Expressibility of linear combination of ansatz circuits

Peng Wang[1] and Ruyu Yang[2, *]

[1]*Department of Mathematics and Physics, North China Electric Power University, 102206 Beijing, People's Republic of China*
[2]*Graduate School of China Academy of Engineering Physics, Beijing 100193, China*

Variational Quantum Eigensolver is considered promising for medium-scale noisy quantum computers. Expressibility is an important metric for measuring the capability of a variational quantum Ansatz circuit. A commonly used method to increase expressibility is to increase the circuit depth. However, increasing the circuit depth also introduces more noise. We propose to use a linear combination of ansatzes to improve the expressibility of variational circuits, thus avoiding the increase of circuit depth. Concurrently, we introduce a novel measurement strategy that circumvents the necessity for the Hadamard test, thereby significantly diminishing the reliance on two-qubit gates, which are presently the predominant contributors to quantum noise. We also provide a corresponding gradient calculation method, which makes it convenient to update the parameters. Compared with the method of increasing the circuit depth, our method of improving expressibility is more practical. Numerical simulations demonstrate the effectiveness of our method.

I. INTRODUCTION

The physical realization of quantum computing has undergone rapid development in recent years, and scientists are now able to create quantum computers with over one hundred qubits in the laboratory [10]. Due to limitations in the number of qubits and the precision of quantum gates, we are currently in the Noisy Intermediate-Scale Quantum (NISQ) era [1, 21, 33], and there is still a way to go before fault-tolerant quantum computing can be achieved. An important example of an algorithm suitable for NISQ quantum computers is Variational Quantum Eigensolver (VQE) [6, 12, 26, 32, 39, 41]. VQE itself is a broad and active research field. From an application perspective, VQE can be used to solve problems in quantum chemistry [5, 8, 25, 31, 34], condensed matter and materials physics [4, 7, 40, 43], and nuclear physics [23, 35, 40]. VQE itself is not a deterministic algorithm. An important factor affecting the performance of VQE is the design of the ansatz circuit. In order to make VQE achieve better performance, various ansatz have been designed [2, 14–16, 19, 30, 42]. In Ref [37], the authors propose the concept of expressibility of parameterized quantum circuits. In addition to this definition method, there are other definitions of expressibility [9, 13]. Intuitively, expressibility measures the ability of a parameterized quantum circuit to generate quantum states. Generally speaking, when little is known about the target state, the higher expressibility the ansatz circuit is, the more likely it is that it covers the potential target state. Only when the target state is covered it can be possible to solve the corresponding problem through VQE. A natural way to improve the expressibility is to increase the depth of the ansatz circuit [9, 37]. However, in the NISQ era, increasing quantum circuit depth will also increase noise and reduce circuit fidelity. This significantly increases the resources required to mitigate these noises [11, 22, 38].

To overcome this difficulty, we propose to improve expressibility through linear combinations of different ansatzes. Intuitively, using the linear combination of ansatzes (LCA) as a new ansatz can greatly increase the number of trainable parameters without increasing the depth of the circuit. Correspondingly, more quantum states can be reached by changing the parameters than single ansatz (SA). Our numerical results provide evidence for that and exhibit LCA has higher expressibility than SA. For a fixed system size, increasing the ansatzes number in linear combination can rapidly improve the expressibility without increasing the circuit depth, until the ansatzes number reaches the threshold that is linear to the system size.

In the past, the implementation of unitary linear combinations often relied on the Hadamard test. However, the Hadamard test relies on the implementation of the Control-U gate (in the LCA algorithm, U is an ansatz circuit). In the circuit that implements Control-U gates, a single-qubit gate in U becomes a two-qubit gate, and a CNOT gate becomes a Toffoli gate. This means that the Hadamard Test circuit requires a large number of two-qubit gates, which are the main source of noise in quantum circuits. To further reduce the impact of noise, we propose a new sub-algorithm, named phase consistent measurement (PCM) for LCA to replace the Hadamard test. PCM avoids the use of Control-U gates, thereby reducing the use of two-qubit gates. PCM relies on the observation that the phase difference between the final states corresponding to different ansatz does not affect the results of VQE. Because we never derive the phase difference between the final states when calculating the average value of the Hamiltonian, the commonly used parameter shift rules cannot be applied directly [3, 17, 20, 28, 36]. Based on our specific question, we propose a corresponding parameter update scheme that works with common ansatz to reduce the number of measurements required to update parameters.

In Sec. II, we revisit the core concept employed in this article. Subsequently, in Sec. III, we numerically inves-

tigate the Expressibility of LCA. Moving on, Sec. IV demonstrates the implementation of LCA without ancilla qubits, highlighting the reduction in the number of CNOT gates required. Lastly, we provide a comprehensive summary of this work in Section V.

II. PRELIMINARIES

A. The linear combination of ansatz

Ansatz is the key part of VQA. It is a parameterized quantum circuit $U(\boldsymbol{\theta})$ generating a trial state $|\psi_{\boldsymbol{\theta}}\rangle = U(\boldsymbol{\theta})|\phi_0\rangle$, where $\boldsymbol{\theta}$ is a parameter vector $\boldsymbol{\theta} = \{\theta_1, \theta_2, ...\}$, $|\phi_0\rangle$ is a reference state. VQA minimizes the cost function $C = \langle\psi_{\boldsymbol{\theta}}| A |\psi_{\boldsymbol{\theta}}\rangle$ by training the parameter vector $\boldsymbol{\theta}$ according to the formula

$$\theta_k \rightarrow \theta_k - \alpha \partial_k C, \tag{1}$$

where $\partial_k C \equiv \partial C/\partial\theta_k$ and α is the learning rate, until the cost function approaches the exact result with acceptable accuracy.

In this article, we introduce the combined ansatz, i.e. the linear combination of several ansatz circuits. Given a set of ansatz circuit $\{U^i(\boldsymbol{\theta}^i), i = 1, \cdots, M\}$, the combined ansatz is

$$\widetilde{U}(\tilde{\boldsymbol{\theta}}) = \frac{1}{\Omega} \sum_{i=0}^{M-1} c_i U^i(\boldsymbol{\theta}^i), \tag{2}$$

where $\boldsymbol{c} = \{c_1, c_2, ..., c_M\}$ is the vector of complex coefficients, $\tilde{\boldsymbol{\theta}} = \{\boldsymbol{\theta}^1, \boldsymbol{\theta}^2, \cdots, \boldsymbol{\theta}^M, c\}$ is the set of parameter vectors. Ω is the normalization coefficient and determined through $\langle\phi_0| (\widetilde{U}(\tilde{\boldsymbol{\theta}}))^{\dagger}\widetilde{U}(\tilde{\boldsymbol{\theta}}) |\phi_0\rangle = 1$. Note that $\text{Re}(c_i)$, $\text{Im}(c_i)$ and $\boldsymbol{\theta}^i$ ($i = 1, 2, ...$) are trainable. The Hardware efficient ansatz (HEA) is a type of ansatzes that are easily implementable and not tailored to specific problems. In this work, we adopt 14 HEAs described in Ref. [37] as U^i to construct the linear combination of ansatzes. The schematic diagram of the 14 HEAs with system size $Q = 4$ is shown in Fig. 1.

B. Expressibility

The concept of expressibility serves as a key metric for evaluating the efficacy of ansatz circuits. An ansatz circuit is deemed "good" for a given problem if the generated ensemble of trial states harbors the true solution. However, in many practical scenarios, access to the solution of the problem might be limited. As an alternative, we use expressibility to evaluate the quality of parameter circuits. Expressibility can be defined as the capability to generate pure states, encompassing the ability to comprehensively and uniformly explore the entire unitary space. In this paper, we use Kullback-Leibler (KL) divergence to quantify expressibility [37],

$$\text{Expr} = D_{\text{KL}}(P_{\text{est}}(F)||P_{\text{Haar}}(F)), \tag{3}$$

which measures the distance between two discrete probability distributions of fidelity. $P_{\text{est}}(F)$ is the estimated probability distribution of the fidelity $F = |\langle\psi_{\phi} |\psi_{\theta}\rangle|^2$, where $|\psi_{\theta}\rangle$ and $|\psi_{\phi}\rangle$ are obtained by independently sampling a pair of parameter vectors in ansatz circuits. $P_{\text{Haar}}(F)$ is the probability distribution of the fidelity F for the Haar-distributed random states, which can be obtained by the analytical form $P_{\text{Haar}}(F) = (N-1)(1-F)^{N-2}$, where $N = 2^Q$ is the dimension of the Hilbert space. The discretization implies splitting the probability distributions into n_{bin} parts, then the value of D_{KL} can be numerically calculated according to the definition $D_{\text{KL}}(P||P') = \sum_i^{n_{\text{bin}}} P_i \ln(P_i/P_i')$. D_{KL} can be visualized using a histogram with bin number of n_{bin}. Fig. 2 exhibits the histogram where P_{est} are generated by ansatz circuit 10 with qubit number $Q = 12$.

The value of D_{KL} in Eq. (3) depends on the bin number of n_{bin}. One can use a fixed bin number such as dividing $[0, 1]$ into n_{bin} parts and assign each part a value according to the probability distribution [37]. However, in multi-qubit systems, the estimated probability distribution is nonzero only in a narrow region, and the zero region does not contribute to D_{KL}. Therefore, we use an unfixed bin number dividing the nonzero region into n_{bin} parts. Fig. 2 and the inset imply that the histogram or D_{KL} with an unfixed bin number more clearly distinguishes the difference between two probability distributions than that with a fixed bin number when the bin size is larger than the width of the nonzero region.

The common method to enhance the expressibility is increasing the depth of the ansatz circuit, such as repeating ansatz $U(\theta)$ of L times. Taking circuit 10 as an example, we shows the numerical results of D_{KL} varying with the repetitions number L and system size Q in Fig. 3 b. For a fixed system size, the expressibility of $[U(\theta)]^L$ tends to stabilize after reaching a threshold value of L. The inset suggests that the threshold $L_{\text{th}} = aQ + b$ is linearly dependent on the system size, where Q is the qubit number, and a and b are constant. However, increasing the number of repetitions L will exponentially increase the noise, making the system unusable due to excessive noise. In the following text, we will show that LCA can improve expressibility with $L = 1$ for any number of qubits.

III. EXPRESSIBILITY OF LCA

In this section, we extend the expressibility defined in Eq. (3) for SA to a linear combination of LCMA. The combined trial state, used for sampling the estimated distribution is obtained by applying Eq. (2) to the reference state:

$$|\psi_{\tilde{\boldsymbol{\theta}}}\rangle = \frac{1}{\Omega} \sum_i^M c_i U^i(\boldsymbol{\theta}^i) |\phi_0\rangle. \tag{4}$$
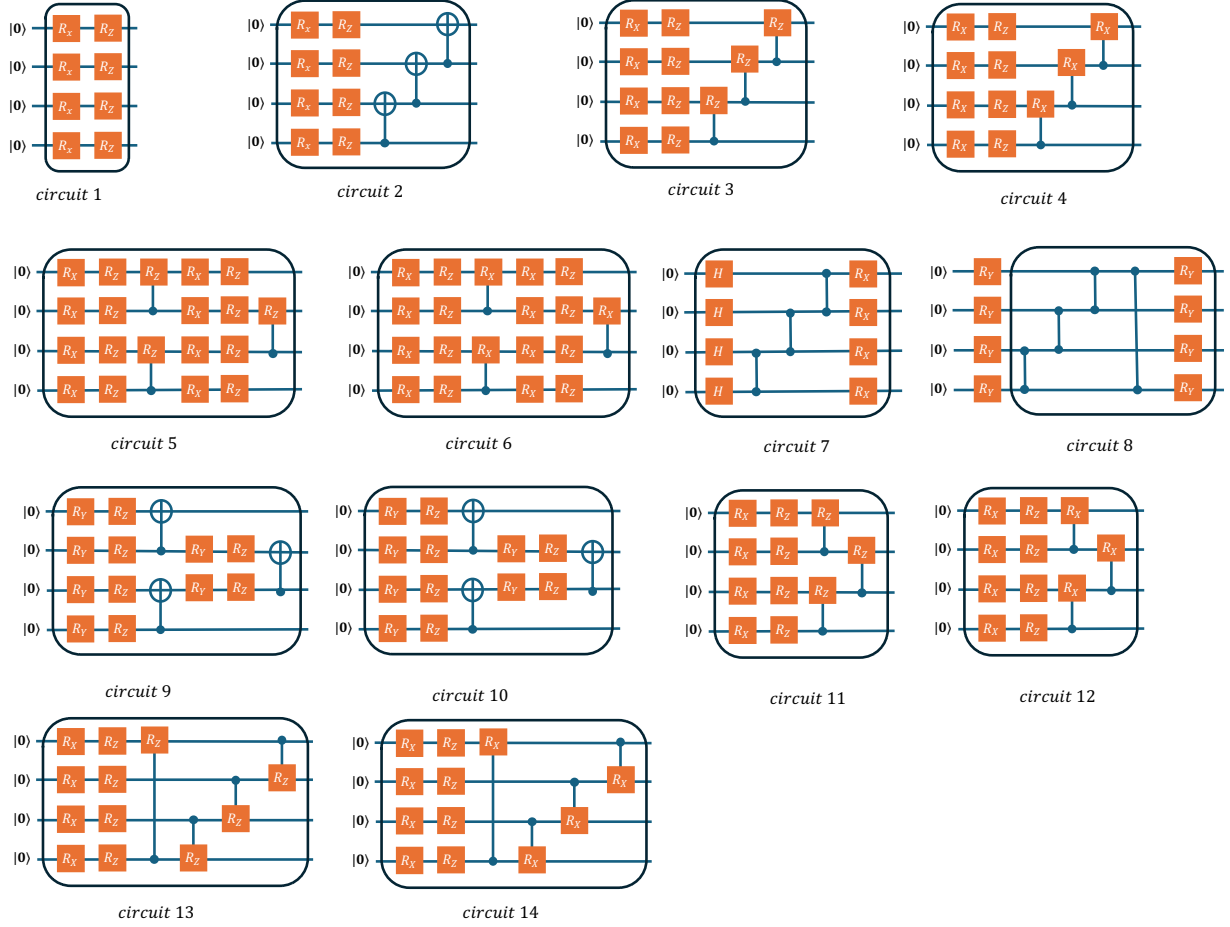
FIG. 1. Schematic diagram of 14 ansatzes of which the linear combination is studied in this work. The box represents a layer of the circuit. These circuits can be generalized to multi-qubit cases.
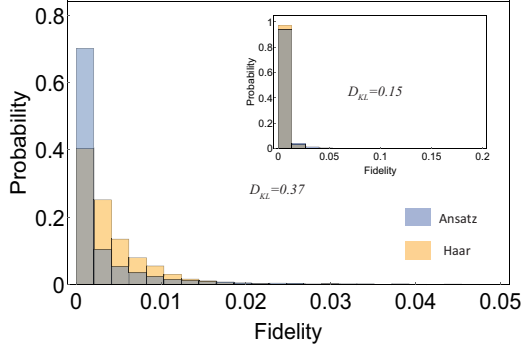

FIG. 2. Histograms of estimated fidelities are shown, overlaid with fidelities of the Haar-distributed ensemble, with the computed KL divergences.

Here, $U^i(\theta^i)$ represents a hardware efficient ansatz. We employ 14 of 19 HEAs studied in Ref. [37]. These ansatzes in Ref. [37] with high expressibility, such as circuit $5, 6, 13, 14$, and unchanged expressibility as increasing repeating layers, such as circuit 15, are ignored. The expressibilities of the chosen 14 ansatzes are shown in Fig. 3(a), calculated with unfixed bin number. For the sake of convenience, the ansatzes involved in linear combinations in Eq. (4) are denoted by the ansatz set $\bar{M} = \{i, j, k...\}$, where i represents the ith HEA and M is the size of ansatz set. In the following subsections, we exhibit the numerical simulation to compare the expressibility of SA and LCMA through the calculation of D_{KL} and ground state energy estimation.in Fig. 3 (b)

A. Demonstration with single-qubit case

In this subsection, we first present a single-qubit case to demonstrate that the expressibility of LCMA is superior to that of SA.

Consider the scenario where $M = 2$ with $U^1(\theta) = R_x(\theta)$ and $U^2(\phi) = R_z(\phi)$. Let the reference state be $|\phi_0\rangle = (\sqrt{2}|0\rangle + |1\rangle)/\sqrt{3}$, depicted as the orange point in Fig. 4. 1000 sample pairs of parameterized states $\{U^1(\theta)|\phi_0\rangle, U^1(\theta')|\phi_0\rangle\}$ and $\{U^2(\phi)|\phi_0\rangle, U^2(\phi')|\phi_0\rangle\}$ drawn in Fig. 4. Notably, these parameterized states
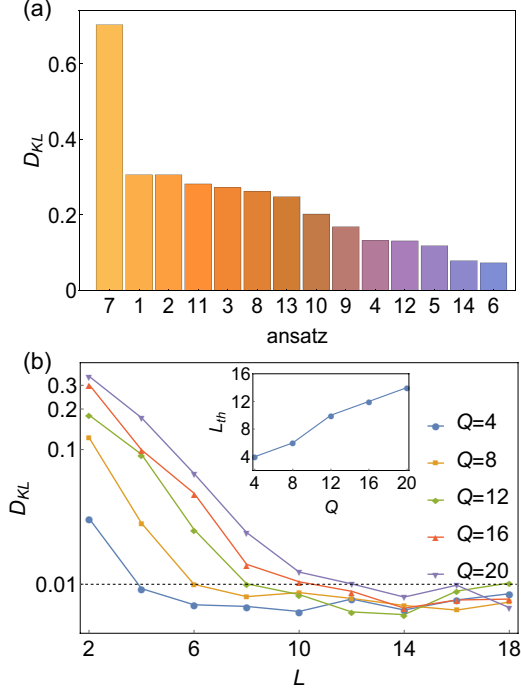
FIG. 3. The numerical results of KL divergence (i.e., $D_{\rm KL}$) with the unfixed bin number. (a) $D_{\rm KL}$ for single ansatz. The x-axis represents the circuit ID. (b) $D_{\rm KL}$ as a function of the circuit layer L for qubit numbers $N = 4, 8, 12, 16$ and 20. The black dotted line serves as a reference line for the threshold of high expressibility. The inset suggests that the threshold value is linear in the qubit number.

form distinct circles (red for U^1 and blue for U^2) on the Bloch sphere, indicating a one-dimensional subspace exploration. We calculate numerically the distribution of 1000 fidelities $|\langle \psi_{\theta'} | \psi_\theta \rangle|^2$ and get $D_{KL} = 2.18$ (for U^1) and 0.34 (for U^2). As for the combined parameterized states, we sample 1000 pairs of parameter vectors $\{\theta, \theta'\}$ and obtain the corresponding 1000 pairs of parameterized states $\{|\psi(\theta)\rangle, |\psi(\theta')\rangle\}$, where $\theta = \{\theta_1, \theta_2\}$, $|\psi(\theta)\rangle = 1/\Omega(U^1(\theta_1) + U^2(\theta_2)) |\phi_0\rangle$ and $1/\Omega$ is a normalization coefficient. The combined parameterized states fill the Hilbert space, which indicates that LCA can explore more space than SA. This coincides with the numerical result $D_{KL} = 0.047$ indicating the LCA has a higher expressibility than SA.

B. The improvement of the expressibility for two ansatzs

This subsection presents a detailed comparison between LC2A (i.e., $M = 2$) and SA based on the numerical results of expressibility and ground state energy estimation.

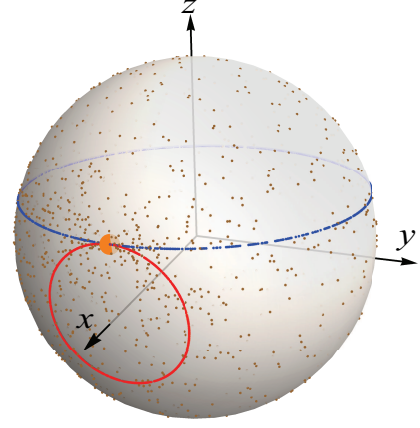We numerically calculate the KL divergence of LC2A

FIG. 4. 2000 parameterized states were plotted on the Bloch sphere using QuESTLink [18]. The red and blue data respectively represent the parameterized states $R_x(\theta)\phi(0)$ and $R_z(\theta')\phi(0)$, where $\phi(0) = (\sqrt{2}|0\rangle + |1\rangle)/\sqrt{3}$. The brown data are the parameterized states $[c_1 R_x(\theta) + c_2 R_z(\theta')]/\Omega\phi(0)$ where Ω is a normalization coefficient.

and SA to compare their expressibilities. The numerical simulation involves 14 common HEAs, with LC2A representing the linear combination of every two HEAs. The expressibility of the ith SA is denoted by D^i_{KL}, while that of LCMA is denoted by $D^{\bar{M}}_{KL}$. The improvement in expressibility is thus defined by

$$R_{\bar{M}} = \frac{D^{\bar{M}}_{KL} - {\rm Min}[\{D^i_{KL}, D^j_{KL}, D^k_{KL}, ...\}]}{{\rm Min}[\{D^i_{KL}, D^j_{KL}, D^k_{KL}, ...\}]}. \quad (5)$$

The term ${\rm Min}[D^i_{KL}, D^j_{KL}, D^k_{KL}, ...]$ denotes the maximum expressibility among the ansatzes combined. Eq. (3) demonstrates that the expressibility of the LCMA for the ansatz set $i, j, k...$ is improved by at least a degree of $R_{\bar{M}}$ over the SA expressibility. For $M = 2$, the numerical values of $R_{\bar{2}}$ are depicted by the blue triangles in Fig. 5. The color block for ansatz i, ansatz j signifies $R_{i,j}$, indicating the improvement in expressibility of the ansatz set i, j over the highest SA expressibility between the ith and jth ansatzes. The magnitude of $R_{\bar{2}}$ is indicated by the color bar.

For instance, $R_{\bar{M}}$ in the first column, where $\bar{M} = \{i, 14\}$ and $i = 1, ...13$, means the expressibility of ansatz 14 has a $R_{\bar{M}} \times 100\%$ improvement when ansatz 14 is linearly combined with other ansatzes. Among the 13 ansatzes, ansatzes $1, 2, 3, 7, 8, 11, 13$ exhibit lower improvements in expressibility than others. This can be explained by the fact that ansatz 14 has relatively high expressibility (see Fig. 3(a)), while the above-mentioned seven ansatzes have lower expressibility ($D_{KL} > 0.2$) than the remaining six ansatzes ($D_{KL} < 0.2$) and ansatz 14 ($D_{KL} < 0.1$). The stronger the expressibility of the ansatz involved in the linear combination, the greater the improvement in the expressibility of ansatz 14. This pattern holds true for $R_{\bar{M}}$ in other columns. These numerical results suggest that the expressibility of LC2A sur-

FIG. 6. Histogram showing the percentages linked to enhancements in expressibility and the estimated ground state energy. Orange and blue rectangles indicate improvements in expressibility and estimated ground state energy, respectively, achieved by the three-ansatz LCA over the SA, quantified by percentages $R_{\bar{M}}$ for expressibility and $L_{\bar{M}}$ for ground state energy.
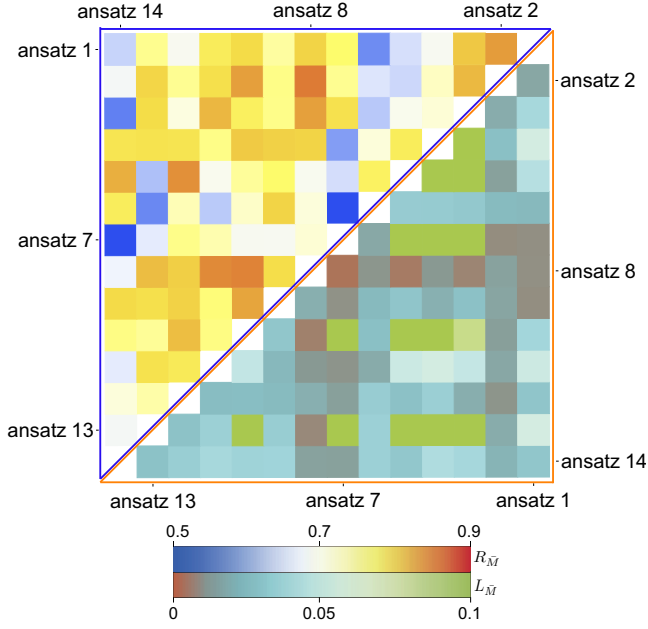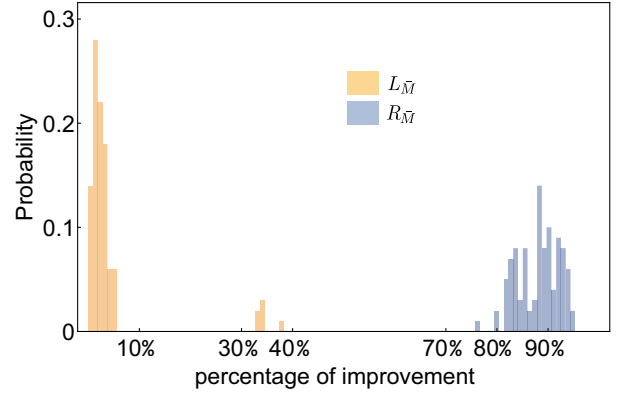
FIG. 5. (Color online) Plot of numerical results demonstrating the superiority of LCA over SA. In the blue triangle, the colored block corresponding to ansatz $\{i,j\}$ represents $R_{\{i,j\}}$. The magnitude of $R_{\bar{M}}$ can be seen in the first color bar and indicates that LCA $\{i,j\}$ has higher expressibility than the highest one among ansatz i and j by a percentage of $R_{\bar{M}}$. Similarly, in the brown triangle, the colored block corresponding to ansatz $\{i,j\}$ represents $L_{\{i,j\}}$. The magnitude of $R_{\bar{M}}$ can be seen in the second color bar and indicates that the ground state energy estimated using LCA $\{i,j\}$ is more accurate than the most accurate using ansatz i or j by a percentage of $L_{\bar{M}}$.

passes that of SA by 50% to 90%, predominantly around 80%.

To demonstrate LC2A's superiority, we compare its calculated ground state energy with that of simulated SA using the transverse-field XY model as a benchmark. The Hamiltonian for this model is defined as

$$H = \sum_{i=1}^{N}(J_{xx}X_iX_{i+1}+J_{yy}Y_iY_{i+1})+J_x\sum_{i=1}^{N}X_i+J_z\sum_{i=1}^{N}Z_i \quad (6)$$

and $X_{N+1}=X_1, Y_{N+1}=Y_1$. Such a Hamiltonian does not have an eigenstate that is easy to prepare, so the previous method of measuring amplitude that does not require auxiliary qubits cannot be used [24].

We define

$$L_{\bar{M}} = \frac{E_{KL}^{\bar{M}} - \text{Min}[\{E_{KL}^i, E_{KL}^j, E_{KL}^k, ...\}]}{\text{Min}[\{E_{KL}^i, E_{KL}^j, E_{KL}^k, ...\}]}, \quad (7)$$

where E_{KL}^i ($E_{KL}^{\bar{M}}$) denotes the estimated ground state energy of Hamiltonian H obtained by the Variational Quantum Eigensolver (VQE) with the i-th ansatz (the linear combination of ansatzes set $\bar{M}=\{i,j,k,...\}$). The notation $\text{Min}[E_{KL}^i, E_{KL}^j, E_{KL}^k, ...]$ signifies the closest approximation to the ground state energy within the

set $E_{KL}^i, E_{KL}^j, E_{KL}^k, \ldots$. The numerical simulation for $L_{\bar{M}}$, as illustrated by the brown triangle in Fig. 2 for $M=2$, indicates that the ground state energy estimation from LC2A are more precise than those from SA by 0.1% to 10%, predominantly around 5%. This suggests that enhancements in the expressibility of ansatz sets corresponds to increased accuracy in ground state energy estimation.

C. The improvement of the expressibility for multi-ansatzs

This subsection evaluates the expressibility enhancement of the LCMA for $M \geqslant 3$. For $M=3$ and $N=4$, our numerical simulations, depicted in Fig. 5, were performed on 100 sets of three ansatzes randomly selected from 14 HEAs. The improvements in expressibility and estimated ground state energy for each set are represented as orange and blue rectangles, respectively, in Fig. 5. The expressibility data, with $R_{\bar{M}}$ averaging around 90%, demonstrates that LC3A's expressibility surpasses that of SA by over 90% and also indicates superior expressibility compared to LC2A. Meanwhile, the $L_{\bar{M}}$ data, averaging around 3%, suggests a modest expressibility enhancement over SA by 3%. Notably, a few $L_{\bar{M}}$ values exceed 30%, attributed to the relatively high ground state energy estimation by SA. Contrarily, the improvements on ground state energy estimation offered by LC3A are not markedly superior to those by LC2A, given the latter's already high accuracy. Increasing M does not lead to a continuous increase in the expressibility of LCMA. D_{KL} of LCMA stabilizes when M increases to the critical value M_c such that the expressibilitiy of LCMA becomes saturated, that is, continuing to increase M will not significantly improve expressibility. The value of M_c

depends on the system size as shown in Fig. 7, where we calculate how the expressibility of systems with qubit numbers of 4, 6, 8, 10, and 12 changes with the number of ansatz. For each fixed-size system, we performed four experiments, marked by different colors in Fig. 7(a)-(e). We randomly selected and added new ansatz to the LCA for each experiment until 14 different ansatz were selected. Fig. 7(f) indicates that M_c increases almost linearly with the number of qubits.

IV. THE EXPERIMENTAL SCHEME OF LCA

A. The implementation of LCA

To update the parameters, we must compute the mean value of the Hamiltonian:

$$\langle H \rangle_{\tilde{\theta}} = \langle \psi_{\tilde{\theta}} | H | \psi_{\tilde{\theta}} \rangle = \frac{1}{\Omega^2} \sum_{i,j=0}^{M-1} c_i^\dagger c_j \langle \phi_0 | U^{i\dagger} H U^j | \phi_0 \rangle, \quad (8)$$

where Ω is the normalization constant

$$\Omega^2 = \sum_{i,j}^{M} c_i^\dagger c_j \langle \phi_0 | U^{i\dagger} U^j | \phi_0 \rangle. \quad (9)$$

The most common method for calculating the cross terms $\langle \phi_0 | U^{i\dagger} H U^j | \phi_0 \rangle$ and $\langle \phi_0 | U^{i\dagger} U^j | \phi_0 \rangle$ where $i \neq j$ is the Hadamard test(HT). HT requires the implementation of control-U^i gate, gates, which typically involve numerous two-qubit gates. In real devices, these contribute significantly to error accumulation. To mitigate this overhead, we propose PCM specifically designed for LCA. PCM eliminates the need for auxiliary qubits and minimizes two-qubit gate usage. We subsequently detail efficient methods for measuring both $\langle \psi_i | \psi_j \rangle$ and $\langle \psi_i | H | \psi_j \rangle$.

1. The algorithm begins by calculating $\text{Tr}(\rho_0 \rho_i)$ for $i = 1, 2, \ldots, M - 1$, where $\rho_i = |\psi_i\rangle\langle\psi_i| = U^{i\dagger}|\phi_0\rangle\langle\phi_0|U^i$. Although $\text{Tr}(\rho_0\rho_i) = |\langle\psi_0|\psi_i\rangle|^2$, we initially assume $\langle\psi_0|\psi_i\rangle = \sqrt{\text{Tr}(\rho_0\rho_i)}$ in this step. In fact, this equation does not necessarily hold, because there may be a relative phase between the two states. This is valid despite potential relative phases between the states, as the next section will justify their irrelevance in this context. The specific implementation circuit diagram is shown in Fig. 8(a).

2. The next step is to measure $\text{Tr}(\rho_0\rho_i\rho_j)$ for $i, j = 1, 2, \ldots, M - 1$ and $i \neq j$ using the circuit in Fig. 8(b). To obtain $\text{Tr}(\rho_0\rho_i\rho_i)$ we implement the circuit consisting of $\{U_0, (U_i)^\dagger, I - x|0\rangle\langle0|, U_i, (U_j)^\dagger\}$, where $x = 1 - e^{i\theta}$ and $I - x|0\rangle\langle0| = e^{i\theta|0\rangle\langle0|}$. The unitary operator applied to ρ is actually $U_i(I - x|0\rangle\langle0|)U_i^\dagger = I - x|\psi_i\rangle\langle\psi_i| =$

$I - x\rho_i$. What the circuit outputs are

$$\text{Tr}((I - x\rho_i)\rho_0(I - x^*\rho_i)\rho_j) \quad (10)$$
$$= \text{Tr}(\rho_0\rho_j - x\rho_i\rho_0\rho_j - x^*\rho_0\rho_i\rho_j + |x|^2\rho_i\rho_0\rho_i\rho_j). \quad (11)$$

Note that $\text{Tr}(\rho_0\rho_j)$ was already obtained and $\text{Tr}(\rho_i\rho_0\rho_i\rho_j) = \text{Tr}(\rho_i\rho_0)\text{Tr}(\rho_i\rho_j)$ so they can be calculated from Fig. 8(b). To extract $\text{Tr}(\rho_0\rho_i\rho_j)$ from the measurements, we use two different x values, treating treat $\text{Tr}(\rho_0\rho_i\rho_j)$ and $\text{Tr}(\rho_i\rho_0\rho_j)$ as unknowns and solving the linear system. In our numerical simulations, we set $x = 2$ and $x = 1 - i$ respectively. The resulting system of equations satisfy the following form:

$$\begin{cases} \text{Tr}(2\rho_0\rho_i\rho_j + 2\rho_i\rho_0\rho_j) = R_1 \\ \text{Tr}((1+i)\rho_0\rho_i\rho_j + (1-i)\rho_i\rho_0\rho_j) = R_2 \end{cases},$$

Where R_1 and R_2 are derived from the measurements. This system has a unique solution. With the extracted $\text{Tr}(\rho_0\rho_i\rho_j)$, we can calculate $\langle\psi_i|\psi_j\rangle$ form the relation

$$\begin{aligned} \langle\psi_i|\psi_j\rangle &= \frac{\langle\psi_0|\psi_i\rangle\langle\psi_i|\psi_j\rangle\langle\psi_j|\psi_0\rangle}{\langle\psi_0|\psi_i\rangle\langle\psi_j|\psi_0\rangle} \\ &= \frac{\text{Tr}(\rho_0\rho_i\rho_j)}{\sqrt{\rho_0\rho_i}\sqrt{\rho_0\rho_j}}. \end{aligned}$$

3. Given the Hamiltonian $H = \sum_s a_s P_s$, where P_s are Hermitian operators, we can estimate $\langle\psi_i|P_s|\psi_j\rangle$ using the circuit in Fig. 8(c). This is similar to the previous step but with the final measurement replaced by P_s. The measurement results yield:

$$\text{Tr}((I - x\rho_j)\rho_i(I - x^*\rho_j)P_s)$$
$$= \text{Tr}(\rho_i P_s - x\rho_j\rho_i P_s - x^*\rho_i\rho_j P_s + |x|^2\rho_j\rho_i\rho_j P_s)$$

Both $\text{Tr}(\rho_i P_s)$ and $\text{Tr}(\rho_j\rho_i\rho_j P_s) = \text{Tr}(\rho_j\rho_i)\text{Tr}(\rho_j P_s)$ can be obtained directly from other quantum circuits. Similar to the previous approach, we use different x values and solve the system of equations:

$$\begin{cases} \text{Tr}(2\rho_j\rho_i P_s + 2\rho_i\rho_j P_s) = R_1' \\ \text{Tr}((1+i)\rho_j\rho_i P_s + (1-i)\rho_i\rho_j P_s) = R_2' \end{cases},$$

Finally we calculate $\langle\psi_i|P_s|\psi_j\rangle$ according to the relation

$$\begin{aligned} \langle\psi_i|P_s|\psi_j\rangle &= \frac{\langle\psi_j|\psi_i\rangle\langle\psi_i|P_s|\psi_j\rangle}{\langle\psi_j|\psi_i\rangle} \\ &= \frac{\rho_j\rho_i P_s}{\langle\psi_j|\psi_i\rangle}, \end{aligned}$$

where $\langle\psi_j|\psi_i\rangle$ is obtainable from the previous step.

B. Unknown phase

In the previous section, we assumed phaseless $\langle\psi_0|\psi_i\rangle$. Here, we justify this assumption. Suppose a method like HT accurately determines the phase of $\langle\psi_0|\psi_i\rangle$. Using
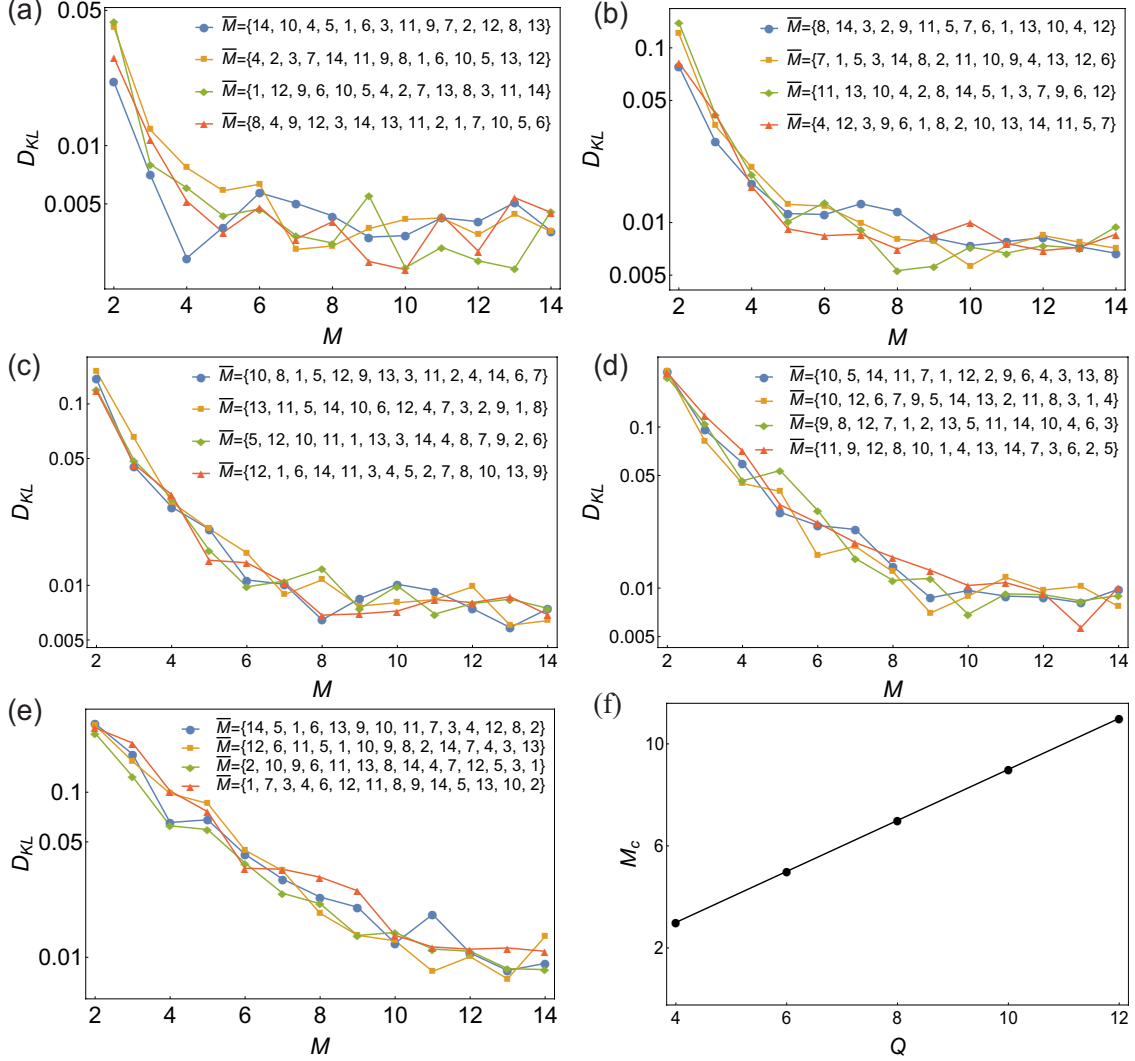
FIG. 7. Expressibility changes with the number of ansatzes under different numbers of qubits. (a), (b), (c), (d), and (e) represent the numerical results for systems with qubit numbers of 4, 6, 8, 10, and 12 respectively. The horizontal axis is the number of ansatzes, and the vertical axis is expressibility. Different colors represent different LCAs. (f) represents the relation between M_c and the number of qubits.

the variational method, we minimize the energy to obtain solutions $\{c_i, \boldsymbol{\theta}^i\}$. Suppose the phase of $\langle\psi_0|\psi_j\rangle$ is $e^{i\alpha_i}$. Ignoring the phase during calculations effectively applies a transformation :

$$|\psi_i\rangle \rightarrow |\psi_i'\rangle = e^{-i\alpha_i}|\psi_i\rangle \qquad (12)$$

The calculated average energy then becomes:

$$\langle H\rangle_{\tilde{\boldsymbol{\theta}}} = \langle\psi_{\tilde{\boldsymbol{\theta}}}'|H|\psi_{\tilde{\boldsymbol{\theta}}}'\rangle = \frac{1}{\Omega^2}\sum_{i,j=0}^{M-1} c_i^\dagger c_j \langle\psi_i'|H|\psi_j'\rangle. \qquad (13)$$

Thus, there still exists a set of solutions $\{c_i e^{i\alpha_i}, \boldsymbol{\theta}^i\}$ that yields the same minimum energy value as the HT method. In this sense, we can consider that the assumption of no phase in $\langle\psi_0|\psi_j\rangle$ has no impact on the algorithm.

C.   Two qubits gate counting

A key advantage of PCM over HT lies in the markedly reduced need for two-qubit gates. In HT, each CNOT gate in the ansatz transforms into a Toffoli gate, which then decomposes into five two-qubit gates. Conversely, in PCM, CNOT gates in the ansatz are preserved, though the number of two-qubit gates for implementing the $G(\theta)$ gate escalates with the bit count. This distinction is critical for NISQ-era hardware.

We take circuit-2 and circuit-15 as an example of LCA and compare the number of two-qubit gates required by the two methods, as shown in Fig. 9(a) and Fig. 9(b). An n-qubit $G(\theta)$ gate is implemented by a series of single-qubit gates and an n-qubit Toffoli gate. This Toffoli gate is implemented using the method in Ref. [27] (Note that

(a) Quantum circuit for first step

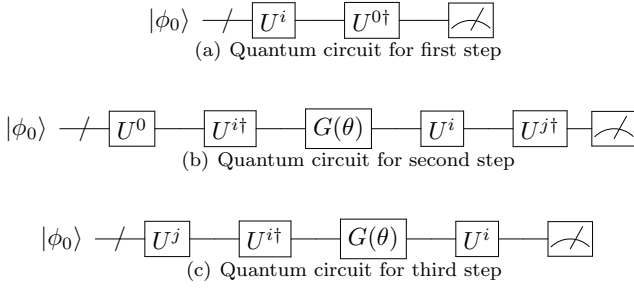(b) Quantum circuit for second step

(c) Quantum circuit for third step

FIG. 8. Quantum circuits for LCA. (a) depicts the circuit for measuring the overlap $\langle\psi_0|\psi_i\rangle$, using a final projection measurement. Similarly, (b) demonstrates the circuit for measuring the overlap $\langle\psi_i|\psi_j\rangle$, again with a final projection measurement. Finally, (c) illustrates the circuit for measuring $\langle\psi_i|H|\psi_j\rangle$, where the final measurement employs Pauli operators P_s.

some other, better implementations of Toffoli gates [29] would bring more advantages to PCM). When the ansatz is shallow and the number of qubits is large, PCM has no advantage over HT. However, this scenario is often less relevant in practical applications. To achieve an accurate approximation of the desired state, the ansatz depth typically scales with the number of qubits. As shown in Fig. 9(a), for a fixed number of qubits, the advantage of PCM over HT becomes increasingly pronounced with increasing ansatz depth. Furthermore, Fig. 9(b) demonstrates that when the ansatz depth exceeds a critical threshold($d \geq 3$), PCM exhibits a significant reduction in two-qubit gates required compared to HT as the number of qubits increases. Although this threshold will change slightly depending on the choice of ansatz, within the regime where the quantum advantage is achievable(number of qubits > 50), the performance of PCM is anticipated to significantly surpass that of HT.

D. Parameter shift rules

A straightforward approach to the optimization of variational algorithms is gradient descent, which requires the estimation of gradients. The gradients for $\boldsymbol{c}$ are relatively simple:

$$\frac{\partial\langle H\rangle}{\partial c_i} = \frac{1}{|\langle\psi_{\tilde{\boldsymbol{\theta}}}|\psi_{\tilde{\boldsymbol{\theta}}}\rangle|^2}\left(\langle\psi_{\tilde{\boldsymbol{\theta}}}|H|\psi_i\rangle + h.c.\right). \qquad (14)$$

The gradients for $\boldsymbol{\theta}$ are relatively complex:

$$\frac{\partial\langle H\rangle}{\partial\theta_{i,l}} = \frac{1}{|\langle\psi_{\tilde{\boldsymbol{\theta}}}|\psi_{\tilde{\boldsymbol{\theta}}}\rangle|^2}\left(\frac{\partial\langle\psi_{\tilde{\boldsymbol{\theta}}}|H|\psi_{\tilde{\boldsymbol{\theta}}}\rangle}{\partial\theta_{i,l}}\langle\psi_{\tilde{\boldsymbol{\theta}}}|\psi_{\tilde{\boldsymbol{\theta}}}\rangle \\ - \langle\psi_{\tilde{\boldsymbol{\theta}}}|H|\psi_{\tilde{\boldsymbol{\theta}}}\rangle\frac{\partial\langle\psi_{\tilde{\boldsymbol{\theta}}}|\psi_{\tilde{\boldsymbol{\theta}}}\rangle}{\partial\theta_{i,l}}\right), \qquad (15)$$
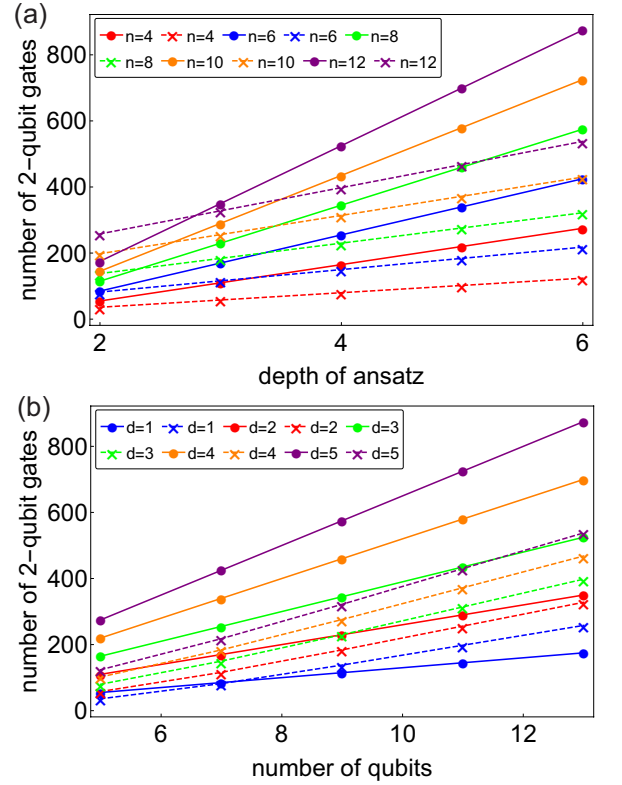

FIG. 9. Comparison of PCM and HT. The solid lines represent data for the HT method, while the dashed lines represent data for the PCM method. (a) The horizontal axis denotes the ansatz depth, defined as the number of repetitions of circuits 2 and 15 within the overall circuit. The vertical axis represents the two-qubit gate count. The number of qubits, n, is indicated by lines of different colors. (b) The number of qubits is shown on the horizontal axis, and the vertical axis displays the two-qubit gate count. The depth of ansatz, d, is represented by lines of different colors.

where

$$\frac{\partial\langle\psi_{\tilde{\boldsymbol{\theta}}}|\psi_{\tilde{\boldsymbol{\theta}}}\rangle}{\partial\theta_{i,l}} = \sum_j c_i^* c_j \frac{\partial\langle\psi_i|}{\partial\theta_{i,l}}|\psi_j\rangle + h.c. \\ = \sum_j c_i^* c_j \langle\phi_0|\frac{\partial U^{i\dagger}}{\partial\theta_{i,l}}U^j|\phi_0\rangle + h.c., \qquad (16)$$

and

$$\frac{\partial\langle\psi_{\tilde{\boldsymbol{\theta}}}|H|\psi_{\tilde{\boldsymbol{\theta}}}\rangle}{\partial\theta_{i,l}} = \sum_j c_i^* c_j \frac{\partial\langle\psi_i|H|}{\partial\theta_{i,l}}|\psi_j\rangle + h.c. \\ = \sum_j c_i^* c_j \langle\phi_0|\frac{\partial U^{i\dagger}}{\partial\theta_{i,l}}HU^j|\phi_0\rangle + h.c.. \qquad (17)$$

As mentioned in the previous subsection, when we estimate the normalization $\langle\psi_{\tilde{\boldsymbol{\theta}}}|\psi_{\tilde{\boldsymbol{\theta}}}\rangle$ average the Hamiltonian $\langle\psi_{\tilde{\boldsymbol{\theta}}}|H|\psi_{\tilde{\boldsymbol{\theta}}}\rangle$, we will change the final state by unknown phases. When estimating Eq. (16) and Eq. (17), we cannot introduce more unknown phases to ensure correct results. This can be achieved when the t-th level of $U^{i\dagger}$ is of

the form $U_t^{i\dagger} = e^{iO_{i,t}\theta_{i,t}}$, where $O_{i,l}$ is a known Hermitian operator, e.g., Pauli operator. Let $U^{i\dagger} = \prod_t U_t^{i\dagger}(\theta_{i,t})$, then the partial differentiation $\partial U_i^\dagger / \partial\theta_{i,l}$ can be written as

$$\frac{\partial U^{i\dagger}}{\partial\theta_{i,l}} = \prod_{t=1}^{l} U_t^{i\dagger}(\theta_{i,t}) O_{i,l} \prod_{t=l+1} U_t^{i\dagger}(\theta_{i,t}). \qquad (18)$$

We first consider the inner product in Eq. (16) in the case where $i = j$ can be expressed as

$$\langle\phi_0|\frac{\partial U^{i\dagger}}{\partial\theta_{i,l}}U^i|\phi_0\rangle = \langle\phi_0|\prod_{t=1}^{l} U_t^{i\dagger}(\theta_{i,t}) O_{i,l} \prod_{t=l}^{1} U_t^{i\dagger}(\theta_{i,t})|\phi_0\rangle$$
$$= \langle\psi_{i,l}|O_{i,l}|\psi_{i,l}\rangle, \qquad (19)$$

which can be estimated directly using quantum circuits. Then we can estimate the inner product in the general case where $i \neq j$ from

$$\langle\phi_0|\frac{\partial U^{i\dagger}}{\partial\theta_{i,l}}|\psi_i\rangle\langle\psi_i|\psi_j\rangle\langle\psi_j|\frac{\partial U^i}{\partial\theta_{i,l}}|\phi_0\rangle, \qquad (20)$$

which can be estimated using Step 2 mentioned in the last section. We use

$$\langle\phi_0|\frac{\partial U^{i\dagger}}{\partial\theta_{i,l}}|\psi_j\rangle\langle\psi_j|H|\frac{\partial U^i}{\partial\theta_{i,l}}|\phi_0\rangle, \qquad (21)$$

to estimate $\langle\psi_j|H|(\partial U^i/\partial\theta_{i,l})|\phi_0\rangle$. Eq. (21) can be estimated using Step 3 mentioned in the last section.

V. SUMMARY

To overcome the limitations of traditional approaches, we develop a novel method for enhancing the expressibility of variational quantum circuits for noisy quantum computers. Our key innovation lies in combining a linear combination of ansatzes with a specialized measurement scheme, effectively sidestepping the detrimental effects of increased circuit depth and noisy Hadamard tests. This translates to reduced dependence on two-qubit gates, leading to potentially more robust quantum computations. Furthermore, the method is readily applicable due to its efficient gradient calculation scheme. Numerical simulations validate the effectiveness of our approach, paving the way for practical applications on current noisy quantum hardware.

[1] Muhammad AbuGhanem and Hichem Eleuch. Nisq computers: A path to quantum supremacy. arXiv preprint arXiv:2310.01431, 2023.

[2] Abhinav Anand, Philipp Schleich, Sumner Alperin-Lea, Phillip WK Jensen, Sukin Sim, Manuel Díaz-Tinoco, Jakob S Kottmann, Matthias Degroote, Artur F Izmaylov, and Alán Aspuru-Guzik. A quantum computing view on unitary coupled cluster theory. Chemical Society Reviews, 51(5):1659–1684, 2022.

[3] Ville Bergholm, Josh Izaac, Maria Schuld, Christian Gogolin, Shahnawaz Ahmed, Vishnu Ajith, M Sohaib Alam, Guillermo Alonso-Linaje, B AkashNarayanan, Ali Asadi, et al. Pennylane: Automatic differentiation of hybrid quantum-classical computations. arXiv preprint arXiv:1811.04968, 2018.

[4] Carlos Bravo-Prieto, Josep Lumbreras-Zarapico, Luca Tagliacozzo, and José I Latorre. Scaling of variational quantum circuit depth for condensed matter systems. Quantum, 4:272, 2020.

[5] Yudong Cao, Jonathan Romero, Jonathan P Olson, Matthias Degroote, Peter D Johnson, Mária Kieferová, Ian D Kivlichan, Tim Menke, Borja Peropadre, Nicolas PD Sawaya, et al. Quantum chemistry in the age of quantum computing. Chemical reviews, 119(19):10856–10915, 2019.

[6] Marco Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, et al. Variational quantum algorithms. Nature Reviews Physics, 3(9):625–644, 2021.

[7] Pierre-Luc Dallaire-Demers, Jonathan Romero, Libor Veis, Sukin Sim, and Alán Aspuru-Guzik. Low-depth circuit ansatz for preparing correlated fermionic states on a quantum computer. Quantum Science and Technology, 4(4):045005, 2019.

[8] Peter Deglmann, Ansgar Schäfer, and Christian Lennartz. Application of quantum calculations in the chemical industry—an overview. International Journal of Quantum Chemistry, 115(3):107–136, 2015.

[9] Yuxuan Du, Zhuozhuo Tu, Xiao Yuan, and Dacheng Tao. Efficient measure for the expressivity of variational quantum algorithms. Physical Review Letters, 128(8):080506, 2022.

[10] Sepehr Ebadi, Tout T Wang, Harry Levine, Alexander Keesling, Giulia Semeghini, Ahmed Omran, Dolev Bluvstein, Rhine Samajdar, Hannes Pichler, Wen Wei Ho, et al. Quantum phases of matter on a 256-atom programmable quantum simulator. Nature, 595(7866):227–232, 2021.

[11] Suguru Endo, Simon C Benjamin, and Ying Li. Practical quantum error mitigation for near-future applications. Physical Review X, 8(3):031027, 2018.

[12] Dmitry A Fedorov, Bo Peng, Niranjan Govind, and Yuri Alexeev. Vqe method: a short survey and recent developments. Materials Theory, 6(1):1–21, 2022.

[13] Lena Funcke, Tobias Hartung, Karl Jansen, Stefan Kühn, and Paolo Stornati. Dimensional expressivity analysis of parametric quantum circuits. Quantum, 5:422, 2021.

[14] Bryan T Gard, Linghua Zhu, George S Barron, Nicholas J Mayhall, Sophia E Economou, and Edwin Barnes. Efficient symmetry-preserving state preparation circuits for the variational quantum eigensolver algorithm. npj Quantum Information, 6(1):10, 2020.

[15] Harper R Grimsley, Daniel Claudino, Sophia E Economou, Edwin Barnes, and Nicholas J Mayhall. Is the trotterized uccsd ansatz chemically well-defined? Journal of chemical theory and computation, 16(1):1–6, 2019.

[16] Harper R Grimsley, Sophia E Economou, Edwin Barnes, and Nicholas J Mayhall. An adaptive variational algorithm for exact molecular simulations on a quantum computer. Nature communications, 10(1):3007, 2019.

[17] Gian Giacomo Guerreschi and Mikhail Smelyanskiy. Practical optimization for hybrid quantum-classical algorithms. arXiv preprint arXiv:1701.01450, 2017.

[18] J Robert Johansson, Paul D Nation, and Franco Nori. Qutip: An open-source python framework for the dynamics of open quantum systems. Computer Physics Communications, 183(8):1760–1772, 2012.

[19] Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M Chow, and Jay M Gambetta. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. Nature, 549(7671):242–246, 2017.

[20] Oleksandr Kyriienko and Vincent E Elfving. Generalized quantum circuit differentiation rules. Physical Review A, 104(5):052417, 2021.

[21] Jonathan Wei Zhong Lau, Kian Hwee Lim, Harshank Shrotriya, and Leong Chuan Kwek. Nisq computing: where are we and where do we go? AAPPS Bulletin, 32(1):27, 2022.

[22] Ying Li and Simon C Benjamin. Efficient variational quantum simulator incorporating active error minimization. Physical Review X, 7(2):021050, 2017.

[23] Hsuan-Hao Lu, Natalie Klco, Joseph M. Lukens, Titus D. Morris, Aaina Bansal, Andreas Ekström, Gaute Hagen, Thomas Papenbrock, Andrew M. Weiner, Martin J. Savage, and Pavel Lougovski. Simulations of subatomic many-body physics on a quantum frequency processor. Phys. Rev. A, 100:012320, Jul 2019.

[24] Sirui Lu, Mari Carmen Banuls, and J Ignacio Cirac. Algorithms for quantum simulation at finite energies. PRX Quantum, 2(2):020321, 2021.

[25] Sam McArdle, Suguru Endo, Alán Aspuru-Guzik, Simon C Benjamin, and Xiao Yuan. Quantum computational chemistry. Reviews of Modern Physics, 92(1):015003, 2020.

[26] Jarrod R McClean, Jonathan Romero, Ryan Babbush, and Alán Aspuru-Guzik. The theory of variational hybrid quantum-classical algorithms. New Journal of Physics, 18(2):023023, 2016.

[27] D. Michael Miller, Robert Wille, and Zahra Sasanian. Elementary quantum gate realizations for multiple-control toffoli gates. In 2011 41st IEEE International Symposium on Multiple-Valued Logic, pages 288–293, 2011.

[28] Kosuke Mitarai, Makoto Negoro, Masahiro Kitagawa, and Keisuke Fujii. Quantum circuit learning. Physical Review A, 98(3):032309, 2018.

[29] Junhong Nie, Wei Zi, and Xiaoming Sun. Quantum circuit for multi-qubit toffoli gate with optimal resource, 2024.

[30] Matthew Otten, Cristian L. Cortes, and Stephen K. Gray. Noise-resilient quantum dynamics using symmetry-preserving ansatzes, 2019.

[31] Carlos Outeiral, Martin Strahm, Jiye Shi, Garrett M Morris, Simon C Benjamin, and Charlotte M Deane. The prospects of quantum computing in computational molecular biology. Wiley Interdisciplinary Reviews: Computational Molecular Science, 11(1):e1481, 2021.

[32] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J Love, Alán Aspuru-Guzik, and Jeremy L O'brien. A variational eigenvalue solver on a photonic quantum processor. Nature communications, 5(1):4213, 2014.

[33] John Preskill. Quantum computing in the nisq era and beyond. Quantum, 2:79, 2018.

[34] Google AI Quantum, Collaborators*†, Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C Bardin, Rami Barends, Sergio Boixo, Michael Broughton, Bob B Buckley, et al. Hartree-fock on a superconducting qubit quantum computer. Science, 369(6507):1084–1089, 2020.

[35] Alessandro Roggero, Andy C. Y. Li, Joseph Carlson, Rajan Gupta, and Gabriel N. Perdue. Quantum computing for neutrino-nucleus scattering. Phys. Rev. D, 101:074038, Apr 2020.

[36] Maria Schuld, Ville Bergholm, Christian Gogolin, Josh Izaac, and Nathan Killoran. Evaluating analytic gradients on quantum hardware. Phys. Rev. A, 99:032331, Mar 2019.

[37] Sukin Sim, Peter D Johnson, and Alán Aspuru-Guzik. Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms. Advanced Quantum Technologies, 2(12):1900070, 2019.

[38] Kristan Temme, Sergey Bravyi, and Jay M. Gambetta. Error mitigation for short-depth quantum circuits. Phys. Rev. Lett., 119:180509, Nov 2017.

[39] Jules Tilly, Hongxiang Chen, Shuxiang Cao, Dario Picozzi, Kanav Setia, Ying Li, Edward Grant, Leonard Wossnig, Ivan Rungger, George H Booth, et al. The variational quantum eigensolver: a review of methods and best practices. Physics Reports, 986:1–128, 2022.

[40] Alexey Uvarov, Jacob D. Biamonte, and Dmitry Yudin. Variational quantum eigensolver for frustrated quantum systems. Phys. Rev. B, 102:075104, Aug 2020.

[41] Dave Wecker, Matthew B Hastings, and Matthias Troyer. Progress towards practical quantum variational algorithms. Physical Review A, 92(4):042303, 2015.

[42] Roeland Wiersema, Cunlu Zhou, Yvette de Sereville, Juan Felipe Carrasquilla, Yong Baek Kim, and Henry Yuen. Exploring entanglement and optimization within the hamiltonian variational ansatz. PRX Quantum, 1:020319, Dec 2020.

[43] Luogen Xu, Joseph T Lee, and JK Freericks. Test of the unitary coupled-cluster variational quantum eigensolver for a simple strongly correlated condensed-matter system. Modern Physics Letters B, 34(19n20):2040049, 2020.