pq}$ preserves the $\hat{S}_z$ and $\hat{N}$ symmetries. Because the difference between $\hat{\tilde{\tau}}_{rs}^{pq}$ and $\hat{\tau}_{rs}^{pq}$ is simply that the former neglects the parities between qubits, we can express it as

$$\hat{\tilde{\tau}}_{rs}^{pq} = \hat{\tau}_{rs}^{pq} + \sum_t c_t\hat{\tau}_{rst}^{pqt} + \sum_{tu} c_{tu}\hat{\tau}_{rstu}^{pqtu} + \cdots, \tag{20}$$

where t, u are the qubit indices appearing as a Z string in Eq. (2), that is, $t, u = [q+1, p-1], [s+1, r-1]$. As a concrete example, it can be verified that

$$\hat{\tilde{\tau}}_{10}^{63} = \hat{\tau}_{10}^{63} - 2\hat{\tau}_{104}^{634} - 2\hat{\tau}_{105}^{635} - 4\hat{\tau}_{1054}^{6354}. \tag{21}$$

Hence, spin projection can be easily combined with QEB-ADAPT, and we call such an algorithm SP-QEB.

Finally, let us briefly consider the tapering-off technique for spin projection. As it is known, because chemical Hamiltonians possess number and point-group symmetries, one can identify unitaries that transform a Hamiltonian such that it has only Pauli operators that trivially act on certain qubits, which can thus be discarded [47,48]. Such unitaries are identified using the $\mathbb{Z}_2$ symmetry, namely, the parities of α and β electron numbers for the number symmetry [47], and the sign changes of the underlying wave function by (Abelian) point-group symmetry operations [48]. However, the spin-rotation operator $\exp(-i\beta\hat{S}_y)$ changes the number parity of α and β electrons as follows:

$$\hat{S}_y = \frac{1}{2i}\sum_p (a_{p_\alpha}^\dagger a_{p_\beta} - a_{p_\beta}^\dagger a_{p_\alpha}). \tag{22}$$

Therefore, the $\mathbb{Z}_2$ symmetry can be exploited only for the total number operator $\hat{N} = \hat{N}_\alpha + \hat{N}_\beta$, but not for $\hat{S}_z = \frac{1}{2}(\hat{N}_\alpha - \hat{N}_\beta)$ because $[\hat{S}_y, \hat{S}_z] \neq 0$, resulting in the reduction of only one qubit instead of two. This is a necessary cost for spin projection to exert its advantages; note that the $\hat{S}^2$ symmetry is not categorized as a $\mathbb{Z}_2$ symmetry, and the tapering-off scheme is not applicable. From a different perspective, spin projection can explore a hidden Hilbert space that is not accessible by standard UCC-like *Ansätze* by deliberately breaking and restoring the number symmetry of each electron spin. Although one cannot use number parity to discard the other one qubit from the simulation, there is no such restriction for the point-group symmetry. In some of our calculations discussed below, we taper qubits to ease the computational cost. However, we will assume the number of CNOT gates

N_{CNOT}, which is central to measuring the circuit complexity in this paper, remains the same. In addition, we will not discuss how gate operations would change by the transformation in the tapering-off algorithm.

III. FIRST-ORDER MOLECULAR PROPERTIES WITH ADAPT-VQE

In molecular systems, one is often interested not only in the total energy but also chemical properties. Time-independent molecular properties are defined as the total energy change with respect to a perturbation x introduced to the Hamiltonian. To obtain properties from quantum computing, several studies have focused on evaluating analytical energy derivatives [49,50]. Although higher-order properties such as polarizability require higher-order derivatives and therefore are difficult to compute, first-order properties $dE[x]/dx$ such as force and dipole moment can easily be obtained in VQE. This is because the quantum state is variational with respect to VQE parameters $\{\theta_k\}$ used in the quantum circuit:

$$\frac{\partial E[\boldsymbol{\theta}]}{\partial \theta_k} = 0. \tag{23}$$

However, in both ADAPT-VQE and SP-ADAPT-VQE, the canonical orbitals of HF commonly employed as a starting point are generally not optimal. The fully parameterized wave function is given by $\hat{P}\exp(\hat{\kappa})|\psi_{\mathrm{ADAPT}}[\boldsymbol{\theta}]\rangle$ ($\hat{P}$ is discarded for standard ADAPT-VQE), where

$$\hat{\kappa} = \sum_{p>q} \kappa_{pq} \hat{\tau}_q^p. \tag{24}$$

Hence, the total energy:

$$E[\boldsymbol{\theta}, \boldsymbol{\kappa}, x] = \frac{\langle\psi_{\mathrm{ADAPT}}[\boldsymbol{\theta}]|\exp(-\hat{\kappa})\hat{H}[x]\hat{P}\exp(\hat{\kappa})|\psi_{\mathrm{ADAPT}}[\boldsymbol{\theta}]\rangle}{\langle\psi_{\mathrm{ADAPT}}[\boldsymbol{\theta}]|\exp(-\hat{\kappa})\hat{P}\exp(\hat{\kappa})|\psi_{\mathrm{ADAPT}}[\boldsymbol{\theta}]\rangle}, \tag{25}$$

is not usually fully stationary with respect to orbital change:

$$\left.\frac{\partial E[\boldsymbol{\theta}, \boldsymbol{\kappa}, x]}{\partial \kappa_{pq}}\right|_{\boldsymbol{\kappa}=\mathbf{0}} \neq 0, \tag{26}$$

unless it is fully converged to the FCI state. In other words, the 1-particle generalized Brillouin condition is not satisfied; note that the left side of Eq. (26) is equivalent to the gradient R_m with $\hat{A}_m = \hat{\tau}_q^p$ defined in Eqs. (6) and (18) for standard ADAPT and SP-ADAPT *Ansätze*, respectively (we will denote this as R_{pq} below). To address this issue, we can compute the changes in molecular orbitals induced by x, by solving the coupled-perturbed HF equation [51,52] or the Z-vector equation [53], similar to previous studies [49,50]. In this paper, we adopt an alternative (but mathematically equivalent) formulation commonly used in quantum chemistry [54,55] and rederive the necessary equations to obtain first-order molecular properties in SP-ADAPT.

Accordingly, we introduce the following Lagrangian:

$$\mathcal{L}[\boldsymbol{\theta}, \boldsymbol{\kappa}, \mathbf{z}, x] = E[\boldsymbol{\theta}, \boldsymbol{\kappa}, x] + \sum_{p>q} z_{pq} F_{pq}, \tag{27}$$

where z_{pq} and F_{pq} are Lagrange multipliers to be determined and off-diagonal elements of the canonical Fock matrix, respectively. We note that the canonical HF orbitals require $F_{pq} = 0$ for $p \neq q$ at each x, and therefore, $\mathcal{L}$ is stationary with respect to z_{pq}. Evidently, for any x, $\mathcal{L}[\boldsymbol{\theta}, \boldsymbol{\kappa}, \mathbf{z}, x]$ reproduces the same value as $E[\boldsymbol{\theta}, \boldsymbol{\kappa}, x]$, and therefore, $d\mathcal{L}/dx \equiv dE/dx$; however, $\mathcal{L}$ has an added advantage in that it can be stationary with respect to all variational parameters: $\boldsymbol{\theta}$, $\boldsymbol{\kappa}$, and $\mathbf{z}$ [note that $\partial\mathcal{L}/\partial\theta_k = 0$ is automatically satisfied by VQE, see Eq. (23)]. The stationary condition for $\boldsymbol{\kappa}$ can be achieved by solving the linear equation to determine $\mathbf{z}$:

$$\frac{\partial \mathcal{L}[\boldsymbol{\theta}, \boldsymbol{\kappa}, \mathbf{z}, x]}{\partial \kappa_{rs}} = R_{rs} + \sum_{p>q} z_{pq} A_{pq,rs} = 0, \tag{28}$$

where

$$A_{pq,rs} = \frac{\partial F_{pq}}{\partial \kappa_{rs}} \tag{29}$$

is the Hessian of the HF energy and can easily be computed by a classical computer. We present the explicit form of $\mathbf{A}$ and the solution to Eq. (28) in Appendix C.

Using the chain rule,

$$\begin{aligned}\frac{dE}{dx} &= \frac{d\mathcal{L}}{dx} \\ &= \frac{\partial \mathcal{L}}{\partial x} + \sum_k \left(\frac{\partial \mathcal{L}}{\partial \theta_k}\frac{d\theta_k}{dx}\right) \\ &\quad + \sum_{p>q}\left(\frac{\partial \mathcal{L}}{\partial \kappa_{pq}}\frac{d\kappa_{pq}}{dx} + \frac{\partial \mathcal{L}}{\partial z_{pq}}\frac{dz_{pq}}{dx}\right) \\ &= \frac{\partial \mathcal{L}}{\partial x},\end{aligned} \tag{30}$$

where the last term is straightforward to evaluate once $\mathbf{z}$ is available. It is noteworthy that this approach is applicable to higher-order derivatives in a mathematically clear way, although we will not go into the details.

The Lagrangian $\mathcal{L}$ can be expressed by the so-called relaxed density matrices $\mathbf{D}^{\mathrm{relax}}$, which incorporate the response correction ascribed to the orbital change by perturbation (Appendix C). It can easily be shown that molecular properties as energy derivatives are computed using the relaxed density matrices, instead of the expectation value of the corresponding observable operator, that is, with the regular density matrices $D_{pq} = \langle\psi|a_p^\dagger a_q|\psi\rangle$. The simple expectation value using the latter is generally less accurate because the orbital response is neglected. To extend the orbital response correction to SP-ADAPT, we can resort to the derivation in Ref. [56]. In Appendix C, we summarize the equations and also review the evaluation of nuclear gradients in the case where HF orbitals are a function of nuclear coordinates.

Although this approach can be applied to any VQE method, in several chemistry-inspired *Ansätze* such as the UCCSD, R_{rs} values are zero or negligible because single excitations may be explicitly treated in the variational optimization. However, we note that the frozen-core (or frozen-virtual) approximation is frequently exercised both in classical and quantum computing, where the lower core (higher virtual) orbitals are considered inert and fixed to those of the reference state that is often HF. Because these frozen orbitals are not explicitly optimized within VQE, a response correction is required. This can easily be achieved by simply expanding the range

TABLE I. Summary of ADAPT protocols.

Method	Pool	Symmetry
Fermionic	Eq. (3)	$\hat{S}_z, \hat{N}$
Spin-dependent fermionic	Eq. (19)	$\hat{S}_z, \hat{N}$
Qubit	$\{X_pY_q, X_pX_qX_rY_s\}$	None
QEB	$\{\hat{\tilde{\tau}}_q^p, \hat{\tilde{\tau}}_{rs}^{pq}\}$	$\hat{S}_z, \hat{N}$

of orbital space p, q in Eq. (27) to include the frozen orbitals. In such cases, matrix elements with frozen orbitals, denoted by I, J, are required. Although the expression for $A_{pq,rI}$ remains the same, the gradient R_{rI} needs to be handled in classical computers because I is not mapped to qubits (refer to Appendix C for comprehensive derivations). Certainly, all these problems would be eliminated if orbitals are optimized [57,58]; however, orbital-optimized VQE itself requires these quantities and extra VQE simulations.

As mentioned above, for both fermionic and SP-fermionic ADAPTs, we have access to $R_{rs}^{(n)}$ at each nth cycle by construction. Therefore, the energy derivative and first-order molecular properties are readily available by processing density matrices in a classical computer. Conversely, for the qubit-based ADAPT, R_{rs} are not available in the course of the ADAPT-VQE protocol, and therefore, one needs to explicitly evaluate these quantities to properly estimate molecular properties, unless the obtained quantum state is an exact eigenstate of the Hamiltonian where $R_{rs} = 0$.

IV. ILLUSTRATIVE CALCULATIONS

A. Computational details

Before discussing our numerical results, we describe the computational details. All simulations were conducted using Quantum Unified Kernel for Emulation (QUKET), developed by us, without the effect of noise [59]. QUKET compiles several open-source libraries to generate a Hamiltonian mapped to a qubit basis and perform quantum simulations with parallel computing. Specifically, it employs PYSCF [60] to generate molecular orbitals and integrals, and OPENFERMION [61] to perform the Jordan-Wigner transformation. To simulate quantum circuits, we utilize QULACS [62]. The energy minimization in VQE uses the Broyden-Fletcher-Goldfarb-Shannon (BFGS) algorithm implemented in Scipy. The geometry optimization was performed by interfacing QUKET with PYBERNY [63].

For fermionic ADAPT, we adopt two different pools, Eqs. (3) and (19), and the controlled-Ry circuit to perform $\exp(\theta\hat{\tau}_{rs}^{pq})$ and $\exp(\theta\hat{\tau}_q^p)$ and estimate the number of CNOT gates N_{CNOT} in Fig. 1(c). We will not consider the effect of noise in this paper but simply evaluate the expected performance.

We summarize the ADAPT protocols used in this paper and the corresponding operator pools in Table I. For spin projection, we always use the pool of either the spin-dependent fermionic excitations or qubit excitations.

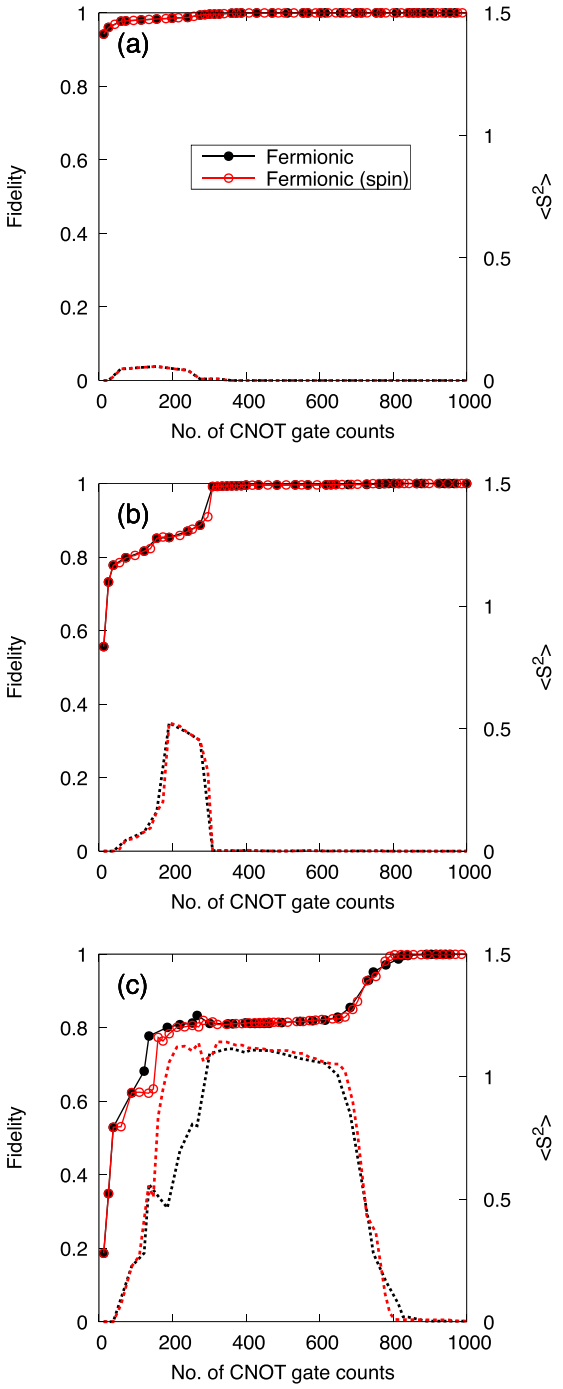

FIG. 2. Fidelity (solid lines) and $\langle\hat{S}^2\rangle$ (dotted lines) of ADAPT-VQE, plotted against N_{CNOT} for N_2 with different bond lengths: (a) 1.098 Å, (b) 1.8 Å, and (c) 2.5 Å. Circles indicate each ADAPT step.

B. Spontaneous symmetry breaking in ADAPT

First, we examine the spontaneous symmetry breaking and its consequence in ADAPT for the N_2 molecule with the STO-3G basis set. We have considered three different bond lengths, $R_{N-N} = 1.098$, 1.8, and 2.5 Å to represent weakly and strongly correlated cases. The solid lines in Fig. 2 depict the fidelity of the fermionic ADAPT-VQE state to the exact FCI state $|\langle\Psi_{\mathrm{FCI}}|\psi_{\mathrm{ADAPT}}\rangle|^2$ as a function of N_{CNOT}, using

either $\mathcal{P}$ or $\mathcal{P}^{\text{spin}}$, labeled as fermionic and fermionic (spin), respectively. In all these cases, ADAPT-VQE approaches the exact state as more operators are added, as expected. The figure indicates that there is no significant difference in convergence between $\mathcal{P}$ and $\mathcal{P}^{\text{spin}}$, although the former, the original algorithm proposed in Ref. [13], is more advantageous in that it requires less VQE parameters, as indicated by the circles. However, the convergence behavior is significantly different depending on the system. In other words, the stronger the correlation of a system, the slower the convergence of ADAPT-VQE to FCI.

This behavior is attributed to the broken spin characterized by $\langle \hat{S}^2 \rangle$, which is also illustrated in Fig. 2 with the dotted lines. In the strongly correlated case of $R_{N-N} = 2.5$ Å [Fig. 2(c)], the HF state exhibits a very small overlap with FCI, with a fidelity of 0.2, which quickly increases to 0.8 after the initial few ADAPT steps with ∼200 CNOTs. However, simultaneously, $\langle \hat{S}^2 \rangle$ (simply the degree of the spin-symmetry breaking) becomes >1, and at this point, the fidelity convergence gets trapped in a plateau.

Such large spin-symmetry breaking occurs initially because, in the VQE step of ADAPT, large VQE parameters are essential for describing strong correlations (i.e., highly excited configurations from the HF reference). Consequently, the α and β electrons are treated quite differently, especially in the absence of the Trotter decomposition [27]. We should mention that the spin symmetry is safely conserved if paired double excitations such as $\hat{\tau}_{q_\alpha q_\beta}^{p_\alpha p_\beta}$ are chosen. In fact, the first few operators selected in ADAPT are of this type, and $\langle \hat{S}^2 \rangle$ remains zero (see Fig. 2). However, paired double excitations do not have a sufficient ability to describe all types of correlation effects, and spin-unpaired excitations are more suitable in several cases. With $\langle \hat{S}^2 \rangle > 1$, the ADAPT state is a mixture of several different spin states, each with a significant weight. Therefore, the derivative approach expressed in Eq. (6) to determine the operator candidate is likely inappropriate because such an operator search is performed in an incorrect symmetry space. Of course, by further processing ADAPT with more operators and CNOT gates, the lost spin symmetry starts to be restored [∼700 CNOTs for Fig. 2(c)]. At this point, the fidelity also quickly increases to one. Similar but less pronounced results can be confirmed for the tests with shorter bond lengths.

Although fermionic ADAPT follows the $\hat{S}_z$ and $\hat{N}$ symmetries, qubit-ADAPT breaks all, and as a result, exhibits a significant slowdown of convergence to FCI. This is illustrated by Fig. 3, where we have plotted the fidelity and symmetry expectation values $\langle \hat{S}^2 \rangle$, $\langle \hat{S}_z \rangle$, and $\langle \hat{N} \rangle$ using the problematic case of N_2 at $R_{N-N} = 2.5$ Å. We note that, in this particular case, both algorithms coincidentally result in similar CNOT gate counts to achieve the FCI state. However, the number of VQE parameters for fermionic ADAPT to achieve a fidelity of 0.99 is 30, while that for qubit-ADAPT is 128, implying that more measurements are required for qubit-ADAPT. Interestingly, the symmetry breaking of qubit-ADAPT in $\hat{S}_z$ and $\hat{N}$ is rather moderate; however, the error in $\langle \hat{S}^2 \rangle$ is substantially larger than that of fermionic ADAPT. Apparently, such a large error in $\langle \hat{S}^2 \rangle$ severely affects the fidelity, and it is evident that the plateaus that emerge in the fidelity and $\langle \hat{S}^2 \rangle$ plots correspond to each other.

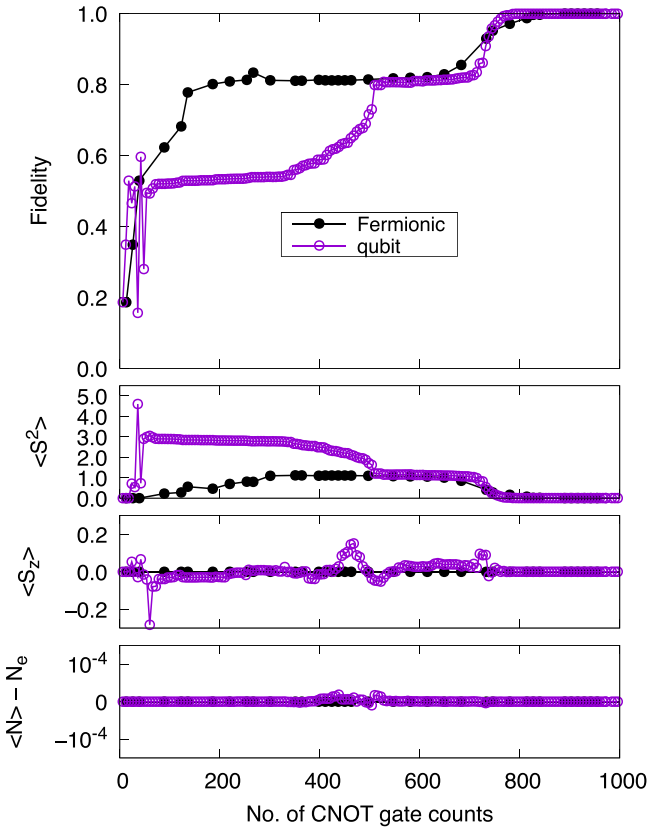

FIG. 3. Comparison between fermionic ADAPT and qubit-ADAPT for N_2 at $R_{N-N} = 2.5$ Å. Circles indicate each ADAPT step.

C. Accuracy and gate efficiency in ADAPT and SP-ADAPT

As demonstrated above, symmetry breaking is associated with strong correlation and occurs spontaneously in the ADAPT algorithms, thereby slowing down the convergence. Therefore, one would naturally expect that maintaining the correct symmetry can mitigate the problem. In this section, we study the effect of spin projection using several systems, which include the N_2 molecule (with $R_{N-N} = 1.098, 2.5$ Å), linear chain of H_6 equally separated by 2 Å, and half-filled one-dimensional periodic Fermi-Hubbard model with six sites and $U = 8$. For the molecular systems, we have used the STO-3G basis set and six electrons in six orbitals. Therefore, all the models considered here are composed of 12 qubits.

We choose the initial state to be the HF determinant for the molecular systems; however, for the Hubbard model, we intentionally set the charge localized state where the first three sites are doubly occupied and the remaining three sites are empty to break the spatial symmetry.

Figure 4 presents the energy error from FCI and the number of CNOT gates in the quantum circuit for several methods, where the shaded area in orange corresponds to chemical accuracy (<1 mHartree error). We first consider the performance of qubit-ADAPT [15]. For the molecular systems [Figs. 4(a)–4(c)], no significant advantage is found

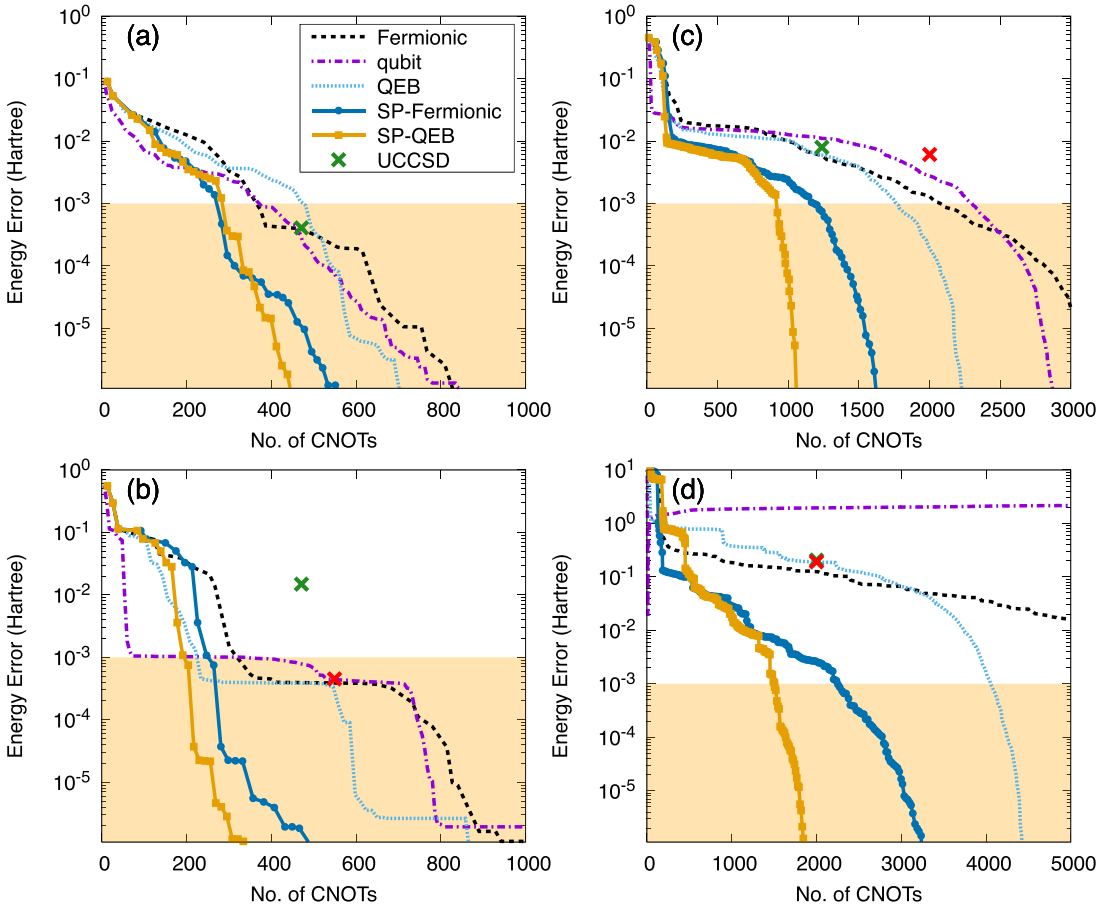

FIG. 4. Energy error from the full configuration interaction (FCI; in Hartree) at each CNOT gate count. (a) N_2 at equilibrium geometry ($R_{N-N} = 1.098$ Å), (b) stretched N_2 ($R_{N-N} = 2.5$ Å), (c) stretched H_6 ($R_{H-H} = 2$ Å), (d) half-filled 6-site Hubbard chain ($U = 8$). The green and red plots indicate the standard unitary coupled cluster with single and double substitutions (UCCSD) and broken-symmetry UCCSD, respectively.

over fermionic ADAPT in terms of CNOT gate counts. This is because the current implementation of fermionic ADAPT relies on the controlled-Ry circuit of Ref. [19], which can eliminate a large number of CNOT gates. In contrast, we find that qubit-ADAPT is disadvantageous in two aspects. First, the energy lowering in each VQE step is considerably small, as plotted in Fig. 5, and it requires several VQE parameters (measurements), which also makes it difficult to converge each VQE step because of a highly nonlinear parameter space. Second, it breaks the electron number symmetry; the application of qubit-ADAPT to ionic cases is not straightforward, as illustrated by the example of the half-filled Hubbard model, where qubit-ADAPT falls into the true ground state of $\langle \hat{N} \rangle = 4$ and $\langle \hat{S}^2 \rangle = 2$.

Now let us focus on the results obtained by spin projection. Evidently, by eliminating the incorrect spin-symmetry components, SP-ADAPT methods (both for fermionic and QEB) require much less CNOT gates to achieve the same accuracy as those without spin-projection. SP-fermionic ADAPT is already effective for the weakly correlated N_2 at equilibrium geometry, as illustrated in Fig. 4(a). The CNOT gate reduction in this system with spin projection (by a factor of 0.6–0.7) is almost solely attributed to the fact that spin-complement excitations (i.e., those that are missing to make a certain excitation spin adapted) are handled by the superposition of spin-rotated states; in contrast, with fermionic ADAPT, one needs to explicitly construct a quantum circuit to perform spin-complement excitations, except for paired doubles. This can be verified by plotting the energy error against the number of VQE parameters (Fig. 5). From Fig. 5(a), the number of VQE parameters required to achieve the same accuracy is quite similar between fermionic and SP-fermionic ADAPT algorithms, which implies the two algorithms yield similar quantum states for each ADAPT iteration.

The efficacy of spin projection becomes more distinct for strongly correlated systems. For the stretched N_2 and H_6 [Figs. 4(b) and 4(c)], the number of CNOT gates required to reach the FCI ground state with SP-ADAPT is less than half of that with the corresponding broken-symmetry ADAPT. Having said that, spin projection may not seem to be advantageous because it employs up to 200 CNOT gates. For instance, QEB-ADAPT exhibits a lower energy than SP-QEB-ADAPT in the stretched N_2 case, and the convergence of qubit-ADAPT is initially even faster. However, this is an artifact due to the unphysical spin contamination effect, and the quality of the quantum state is far from satisfactory. A large spin

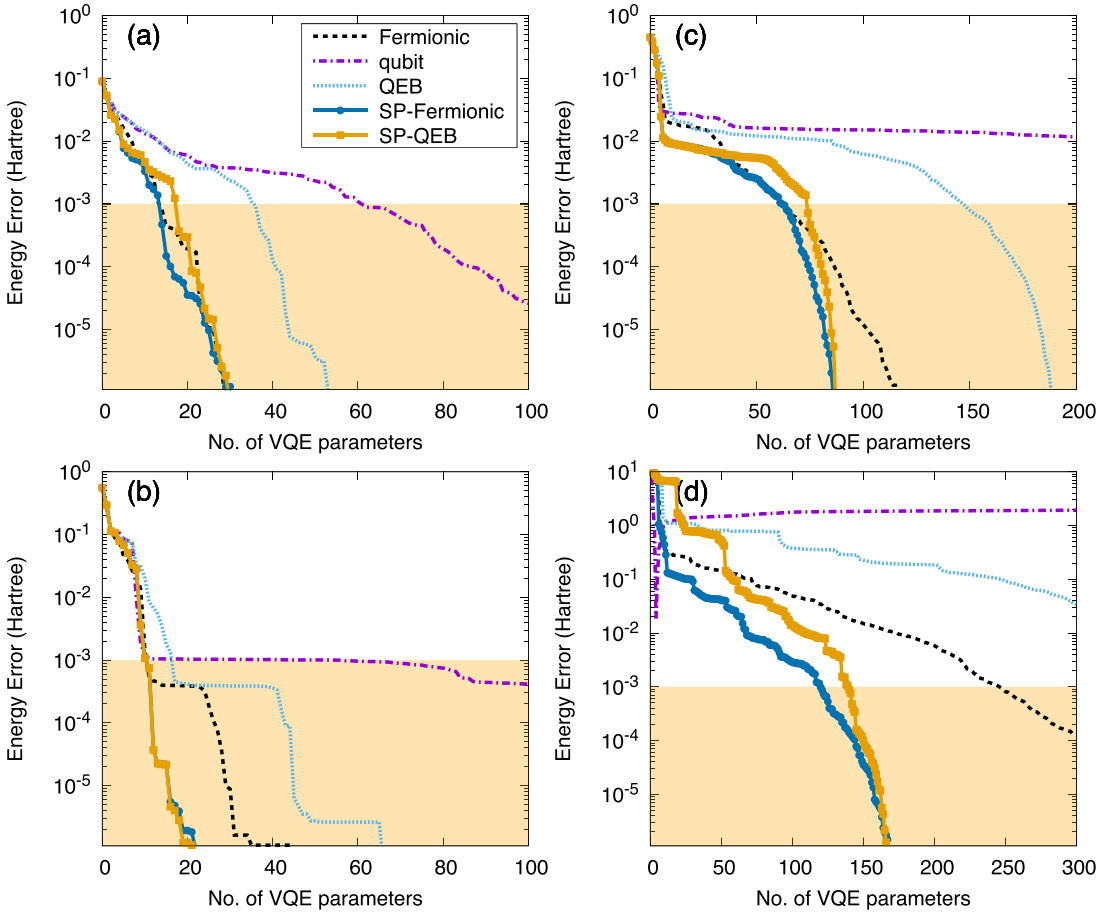

FIG. 5. Same as Fig. 4 but as a function of the number of parameters.

contamination also leads to incorrect properties and state assignment. Furthermore, as seen in Fig. 3, again, it often causes the plateau.

The number of required parameters in SP-ADAPT (Fig. 5) is less than half of standard ADAPT, as expected. As illustrated in Fig. 6, the scaling of CNOT gate counts with the number of used parameters is essentially identical between ADAPT and SP-ADAPT when the same operator pool is used. Therefore, a lower number of parameters in the latter simply implies a more gate-efficient circuit at each energy accuracy. SP-fermionic ADAPT converges slightly faster than SP-QEB-ADAPT; this may simply be because the former contains more operators in the pool. However, it is interesting to observe that both the SP-fermionic and SP-QEB algorithms require almost the same number of parameters to represent the exact FCI. This suggests we can further reduce the number of operators to form a pool, which can reduce the measurements for derivative evaluations, while expecting an unchanged fine convergence behavior.

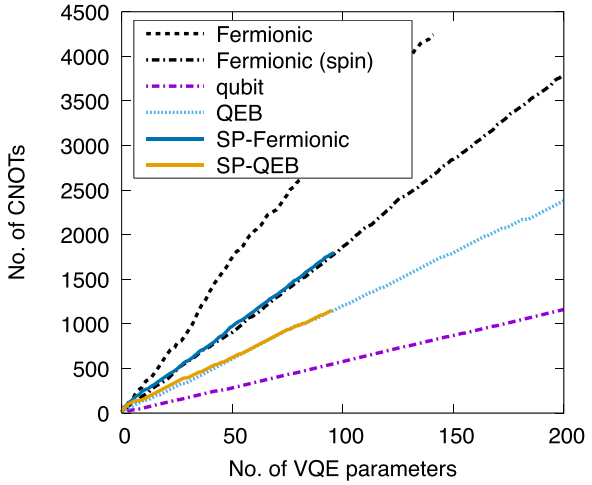

FIG. 6. Increase in the number of CNOT gates with the number of parameters for H_6.

Finally, we also performed the UCCSD to evaluate the advantages of ADAPT and SP-ADAPT. In our UCCSD implementation, only the gates with nonzero parameters are constructed to minimize the gate depth. Because the molecular systems we consider have a high spatial symmetry, most UCCSD excitations are symmetry forbidden, thereby permitting gate-efficient circuits. With such a treatment, the UCCSD is already as powerful as ADAPT for N_2 at equilibrium and H_6 [see the green plots in Figs. 4(a) and 4(c)]. For the stretched N_2, ADAPT appears to be significantly more efficient and accurate; however, this may well be attributed to the spin-symmetry breaking. For a fair comparison, we also present the broken-symmetry UCCSD [27] in red, which provides a fully variational solution at the cost of large spin

contamination. In addition, note that the full UCCSD, which performs symmetry-forbidden gate operations, that is, even if the amplitudes are zero, can be prepared with 2001 CNOT gates for all cases. For the Hubbard model, the UCCSD naturally requires all excitations, and thus, 2001 CNOT gates, as we have initiated the calculation with no spatial symmetry; the performance of the UCCSD is similar to QEB-ADAPT. In conclusion, although ADAPT is imminently more flexible than the fixed *Ansatz* of the UCCSD, fermionic ADAPT passes through a UCCSD-like state in the process of converging to the FCI state; this is quite logical because the UCCSD is usually considered an accurate method that is both chemically and theoretically well established.

D. Comparison between different operator pools

The previous section suggested that not only is the number of CNOT gates required to obtain the FCI state constantly smaller for QEB-ADAPT than for fermionic ADAPT, but also the number of required variational parameters is very similar with spin-projection. We also performed calculations with spin-dependent fermionic ADAPT using $\mathcal{P}^{\text{spin}}$ defined in Eq. (19) and arrived at the same conclusion for broken-symmetry standard ADAPT algorithms.

The authors of Ref. [15] verified that only $2n-2$ Pauli operators can span the n-qubit real Fock space. Although this theorem appears very promising, much more operators are usually required in the ADAPT algorithm because it chooses only one operator at a time; thus, the pool has to be physical in some sense (i.e., it should contain fermionic excitations higher than singles to examine the k-particle generalized Brillouin theorem). Nevertheless, the theorem suggests the number of operators in the QEB pool can be further reduced, thereby providing an opportunity to save the computational efforts to choose the next operator. Here, we propose the following three additional qubit-excitation pools and compare their performances:

(1) Scheme 1. $\hat{\tilde{\tau}}^{p_\alpha}_{q_\alpha}$, $\hat{\tilde{\tau}}^{p_\beta q_\alpha}_{r_\beta s_\alpha}$ (β singles and doubles with the same spins are omitted).

(2) Scheme 2. Same as Scheme 1 but with $p \geqslant q$.

(3) Scheme 3. Same as Scheme 2 but with $r \geqslant s$.

For the 12-qubit systems, the number of operators in each pool is 855, 555, 400, 190, and 155 for spin-dependent fermionic, QEB, and schemes 1–3, respectively. We have plotted the energy convergence of the earlier H_6 example against CNOT gates in Fig. 7 and parameter numbers in Fig. 8. From these figures, it can be observed that, at convergence (an energy error of 10^{-6}), all QEB schemes perform similarly in terms of both the numbers of CNOT gates and VQE parameters. However, each plot behaves slightly differently. For broken-symmetry ADAPT algorithms (indicated by the lines without points), full QEB provides a slightly better description than Scheme 3 before the convergence. This could be because the entire operator pool in QEB contains more choices than Scheme 3. In particular, the latter does not contain spin-complement operators, which are subsequently chosen in several instances because spin-complement excitations result in similar energy gradients. It is also clear from Fig. 8 that spin-dependent fermionic ADAPT generally exhibits a better energy convergence for a given number of VQE parameters,

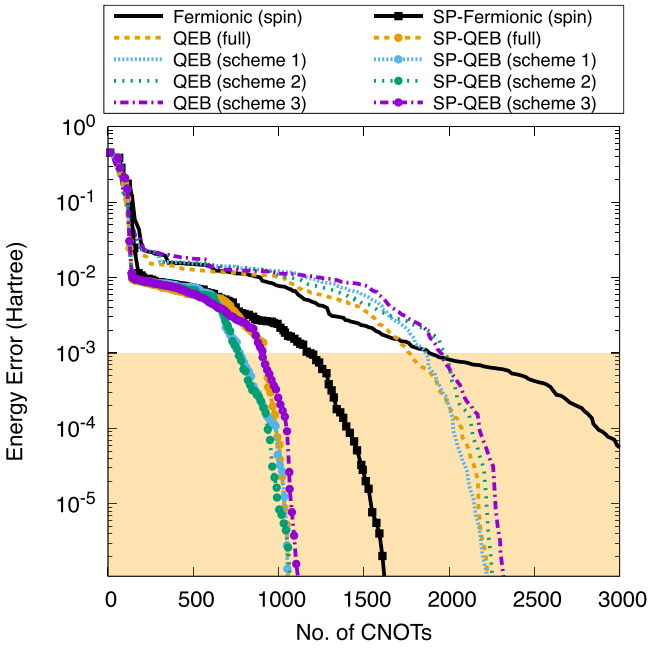

FIG. 7. Comparison of the CNOT performance between different operator pools in qubit-excitation-based (QEB) and spin-projected QEB (SP-QEB).

although it is inefficient in terms of CNOT gate counts. For example, with 100 parameters, the energy obtained by fermionic ADAPT is one order of magnitude more accurate than those by QEB schemes.

With spin projection, the performances of different QEB schemes are almost equivalent with a certain CNOT gate count and parameter number. In particular, its accuracy is comparable with SP-fermionic ADAPT when the number

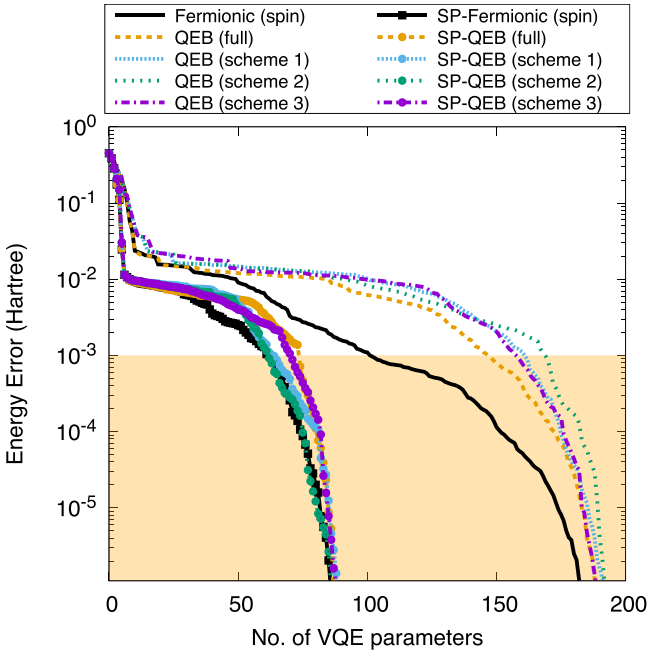

FIG. 8. Same as Fig. 7 but as a function of the number of parameters.

of parameters is fixed because spin projection can partially replicate the spin-complement effects. However, SP-fermionic ADAPT is almost always more accurate than SP-QEB-ADAPT, although slightly. In the particular case of H_6, their differences are marginal and may be rather attributed to the fact that the current approach to determining the next operator is not optimal; operators that are not chosen could identify a lower energy than the chosen one when VQE is performed. However, this behavior is general, as is clearly observed in Fig. 5, and can seemingly become more significant for larger systems, as will be demonstrated later. Nevertheless, the improved similarity between the SP-fermionic and SP-QEB algorithms over standard ADAPT is encouraging for practical applications where the actual operator number (or CNOT gate counts) that can be implemented is assumed to be limited.

E. Molecular properties

In this section, we study the molecular properties computed by several ADAPT-based algorithms using the method discussed in Sec. III. In a practical application of the ADAPT algorithm, we cannot expect to obtain FCI accuracy but must make a tradeoff between accuracy and either the circuit depth or the number of measurements. To benchmark their accuracy, we compare the results obtained with the fixed number of VQE parameters N_{param} or the number of CNOT gates N_{CNOT}.

1. *Dipole moment of* H_2O

The dipole moment is an important first-order property of a molecule, providing a measure of polarity of the system. To assess the accuracy of ADAPT in computing this quantity, we have employed the H_2O molecule with a fixed angle of 104.5° and varied the OH bond length symmetrically. Figure 9 summarizes the dipole moment error from the FCI value calculated at different ADAPT cycles ($N_{\text{param}} = 6$, 9, and 12). The results of spin-dependent fermionic ADAPT without and with the orbital response correction are presented in Figs. 9(a) and 9(b), respectively. As can be observed from the figures, the orbital response correction is important for describing the dipole moment correctly at short bond lengths (∼1.5 Å). Note that the correction effect is almost absent in the UCCSD because it takes care of single excitations explicitly in VQE by playing the role of orbital relaxation. For the larger bond distance, the orbital response correction sometimes worsens the result; however, this means the ADAPT state is not stationary with respect to orbital rotation, and the good dipole moments obtained in the unrelaxed results [Fig. 9(a)] are fortuitous. Having said that, because we are using the HF canonical orbitals in ADAPT, the orbital Hessian matrix $\mathbf{A}$ in Eq. (28) can be nearly singular at a certain bond length (∼1.9 Å), where the first-order relaxed density matrix of ADAPT is overcorrected. This singular behavior is circumvented by increasing the number of operators such that the residual R_{rs} is close enough to zero.

When spin projection is applied, we essentially observe a similar trend [see Figs. 9(c) and 9(d)]. Although spin-symmetry breaking in ADAPT does not affect the estimated dipole moment because both the singlet and contaminated triplet states give almost zero dipole moments at stretched bond lengths, SP-fermionic ADAPT shows promising results,

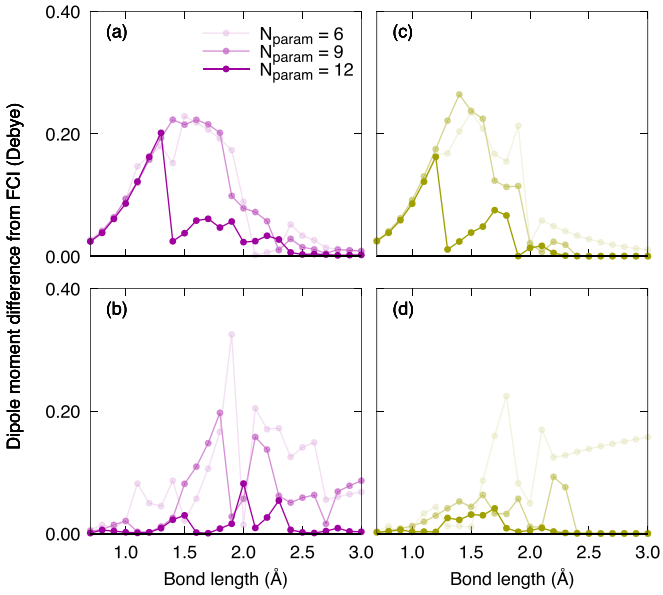

FIG. 9. Error from the full configuration interaction (FCI) in the calculated dipole moment of H_2O by altering the number of operators in ADAPT. (a) Spin-dependent fermionic ADAPT without orbital response, (b) spin-dependent fermionic ADAPT with orbital response, (c) SP-fermionic ADAPT without orbital response, and (d) SP-fermionic ADAPT with orbital response.

yielding better convergence with respect to N_{param}, especially in a more strongly correlated region [Fig. 9(d)].

2. *Geometry and potential curve of* O_3

It is difficult to investigate the accuracy in estimating geometrical parameters of strongly correlated systems, as we are restricted to small molecules or toy models because of a limited computational budget, and most small molecules are only weakly correlated at equilibrium. In fact, previous studies have considered such simple molecules [49,50,64], which do not allow one to properly evaluate the potential of quantum computers. Another issue is that, for adaptive algorithms like ADAPT, the potential energy surface is not smooth, and determining the minimum can be challenging. However, it is interesting to benchmark how effective the geometry optimization in ADAPT methods can be. We think it is also pedagogical to compare the results of spin-dependent fermionic and QEB ADAPT algorithms in predicting potential energy surfaces when the same quantum resource is available.

Accordingly, we choose the ozone molecule with the STO-3G basis as our test case, using an active space comprising 9 orbitals and 12 electrons, resulting in an 18-qubit system. To make a direct comparison between the accuracy of each ADAPT method, we have assumed that the number of CNOT gates that one can handle for each calculation is limited (except for the UCCSD, where 2534 CNOT gates were required). Namely, all ADAPT results were obtained with less than a certain N_{CNOT} ranging from 300 to 1000. In all simulations, we tapered qubits to reduce the computational cost but assumed the total number of CNOT gates remain unchanged.

Figures 10(a) and 10(b) depict the optimized bond length (Å) and angle (°), where the FCI result is depicted by the red

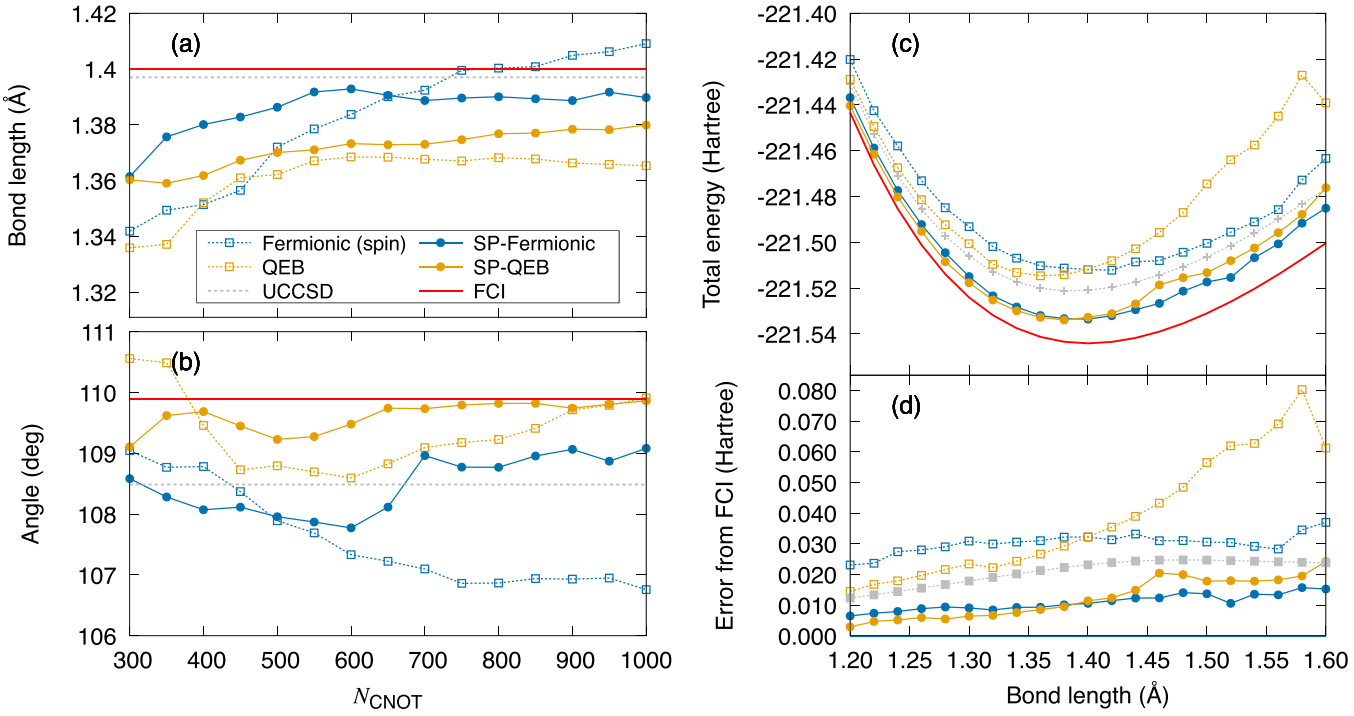

FIG. 10. (a) Optimized bond length of O_3 as a function of the maximum number of CNOTs used in ADAPT. (b) Same as (a) but for the angle. (c) Potential energy curve of O_3 in Hartree, computed with $N_{CNOT} = 1000$. (d) Energy error from the full configuration interaction (FCI).

line ($R_{O-O} = 1.400$ Å and $\angle OOO = 109.9°$). Because O_3 is a two-determinant system, the UCCSD is reasonably accurate, as indicated by the dotted gray line. Thus, symmetry breaking in ADAPT is not so significant for both the fermionic and QEB states. Our calculations indicate that $\langle \hat{S}^2 \rangle$ in ADAPT is ∼0.1 at most, at the equilibrium geometry estimated by FCI. From the figures, the bond length is more accurately predicted by fermionic ADAPT than by QEB ADAPT, with or without spin projection, while the results for the angle exhibit the opposite trend. Without spin projection, the two methods yielded significantly different results for geometry optimization, especially at a larger N_{CNOT}. Increasing N_{CNOT} does not always lead to more accurate results, at least up to $N_{CNOT} = 1000$. However, by applying spin projection, it appears the geometries predicted by the SP-fermionic and SP-QEB algorithms are improved with N_{CNOT}. They are also similar to each other, implying the two methods also result in similar states.

To further investigate these results, we have plotted in Fig. 10(c) the potential energy curve of symmetric bond dissociation of O_3 with a fixed angle of 109.9°, using $N_{CNOT} = 1000$. In Fig. 10(d), we have also presented the energy error from FCI in Hartree. In all the methods, to a certain degree, the energy is more accurate at a shorter (weakly correlated) distance and becomes worse as the bond is stretched (strongly correlated). While this result is quite reasonable, we found that the change in QEB energy is rather significant compared with that in the fermionic methods, especially without spin projection. When comparing the fermionic and QEB results, their energy accuracy is inverted at 1.40 Å. We took the nonparallelity error (NPE) of the potential energy curve, which is defined as the difference between the maximum and minimum errors from FCI throughout the potential curve, and tabulated it in Table II. As this table shows, fermionic ADAPT provides a more parallel curve to that of FCI than QEB-ADAPT. This explains why the latter is less accurate than the former in predicting the bond length, as illustrated in Fig. 10(a).

At this point, we are uncertain whether this failure of QEB can be attributed to its fundamental deficiency or the possibly inappropriate choice of operators in the ADAPT algorithm. Because the deviation between fermionic and QEB-ADAPT results is significantly mitigated by spin projection [Figs. 10(c) and 10(d)], we suspect symmetry breaking to be behind the different behaviors. Therefore, in Fig. 11, we have summarized the $\langle \hat{S}^2 \rangle$ values obtained from the calculations in Fig. 10(c). In fact, we find that, although fermionic ADAPT retains the same degree of symmetry breaking throughout all bond distances, QEB is more spin contaminated when the molecule is stretched. As we have seen several times, symmetry breaking can slow down the convergence in ADAPT; therefore, the degree of spin contamination is directly related to the energy accuracy with a fixed N_{CNOT}. Hence, spin projection plays an important role in equalizing the two schemes to some extent.

V. CONCLUSIONS

One of the unsolved problems in quantum chemistry is a consensus way of treating strong correlations. Quantum

TABLE II. NPE in Hartree.

Fermionic	QEB	SP-fermionic	SP-QEB	UCCSD
0.014	0.066	0.009	0.021	0.012

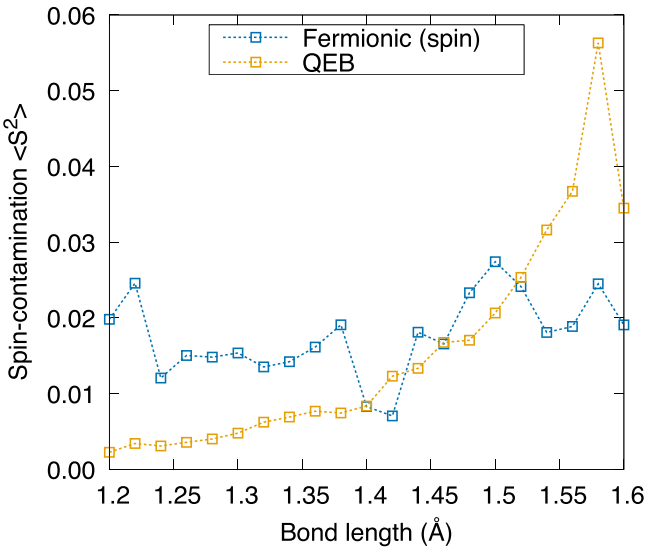

FIG. 11. Spin contamination of fermionic and QEB-ADAPT in O_3.

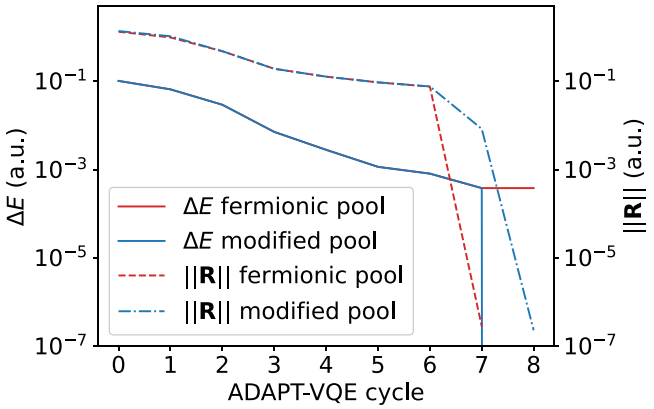

FIG. 12. Convergence profile of fermionic ADAPT-VQE using single and double excitation operators (fermionic pool). The modified pool additionally contains $\hat{\tau}^{\pi_{uy,\alpha}\pi_{ux,\alpha}\pi_{ux,\beta}}_{\pi_{uy,\alpha}\pi_{gx,\alpha}\pi_{gx,\beta}}$.

computing is expected to provide a solution to this problem because it can, in principle, handle entangled quantum states directly mapped onto qubits. The recently proposed ADAPT algorithm has paved the way to construct an efficient quantum circuit adaptively; however, for fermionic and QEB-ADAPT pools, its convergence is slow for strongly correlated systems because of significant symmetry breaking.

To address this issue, we introduced spin projection to ADAPT. We demonstrated that spin projection can be quite effective when combined with ADAPT, offering shallower circuits with fewer CNOT gates and VQE parameters to achieve the same accuracy at the cost of increased measurements. This is especially the case if a gate-efficient circuit is adopted for fermionic and qubit excitations, whose performances are often comparable. Our calculations also indicated that the pool of QEB-ADAPT can be reduced by discarding certain classes of spin-dependent qubit excitations. However, we concluded that it is not worthwhile to apply symmetry projection to qubit-ADAPT because it would require the restoration of lost $\hat{S}_z$ (and number) symmetry, which considerably increases the number of measurements in evaluating the energy.

We have also derived the first-order energy derivative in the presence of spin projection, which enabled the calculation of dipole moment and geometry optimization with SP-ADAPT. It was demonstrated that the orbital response correction is important both for ADAPT and SP-ADAPT in the calculation of the dipole moment because they are far from stationary with respect to orbital changes, unlike the UCCSD, which is less sensitive to orbital rotation because of the presence of explicit single excitations. Furthermore, the method was applied to the geometry optimization of O_3 to quantify the capabilities of fermionic and QEB schemes. We found that QEB-ADAPT is less stable in achieving a constant accuracy throughout the potential energy surface and thus is less predictable than fermionic ADAPT in terms of optimized geometry. The deviation between the two schemes was caused by a larger spin contamination in QEB-ADAPT and could be largely mitigated by performing spin projection. However, the reason behind this unfavorable behavior of QEB-ADAPT remains unclear, and further studies will be required to elucidate its cause.

Note added. After this paper was submitted, a relevant preprint concerning spin symmetry breaking in ADAPT-VQE appeared [65].

ACKNOWLEDGMENTS

This paper was supported by JST, PRESTO Grant No. JPMJPR2016, Japan and also by JSPS KAKENHI, Grants No. JP20K15231, No. JP18H03900, and No. JP22H00316. We are grateful for the computational resources provided by ECCSE, Kobe University.

APPENDIX A: NUMERICAL EXAMPLE OF THE *k*-PARTICLE GENERALIZED BRILLOUIN CONDITION TEST WITH FERMIONIC ADAPT-VQE

As mentioned in the main text, fermionic ADAPT-VQE that uses only single and double excitation operators as a pool satisfies the 1- and 2-particle generalized Brillouin conditions but not necessarily higher rank conditions. If this happens, then it means ADAPT-VQE is trapped in a local minimum. To see this, let us consider the N_2 molecule with π_{ux}, π_{uy}, π_{gx}, and π_{gy}. In the initial HF state, the first two orbitals are doubly occupied, and the rest are vacant. Using the same equilibrium geometry and the basis set (STO-3G) as the main text, the exact energy is -107.59850562 Hartree. In contrast, we obtained -107.5981221 Hartree with ADAPT-VQE, which is 0.4 mHartree higher than the exact energy. The convergence profile is plotted in Fig. 12. We note the generalized Brillouin conditions are fulfilled for $k = 1, 2$ ($||\mathbf{R}|| < 10^{-7}$) at the seventh cycle. We tested the 3-particle condition against this converged $|\psi\rangle$ and observed that this state fails to satisfy the condition for some triple excitations; the largest nonzero gradient is >0.004. These excitations are not purely triple excitations, meaning that some of the creation operators have the same labels as the annihilation operators, e.g., $\hat{\tau}^{\pi_{uy,\alpha}\pi_{ux,\alpha}\pi_{ux,\beta}}_{\pi_{uy,\alpha}\pi_{gx,\alpha}\pi_{gx,\beta}}$. Here, the excitation is a controlled double excitation because it performs an excitation between $|\pi_{ux,\alpha}\pi_{ux,\beta}\rangle$ and $|\pi_{gx,\alpha}\pi_{gx,\beta}\rangle$ if $|\pi_{uy,\alpha}\rangle = |1\rangle$. If we add this excitation in the original pool

(called modified pool), ADAPT-VQE now converges to the exact FCI energy at the eighth cycle, fulfilling all the conditions including the 4-particle Brillouin condition.

We should stress that Eq. (1) is rigorous with single and double excitations in principle [32]. However, this example critically illustrates the insufficiency of a pool with only single and double excitations in the ADAPT-VQE algorithm.

APPENDIX B: GAUSS-LEGENDRE QUADRATURE FOR NUMERICAL INTEGRATION

When evaluating the projected overlap (and Hamiltonian coupling) in Eq. (16), we assume the quantum state $|\psi\rangle$ is an eigenstate of $\hat{S}_z$ with an eigenvalue of m_s, resulting in

$$\begin{aligned}\langle\psi|\hat{P}|\psi\rangle &= \frac{2s+1}{8\pi^2}\int_0^{2\pi} d\alpha \int_0^{\pi} \sin\beta d\beta \int_0^{2\pi} d\gamma D^{s*}_{m_s m_s}(\Omega) \\ &\quad\times \langle\psi|\exp(-i\alpha\hat{S}_z)\exp(-i\beta\hat{S}_y)\exp(-i\gamma\hat{S}_z)|\psi\rangle \\ &= \frac{2s+1}{2}\int_0^{\pi} \sin\beta\, d^{s*}_{m_s m_s}(\beta)\langle\psi|\exp(-i\beta\hat{S}_y)|\psi\rangle d\beta,\end{aligned} \tag{B1}$$

where we have used

$$\begin{aligned}D^s_{m_s m_s} &= \langle s; m_s|\hat{U}_g|s; m_s\rangle \\ &= \exp(-i\alpha m_s) d^s_{m_s m_s}(\beta)\ \exp(-i\gamma m_s),\end{aligned} \tag{B2}$$

and $\exp(-i\gamma\hat{S}_z)|\psi\rangle = \exp(-i\gamma m_s)|\psi\rangle$, etc. Here, $d^s_{m_s m_s}(\beta)$ is Wigner's small d function.

To comply with the integration range $[-1, 1]$ for Gauss-Legendre quadrature, we set

$$x = -\cos(\beta), \tag{B3}$$

which yields

$$\langle\psi|\hat{P}|\psi\rangle = \frac{2s+1}{2}\int_{-1}^{1} f(x)dx, \tag{B4}$$

$$f(x) = \left\{d^s_{m_s m_s}[\arccos(x)]\right\}\langle\psi|\exp[i\arccos(x)\hat{S}_y]|\psi\rangle, \tag{B5}$$

and then obtain the weights and quadratures.

APPENDIX C: FIRST-ORDER ENERGY DERIVATIVE OF VQE

1. General strategy

Here, we discuss the most general formulation for the first-order derivative of energy to obtain the relaxed density matrix in variational methods. Although the derivation is known, we believe it is still beneficial to rederive and summarize the equations for nonexperts. As shown below, we adopt spin molecular orbitals instead of spatial molecular orbitals to ensure that our derivation is as general as possible. With the canonical HF orbitals, the Fock matrix:

$$F_{pq} = h_{pq} + \sum_{rs}\langle pr||qs\rangle D^{\mathrm{HF}}_{rs}, \tag{C1}$$

is diagonal, where we define the idempotent HF density matrix:

$$D^{\mathrm{HF}}_{rs} = \langle\Phi^{\mathrm{HF}}|a^\dagger_r a_s|\Phi^{\mathrm{HF}}\rangle = \begin{cases}1 & (r = s \in occ)\\ 0 & (r \neq s)\end{cases}, \tag{C2}$$

and the one-and two-body integrals of the Hamiltonian:

$$h_{pq} = \int d\mathbf{r}\phi^*_p(\mathbf{r})\left(-\frac{1}{2}\nabla^2 - \sum_A \frac{Z_A}{|\mathbf{r}-\mathbf{r}_A|}\right)\phi_q(\mathbf{r}), \tag{C3}$$

$$\langle pq||rs\rangle = \langle pq|rs\rangle - \langle pq|sr\rangle, \tag{C4}$$

$$\langle pq|rs\rangle = \int d\mathbf{r}_1 d\mathbf{r}_2 \frac{\phi^*_p(\mathbf{r}_1)\phi^*_q(\mathbf{r}_2)\phi_r(\mathbf{r}_1)\phi_s(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}. \tag{C5}$$

Therefore, it is convenient to use off-diagonal elements to constrain the Lagrangian in Eq. (27). Because canonical HF orbitals have occupied and virtual spaces, we decompose the Fock elements in $\mathcal{L}$ for convenience as

$$\begin{aligned}\mathcal{L}[\boldsymbol{\theta}, \boldsymbol{\kappa}, \mathbf{z}, x] = E[\boldsymbol{\theta}, \boldsymbol{\kappa}, x] &+ \sum_a^{vir}\sum_i^{occ} z_{ai}F_{ai} \\ &+ \sum_{i>j}^{occ} z_{ij}F_{ij} + \sum_{a>b}^{vir} z_{ab}F_{ab},\end{aligned} \tag{C6}$$

where i, j, k, l denote occupied orbitals, and a, b, c, d denote virtual orbitals. The requirements for the orbital derivative are

$$\frac{\partial\mathcal{L}}{\partial\kappa_{ck}} = 0, \tag{C7a}$$

$$\frac{\partial\mathcal{L}}{\partial\kappa_{kl}} = 0, \tag{C7b}$$

$$\frac{\partial\mathcal{L}}{\partial\kappa_{cd}} = 0, \tag{C7c}$$

and therefore, we compute the derivative of Fock matrix $A_{pq,rs} = \partial F_{pq}/\partial\kappa_{rs}$ in each orbital sector. The results are

$$\begin{aligned}A_{ai,ck} &= (F_{aa} - F_{ii})\delta_{ac}\delta_{ik} + \langle ak||ic\rangle \\ &\quad + \langle ac||ik\rangle,\end{aligned} \tag{C8a}$$

$$A_{ij,ck} = \langle ic||jk\rangle + \langle ik||jc\rangle, \tag{C8b}$$

$$A_{ab,ck} = \langle ac||bk\rangle + \langle ak||bc\rangle, \tag{C8c}$$

$$A_{ij,kl} = (F_{ii} - F_{jj})(\delta_{ik}\delta_{jl} - \delta_{il}\delta_{jk}), \tag{C8d}$$

$$A_{ab,cd} = (F_{aa} - F_{bb})(\delta_{ac}\delta_{bd} - \delta_{ad}\delta_{bc}), \tag{C8e}$$

and all other terms are strictly zero.

We first solve Eq. (C7b). Because $i > j$ and $k > l$, δ_{il} and δ_{jk} are always zero in Eq. (C8b); hence,

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial\kappa_{kl}} &= \frac{\partial E}{\partial\kappa_{kl}} + \sum_{i>j}^{occ} z_{ij}A_{ij,kl} \\ &= R_{kl} + \sum_{i>j}^{occ} z_{ij}(F_{ii} - F_{jj})\delta_{ik}\delta_{jl} = 0,\end{aligned} \tag{C9}$$

which leads to

$$z_{ij} = -\frac{R_{ij}}{F_{ii} - F_{jj}}. \tag{C10}$$

Similarly, solving Eq. (C7c) results in

$$z_{ab} = -\frac{R_{ab}}{F_{aa} - F_{bb}}. \tag{C11}$$

It should be mentioned that, if orbitals are (nearly) degenerate, the denominator can become numerically zero; however, in

such a case, we expect the numerator to be ∼0, meaning the rotation is redundant and can be discarded. We have not faced such a numerically challenging situation.

Now, from Eq. (C7a), we find

$$\frac{\partial \mathcal{L}}{\partial \kappa_{ck}} = R_{ck} + \sum_{a}^{vir}\sum_{i}^{occ} z_{ai}A_{ai,ck} + \sum_{i>j} z_{ij}A_{ij,ck} + \sum_{a>b} z_{ab}A_{ab,ck}$$
$$= R_{ck} + R^{occ}_{ck} + R^{vir}_{ck} + \sum_{a}^{vir}\sum_{i}^{occ} z_{ai}A_{ai,ck}$$
$$= 0, \tag{C12}$$

where

$$R^{occ}_{ck} = -\sum_{i>j} \frac{R_{ij}}{F_{ii} - F_{jj}}(\langle ic||jk\rangle + \langle ik||jc\rangle), \tag{C13}$$

$$R^{vir}_{ck} = -\sum_{a>b} \frac{R_{ab}}{F_{aa} - F_{bb}}(\langle ac||bk\rangle + \langle ak||bc\rangle). \tag{C14}$$

Therefore, we solve

$$\sum_{a}^{vir}\sum_{i}^{occ} z_{ai}A_{ai,ck} + \tilde{R}_{ck} = 0, \tag{C15}$$

with

$$\tilde{R}_{ck} \equiv R_{ck} + R^{occ}_{ck} + R^{vir}_{ck}, \tag{C16}$$

to determine z_{ai}.

Finally, the relaxed density matrices $D^{\rm relax}_{pq}$ and $D^{\rm relax}_{pq,rs}$ are obtained by comparing the one-body and two-body terms in $\mathcal{L}$:

$$\mathcal{L} = \sum_{pq} h_{pq}\langle\psi|a^\dagger_p a_q|\psi\rangle + \frac{1}{4}\sum_{pqrs}\langle pq||rs\rangle\langle\psi|a^\dagger_p a^\dagger_q a_s a_r|\psi\rangle$$
$$+ \sum_{p>q} z_{pq}(h_{pq} + \langle pr||qs\rangle\langle\Phi^{\rm HF}|a^\dagger_r a_s|\Phi^{\rm HF}\rangle)$$
$$= \sum_{pq} h_{pq}D^{\rm relax}_{pq} + \frac{1}{4}\sum_{pqrs}\langle pq||rs\rangle D^{\rm relax}_{pq,rs}, \tag{C17}$$

where $|\psi\rangle$ represents the VQE wave function, and

$$D^{\rm relax}_{pq} = D_{pq} + \frac{1}{2}z_{pq}, \tag{C18}$$

$$D^{\rm relax}_{pq,rs} = D_{pq,rs} + \frac{1}{2}\Big(z_{pr}D^{\rm HF}_{qs} - z_{qr}D^{\rm HF}_{ps}$$
$$- z_{ps}D^{\rm HF}_{qr} + z_{qs}D^{\rm HF}_{pr}\Big), \tag{C19}$$

$$D_{pq} = \langle\psi|a^\dagger_p a_q|\psi\rangle, \tag{C20}$$

$$D_{pq,rs} = \langle\psi|a^\dagger_p a^\dagger_q a_s a_r|\psi\rangle. \tag{C21}$$

Note that $D^{\rm relax}_{pq,rs}$ is explicitly antisymmetrized such that $D^{\rm relax}_{pq,rs} = -D^{\rm relax}_{qp,rs}$, and so on, for convenience.

2. Frozen-core approximation

When the frozen-core approximation is exercised, the VQE energy derivative with respect to the mixing between the frozen-core I and active p orbitals (i.e., those mapped to qubit), R_{pI}, cannot be directly evaluated by qubit measurements. Nevertheless, we can compute these values using the density matrices within the active space as follows:

$$R_{pI} = 2\Bigg(\tilde{h}_{Ip} - \sum_{q}^{ACT} D_{qp}\tilde{h}_{qI} + \sum_{qr}^{ACT}\langle Iq||pr\rangle D_{qr}$$
$$- \frac{1}{2}\sum_{qrs}^{ACT}\langle qr||sI\rangle D_{qr,sp}\Bigg), \tag{C22}$$

where we have defined the core Hamiltonian:

$$\tilde{h}_{PQ} = h_{PQ} + \sum_{I}^{\text{Frozen-core}} \langle PI||QI\rangle, \tag{C23}$$

with general spin orbitals P, Q.

The Fock derivatives and multipliers with frozen-core indices, such as $A_{aI,ck}$ and z_{cI}, are computed with the same equations as above, simply by expanding i, j, k, l to include the frozen-core orbitals.

For the frozen-virtual approximation, we must consider the contribution from the frozen-virtual orbitals A:

$$R_{Ap} = 2\Bigg(\sum_{q}^{ACT} D_{pA}\tilde{h}_{qA} + \frac{1}{2}\sum_{qrs}^{ACT}\langle qr||sA\rangle D_{qr,sp}\Bigg), \tag{C24}$$

$$R_{AI} = 2\Bigg(\tilde{h}_{IA} + \sum_{pq}^{ACT}\langle pI||qA\rangle D_{qp}\Bigg). \tag{C25}$$

3. Nuclear gradients

In several cases, HF orbitals are expressed as a linear combination of atomic orbitals (AOs) that are functions of atomic coordinates. If AOs depend on the perturbation x (nuclear displacement), the electronic energy derivative includes a contribution from the AO overlap derivative $S^{(x)} = \partial\langle\phi_\mu|\phi_\nu\rangle/\partial x$. The general expression is

$$\frac{dE}{dx} = \sum_{\mu\nu} h^{(x)}_{\mu\nu}D^{\rm relax}_{\mu\nu} + \frac{1}{4}\sum_{\mu\nu\lambda\sigma}\langle\mu\nu||\lambda\sigma\rangle^{(x)}D^{\rm relax}_{\mu\nu,\lambda\sigma}$$
$$+ \sum_{\mu\nu} S^{(x)}_{\mu\nu}W_{\mu\nu}, \tag{C26}$$

where the superscript (x) represents the (explicit) partial derivative, and $W_{\mu\nu}$ denotes the so-called energy weighted density matrix:

$$W_{\mu\nu} = -\Bigg(\sum_{\lambda} h_{\mu\lambda}D^{\rm relax}_{\nu\lambda} + \frac{1}{2}\sum_{\lambda}\langle\mu\kappa||\lambda\sigma\rangle D^{\rm relax}_{\nu\kappa,\lambda\sigma}\Bigg). \tag{C27}$$

4. Spin projection

When the spin-projection operator $\hat{P}$ is present, the same procedure is followed, but with Eq. (18) instead of Eq. (6) for R_{pq}. As a caveat, for spin-dependent properties such

as hyperfine coupling constants, the relaxed density matrices need to be slightly modified. Our SP-VQE energy is expressed as

$$E = \frac{1}{\langle\psi|\hat{P}|\psi\rangle}\left(\sum_{pq} h_{pq}\langle\psi|a_p^\dagger a_q\hat{P}|\psi\rangle + \frac{1}{4}\sum_{pqrs}\langle pq||rs\rangle\langle\psi|a_p^\dagger a_q^\dagger a_s a_r\hat{P}|\psi\rangle\right). \tag{C28}$$

Here, notice that $\langle\psi|a_p^\dagger a_q\hat{P}|\psi\rangle/\langle\psi|\hat{P}|\psi\rangle$ does not correspond to the genuine density matrix of SP-VQE, that is, $\langle\psi|\hat{P}^\dagger a_p^\dagger a_q\hat{P}|\psi\rangle \neq \langle\psi|a_p^\dagger a_q\hat{P}|\psi\rangle$. In fact, the half-projected density matrix and its response correction are not spin adapted and therefore lead to an incorrect spin density. To avoid this issue, the Wigner-Eckart theorem can be adopted, which is expressed as

$$\hat{P}^s_{m,m}(a^\dagger_{p\alpha}a_{q\alpha} + a^\dagger_{p\beta}a_{q\beta})\hat{P}^s_{m,m} = \langle s\ m\ 0\ 0|s\ m\rangle^2(a^\dagger_{p\alpha}a_{q\alpha} + a^\dagger_{p\beta}a_{q\beta})\hat{P}^s_{m,m}, \tag{C29}$$

$$\begin{aligned}\hat{P}^s_{m,m}(a^\dagger_{p\alpha}a_{q\alpha} + a^\dagger_{p\beta}a_{q\beta})\hat{P}^s_{m,m} &= \langle s\ m\ 1\ 0|s\ m\rangle[\langle s\ m\ 1\ 0|s\ m\rangle(a^\dagger_{p\alpha}a_{q\alpha} - a^\dagger_{p\beta}a_{q\beta})\hat{P}^s_{m,m} \\ &+ \langle s\ m\ 1\ -1|s\ m\rangle a^\dagger_{p\alpha}a_{q\beta}\hat{P}^s_{m-1,m} \\ &- \langle s\ m\ 1\ 1|s\ m\rangle(a^\dagger_{p\beta}a_{q\alpha})\hat{P}^s_{m+1,m}],\end{aligned} \tag{C30}$$

where $\langle s_1\ m_1\ s_2\ m_2|S\ M\rangle$ represents the Clebsch-Gordan coefficient, and $\hat{P}^s_{m,k} = |s;m\rangle\langle s;k|$ denotes the transfer operator, which is a general form of the spin-projection operator $\hat{P} \equiv \hat{P}^s_{m,m}$.

[1] J. Preskill, Quantum computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

[2] 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 *et al.*, Quantum supremacy using a programmable superconducting processor, Nature (London) **574**, 505 (2019).

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

[4] 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).

[5] 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).

[6] W. Kutzelnigg, Quantum chemistry in Fock space. I. The universal wave and energy operators, J. Chem. Phys. **77**, 3081 (1982).

[7] W. Kutzelnigg, Quantum chemistry in Fock space. III. Particle-hole formalism, J. Chem. Phys. **80**, 822 (1984).

[8] R. J. Bartlett, S. A. Kucharski, and J. Noga, Alternative coupled-cluster *Ansätze*. II. The unitary coupled-cluster method, Chem. Phys. Lett. **155**, 133 (1989).

[9] J. Romero, R. Babbush, J. R. McClean, C. Hempe, P. J. Love, and A. Aspuru-Guzik, Strategies for quantum computing molecular energies using the unitary coupled cluster *Ansatz*, Quantum Sci. Technol. **4**, 014008 (2019).

[10] I. O. Sokolov, P. K. Barkoutsos, P. J. Ollitrault, D. Greenberg, J. Rice, M. Pistoia, and I. Tavernelli, Quantum orbital-optimized unitary coupled cluster methods in the strongly correlated regime: can quantum algorithms outperform their classical equivalents? J. Chem. Phys. **152**, 124107 (2020).

[11] J. Lee, W. J. Huggins, M. Head-Gordon, and K. B. Whaley, Generalized unitary coupled cluster wave functions for quantum computation, J. Chem. Theory Comput. **15**, 311 (2019).

[12] D. Wecker, M. B. Hastings, N. Wiebe, B. K. Clark, C. Nayak, and M. Troyer, Solving strongly correlated electron models on a quantum computer, Phys. Rev. A **92**, 062318 (2015).

[13] H. R. Grimsley, S. E. Economou, E. Barnes, and N. J. Mayhall, An adaptive variational algorithm for exact molecular simulations on a quantum computer, Nat. Commun. **10**, 3007 (2019).

[14] I. G. Ryabinkin, R. A. Lang, S. N. Genin, and A. F. Izmaylov, Iterative qubit coupled cluster approach with efficient screening of generators, J. Chem. Theory Comput. **16**, 1055 (2020).

[15] H. L. Tang, V. O. Shkolnikov, G. S. Barron, H. R. Grimsley, N. J. Mayhall, E. Barnes, and S. E. Economou, Qubit-ADAPT-VQE: An Adaptive Algorithm for Constructing Hardware-Efficient *Ansätze* on a Quantum Processor, Phys. Rev. X Quantum **2**, 020310 (2021).

[16] N. Gomes, A. Mukherjee, F. Zhang, T. Iadecola, C.-Z. Wang, K.-M. Ho, P. P. Orth, and Y.-X. Yao, Adaptive variational quantum imaginary time evolution approach for ground state preparation, Adv. Quantum Technol. **4**, 2100114 (2021).

[17] Y. S. Yordanov, V. Armaos, C. H. W. Barnes, and D. R. M. Arvidsson-Shukur, Qubit-excitation-based adaptive variational quantum eigensolver, Commun. Phys. **4**, 228 (2021).

[18] V. O. Shkolnikov, N. J. Mayhall, S. E. Economou, and E. Barnes, Avoiding symmetry roadblocks and minimizing the measurement overhead of adaptive variational quantum eigensolvers, arXiv:2109.05340.

[19] Y. S. Yordanov, D. R. M. Arvidsson-Shukur, and C. H. W. Barnes, Efficient quantum circuits for quantum computational chemistry, Phys. Rev. A **102**, 062612 (2020).

[20] N. Handy, P. Knowles, and K. Somasundram, On the convergence of the Møller-Plesset perturbation series, Theor. Chim. Acta **68**, 87 (1985).

[21] J. S. Andrews, D. Jayatilaka, R. G. Bone, N. C. Handy, and R. D. Amos, Spin contamination in single-determinant wavefunctions, Chem. Phys. Lett. **183**, 423 (1991).

[22] T. Tsuchimochi and G. E. Scuseria, ROHF theory made simple, J. Chem. Phys. **133**, 141102 (2010).

[23] P. K. Barkoutsos, J. F. Gonthier, I. Sokolov, N. Moll, G. Salis, A. Fuhrer, M. Ganzhorn, D. J. Egger, M. Troyer, A. Mezzacapo *et al.*, Quantum algorithms for electronic structure calculations: particle-hole Hamiltonian and optimized wave-function expansions, Phys. Rev. A **98**, 022322 (2018).

[24] J.-G. Liu, Y.-H. Zhang, Y. Wan, and L. Wang, Variational quantum eigensolver with fewer qubits, Phys. Rev. Research **1**, 023025 (2019).

[25] K. Seki, T. Shirakawa, and S. Yunoki, Symmetry-adapted variational quantum eigensolver, Phys. Rev. A **101**, 052340 (2020).

[26] B. T. Gard, L. Zhu, G. S. Barron, N. J. Mayhall, S. E. Economou, and E. Barnes, Efficient symmetry-preserving state preparation circuits for the variational quantum eigensolver algorithm, npj Quantum Inf. **6**, 10 (2020).

[27] T. Tsuchimochi, Y. Mori, and S. L. Ten-no, Spin-projection for quantum computation: A low-depth approach to strong correlation, Phys. Rev. Research **2**, 043142 (2020).

[28] A. Khamoshi, F. A. Evangelista, and G. E. Scuseria, Correlating AGP on a quantum computer, Quantum Sci. Technol. **6**, 014004 (2021).

[29] P. Siwach and D. Lacroix, Filtering states with total spin on a quantum computer, Phys. Rev. A **104**, 062435 (2021).

[30] K. Seki and S. Yunoki, Spatial, spin, and charge symmetry projections for a Fermi-Hubbard model on a quantum computer, Phys. Rev. A **105**, 032419 (2022).

[31] A. M. Romero, J. Engel, H. L. Tang, and S. E. Economou, Solving nuclear structure problems with the adaptive variational quantum algorithm, Phys. Rev. C **105**, 064317 (2022).

[32] F. A. Evangelista, G. K.-L. Chan, and G. E. Scuseria, Exact parameterization of fermionic wave functions via unitary coupled cluster theory, J. Chem. Phys. **151**, 244112 (2019).

[33] W. Kutzelnigg, Generalized k-particle Brillouin conditions and their use for the construction of correlated electronic wavefunctions, Chem. Phys. Lett. **64**, 383 (1979).

[34] D. Claudino, J. Wright, A. J. McCaskey, and T. S. Humble, Benchmarking adaptive variational quantum eigensolvers, Front. Chem. **8**, 606863 (2020).

[35] N. Gomes, F. Zhang, N. F. Berthusen, C.-Z. Wang, K.-M. Ho, P. P. Orth, and Y. Yao, Efficient step-merged quantum imaginary time evolution algorithm for quantum chemistry, J. Chem. Theory Comput. **16**, 6256 (2020).

[36] R. Xia and S. Kais, Qubit coupled cluster singles and doubles variational quantum eigensolver *Ansatz* for electronic structure calculations, Quantum Sci. Technol. **6**, 015001 (2021).

[37] F. Zhang, N. Gomes, N. F. Berthusen, P. P. Orth, C.-Z. Wang, K.-M. Ho, and Y.-X. Yao, Shallow-circuit variational quantum eigensolver based on symmetry-inspired Hilbert space partitioning for quantum chemical calculations, Phys. Rev. Research **3**, 013039 (2021).

[38] P.-O. Löwdin, Quantum theory of many-particle systems. I. Physical interpretations by means of density matrices, natural spin-orbitals, and convergence problems in the method of configurational interaction, Phys. Rev. **97**, 1474 (1955).

[39] P. Ring and P. Schuck, *The Nuclear Many-Body Problem* (Springer-Verlag, Berlin, 1980).

[40] G. E. Scuseria, C. A. Jiménez-Hoyos, T. M. Henderson, K. Samanta, and J. K. Ellis, Projected quasiparticle theory for molecular electronic structure, J. Chem. Phys. **135**, 124108 (2011).

[41] C. A. Jiménez-Hoyos, T. M. Henderson, T. Tsuchimochi, and G. E. Scuseria, Projected Hartree-Fock theory, J. Chem. Phys. **136**, 164109 (2012).

[42] T. Tsuchimochi and S. Ten-no, Communication: Configuration interaction combined with spin-projection for strongly correlated molecular electronic structures, J. Chem. Phys. **144**, 011101 (2016).

[43] T. Tsuchimochi and S. L. Ten-no, Orbital-invariant spin-extended approximate coupled-cluster for multi-reference systems, J. Chem. Phys. **149**, 044109 (2018).

[44] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, Evaluating analytic gradients on quantum hardware, Phys. Rev. A **99**, 032331 (2019).

[45] J. S. Kottmann, A. Anand, and A. Aspuru-Guzik, A feasible approach for automatically differentiable unitary coupled-cluster on quantum computers, Chem. Sci. **12**, 3497 (2021).

[46] A. F. Izmaylov, R. A. Lang, and T.-C. Yen, Analytic gradients in variational quantum algorithms: Algebraic extensions of the parameter-shift rule to general unitary transformations, Phys. Rev. A **104**, 062443 (2021).

[47] S. Bravyi, J. M. Gambetta, A. Mezzacapo, and K. Temme, Tapering off qubits to simulate fermionic Hamiltonians, arXiv:1701.08213.

[48] K. Setia, R. Chen, J. E. Rice, A. Mezzacapo, M. Pistoia, and J. D. Whitfield, Reducing qubit requirements for quantum simulations using molecular point group symmetries, J. Chem. Theory Comput. **16**, 6091 (2020).

[49] T. E. O'Brien, B. Senjean, R. Sagastizabal, X. Bonet-Monroig, A. Dutkiewicz, F. Buda, L. DiCarlo, and L. Visscher, Calculating energy derivatives for quantum chemistry on a quantum computer, npj Quantum Inf. **5**, 113 (2019).

[50] K. Mitarai, Y. O. Nakagawa, and W. Mizukami, Theory of analytical energy derivatives for the variational quantum eigensolver, Phys. Rev. Research **2**, 013129 (2020).

[51] J. Gerratt and I. M. Mills, Force constants and dipole-moment derivatives of molecules from perturbed Hartree-Fock calculations. I, J. Chem. Phys. **49**, 1719 (1968).

[52] J. A. Pople, R. Krishnan, H. B. Schlegel, and J. S. Binkley, Derivative studies in Hartree-Fock and Møller-Plesset theories, Int. J. Quantum Chem. **16**, 225 (1979).

[53] N. C. Handy and H. F. Schaefer III, On the evaluation of analytic energy derivatives for correlated wave functions, J. Chem. Phys. **81**, 5031 (1984).

[54] T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic-Structure Theory* (Wiley, Chichester, 2000).

[55] T. Helgaker, S. Coriani, P. Jørgensen, K. Kristensen, J. Olsen, and K. Ruud, Recent advances in wave function-based methods of molecular-property calculations, Chem. Rev. **112**, 543 (2012).

[56] T. Tsuchimochi and S. Ten-no, General technique for analytical derivatives of post-projected Hartree-Fock, J. Chem. Phys. **146**, 074104 (2017).

[57] T. Takeshita, N. C. Rubin, Z. Jiang, E. Lee, R. Babbush, and J. R. McClean, Increasing the Representation Accuracy of Quantum Simulations of Chemistry without Extra Quantum Resources, Phys. Rev. X **10**, 011004 (2020).

[58] W. Mizukami, K. Mitarai, Y. O. Nakagawa, T. Yamamoto, T. Yan, and Y.-y. Ohnishi, Orbital optimized unitary coupled cluster theory for quantum computer, Phys. Rev. Research **2**, 033421 (2020).

[59] T. Tsuchimochi, Y. Mori, Y. Shimomoto, T. Nishimaki, Y. Ryo, M. Taii, T. Yoshikura, T. S. Chung, K. Sasasako, and

K. Yoshimura, QUKET: The comprehensive quantum simulator for quantum chemistry (2022), https://github.com/quket/quket.

[60] 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).

[61] 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 *et al.*, OPENFERMION: The electronic structure package for quantum computers, Quantum Sci. Technol. **5**, 034014 (2020).

[62] Y. Suzuki, Y. Kawase, Y. Masumura, Y. Hiraga, M. Nakadai, J. Chen, K. M. Nakanishi, K. Mitarai, R. Imai, S. Tamiya *et al.*, QULACS: A fast and versatile quantum circuit simulator for research purpose, Quantum **5**, 559 (2021).

[63] J. Hermann, https://github.com/jhrmnn/pyberny (2020).

[64] A. Delgado, J. M. Arrazola, S. Jahangiri, Z. Niu, J. Izaac, C. Roberts, and N. Killoran, Variational quantum algorithm for molecular geometry optimization, Phys. Rev. A **104**, 052402 (2021).

[65] L. W. Bertels, H. R. Grimsley, S. E. Economou, E. Barnes, and N. J. Mayhall, Symmetry breaking slows convergence of the adapt variational quantum eigensolver, arXiv:2207.03063.