

Accelerated variational quantum eigensolver with joint Bell measurement

Chenfeng Cao ^{1,2,*} Hiroshi Yano,^{3,2} and Yuya O. Nakagawa ^{2,†}¹*Department of Physics, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China*²*QunaSys Inc., Aqua Hakusan Building 9F, 1-13-7 Hakusan, Bunkyo, Tokyo 113-0001, Japan*³*Department of Applied Physics and Physico-Informatics, Keio University, Hiyoshi 3-14-1, Kohoku, Yokohama 223-8522, Japan*

(Received 20 June 2023; accepted 2 February 2024; published 26 February 2024)

The variational quantum eigensolver (VQE) stands as a prominent quantum-classical hybrid algorithm for near-term quantum computers to obtain the ground states of molecular Hamiltonians in quantum chemistry. However, due to the noncommutativity of the Pauli operators in the Hamiltonian, the number of measurements required on quantum computers increases significantly as the system size grows, which may hinder practical applications of VQE. In this work, we present a protocol termed *joint Bell measurement VQE* (JBM-VQE) to reduce the number of measurements and speed up the VQE algorithm. Our method employs joint Bell measurements, enabling the simultaneous measurement of the absolute values of all expectation values of Pauli operators present in the Hamiltonian. In the course of the optimization, JBM-VQE estimates the absolute values of the expectation values of the Pauli operators for each iteration by the joint Bell measurement, while the signs of them are measured less frequently by the conventional method to measure the expectation values. Our approach is based on the empirical observation that the signs do not often change during optimization. We illustrate the speed-up of JBM-VQE compared to conventional VQE by numerical simulations for finding the ground states of molecular Hamiltonians of small molecules, and the speed-up of JBM-VQE at the early stage of the optimization becomes increasingly pronounced in larger systems. Our approach based on the joint Bell measurement is not limited to VQE and can be utilized in various quantum algorithms whose cost functions are expectation values of many Pauli operators.

DOI: [10.1103/PhysRevResearch.6.013205](https://doi.org/10.1103/PhysRevResearch.6.013205)

I. INTRODUCTION

Noisy intermediate-scale quantum (NISQ) devices [1] have attracted considerable interest as they hold the potential to solve certain computational tasks faster than classical computers [2–4]. These devices typically have a relatively small number of qubits (usually between 50 and a few hundred) and are subject to hardware noise, so the computational results obtained from them may not be completely reliable. Despite these limitations, NISQ devices are expected to be capable of performing calculations that are beyond the capabilities of classical computers, which makes them exciting tools in the near future. The research community has made substantial progress in developing NISQ-friendly variational quantum optimization algorithms for a variety of applications, including quantum machine learning [5], fidelity estimation [6,7], and quantum error-correcting code discovery [8]. The variational quantum eigensolver (VQE) is considered a flagship algorithm within this field, utilizing parameterized quantum circuits and the variational principle to prepare the ground or excited states of quantum many-body systems [9,10].

VQE has already been experimentally realized on actual quantum hardware to solve small-sized problems in quantum chemistry and material calculation [9,11]. However, the scalability of VQE to larger system sizes remains a challenge. One of the main reasons, especially in the application to quantum chemistry, is the large number of measurements required during the optimization process. For the molecular Hamiltonians in quantum chemistry, there are usually $\mathcal{O}(n^4)$ Pauli operators for an n -qubit system, and estimation of their expectation values is required in each iteration of VQE. For example, it was estimated that a single evaluation of the expectation value of the Hamiltonian for analyzing the combustion energies of some organic molecules requires $\sim 10^9$ measurement shots and may take as long as several days [12]. To tackle the problem of scalability, various methods have been proposed to reduce the number of measurements in the optimization of VQE. One possible strategy is to divide the $\mathcal{O}(n^4)$ Pauli operators in the Hamiltonian into the groups of simultaneously measurable operators, and there are methods that realize $\mathcal{O}(n^3)$ groups [13,14] and even $\mathcal{O}(n^2)$ groups [15,16]. The reduction of the number of groups can result in the reduction of the number of measurements to estimate their expectation values, which leads to the alleviation of the scalability problem of VQE. Nonetheless, it is still highly demanded to develop a method to reduce the number of measurements in the whole optimization process of VQE.

Although it is impossible to perform the projective measurement simultaneously on the noncommuting Pauli operators to estimate their expectation values, the *absolute*

*chenfeng.cao@connect.ust.hk

†nakagawa@qunasys.com

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

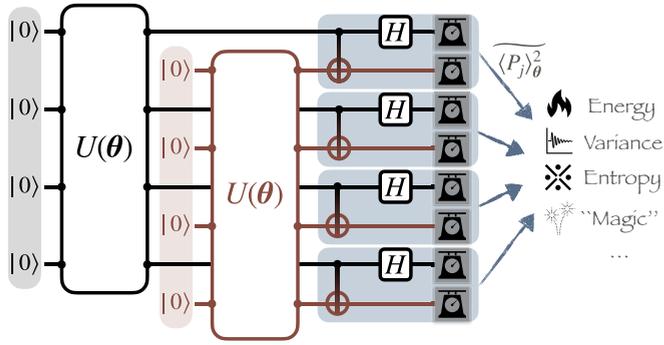


FIG. 1. Schematic picture of the joint Bell measurement and JBM-VQE. For an n -qubit quantum state $|\psi(\theta)\rangle = U(\theta)|0\rangle$, the $2n$ -qubit quantum states $|\psi(\theta)\rangle \otimes |\psi(\theta)\rangle$ is prepared and measured in the Bell basis. The measurement results can give the estimates of all expectation values $\langle P_j \rangle_\theta^2 = (\langle \psi(\theta) | P_j | \psi(\theta) \rangle)^2$ in the original n -qubit systems, and the state information (energy in JBM-VQE) is estimated by these estimates as well as the pre-given signs of the expectation values.

values of the expectation values can be estimated simultaneously by using the so-called joint Bell measurement in a doubled system consisting of $2n$ qubits (see Sec. II B for detailed explanations). This is because for any n -qubit Pauli operators P_1, P_2, \dots, P_M acting on the original system of n qubit, the operators $P_1 \otimes P_1, P_2 \otimes P_2, \dots, P_M \otimes P_M$ acting on the doubled system of $2n$ qubits commute with each other. Then the expectation values of the doubled state $(\langle \psi | \otimes \langle \psi |)(P_j \otimes P_j)(|\psi\rangle \otimes |\psi\rangle) = (\langle \psi | P_j | \psi \rangle)^2$ can be estimated simultaneously for all $j = 1, \dots, M$, where $|\psi\rangle$ is a state in the original n -qubit system. The joint Bell measurement is pivotal in numerous domains of quantum information science, including quantum computing, quantum communications, and quantum metrology. This technique was utilized to show an exponential advantage of quantum computers over classical ones in predicting properties of physical systems [17,18], estimate the reduced density matrices [19], and learn the non-stabilizerness of a quantum state [20]. In photonic systems, the joint Bell measurement has facilitated the implementation of an optimal orienteering protocol [21] as well as breaking the classical limit in quantum target identification [22].

It is noteworthy to point out that the reduction of the number of measurements by a constant factor was achieved in Ref. [23] through the use of Bell measurement. While the classical shadow technique [24] also seeks to predict the expectation values of numerous operators simultaneously, the measurement overhead in its protocol with random Pauli measurements scales exponentially with the operator's locality. This makes its application to quantum chemistry Hamiltonians with nonlocal Pauli operators less straightforward (see also Ref. [25]).

In this study, we introduce a method to reduce measurement overhead in VQE by employing joint Bell measurements (JBM), referred to as JBM-VQE (a schematic illustration is provided in Fig. 1). Our approach takes advantage of the correlation between Pauli expectation values across successive iterations of VQE. More concretely, we observe that the signs of expectation values of the Pauli operators in

the Hamiltonian change infrequently during the optimization of VQE while their absolute values change frequently. In JBM-VQE, one measures the absolute values by the joint Bell measurement, which requires only one circuit to measure, during every iteration. On the other hand, the signs of expectation values of the Pauli operators are measured once in the fixed number of iterations by the conventional measurement method (using naively $\mathcal{O}(n^4)$, at least $\mathcal{O}(n^2)$, distinct circuits). The expectation value of the Hamiltonian is subsequently constructed by combining the estimated absolute values and signs with assuming that the latest estimation of the signs is valid for the following iterations. We can expect a reduction in the measurement cost during the optimization by using this protocol. To exemplify this, we numerically compare JBM-VQE to the conventional VQE for molecular Hamiltonians of small molecules under the reasonable condition that the statistical fluctuations of the energy expectation values in both methods are almost the same. JBM-VQE requires fewer shots to approach the vicinity of the exact ground state compared to conventional VQE, with this trend becoming more pronounced for larger molecules. Our proposal is applicable to any molecular Hamiltonian in quantum chemistry and is expected to expedite the practical utilization of VQE. Furthermore, the protocols in JBM-VQE can be utilized in various variational quantum algorithms which optimize the expectation values of many Pauli operators, other than VQE.

The paper is organized as follows. We present the JBM-VQE algorithm in Sec. II. The required number of measurements to estimate the expectation values of the Pauli operators at certain precision is discussed for JBM-VQE and the conventional VQE in Sec. III. The numerical comparison between our proposed JBM-VQE and the conventional VQE on various molecular Hamiltonians is presented in Sec. IV, demonstrating the acceleration of JBM-VQE. We discuss several aspects of JBM-VQE in Sec. V. Finally, we summarize the paper and provide an outlook in Sec. VI.

II. ALGORITHM

In this section, we describe the algorithm of JBM-VQE. We first explain our target Hamiltonian and the joint Bell measurement. We then explain the algorithm of JBM-VQE.

A. Setup

We focus on the quantum chemistry Hamiltonian in the second-quantized form, given by

$$H_f = \sum_{i,j=1}^n h_{ij} c_i^\dagger c_j + \frac{1}{2} \sum_{i,j,k,l=1}^n V_{ijkl} c_i^\dagger c_j^\dagger c_l c_k, \quad (1)$$

where c_i (c_i^\dagger) is an annihilation (creation) operator of an electron labeled by $i = 1, 2, \dots, n$ satisfying the canonical anticommutation relation $\{c_i, c_j\} = \{c_i^\dagger, c_j^\dagger\} = 0$, $\{c_i, c_j^\dagger\} = \delta_{ij}$, and h_{ij} (V_{ijkl}) is the scalar related to the so-called one-electron (two-electron) integrals [26,27]. This Hamiltonian is mapped to a qubit representation by fermion-qubit mappings such Jordan-Wigner mapping [28], parity mapping [29], or

Bravyi-Kitaev mapping [30], as follows:

$$H = \sum_{j=1}^M \lambda_j P_j, \quad (2)$$

where $\lambda_j \in \mathbb{R}$ is a coefficient, P_j is an n -qubit Pauli operators $P_j \in \{I, X, Y, Z\}^{\otimes n}$, and $M = \mathcal{O}(n^4)$ is the number of the Pauli operators. Throughout this paper, we omit the identity term $I^{\otimes n}$ in the Hamiltonian and assume $P_j \neq I^{\otimes n}$. Similar to the conventional VQE, JBM-VQE is based on the ansatz quantum state:

$$|\psi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})|0\rangle, \quad (3)$$

where $U(\boldsymbol{\theta})$ is a parameterized quantum circuit with parameters $\boldsymbol{\theta} = (\theta_1, \dots, \theta_{N_\theta})$. Our objective is to minimize the energy expectation value:

$$E(\boldsymbol{\theta}) := \langle H \rangle_{\boldsymbol{\theta}} = \sum_{j=1}^M \lambda_j \langle P_j \rangle_{\boldsymbol{\theta}}, \quad (4)$$

where $\langle \dots \rangle_{\boldsymbol{\theta}} := \langle \psi(\boldsymbol{\theta}) | \dots | \psi(\boldsymbol{\theta}) \rangle$, with respect to the parameters $\boldsymbol{\theta}$.

B. Joint Bell measurement

The joint Bell measurement is an entangling operation in quantum information processing that allows one to distinguish between the four maximally entangled Bell states. It has profound implications for quantum computation, e.g., enabling the determination of the absolute values of expectation values of all 4^n Pauli operators for an n qubit state [17–19]. It requires a $2n$ -qubit system comprising two identical n qubit systems, denoted as A and B . We prepare a $2n$ qubit state,

$$|\Psi(\boldsymbol{\theta})\rangle := |\psi(\boldsymbol{\theta})\rangle_A \otimes |\psi(\boldsymbol{\theta})\rangle_B = (U(\boldsymbol{\theta}) \otimes U(\boldsymbol{\theta}))|0\rangle_A \otimes |0\rangle_B, \quad (5)$$

and apply a Controlled-NOT (CNOT) gate with one qubit, A , as the control and the other, B , as the target. Then we apply a Hadamard gate on the control qubit and eventually measure both qubits in the computational basis (see Fig. 1). For each pair of qubits, measuring the state in the computational basis results in the projective measurement onto the Bell basis,

$$\begin{aligned} |\Phi_1^\pm\rangle &= \frac{1}{\sqrt{2}}(|0\rangle_A \otimes |0\rangle_B \pm |1\rangle_A \otimes |1\rangle_B), \\ |\Phi_2^\pm\rangle &= \frac{1}{\sqrt{2}}(|0\rangle_A \otimes |1\rangle_B \pm |1\rangle_A \otimes |0\rangle_B). \end{aligned} \quad (6)$$

These basis states are common eigenstates of Pauli operators $X_A \otimes X_B, Y_A \otimes Y_B, Z_A \otimes Z_B$. Therefore the measurement for all $2n$ qubits in Fig. 1 constitutes the projective measurement on simultaneous eigenstates of all $2n$ qubit Pauli operators in the form of $P_j \otimes P_j$ ($P_j \in \{I, X, Y, Z\}^{\otimes n}$), whose total number is 4^n . Consequently, for any n qubit Pauli operators P_j , we can estimate

$$\langle \Psi(\boldsymbol{\theta}) | (P_j \otimes P_j) | \Psi(\boldsymbol{\theta}) \rangle = \langle P_j \rangle_{\boldsymbol{\theta}}^2 \quad (7)$$

from the measurement outcomes of the single quantum circuit in Fig. 1. We denote the estimated value as $\widetilde{\langle P_j \rangle_{\boldsymbol{\theta}}}$. The absolute

value of $|\langle P_j \rangle_{\boldsymbol{\theta}}|$ is then estimated as

$$|\widetilde{\langle P_j \rangle_{\boldsymbol{\theta}}}| = \sqrt{\max\{0, \widetilde{\langle P_j \rangle_{\boldsymbol{\theta}}^2}\}}. \quad (8)$$

We refer to this protocol to estimate the absolute values of expectation values of all 4^n Pauli operators as the *joint Bell measurement*. We observe that the estimate given by Eq. (8) is biased for a finite number of measurement shots m due to the nonlinearity of the square root and the maximum functions although the bias vanishes as m grows. As shown in Appendix A, the bias scales as

$$\begin{aligned} B(|\widetilde{\langle P_j \rangle_{\boldsymbol{\theta}}}|) &= \mathbb{E}^{(m)}(|\widetilde{\langle P_j \rangle_{\boldsymbol{\theta}}}|) - |\langle P_j \rangle_{\boldsymbol{\theta}}| \\ &\sim \begin{cases} m^{-1/4} & (\langle P_j \rangle_{\boldsymbol{\theta}} = 0) \\ -m^{-1} & (\langle P_j \rangle_{\boldsymbol{\theta}} \neq 0) \end{cases}, \end{aligned} \quad (9)$$

where $\mathbb{E}^{(m)}(\dots)$ signifies the expectation value over m shots measurement.

C. JBM-VQE algorithm

In the JBM-VQE algorithm, we estimate the energy $E(\boldsymbol{\theta})$ and its gradient $\frac{\partial E(\boldsymbol{\theta})}{\partial \theta_l}$ by decomposing the expectation value $\langle P_j \rangle_{\boldsymbol{\theta}}$ into its sign,

$$s_j(\boldsymbol{\theta}) := \text{sgn}(\langle P_j \rangle_{\boldsymbol{\theta}}) := \begin{cases} +1 & (\text{if } \langle P_j \rangle_{\boldsymbol{\theta}} \geq 0) \\ -1 & (\text{if } \langle P_j \rangle_{\boldsymbol{\theta}} < 0) \end{cases}, \quad (10)$$

and its absolute value $|\langle P_j \rangle_{\boldsymbol{\theta}}|$. We employ the following two subroutines to estimate the energy and gradient in the algorithm.

Subroutine 1. The first subroutine takes the parameters $\boldsymbol{\theta}$ as its inputs. In this subroutine, we first estimate the signs $s_j(\boldsymbol{\theta})$ by evaluating the expectation values $\langle P_j \rangle_{\boldsymbol{\theta}}$ themselves with the standard measurement strategy using n qubits, as done in the conventional VQE. One naive way to estimate the signs is to perform the projective measurement of each P_j with $\mathcal{O}(n^4)$ distinct measurement circuits, subsequently estimating the sign $s_j(\boldsymbol{\theta})$ via the majority vote of its ± 1 result. The estimated sign is denoted as $\widetilde{s}_j(\boldsymbol{\theta})$, and the total number of shots (repetitions of quantum circuit executions) to estimate all $s_j(\boldsymbol{\theta})$ is represented as m_s^{tot} . Following this, the joint Bell measurement using $2n$ qubits is utilized to approximate the absolute values of the expectation values $|\langle P_j \rangle_{\boldsymbol{\theta}}|$, resulting in their estimates $|\widetilde{\langle P_j \rangle_{\boldsymbol{\theta}}}|$. The number of shots for the joint Bell measurement is denoted as m . The estimate of the energy (expectation value of the Hamiltonian) is constructed as

$$\widetilde{E}(\boldsymbol{\theta}) = \sum_j \lambda_j \widetilde{s}_j(\boldsymbol{\theta}) |\widetilde{\langle P_j \rangle_{\boldsymbol{\theta}}}|. \quad (11)$$

Additionally, we estimate the gradient of $E(\boldsymbol{\theta})$ by using the so-called parameter shift rule [5,31], mathematically expressed in the simplest case as

$$\frac{\partial \langle P_j \rangle_{\boldsymbol{\theta}}}{\partial \theta_l} = \frac{1}{2 \sin \alpha} (\langle P_j \rangle_{\boldsymbol{\theta}_+^{(l)}} - \langle P_j \rangle_{\boldsymbol{\theta}_-^{(l)}}), \quad (12)$$

where $\boldsymbol{\theta}_\pm^{(l)} = \boldsymbol{\theta} \pm \alpha \boldsymbol{\delta}_l$, $\boldsymbol{\delta}_l$ is a unit vector with only the l th component nonzero, and $\alpha \in \mathbb{R}$ is a fixed constant. We take $\alpha = \pi/4$ in the numerical calculation in Sec. IV. Analogous to the energy estimation, we use the standard measurement

strategy to estimate the signs $s_j(\theta_{\pm}^{(l)})$ using the quantum states $|\psi(\theta_{\pm}^{(l)})\rangle$, which may require $\mathcal{O}(n^4)$ distinct quantum circuits in a naive way. The absolute values $|\langle P_j \rangle_{\theta_{\pm}^{(l)}}|$ are then estimated with the joint Bell measurement for the states $|\psi(\theta_{\pm}^{(l)})\rangle$. The gradient of the energy is estimated by

$$\frac{\partial \widetilde{E}(\boldsymbol{\theta})}{\partial \theta_l} = \sum_j \frac{\lambda_j}{2 \sin \alpha} (\widetilde{s_j(\theta_+^{(l)})} |\widetilde{\langle P_j \rangle_{\theta_+^{(l)}}}| - \widetilde{s_j(\theta_-^{(l)})} |\widetilde{\langle P_j \rangle_{\theta_-^{(l)}}}|). \quad (13)$$

The estimates of the energy (11) and the gradient (13) are the outputs of this subroutine.

Subroutine 2. The second subroutine takes the parameters $\boldsymbol{\theta}$ and a set of guessed signs

$$\begin{aligned} & \{t_j\}_{j=1}^M, \{t_{j,+}^{(l=1)}\}_{j=1}^M, \dots, \{t_{j,+}^{(l=N_\theta)}\}_{j=1}^M, \\ & \{t_{j,-}^{(l=1)}\}_{j=1}^M, \dots, \{t_{j,-}^{(l=N_\theta)}\}_{j=1}^M, \end{aligned} \quad (14)$$

as inputs ($t_j, t_{j,\pm}^{(l)} = \pm 1$). In this subroutine, we estimate only the absolute values ($|\langle P_j \rangle_{\theta}|$, $|\langle P_j \rangle_{\theta_+^{(l)}}|$ and $|\langle P_j \rangle_{\theta_-^{(l)}}|$) by the joint Bell measurement. The energy and the gradient are estimated by

$$\begin{aligned} \widetilde{E}(\boldsymbol{\theta}) &= \sum_j \lambda_j t_j |\widetilde{\langle P_j \rangle_{\theta}}|, \quad (15) \\ \frac{\partial \widetilde{E}(\boldsymbol{\theta})}{\partial \theta_l} &= \sum_j \frac{\lambda_j}{2 \sin \alpha} (t_{j,+}^{(l)} |\widetilde{\langle P_j \rangle_{\theta_+^{(l)}}}| - t_{j,-}^{(l)} |\widetilde{\langle P_j \rangle_{\theta_-^{(l)}}}|). \quad (16) \end{aligned}$$

These two estimates are outputs of the second subroutine.

The JBM-VQE algorithm is described in algorithm 1. Let us assume that the parameters are $\boldsymbol{\theta}$ in the n_{iter} th iteration ($n_{\text{iter}} = 0, 1, \dots$) of the algorithm. When n_{iter} is a multiple of T_S , we invoke subroutine 1 to estimate the energy and gradient, $\widetilde{E}(\boldsymbol{\theta})$ [Eq. (11)] and $\frac{\partial \widetilde{E}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ [Eq. (13)], respectively. Importantly, we also record the estimates of the signs $\widetilde{s_j(\boldsymbol{\theta})}$, $\widetilde{s_{j,\pm}^{(l)}(\boldsymbol{\theta})}$ for $j = 1, \dots, M$ and $l = 1, \dots, N_\theta$. When n_{iter} is not a multiple of T_S , we invoke subroutine 2 with utilizing the prerecorded signs (obtained at some past iteration) as the guessed signs [Eq. (14)]. In other words, we estimate the energy and the gradient by Eqs. (15) and (16) with performing only the joint Bell measurement that estimates the absolute values of the Pauli expectation values. Then we update the parameters by the gradient decent method $\boldsymbol{\theta}' = \boldsymbol{\theta} - \eta \frac{\partial \widetilde{E}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$, where η is a learning rate. It is worth noting that the gradient descent is not the only choice in the JBM-VQE and other sophisticated optimization algorithms can be employed (see the discussion in Sec. V).

Several remarks regarding our JBM-VQE algorithm are in order. Firstly, this algorithm relies on the expectation that the signs of the Pauli expectation values ($\langle P_j \rangle_{\theta}$ and $\langle P_j \rangle_{\theta_{\pm}^{(l)}}$) do not change frequently during the optimization process. Subroutine 2 consists of the joint Bell measurement that uses only $(2N_\theta + 1)$ quantum circuits and may typically require fewer measurement shots to estimate the absolute values of the Pauli expectation values than the conventional VQE. For this reason, we expect a reduction in the total number of shots in JBM-VQE compared with the conventional VQE.

Secondly, the joint Bell measurement has a bias on its estimates, causing the energy and gradient estimated in both subroutines 1 and 2 to exhibit bias. Nevertheless, this bias vanishes with the increase of the number of measurement shots and remains a minor concern since it is relatively small compared to the required energy precision, especially during the early optimization stages. (This topic is further elaborated upon in Sec. V and Appendix A). The rough criteria of the number of shots for realizing a certain precision of the estimated expectation values are discussed in Sec. III. Thirdly, if we use a $2n$ -qubit system just as two independent copies of the original n -qubit system and conduct the conventional VQE, m executions of the circuits are equivalent to $2m$ shots in the original system. Consequently, our JBM-VQE algorithm must surpass the conventional VQE by at least a factor of two concerning the number of shots, and it is actually realized in the numerical simulation in Sec. IV. Lastly, the sign-updating period T_S influences the efficiency of JBM-VQE. A larger period leads to fewer shots required for optimization, albeit with the trade-off of less accurate energy and gradient estimates. While there is no a priori criterion for determining T_S , it can be set manually or adaptively by monitoring the optimization history.

III. SHOT THRESHOLDS

In both JBM-VQE and the conventional VQE, the energy estimate exhibits finite statistical fluctuation due to the limited number of measurement shots. This occurs even without any noise present in quantum devices. To facilitate a fair comparison between JBM-VQE and the conventional VQE, it is essential to establish a common criterion ensuring that both methods display the same level of fluctuation. In this section, we discuss such a criterion by investigating the number of shots required to estimate an expectation value of a single Pauli operator with a fixed level of accuracy. We formulate the number of shots to estimate the expectation value with certain accuracy and probability, and numerically calculate the actual numbers. The number established here is utilized in Sec. IV, where numerical demonstrations of JBM-VQE are performed for quantum chemistry Hamiltonians of small molecules.

A. Shot threshold for the conventional VQE

Let us define the number of shots to estimate the expectation value of a single Pauli operator with the projective measurement, which is the standard measurement strategy for the conventional VQE. For a given n -qubit state $|\psi\rangle$ and an n -qubit Pauli operator P , the probability of estimating $\langle P \rangle := \langle \psi | P | \psi \rangle$ within an additive error τ_{th} under the m -shot projective measurement of P is given by

$$\begin{aligned} p(m, \tau_{\text{th}}, \langle P \rangle) &= \sum_{x=x_{\min}}^{x_{\max}} \binom{m}{x} \left(\frac{1 + \langle P \rangle}{2} \right)^x \left(\frac{1 - \langle P \rangle}{2} \right)^{m-x}, \\ x_{\min} &= \max \left\{ 0, \left\lceil \frac{m(1 + \langle P \rangle - \tau_{\text{th}})}{2} \right\rceil \right\}, \\ x_{\max} &= \min \left\{ m, \left\lfloor \frac{m(1 + \langle P \rangle + \tau_{\text{th}})}{2} \right\rfloor \right\}, \end{aligned} \quad (17)$$

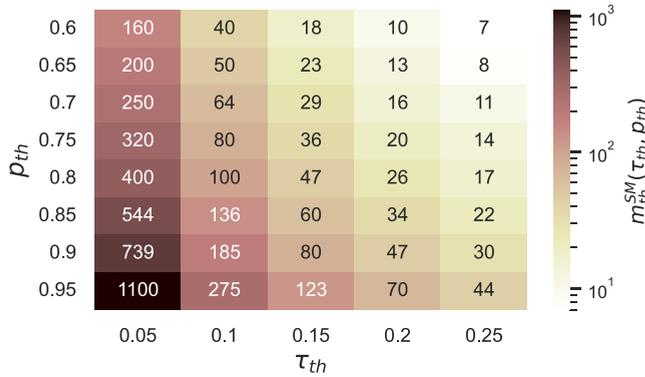


FIG. 2. Values of the standard measurement (SM) shot threshold $m_{\text{th}}^{\text{SM}}(\tau_{\text{th}}, p_{\text{th}})$ for different estimation errors τ_{th} and probabilities p_{th} .

where $\lceil \dots \rceil$ ($\lfloor \dots \rfloor$) is the ceiling (floor) function of integers. Here the expectation value $\langle P \rangle$ itself only matters to calculate the threshold and there is no dependence on the number of qubits. This is partly because we consider the single Pauli operator to define the threshold. Since there are various Pauli operators included in the Hamiltonian, we consider the averaged probability for estimating the expectation value within an additive error τ_{th} ,

$$p^{(\text{av})}(m, \tau_{\text{th}}) = \frac{1}{2} \int_{-1}^1 dy p(m, \tau_{\text{th}}, y). \quad (18)$$

We then define the standard measurement (SM) shot threshold as follows:

Definition 1. SM shot threshold The standard measurement (SM) shot threshold $m_{\text{th}}^{\text{SM}}(\tau_{\text{th}}, p_{\text{th}})$ is defined as

$$m_{\text{th}}^{\text{SM}}(\tau_{\text{th}}, p_{\text{th}}) := \min\{m \in \mathbb{Z}^+ \mid p^{(\text{av})}(m, \tau_{\text{th}}) \geq p_{\text{th}}\}. \quad (19)$$

The SM shot threshold $m_{\text{th}}^{\text{SM}}(\tau_{\text{th}}, p_{\text{th}})$ indicates the minimum number of the shots of the projective measurement of P to estimate $\langle P \rangle$ within an additive error τ_{th} with a probability at least p_{th} , where the expectation value $\langle P \rangle$ is averaged in the uniform distribution for $[-1, 1]$. We leverage $m_{\text{th}}^{\text{SM}}(\tau_{\text{th}}, p_{\text{th}})$ to determine the number of shots in numerical simulation of the conventional VQE in Sec. IV. We note that the value of $\langle P \rangle$ may cluster around 0 if we consider random states in the Hilbert space, e.g., Haar random states, but we employ the uniform distribution because the ground states of quantum chemistry Hamiltonians are not random states and various values of $\langle P \rangle$ may appear. Thus, while the threshold would differ for other distributions, we have chosen the uniform distribution for its relevance and simplicity as a baseline for this study. See Appendix C for discussions on another distribution of $\langle P \rangle$.

Finally, we evaluate actual numerical values of $m_{\text{th}}^{\text{SM}}(\tau_{\text{th}}, p_{\text{th}})$ for various τ_{th} and p_{th} . For a given m , we calculate $p^{\text{ave}}(m, \tau_{\text{th}})$ by approximating the integral through numerical integration, taking 2000 uniformly spaced points of $\langle P \rangle$ in the interval $[-1, 1]$. The results are presented in Fig. 2.

Algorithm 1: JBM-VQE.

Input: Hamiltonian $H = \sum_j \lambda_j P_j$, variational circuit $U(\theta)$, number of shots for the joint Bell measurement m , sign-updating period T_S , number of shots for sign-updating m_S^{tot} , and learning rate η .

Output: Parameters θ_{opt} which approximate the ground state of H by $|\psi(\theta_{\text{opt}})\rangle = U(\theta_{\text{opt}})|0\rangle$ and the estimated optimal energy $\widetilde{E}(\theta)$.

- 1 Initialize θ and set $n_{\text{iter}} = 0$;
 - 2 **while** energy estimate $\widetilde{E}(\theta)$ has not converged **do**
 - 3 **if** $n_{\text{iter}} \bmod T_S = 0$ **then**
 - 4 Call subroutine 1 with input θ , and obtain the estimates of the energy [Eq. (11)] and the gradient [Eq. (13)];
 - 5 Record the signs $\widetilde{s}_j(\theta)$, $\widetilde{s}_{j,\pm}^{(l)}(\theta)$ as $t_j, t_{j,\pm}^{(l)}$;
 - 6 **end**
 - 7 **if** $n_{\text{iter}} \bmod T_S \neq 0$ **then**
 - 8 Call subroutine 2 with input θ and the guessed signs $t_j, t_{j,\pm}^{(l)}$ at some past iteration, and obtain the estimates of the energy [Eq. (15)] and the gradient [Eq. (16)];
 - 9 **end**
 - 10 $\theta \leftarrow \theta - \eta \frac{\partial \widetilde{E}(\theta)}{\partial \theta}$;
 - 11 $n_{\text{iter}} \leftarrow n_{\text{iter}} + 1$;
 - 12 **end**
 - 13 Return θ and $\widetilde{E}(\theta)$;
-

B. Shot threshold for JBM-VQE

In JBM-VQE, the estimation of the absolute value $\langle P \rangle$ and the sign $s := \text{sign}(\langle P \rangle)$ for a given state $|\psi\rangle$ and a Pauli operator P is performed differently. The absolute value is estimated using joint Bell measurements, while the sign is typically determined through the projective measurement of

P . We consider the number of shots to estimate each of them within certain error and probability.

First, let us define the number of shots necessary for accurately estimating the absolute value $\langle P \rangle$. Since the joint Bell measurement provides the expectation values of $P \otimes P$, the probability of determining $|\langle P \rangle|$ within an error τ_{th} is



FIG. 3. Values of the joint Bell measurement (JBM) shot threshold $m_{th}^{JBM}(\tau_{th}, p_{th})$ for different estimation errors τ_{th} and probabilities p_{th} .

calculated as follows:

$$q(m, \tau_{th}, \langle P \rangle) = \sum_{x \in \mathbb{X}} \binom{m}{x} \left(\frac{1 + \langle P \rangle^2}{2} \right)^x \left(\frac{1 - \langle P \rangle^2}{2} \right)^{m-x}, \quad (20)$$

where \mathbb{X} is a set of integers in $0 \leq x \leq m$ that satisfies

$$\left| \sqrt{\max \left\{ 0, \frac{2x}{m} - 1 \right\}} - |\langle P \rangle| \right| \leq \tau_{th} \quad (21)$$

[see Eq. (8)]. Similar to SM shot threshold, we average the probability $q(m, \tau_{th}, \langle P \rangle)$ as

$$q^{(av)}(m, \tau_{th}) = \frac{1}{2} \int_{-1}^1 dy q(m, \tau_{th}, y). \quad (22)$$

We can now define JBM shot threshold:

Definition 2 (JBM shot threshold). JBM shot threshold $m_{th}^{JBM}(\tau_{th}, p_{th})$ is defined as

$$m_{th}^{JBM}(\tau_{th}, p_{th}) := \min\{m \in \mathbb{Z}^+ \mid p^{(av)}(m, \tau_{th}) \geq p_{th}\}. \quad (23)$$

This means that the JBM shot threshold $m_{th}^{JBM}(\tau_{th}, p_{th})$ indicates the minimal number of shots needed to estimate the absolute value of the expectation value $\langle P \rangle$ within an additive error τ_{th} with probability p_{th} . Numerical values of $m_{th}^{JBM}(\tau_{th}, p_{th})$ are again calculated by 2000 points of $\langle P \rangle$ that are uniformly spaced in $[-1, 1]$ and summarized in Fig. 3.

Next, we examine the estimation of the sign s in JBM-VQE that is performed through a majority vote of the results of the projective measurement of P . When the number of shots is an odd integer m_S , the probability of estimating s correctly is

$$p_S(m_S, \langle P \rangle) = \sum_{x \in \mathbb{X}_S} \binom{m_S}{x} \left(\frac{1 + \langle P \rangle}{2} \right)^x \left(\frac{1 - \langle P \rangle}{2} \right)^{m_S-x}, \quad (24)$$

$$\mathbb{X}_S = \begin{cases} \left\{ \frac{m_S+1}{2}, \frac{m_S+3}{2}, \dots, m_S \right\} & (\langle P \rangle \geq 0) \\ \left\{ 0, 1, \dots, \frac{m_S-1}{2} \right\} & (\langle P \rangle < 0) \end{cases}. \quad (25)$$

The numerical values of $p_S(m_S, \langle P \rangle)$ are presented in Fig. 4. We observe that even for a relatively small number of shots such as $m_S = 17$, the probability p' is as high as ≥ 0.8 for

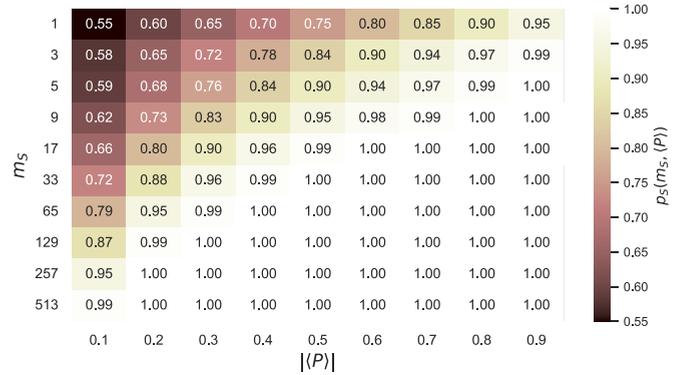


FIG. 4. Values of the probability $p_S(m_S, \langle P \rangle)$ for different measurement shots m_S and absolute values of Pauli expectation values $|\langle P \rangle|$.

$\langle P \rangle \geq 0.2$. In our numerical simulations presented in the next section, we take these values into account for determining the value of m_S in the JBM-VQE algorithm.

Finally, it is worthwhile to point out that m_{th}^{SM} and m_S considered here are defined for a single Pauli operator P , so if there are M Pauli operators, it would require $M \cdot m_{th}^{SM}$ and $M \cdot m_S$ shots to estimate all expectation values and their sign. Similarly, it is shown [18] that $\mathcal{O}(\ln(M)/\epsilon^4)$ measurements are need to estimate $|\langle P_1 \rangle|, \dots, |\langle P_M \rangle|$ within the error ϵ simultaneously, so it would require $\ln(M) \cdot m_{th}^{JBM}$ to estimate all absolute values of them.

IV. NUMERICAL DEMONSTRATION

In this section, we present numerical demonstrations of JBM-VQE for finding the ground states of quantum chemistry Hamiltonians. We first examine an example where the signs of Pauli expectation values, $s_j(\theta)$, do not undergo frequent flipping during the course of standard VQE optimization. This observation leads to the anticipation that JBM-VQE can reduce the number of shots to optimize the parameters with a specified level of accuracy. We then compare JBM-VQE and the conventional VQE by taking various small molecules as examples. The parameter optimization in JBM-VQE proceeds more rapidly than in the conventional VQE, according to the metric we have established for the early stage of the optimization. The advantage of JBM-VQE becomes increasingly apparent as system sizes (the number of qubits) grow larger.

A. Signs of Pauli expectation values during parameter optimization

We conduct a numerical simulation of the conventional VQE to prepare the ground state of the H_2 molecular Hamiltonian whose bond distance is 0.74 Å. The quantum chemistry Hamiltonian of the form (1) is constructed by using the Hartree-Fock molecular orbitals with the STO-3G basis set. The Hamiltonian is then mapped to the qubit form (2) via the Jordan-Wigner transformation. The resulting qubit Hamiltonian consists of $n = 4$ qubits and 14 nonidentity Pauli operators. The construction of the Hamiltonian was implemented using the numerical libraries PYSCF [32] and OPENFERMION [33]. For the VQE ansatz state, we employ

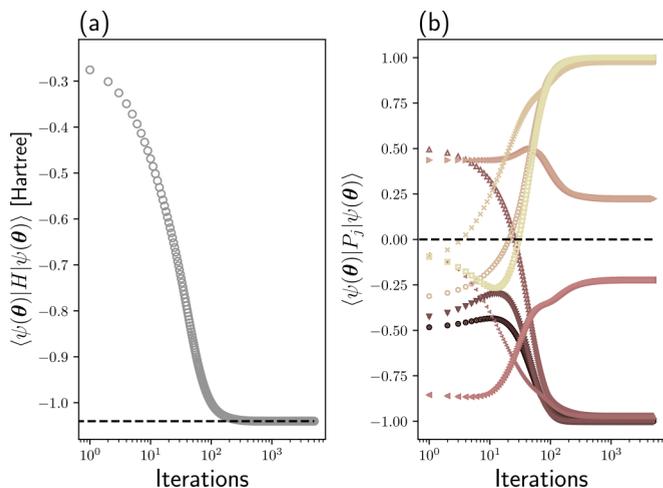


FIG. 5. Expectation values of (a) the Hamiltonian H and (b) the Pauli operators P_j included in the Hamiltonian during the VQE optimization for finding the ground state of the H_2 molecular Hamiltonian. There are 14 nonidentity Pauli operators included in the Hamiltonian H .

the symmetry-preserving ansatz [34,35] $|\psi(\theta)\rangle = U(\theta)|n_e\rangle$, where $U(\theta)$ represents a variational quantum circuit depicted in Fig. 7 of Appendix B and $|n_e\rangle$ is a computational basis state $|0 \dots 01 \dots 1\rangle$ with $n - n_e$ “0”s and n_e “1”s with n_e denoting the number of electrons (for H_2 molecule, $n_e = 2$). It has eight parameters in total, with initial values sampled randomly from $[0, \pi/5]$. Here we choose to start from a random initial state instead of the Hartree-Fock state. This decision is grounded in the understanding that the Hartree-Fock initialization might limit the VQE’s ability to explore more correlated states and lead VQE to converge to a local minimum. Furthermore, it is recognized that for certain molecules such as diatomic molecules (e.g., the nitrogen N_2) with large bond distances, the HF state is not a good approximation of the true ground state. The circuit parameters are updated using the gradient descent method with a learning rate $\eta = 0.02$. We simulate the energy expectation values $E(\theta) = \langle \psi(\theta) | H | \psi(\theta) \rangle$ exactly without assuming any statistical error and noise sources by the numerical library QULACS [36].

The result is presented in Fig. 5. After 3000 iterations, the VQE algorithm successfully finds the exact ground state. Importantly, we observe that the signs of the expectation values of the most Pauli operators $\langle \psi(\theta) | P_j | \psi(\theta) \rangle$ either remain unchanged or exhibit only a single change during the VQE optimization. This observation motivates us to accelerate VQE by estimating only the absolute values of $\langle \psi(\theta) | P_j | \psi(\theta) \rangle$ in the majority of iterations. We note that expectation values of some Pauli operators (e.g., $\langle X_1 \otimes Y_2 \otimes Y_3 \otimes X_4 \rangle$ and $\langle Y_1 \otimes X_2 \otimes X_3 \otimes Y_4 \rangle$) are the same throughout the optimization process due to the ansatz symmetry.

B. Comparison of JBM-VQE with the conventional VQE for various small molecules

Here we present a numerical comparison of the measurement cost between JBM-VQE and the conventional VQE. We consider eight molecular systems: H_2 , H_3^+ , H_4 , H_5^+ , LiH, H_2O ,

NH_3 , and BeH_2 . The geometries of these molecules are provided in Table I of Appendix B. For the latter four molecules (LiH, H_2O , NH_3 , and BeH_2), the active space approximation of four orbitals are taken so that the Hamiltonians become eight-qubit ones.

Similar to the calculations in the previous subsection, we construct quantum chemistry Hamiltonians of the form (1) using Hartree-Fock molecular orbitals with the STO-3G basis set for the eight molecules under consideration. The Jordan-Wigner transformation is employed to obtain the qubit Hamiltonian. The symmetry-preserving ansatz (described in Fig. 7 of Appendix B) is employed for both JBM-VQE and the conventional VQE. The ansatz depths for molecules H_2 , H_3^+ , H_4 , H_5^+ , LiH, H_2O , NH_3 , and BeH_2 are 2, 3, 8, 14, 3, 5, 6, and 5, respectively. Initial parameters are uniformly sampled from $[0, \pi/5]$ for all cases.

We evaluate the expectation value $E(\theta) = \langle \psi(\theta) | H | \psi(\theta) \rangle$ and its gradient as follows. In JBM-VQE, the signs $\{s_j(\theta)\}_{j=1}^M$ and $\{s_j(\theta_{\pm}^{(l)})\}_{j=1}^M$ are estimated by the projective measurement of the Pauli operators P_j included in the Hamiltonian. We employ the qubitwise commuting (QWC) grouping [11,37] to make groups of simultaneously measurable Pauli operators, which requires additional $\mathcal{O}(n)$ one-qubit gates and reduces the number of the total shots for estimation (see Appendix B for details). We allocate the same number of shots for all generated groups, denoting the number of shots for each group as m_S . The absolute values $\{|\langle P_j \rangle_{\theta}|\}_{j=1}^M$ are estimated using the joint Bell measurement with $2n$ qubits, as described in Sec. II. The number of shots for the joint Bell measurement is denoted by m . In the conventional VQE, we estimate the energy expectation value by directly estimating $\langle \psi(\theta) | P_j | \psi(\theta) \rangle$ with the projective measurement of P_j . We also employ QWC grouping to reduce the number of shots and allocate the same amount of shots for all groups. The number of shots for each group is denoted by m_{VQE} . For both JBM-VQE and the conventional VQE, all outcomes of the measurement shots (bitstrings) are simulated without considering any noise sources using the numerical library QULACS [36].

The learning parameter for the gradient descent is set to $\eta = 0.02$ for both JBM-VQE and the conventional VQE. As mentioned in Sec. II, we set $\alpha = \pi/4$ in the parameter shift rule (12) to estimate the gradient because of the reason described in Appendix B. We consider the threshold of $p_{th} = 0.9$ and $\tau_{th} = 0.05$ for estimating the Pauli expectation values. According to Figs. 2–4, the numbers of the shots are taken as $m = 4159$, $m_S = 513$, and $m_{VQE} = 739$. Note that the standard deviations (or fluctuations) in estimating the energy expectation values are not strictly the same between JBM-VQE and the conventional VQE because the values of the thresholds discussed in Sec. III are for a single Pauli operator. The sign-updating period of JBM-VQE is fixed at $T_S = 30$ for all simulations. For every 200 iterations, the mean estimated energy is computed. Should the energy fail to decrease by a minimum of 0.001 following 200 iterations, the optimization process is terminated.

The results are shown in Fig. 6. For all molecules, JBM-VQE exhibits a faster decrease of energy with fewer measurement shots. The sudden jump of the estimated energy in JBM-VQE appears to be due to the update of the signs. The observed bias in our numerical experiment is not so

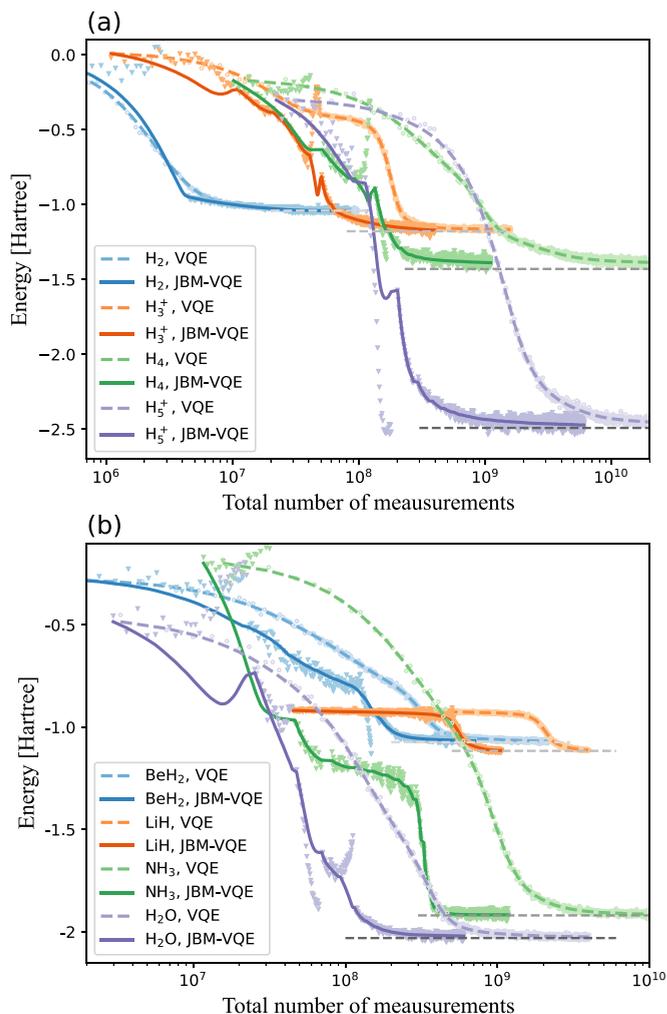


FIG. 6. Comparison of optimization curves for the conventional VQE and the JBM-VQE algorithms applied to molecular Hamiltonians of (a) H_2 , H_3^+ , H_4 , H_5^+ and (b) LiH , H_2O , NH_3 , and BeH_2 . The dashed lines and circles denote the actual energy (exact energy expectation value for the parameters at each iteration) and the estimated energy in the conventional VQE, respectively. The solid lines and triangles represent the actual energy and the estimated energy in JBM-VQE, respectively.

significant to initiate the optimization. It is also observed that the advantage of JBM-VQE becomes more evident as the system size increases from four qubits (H_2) to ten qubits (H_5^+). To quantify the improvement of the measurement cost of JBM-VQE over the conventional VQE, we count the number of the shots required to achieve a result with energy lower than the Hartree-Fock energy. The ratio between such shots for the conventional VQE over JBM-VQE is calculated by averaging the results of at least 50 independent numerical simulations with different initial parameters, which yields (n denotes the number of qubits)

$$\begin{aligned}
 H_2(n=4) : 1.44, & \quad H_3^+(n=6) : 3.45, \\
 H_4(n=8) : 9.00, & \quad H_5^+(n=10) : 9.24, \\
 BeH_2(n=8) : 3.70, & \quad LiH(n=8) : 4.60, \\
 NH_3(n=8) : 11.60, & \quad H_2O(n=8) : 2.24. \quad (26)
 \end{aligned}$$

These numbers illustrate the reduction of the measurement cost of JBM-VQE.

Finally, we comment on the choice of the Hartree-Fock energy to define the number of shots for optimizing the parameters of the ansatz although the Hartree-Fock energy can be obtained by the initial state of the ansatz without applying $U(\theta)$. This is because the purpose of this numerical illustration is to simply show the reduction of the number of shots in the optimization process of (JBM-)VQE. Furthermore, it is important to acknowledge that the ansatz selected in this study does not guarantee the accurate representation of the ground-state for the given Hamiltonian. This limitation arises due to the expressibility constraints of the ansatz and make it difficult to define the number of shots for the optimization by using the exact ground state energy.

V. DISCUSSION

In this section, we discuss several aspects of the JBM-VQE method.

First, JBM-VQE can be expected to present a significant advantage over conventional VQE as the number of qubits n increases, as demonstrated in the hydrogen chain examples from the previous section. This is because the number of distinct quantum circuits to evaluate the energy expectation value scales at least $\mathcal{O}(n^2)$ (naively $\mathcal{O}(n^4)$) in the conventional VQE while that of JBM-VQE does $\mathcal{O}(1)$ when we skip the evaluation of the sign of the Pauli expectation values. Although a small number of distinct quantum circuits does not directly imply the efficiency of the evaluation, we anticipate a reduction of the total number of shots in JBM-VQE, as illustrated in the previous section.

Second, we propose using JBM-VQE as an initial optimizer when the accuracy of the energy estimate is not as high, as demonstrated in the numerical simulation in Sec. IV, because of the following two reasons. The first reason is that the estimate of the energy in JBM-VQE is biased, i.e., the average of $\tilde{E}(\theta)$ [Eqs. (11) and (15)] is not equal to the true value $E(\theta)$ (note that the estimate $\tilde{E}(\theta)$ is a random variable whose probability distribution is determined by that of the outcomes of the joint Bell measurement). The bias has already been seen in the numerical simulation (Fig. 6), where the distribution of the dots (the energy estimates) is not centered at the corresponding line (the exact value of the energy at those parameters) for several molecules. The second reason is the scaling of the number of shots in the joint Bell measurement with respect to the estimation error. As shown in Ref. [18], the number of shots required to estimate all M absolute values $|\langle P_1 \rangle|, \dots, |\langle P_M \rangle|$ with the error ϵ by the joint Bell measurement is $\mathcal{O}(\ln(M)/\epsilon^4)$. This ϵ^{-4} scaling can be problematic when we take small ϵ . These two properties of JBM-VQE can pose challenges when optimizing the ansatz parameters with high accuracy required quantum chemistry, such as the so-called chemical accuracy 10^{-3} Hartree.

Thirdly, the JBM-VQE method is specifically designed for NISQ devices where noise is an inevitable factor. The JBM approach inherently generates entanglement between two state copies, necessitating twice the number of gates compared to conventional methods. Considering an error rate per gate denoted as p and a total gate count of K , the application

of the probabilistic error cancellation [38], a quantum error mitigation strategy, incurs a sampling cost expressed as e^{4Kp} [39]. In the context of the JBM-VQE, where the gate number K is nearly doubled, this cost scales to e^{8Kp} . Consequently, the ratio of these costs is e^{4Kp} . Practically, for the usual VQE with the probabilistic error mitigation to work, the value of Kp is limited to at most $\mathcal{O}(1)$, implying that the escalated sampling cost of probabilistic error cancellation in JBM-VQE compared to conventional VQE methods, quantified as e^{4Kp} , does not scale with the qubit count n .

Lastly, we point out that the JBM-VQE protocol is flexible and can be combined with various variational algorithms which require the evaluation of many Pauli expectation values to optimize the parameters. For example, sophisticated optimizers like Adam [40] or the ones more tailored for VQE [41,42] can be used in JBM-VQE although we employ the plain-vanilla gradient decent in the numerical simulations. One can simply skip the evaluation of the signs of Pauli expectation values at some iterations (parameter updates) and perform the joint Bell measurement to estimate the energy and its gradient that are fed into the optimizers.

VI. SUMMARY AND OUTLOOK

In this study, we introduce a protocol designed to accelerate the Variational Quantum Eigensolver (VQE) algorithm for determining the ground state of molecular Hamiltonians. Our approach employs the joint Bell measurement to estimate the energy expectation value and its gradient of n qubit systems with $\mathcal{O}(n^0) = \mathcal{O}(1)$ distinct quantum circuits in the majority of iterations, under the assumption that the signs of Pauli expectation values do not frequently change during optimization. In contrast, the conventional VQE necessitates at least $\mathcal{O}(n^2)$ distinct quantum circuits for estimating the energy and gradient in each iteration. We conducted numerical simulations of various small molecular Hamiltonians, demonstrating that our proposed protocol effectively reduces the number of measurement shots required to optimize the ansatz parameters to a certain level. JBM-VQE holds promise for application in a broad range of near-term quantum algorithms that depend on Pauli expectation value estimations.

In future work, it is interesting to apply our protocol to various NISQ algorithms other than the simple VQE presented in this study. For example, it is possible to combine JBM-VQE with the variants of VQE for excited states [43,44], quantum imaginary-time evolution [45–48], and algorithmic error mitigation schemes [49,50]. There is reason to believe that in protocols involving estimating quantum computed moments $\langle \psi | H^k | \psi \rangle$ like the Lanczos-inspired error mitigation scheme [50], variance-VQE (where $k = 2$) [51], and variance extrapolation [49], the acceleration from joint Bell measurement become more significant since the number of Pauli terms to be evaluated is of order $\mathcal{O}(n^{4k})$. Furthermore, it would be beneficial to explore the integration of our protocol with two notable strategies - the α -VQE [52] and the parallelized VQE [53]. Both of these approaches have demonstrated promising results in improving the efficiency of VQE optimization.

ACKNOWLEDGMENT

We thank Ridwan Sakidja, Tiglet Besara, Daniel Miller, and Aonan Zhang for helpful discussions.

APPENDIX A: BIAS IN JBM-VQE

In this section, we discuss the bias in estimating $\langle P_j \rangle_\theta = \langle \psi(\theta) | P_j | \psi(\theta) \rangle$ by the joint Bell measurement for a given state $|\psi(\theta)\rangle$ and Pauli operator P_j . The joint Bell measurement yields measurement outcomes that obey the binomial distribution whose mean is $\langle P_j \rangle_\theta^2$. When the number of shots m for the joint Bell measurement is large, the bias of $\langle P_j \rangle_\theta$ can be assessed by approximating the binomial distribution by the normal distribution:

$$\begin{aligned} B(\langle P_j \rangle_\theta) &= \mathbb{E}^{(m)}(\langle P_j \rangle_\theta) - |\langle P_j \rangle_\theta| \\ &\approx \frac{1}{\sigma \sqrt{2\pi}} \int_0^1 dy \sqrt{y} \exp \\ &\quad \times \left(-\frac{(y - \langle P_j \rangle_\theta^2)^2}{2\sigma^2} \right) - |\langle P_j \rangle_\theta|, \end{aligned} \quad (\text{A1})$$

where $\mathbb{E}^{(m)}(\dots)$ denotes the expectation value for m shots measurement and $\sigma = \sqrt{(1 - \langle P_j \rangle_\theta^4)/m}$ is the standard deviation of the measurement outcomes. The standard deviation gets small as m grows so that the bias will vanish. For a sufficiently large m , we further approximate the normal distribution by the uniform distribution across the interval $[\langle P_j \rangle_\theta^2 - \sigma, \langle P_j \rangle_\theta^2 + \sigma]$ to discuss the scaling of the bias. When $\langle P_j \rangle_\theta = 0$, we have

$$B(\langle P_j \rangle_\theta) \approx \int_0^\sigma dy \frac{\sqrt{y}}{2\sigma} = \frac{\sqrt{\sigma}}{3} \sim m^{-1/4}. \quad (\text{A2})$$

When $\langle P_j \rangle_\theta \neq 0$ and m is large enough to satisfy $\langle P_j \rangle_\theta^2 > \sigma$, we have

$$\begin{aligned} B(\langle P_j \rangle_\theta) &\approx \int_{\langle P_j \rangle_\theta^2 - \sigma}^{\langle P_j \rangle_\theta^2 + \sigma} dy \frac{\sqrt{y}}{2\sigma} - |\langle P_j \rangle_\theta| \\ &\approx - \int_{\langle P_j \rangle_\theta^2 - \sigma}^{\langle P_j \rangle_\theta^2 + \sigma} dy \frac{(y - \langle P_j \rangle_\theta^2)^2}{16|\langle P_j \rangle_\theta|^3 \sigma} \\ &= -\frac{\sigma^2}{24|\langle P_j \rangle_\theta|^3} \sim -m^{-1}, \end{aligned} \quad (\text{A3})$$

where we used the Taylor expansion of \sqrt{y} at $y = \langle P_j \rangle_\theta^2 > 0$:

$$\sqrt{y} \approx |\langle P_j \rangle_\theta| + \frac{1}{2|\langle P_j \rangle_\theta|} (y - \langle P_j \rangle_\theta^2) - \frac{1}{8|\langle P_j \rangle_\theta|^3} (y - \langle P_j \rangle_\theta^2)^2. \quad (\text{A4})$$

These equations lead to Eq. (9). To mitigate this bias, one prospective approach entails choosing multiple finite values of m and subsequently extrapolating m to $+\infty$, guided by the decay rate of m^{-1} (or $m^{-1/4}$). We leave it for future investigation.

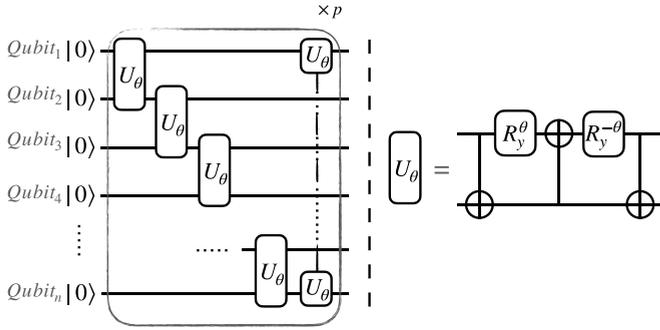


FIG. 7. (Left) The symmetry preserving real-valued ansatz with n qubits and p layers for constructing the ground state of a molecular Hamiltonian. (Right) The $SO(4)$ block with three CNOT gates and two R_y rotations with opposite rotation angles.

APPENDIX B: DETAILS OF NUMERICAL CALCULATION

1. Details of ansatz and molecules

In the numerical simulations presented in Sec. IV, we employ the symmetry-preserving real-valued ansatz depicted in Fig. 7. We use molecules with geometries and active spaces listed in Table I. Some of these geometries are chosen as the stable structures at the level of Hartree-Fock/STO-3G, as referenced from the CCCBDB database.

2. Qubit wise commuting grouping

For the numerical simulations shown in Fig. 6, we employ the qubitwise commuting (QWC) grouping [11,37] to evaluate the expectation values of the Hamiltonian H . Consider two n -qubit Pauli operators, $Q_1 = \sigma_1^{(1)} \otimes \sigma_1^{(2)} \otimes \cdots \otimes \sigma_1^{(n)}$ and $Q_2 = \sigma_2^{(1)} \otimes \sigma_2^{(2)} \otimes \cdots \otimes \sigma_2^{(n)}$, where $\sigma_{1,2}^{(j)}$ is a single qubit Pauli operator I, X, Y, Z acting on j th qubit. We define that Q_1 and Q_2 are qubitwise commuting if and only if $\sigma_1^{(j)} \sigma_2^{(j)} = \sigma_2^{(j)} \sigma_1^{(j)}$ for all $j \in \{1, 2, \dots, n\}$. The Pauli operators $\{Q_1, Q_2, \dots\}$ that are mutually qubitwise commuting can be simultaneously measured by applying an additional quantum circuit consisting of $O(n)$ one-qubit gates to the state $|\psi\rangle$. Therefore we can reduce the number of distinct quantum circuit to evaluate the expectation value of the Hamiltonian $H = \sum_{j=1}^M \lambda_j P_j$ by dividing the Pauli operators $\{P_j\}_{j=1}^M$ into groups of mutually qubitwise commuting operators. It should be noted that various grouping methods have been explored in the literature, including those considering usual

commutativity and anticommutativity of the Pauli operators [14,54–57]. The QWC grouping method is adopted in our simulation because it does not require deep and complicated quantum circuits to perform simultaneous measurements of the operators in each group.

The greedy search with sorting the Pauli operators [58] is employed when grouping the Pauli operators included in the Hamiltonian. We sort the M Pauli operators $\{P_j\}_{j=1}^M$ in the Hamiltonian by descending order of the absolute values of the coefficients $|\lambda_j|$. The sorted operators are denoted $\{P'_j\}_{j=1}^M$. We assign P'_1 to the first group. For $j = 2, 3, \dots, k$, if P'_j qubitwise commutes with all Pauli operators in an existing group, it is assigned to that group. If P'_j does not qubitwise commute with any Pauli operators in an existing group, a new group is created to house P'_j . This procedure is repeated until all Pauli operators are assigned to a group.

3. Choice of α in parameter-shift rule

Here, we explain the choice of $\alpha = \pi/4$ in the parameter shift rule (12) in our numerical calculation. The parameter shift rule is related to the fact [59] that the functional form of the Pauli expectation value with respect to θ_l is a trigonometric function when other circuit parameters remain constant: in the simplest cases where the parameter θ_l in the ansatz $|\psi(\theta)\rangle$ is an angle of some Pauli rotation gate $e^{-i\frac{\theta_l}{2}Q_l}$ satisfying $Q_l^2 = I$, we have

$$\langle P_j \rangle_{\theta} = a_l \cos(\theta_l - \phi_l) + c_l, \quad (\text{B1})$$

where a_l , ϕ_l , and c_l are real coefficients depending on $\theta_1, \dots, \theta_{l-1}, \theta_{l+1}, \dots, \theta_{N_\theta}$. The parameter shift rule can improve the signal-to-noise ratio when we take larger α . For example, if we use small α like $\alpha = 0.01$, the values $\langle P_j \rangle_{\theta+\alpha\delta_l}$ and $\langle P_j \rangle_{\theta-\alpha\delta_l}$ become almost the same so that a lot of measurement shots are required to estimate the gradient ($\propto \langle P_j \rangle_{\theta+\alpha\delta_l} - \langle P_j \rangle_{\theta-\alpha\delta_l}$) with high accuracy. This is why $\alpha = \pi/2$, the largest α considering the periodicity of the function (B1), is typically used in the literature [5]. In our numerical calculation, we observed that some Pauli terms exhibit expectation values approaching $\langle P_j \rangle_{\theta} \approx \pm 1$ in the late stages of optimization, or at the vicinity of the exact ground state. For these Pauli terms, $|\langle P_j \rangle_{\theta \pm (\pi/2)\delta_l}|$ sometimes becomes close to zero [e.g., when $a_l = 1$, $c_l = 0$ in (B1)] and the estimation of it by the joint Bell measurement requires a large number of shots (see Table 3). Consequently, to strike a balance, we adopt $\alpha = \pi/4$ for our numerical computations presented

TABLE I. Geometries and active spaces (if any) of molecules. n is the number of qubits of the molecular Hamiltonian. “(X, (x, y, z))” denotes three dimensional coordinates x, y, z of atom X in units of Å.

Molecule (active space)	n	Geometry
H ₂	4	(H, (0, 0, 0)), (H, (0, 0, 0.74))
H ₃ ⁺	6	(H, (0, 0, 0)), (H, (0, 0, 0.85)), (H, (0, 0.74, 0.43))
H ₄	8	(H, (0, 0, 0)), (H, (0, 0, 1.2)), (H, (0, 0, 2.4)), (H, (0, 0, 3.6))
H ₅ ⁺	10	(H, (0, 0, 0)), (H, (0, 0.74, 0.43)), (H, (0, 0.74, -0.43)), (H, (-0.74, 0, 0.43)), (H, (-0.74, 0, -0.43))
LiH(4o,2e)	8	(Li, (0, 0, 0)), (H, (0, 0, 1.59))
H ₂ O(4o,4e)	8	(O, (0, 0, 0.137)), (H, (0, 0.76, 0.50)), (H, (0, -0.76, -0.50))
NH ₃ (4o,4e)	8	(N, (0, 0, 0)), (H, (0, 0, 1.01)), (H, (0.95, 0, -0.34)), (H, (-0.48, -0.82, -0.34))
BeH ₂ (4o,4e)	8	(Be, (0, 0, 0)), (H, (0, 0, 1.33)), (H, (0, 0, -1.33))

in Sec. IV. In fact, for Pauli terms satisfying $\langle P_j \rangle_\theta = \pm 1$, $|\langle P_j \rangle_{\theta \pm \pi/4\delta_l}|$ is farther away from zero than $|\langle P_j \rangle_{\theta \pm \pi/2\delta_l}|$ and guaranteed to be larger than $\frac{\sqrt{2}}{2}$,

$$|\langle P_j \rangle_{\theta \pm \pi/2\delta_l}| = |a_l \cdot \frac{\sqrt{2}}{2} + c_l| \geq \frac{\sqrt{2}}{2}, \quad (\text{B2})$$

because $|a_l| \leq 1$ and $|a_l| + |c_l| \leq 1$ hold due to $|\langle P_j \rangle_\theta| \leq 1$ for any θ .

APPENDIX C: COMPARISON OF JBM-VQE AND VQE FOR THE GROUND-STATE DISTRIBUTION OF $\langle P_j \rangle$

The shot thresholds defined in Sec. III and utilized in Sec. IV was based on the assumption that $\langle P_j \rangle$ adheres to a uniform distribution within the range of -1 and 1 , chosen for its simplicity and relevance to the study. The numerical results in Sec. IV revealed that JBM-VQE significantly accelerates the optimization process compared to the conventional VQE approach. However, it is important to note that in practical scenarios, a variety of factors—such as different ansatzes, initial states, and optimization landscapes—can lead to nonuniform distributions. Consequently, the thresholds for the number of quantum measurement shots may vary. Despite these potential variations, the superiority of our proposed protocol remains evident across any distribution of $\langle P_j \rangle$, particularly when dealing with a larger number of Pauli operators and seeking less precise estimations.

To illustrate this, we explore a case study involving the H_4 molecule, adopting a more realistic distribution scenario. Here, we consider the actual distribution of the Pauli expectation values for the exact ground state ($|\psi_g\rangle$). These expectation values are represented as $\lambda_1 = \langle \psi_g | P_1 | \psi_g \rangle$, $\lambda_2 = \langle \psi_g | P_2 | \psi_g \rangle$, and so on, up to $\lambda_{184} = \langle \psi_g | P_{184} | \psi_g \rangle$. For both the conventional VQE and JBM-VQE, the average probabilities of estimating the expectation value within an additive error τ_{th} are calculated using the following formulas:

$$p^{(\text{av})}(m, \tau_{\text{th}}) = \frac{1}{184} \sum_j p(m, \tau_{\text{th}}, \lambda_j) \quad (\text{C1})$$

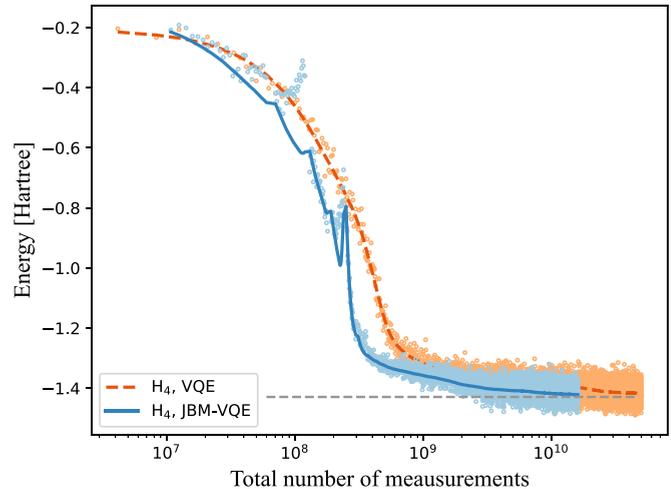


FIG. 8. Comparison between the conventional VQE and the JBM-VQE algorithms applied to the molecular Hamiltonian of H_4 for the ground-state distribution of $\langle P_j \rangle$. (a) Total number of measurement shots versus the number of iterations. (b) Optimization curves. The dashed lines and circles denote the actual energy (exact energy expectation value for the parameters at each iteration) and the estimated energy in the conventional VQE, respectively. The solid lines and triangles represent the actual energy and the estimated energy in JBM-VQE, respectively.

and

$$q^{(\text{av})}(m, \tau_{\text{th}}) = \frac{1}{184} \sum_j q(m, \tau_{\text{th}}, \lambda_j), \quad (\text{C2})$$

where $p(m, \tau_{\text{th}}, \lambda_j)$ and $q(m, \tau_{\text{th}}, \lambda_j)$ are defined as per Eqs. (17) and (20), respectively. For $\tau_{\text{th}} = 0.1$ and $p_{\text{th}} = 0.9$, the shot thresholds, as defined in Definitions 1 and 2, are calculated to be 235 and 6576, respectively. In this experiment, we continue to sample initial circuit parameters from the interval $[0, \pi/5]$ and employ a gradient descent optimization method with a learning rate of $\eta = 0.02$. Figure 8 depicts the optimization energy against the number of measurements. The results demonstrate that, for the actual ground-state distribution of $\langle P_j \rangle$, JBM-VQE maintains a noticeable speed advantage over the conventional VQE approach.

- [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. Brandao, D. A. Buell *et al.*, Quantum supremacy using a programmable superconducting processor, *Nature (London)* **574**, 505 (2019).
- [3] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu *et al.*, Quantum computational advantage using photons, *Science* **370**, 1460 (2020).
- [4] L. S. Madsen, F. Laudenbach, M. F. Askarani, F. Rortais, T. Vincent, J. F. Bulmer, F. M. Miatto, L. Neuhaus, L. G. Helt, M. J. Collins *et al.*, Quantum computational advantage with a programmable photonic processor, *Nature (London)* **606**, 75 (2022).

- [5] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, Quantum circuit learning, *Phys. Rev. A* **98**, 032309 (2018).
- [6] M. Cerezo, A. Poremba, L. Cincio, and P. J. Coles, Variational quantum fidelity estimation, *Quantum* **4**, 248 (2020).
- [7] R. Chen, Z. Song, X. Zhao, and X. Wang, Variational quantum algorithms for trace distance and fidelity estimation, *Quantum Sci. Technol.* **7**, 015019 (2022).
- [8] C. Cao, C. Zhang, Z. Wu, M. Grassl, and B. Zeng, Quantum variational learning for quantum error-correcting codes, *Quantum* **6**, 828 (2022).
- [9] 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).

- [10] J. Tilly, H. Chen, S. Cao, D. Picozzi, K. Setia, Y. Li, E. Grant, L. Wossnig, I. Rungger, G. H. Booth, and J. Tennyson, The variational quantum eigensolver: A review of methods and best practices, *Phys. Rep.* **986**, 1 (2022).
- [11] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, *Nature (London)* **549**, 242 (2017).
- [12] J. F. Gonthier, M. D. Radin, C. Buda, E. J. Duskocil, C. M. Abuan, and J. Romero, Measurements as a roadblock to near-term practical quantum advantage in chemistry: Resource analysis, *Phys. Rev. Res.* **4**, 033154 (2022).
- [13] P. Gokhale, O. Angiuli, Y. Ding, K. Gui, T. Tomesh, M. Suchara, M. Martonosi, and F. T. Chong, $O(N^3)$ measurement cost for variational quantum eigensolver on molecular hamiltonians, *IEEE Trans. Quantum Eng.* **1**, 1 (2020).
- [14] A. Zhao, A. Tranter, W. M. Kirby, S. F. Ung, A. Miyake, and P. J. Love, Measurement reduction in variational quantum algorithms, *Phys. Rev. A* **101**, 062322 (2020).
- [15] X. Bonet-Monroig, R. Babbush, and T. E. O'Brien, Nearly optimal measurement scheduling for partial tomography of quantum states, *Phys. Rev. X* **10**, 031064 (2020).
- [16] W. Inoue, K. Aoyama, Y. Teranishi, K. Kanno, Y. O. Nakagawa, and K. Mitarai, Almost optimal measurement scheduling of molecular Hamiltonian via finite projective plane, *Phys. Rev. Res.* **6**, 013096 (2024)
- [17] H.-Y. Huang, M. Broughton, J. Cotler, S. Chen, J. Li, M. Mohseni, H. Neven, R. Babbush, R. Kueng, J. Preskill, and J. R. McClean, Quantum advantage in learning from experiments, *Science* **376**, 1182 (2022).
- [18] H.-Y. Huang, R. Kueng, and J. Preskill, Information-theoretic bounds on quantum advantage in machine learning, *Phys. Rev. Lett.* **126**, 190505 (2021).
- [19] Z. Jiang, A. Kalev, W. Mruczkiewicz, and H. Neven, Optimal fermion-to-qubit mapping via ternary trees with applications to reduced quantum states learning, *Quantum* **4**, 276 (2020).
- [20] T. Haug and M. S. Kim, Scalable measures of magic resource for quantum computers, *PRX Quantum* **4**, 010301 (2023).
- [21] J.-F. Tang, Z. Hou, J. Shang, H. Zhu, G.-Y. Xiang, C.-F. Li, and G.-C. Guo, Experimental optimal orienteering via parallel and antiparallel spins, *Phys. Rev. Lett.* **124**, 060502 (2020).
- [22] F. Xu, X.-M. Zhang, L. Xu, T. Jiang, M.-H. Yung, and L. Zhang, Experimental quantum target detection approaching the fundamental helstrom limit, *Phys. Rev. Lett.* **127**, 040504 (2021).
- [23] I. Hamamura and T. Imamichi, Efficient evaluation of quantum observables using entangled measurements, *npj Quantum Inf.* **6**, 56 (2020).
- [24] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, *Nat. Phys.* **16**, 1050 (2020).
- [25] A. Zhao, N. C. Rubin, and A. Miyake, Fermionic partial tomography via classical shadows, *Phys. Rev. Lett.* **127**, 110504 (2021).
- [26] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory* (Courier Corporation, 2012).
- [27] T. Helgaker, P. Jorgensen, and J. Olsen, *Molecular Electronic Structure Theory* (Wiley, Hoboken, NJ, 2013).
- [28] P. Jordan and E. Wigner, Über das paulische äquivalenzverbot, *Z. Phys.* **47**, 631 (1928).
- [29] J. T. Seeley, M. J. Richard, and P. J. Love, The bravyyi-kitaev transformation for quantum computation of electronic structure, *J. Chem. Phys.* **137**, 224109 (2012).
- [30] S. B. Bravyi and A. Y. Kitaev, Fermionic quantum computation, *Ann. Phys.* **298**, 210 (2002).
- [31] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, Evaluating analytic gradients on quantum hardware, *Phys. Rev. A* **99**, 032331 (2019).
- [32] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters, and G. K.-L. Chan, Pyscf: the python-based simulations of chemistry framework, *WIREs Comput. Mol. Sci.* **8**, e1340 (2018).
- [33] J. R. McClean, N. C. Rubin, K. J. Sung, I. D. Kivlichan, X. Bonet-Monroig, Y. Cao, C. Dai, E. S. Fried, C. Gidney, B. Gimby, P. Gokhale, T. Häner, T. Hardikar, V. Havlíček, O. Higgott, C. Huang, J. Izaac, Z. Jiang, X. Liu, S. McArdle *et al.*, Openfermion: the electronic structure package for quantum computers, *Quantum Sci. Technol.* **5**, 034014 (2020).
- [34] 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).
- [35] Y. Ibe, Y. O. Nakagawa, N. Earnest, T. Yamamoto, K. Mitarai, Q. Gao, and T. Kobayashi, Calculating transition amplitudes by variational quantum deflation, *Phys. Rev. Res.* **4**, 013173 (2022).
- [36] Y. Suzuki, Y. Kawase, Y. Masumura, Y. Hiraga, M. Nakadai, J. Chen, K. M. Nakanishi, K. Mitarai, R. Imai, S. Tamiya, T. Yamamoto, T. Yan, T. Kawakubo, Y. O. Nakagawa, Y. Ibe, Y. Zhang, H. Yamashita, H. Yoshimura, A. Hayashi, and K. Fujii, QULACS: a fast and versatile quantum circuit simulator for research purpose, *Quantum* **5**, 559 (2021).
- [37] V. Verteletskyi, T.-C. Yen, and A. F. Izmaylov, Measurement optimization in the variational quantum eigensolver using a minimum clique cover, *J. Chem. Phys.* **152**, 124114 (2020).
- [38] K. Temme, S. Bravyi, and J. M. Gambetta, Error mitigation for short-depth quantum circuits, *Phys. Rev. Lett.* **119**, 180509 (2017).
- [39] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O'Brien, Quantum error mitigation, *Rev. Mod. Phys.* **95**, 045005 (2023).
- [40] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- [41] J. M. Kübler, A. Arrasmith, L. Cincio, and P. J. Coles, An adaptive optimizer for measurement-frugal variational algorithms, *Quantum* **4**, 263 (2020).
- [42] A. Arrasmith, L. Cincio, R. D. Somma, and P. J. Coles, Operator sampling for shot-frugal optimization in variational algorithms, [arXiv:2004.06252](https://arxiv.org/abs/2004.06252) [quant-ph].
- [43] K. M. Nakanishi, K. Mitarai, and K. Fujii, Subspace-search variational quantum eigensolver for excited states, *Phys. Rev. Res.* **1**, 033062 (2019).
- [44] R. M. Parrish, E. G. Hohenstein, P. L. McMahan, and T. J. Martínez, Quantum computation of electronic transitions using a variational quantum eigensolver, *Phys. Rev. Lett.* **122**, 230401 (2019).
- [45] M. Motta, C. Sun, A. T. K. Tan, M. J. O'Rourke, E. Ye, A. J. Minnich, F. G. S. L. Brandão, and G. K.-L. Chan, Determining eigenstates and thermal states on a quantum computer

- using quantum imaginary time evolution, *Nat. Phys.* **16**, 205 (2020).
- [46] S. McArdle, T. Jones, S. Endo, Y. Li, S. C. Benjamin, and X. Yuan, Variational ansatz-based quantum simulation of imaginary time evolution, *npj Quantum Inf.* **5**, 75 (2019).
- [47] S.-N. Sun, M. Motta, R. N. Tazhigulov, A. T. Tan, G. K.-L. Chan, and A. J. Minnich, Quantum computation of finite-temperature static and dynamical properties of spin systems using quantum imaginary time evolution, *PRX Quantum* **2**, 010317 (2021).
- [48] C. Cao, Z. An, S.-Y. Hou, D. L. Zhou, and B. Zeng, Quantum imaginary time evolution steered by reinforcement learning, *Commun. Phys.* **5**, 57 (2022).
- [49] C. Cao, Y. Yu, Z. Wu, N. Shannon, B. Zeng, and R. Joynt, Mitigating algorithmic errors in quantum optimization through energy extrapolation, *Quantum Sci. Technol.* **8**, 015004 (2023).
- [50] P. Suchsland, F. Tacchino, M. H. Fischer, T. Neupert, P. K. Barkoutsos, and I. Tavernelli, Algorithmic error mitigation scheme for current quantum processors, *Quantum* **5**, 492 (2021).
- [51] D.-B. Zhang, B. L. Chen, Z.-H. Yuan, and T. Yin, Variational quantum eigensolvers by variance minimization, *Chinese Phys. B* **31**, 120301 (2022).
- [52] D. Wang, O. Higgott, and S. Brierley, Accelerated variational quantum eigensolver, *Phys. Rev. Lett.* **122**, 140504 (2019).
- [53] L. Mineh and A. Montanaro, Accelerating the variational quantum eigensolver using parallelism, *Quantum Sci. Technol.* **8**, 035012 (2023).
- [54] 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).
- [55] A. Jena, S. Genin, and M. Mosca, Pauli partitioning with respect to gate sets, [arXiv:1907.07859](https://arxiv.org/abs/1907.07859).
- [56] A. F. Izmaylov, T.-C. Yen, R. A. Lang, and V. Verteletskyi, Unitary partitioning approach to the measurement problem in the variational quantum eigensolver method, *J. Chem. Theory Comput.* **16**, 190 (2020).
- [57] T.-C. Yen and A. F. Izmaylov, Cartan subalgebra approach to efficient measurements of quantum observables, *PRX Quantum* **2**, 040320 (2021).
- [58] O. Crawford, B. van Straaten, D. Wang, T. Parks, E. Campbell, and S. Brierley, Efficient quantum measurement of Pauli operators in the presence of finite sampling error, *Quantum* **5**, 385 (2021).
- [59] K. M. Nakanishi, K. Fujii, and S. Todo, Sequential minimal optimization for quantum-classical hybrid algorithms, *Phys. Rev. Res.* **2**, 043158 (2020).