Maximizing quantum-computing expressive power through randomized circuits

Yingli Yang [1,*] Zongkang Zhang [2,*] Anbang Wang [2,*] Xiaosi Xu [2] Xiaoting Wang [1,†] and Ying Li [2,‡]

[1]*Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu 610051, China*
[2]*Graduate School of China Academy of Engineering Physics, Beijing 100193, China*

In the noisy intermediate-scale quantum era, variational quantum algorithms (VQAs) have emerged as a promising avenue to obtain quantum advantage. However, the success of VQAs depends on the expressive power of parametrized quantum circuits, which is constrained by the limited gate number and the presence of barren plateaus. In this paper, we propose and numerically demonstrate an approach for VQAs, utilizing randomized quantum circuits to generate the variational wave function. We parametrize the distribution function of these random circuits using artificial neural networks and optimize it to find the solution. This random-circuit approach presents a trade-off between maximizing the expressive power of the variational wave function and minimizing the associated time cost, specifically the sampling cost of quantum circuits. Given a fixed gate number, we can systematically increase the expressive power by extending the quantum-computing time. With a sufficiently large permissible time cost, the variational wave function can approximate any quantum state with arbitrary accuracy. Furthermore, we establish explicit relationships between expressive power, time cost, and gate number for variational quantum eigensolvers. These results highlight the promising potential of the random-circuit approach in achieving a high expressive power in quantum computing.

I. INTRODUCTION

In quantum computing, a central issue is to solve problems using as few quantum gates as possible. The reason is that the accuracy of quantum gates is inevitably degraded by noise, causing the possibility of errors in each gate operation. Consequently, when the gate number increases, quantum computing becomes more and more unreliable, ultimately resulting in catastrophic failure. Therefore, the number of gates must be minimized for realizing any quantum computing application. This issue is particularly relevant for noisy intermediate-scale quantum (NISQ) computers, in which large-scale quantum error correction is unavailable [1].

Over the past decade, tremendous effort has been put forward to develop quantum applications while taking the gate number as the primary consideration. Among such applications, variational quantum algorithms (VQAs) [2] have attracted the most interest. The key idea behind VQAs is constructing an *Ansatz* quantum circuit $U(\theta)$ with parametrized quantum gates. Gate parameters θ are optimized in a feedback loop between quantum and classical computing. See the schematic diagram in Fig. 1(a). The circuit generates a variational wave function $|\phi(\theta)\rangle = U(\theta)|0\rangle^{\otimes n}$. Ideally, once the optimal parameters are found, $|\phi(\theta)\rangle$ is the answer to the given problem. In this way, VQAs can solve certain problems keeping the gate count minimal. Within this variational framework, a diverse range of applications has been developed, such as the variational quantum eigensolver (VQE) [3–10], variational quantum simulators [11–15], the quantum approximate optimization algorithm [16–21], quantum neural networks [22–26], etc. [27–32]. Notably, VQAs are well suited for the NISQ technologies of today, leading to rapid progress in their experimental implementations [2,4,33–35].

The success of VQAs depends on the expressive power of the variational wave function [36–40]. Since the variational wave function can only express a subset of all possible quantum states, a fundamental assumption is that the target quantum state is within this subset. Therefore, it is desirable to use a variational wave function with higher expressive power, capable of exploring a larger portion of the entire state space. By employing such a variational wave function, VQA has a better chance of finding the solution or attaining higher accuracy, as depicted in Fig. 2. To maximize the expressive power, we confront two challenges. First, regarding the practical implementation, the gate number limits the circuit size and the number of parameters. Second, highly expressive circuits usually resemble the unitary t design, which can lead to difficulties during the training process due to the problem of vanishing gradients [41–44]. Despite the challenges, much attention has been devoted to designing circuits [45–53]. However, it remains largely unexplored that the expressive power can be improved by optimizing the strategy of using circuits in the quantum-classical feedback loop.

---

*These authors contributed equally.

†xiaoting@uestc.edu.cn

‡yli@gscaep.ac.cn

(a)

FIG. 2. The subset of variational wave functions. In variational quantum algorithms (VQAs), the solution to a problem is represented by a quantum state. If the solution state is in the subset (e.g., state a) or close to the subset (e.g., state b) with a tolerable error ϵ, VQA can successfully solve the problem (up to a proper optimizer and other practical issues). If the solution state is far from the subset (e.g., state c), VQA fails.
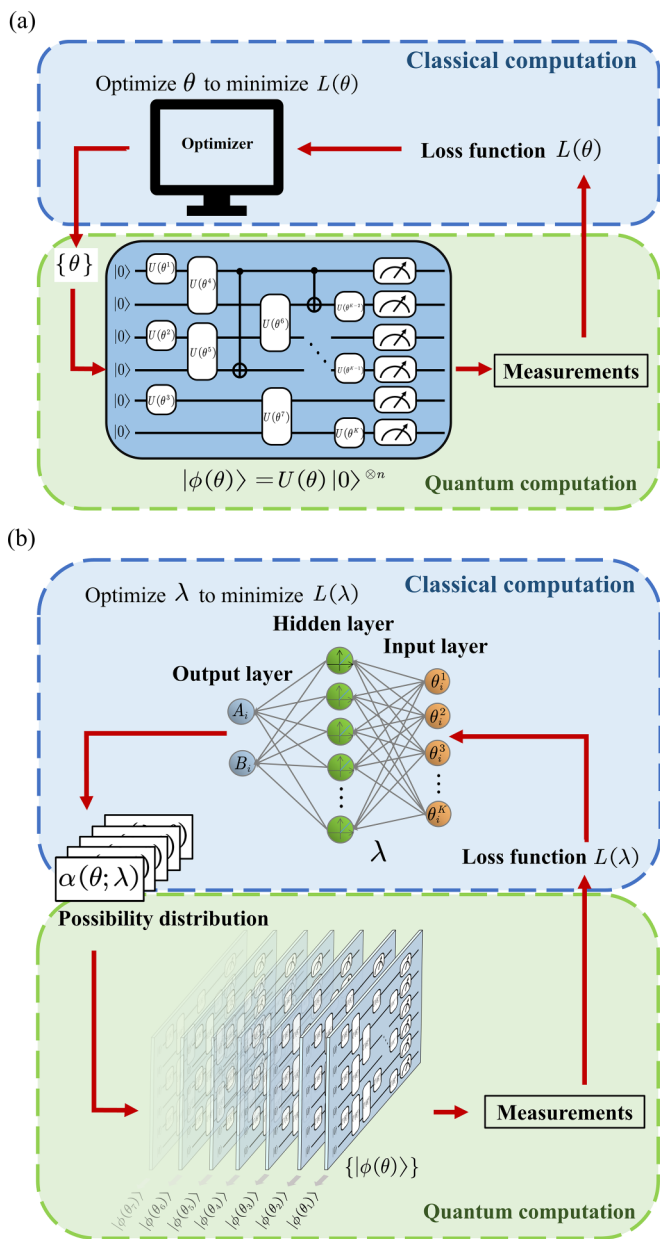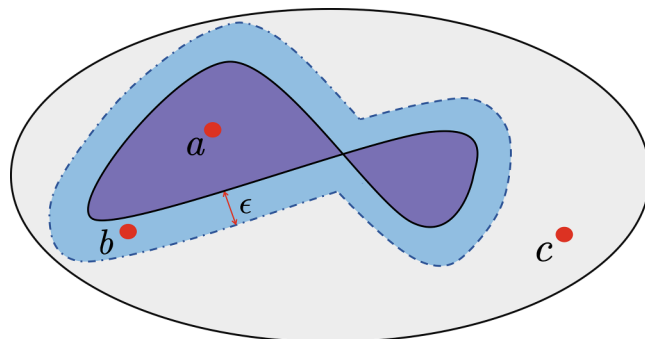
FIG. 1. (a) The schematic diagram of variational quantum algorithms (VQAs) using deterministic circuits. The problem is solved by finding the optimal circuit parameters θ to minimize loss function $L(\theta)$. (b) The schematic diagram of VQAs using random circuits. Circuits are sampled according to a parametrized guiding function $\alpha(\theta; \lambda)$, which determines the possibility distribution. Instead of θ, we optimize the parametrized guiding function $\alpha(\theta; \lambda)$ to solve a problem. See Sec. II for details.

In this paper, we develop a paradigm of utilizing quantum circuits in VQAs, within which we can systematically increase the expressive power while the gate number stays the same. Our method is called variational quantum-circuit Monte Carlo (VQCMC), which possesses greater expressive power than conventional VQA methods. Instead of a deterministic quantum circuit, we consider a variational wave function that is the average of wave functions generated by random circuits, in the form $|\psi(\lambda)\rangle = \mathrm{E}[e^{i\gamma(\theta)}|\phi(\theta)\rangle]$. In

contrast with deterministic-circuit VQAs, the distribution and phase functions are parametrized and optimized. Specifically, we propose parametrizing the distribution and phase with artificial neural networks (ANNs) [54]. In the spirit of VQAs, we solve problems by optimizing the parametrized distribution and phase functions, as illustrated in Fig. 1(b).

The key feature of the random-circuit approach is the trade-off between the expressive power and the time cost. When the gate number is fixed, we can increase the expressive power with an enlarged time cost. This feature is because of the statistical error caused by introducing randomness. To suppress the statistical error, we must repeat the measurement more times than in the deterministic-circuit approach, resulting in an enlarged time cost. We propose methods to control this time cost. When we apply a strong constraint on the time cost, the distribution is close to a δ function. In this case, the time cost and expressive power of $|\psi(\lambda)\rangle$ are almost the same as $|\phi(\theta)\rangle$. When we apply a weak constraint on the time cost, distributions with more randomness are allowed. Then the time cost is larger, and the expressive power becomes higher.

In the first part of this paper, we describe a general framework of VQAs using random circuits. In addition to the framework, we also analyze the statistical error and introduce two methods for controlling the time cost. In the second part, we present a simple numerical demonstration taking the ground-state problem (i.e., the VQE algorithm) as an example. The numerical result illustrates the trade-off between expressive power and time cost. In the third part, we give a set of theorems to justify the performance and potential of the random-circuit approach.

The rigorous theoretical results are listed as follows. Firstly, in the low-cost limit, the expressive power of $|\psi(\lambda)\rangle$ is not lower than $|\phi(\theta)\rangle$. This result is obtained using a rectifier ANN [55–57] to parametrize the distribution function. Secondly, to justify the power-cost trade-off, we show that increasing the time cost always enlarges the expressive power. Thirdly, in the high-cost limit, we prove the universal approximation theorem of the random-circuit approach: For many finite-size *Ansatz* circuits, we can approximate an arbitrary

state to arbitrary accuracy if the time cost is sufficiently large. Fourthly, to have a concrete discussion about this trade-off feature, we specifically focus on the ground-state problem and take the Hamiltonian *Ansatz* circuit [58] as an example. We obtain an upper bound of the error in approximating the ground state. We find that the error upper bound is a monotonic decreasing function of the time cost. Subject to a proper initial state, the error upper bound always vanishes in the limit of large time cost. Lastly, although VQAs are mainly developed for NISQ devices, we show the long-term potential of the random-circuit approach by analyzing its performance with an increasing gate number. We prove that, given a finite time cost, the error upper bound decreases with the gate number in the circuit. The gate number scales with the permissible error ϵ as $O(1/\epsilon^2)$.

II. VQCMC

In this section, firstly, we introduce concepts necessary for understanding the random-circuit approach. Secondly, we give a general formalism of the approach. Finally, we analyze the statistical error and relate it to a quantity that can be evaluated with a quantum computer, and we introduce the methods for controlling the time cost.

A. Preliminaries

Our random-circuit approach, called VQCMC, is inspired by quantum Monte Carlo in classical computing [59,60]. Quantum Monte Carlo methods combined with quantum computing have recently been proposed, including auxiliary-field Monte Carlo [61,62], Green's function Monte Carlo [62,63], variational Monte Carlo [62,64,65], full configuration interaction Monte Carlo [66,67], and stochastic series expansion Monte Carlo [68]. VQCMC consists of two main components: *sample space* and *guiding function*. In this section, we introduce the two components and present a way of parametrizing the guiding function using an ANN.

1. Sample space

The sample space Ω is a set of quantum states from which we draw random samples. A quantum computer can prepare states in the form $|\phi(\theta)\rangle = U(\theta)|0\rangle^{\otimes n}$, where $|0\rangle^{\otimes n}$ is the initial state of n qubits, $U(\theta)$ is the unitary operator of a parametrized quantum circuit, and θ is a vector of parameters. An example of the parametrized circuit is illustrated in Fig. 1(b). For those familiar with conventional deterministic-circuit VQAs, $U(\theta)$ could be one of the prominent *Ansatz* circuits, e.g., the unitary coupled cluster *Ansatz* [3], hardware-efficient *Ansatz* [4], Hamiltonian *Ansatz* [58], or ADAPT-VQE [69]. All these states generated by the circuit form the sample space $\Omega = \{|\phi(\theta)\rangle|\theta \in \Theta\}$, where Θ denotes the space of parameters ($\Theta \subseteq \mathbb{R}^K$ when the circuit has K real parameters). More generally, the sample space can be generated by several different *Ansätze*, which means that the circuit structure can also be randomly sampled.

2. Guiding function

The guiding function $\alpha(\theta; \lambda)$ determines the distribution of states in sample space. It also determines the phase factor $e^{i\gamma}$.

In this paper, we parametrize the guiding function as ANN, as shown in Fig. 1(b). ANNs, consisting of interconnected neurons, have the ability to approximate a function with arbitrary accuracy [70–72]. Recently, they have been used to represent wave functions in classical computing [73–78]. In our case, we use an ANN to parametrize the guiding function $\alpha(\theta; \lambda)$. The input to the network is the circuit parameter vector θ, and the output is a complex number α. The map from input to output depends on network parameters, represented by λ. Details of ANN are given in Sec. II A 3. We solve problems by optimizing λ.

The distribution of states $|\phi(\theta)\rangle$ is described by a probability density function that is proportional to the guiding function, i.e.,

$$P(\theta; \lambda) = \frac{|\alpha(\theta; \lambda)|}{C(\lambda)}, \tag{1}$$

where

$$C(\lambda) = \int d\theta |\alpha(\theta; \lambda)| \tag{2}$$

is the normalization factor. Note that θ is the random variable, and λ represents the parameters of the distribution. Given the guiding function, we can sample θ from the distribution using Markov chain Monte Carlo methods, e.g., the Metropolis-Hastings algorithm [79]. In addition, the phase factor is

$$e^{i\gamma(\theta;\lambda)} = \frac{\alpha(\theta; \lambda)}{|\alpha(\theta; \lambda)|}. \tag{3}$$

3. Rectifier ANN

We take the rectifier ANN as an example. The rectifier ANN is a feed-forward neural network with the rectifier activation function $\text{ReLU}(z) = \max\{0, z\}$ [80]. The activation function is an essential part of ANNs, giving them nonlinear expression ability. Activation functions include the rectifier activation function, sigmoid activation function, andtanh activation function. Rectifier feed-forward neural networks comprise one of the most generally used activation functions. Rectifier feed-forward neural networks, one of the most commonly used activation functions, stand out as a potent tool in various applications due to their exceptional attributes such as rapid convergence rate, swift learning speed, and streamlined expression for simplified calculations.

In the theoretical analysis, we take the rectifier ANN with only one hidden layer to parametrize the guiding function, see Fig. 3. It is straightforward to generalize to the case of multiple hidden layers and other activation functions. In the numerical experiment, we use a neural network with two hidden layers to have better training results. Without loss of generality, we suppose that the circuit parameter $\theta = (\vartheta_1, \vartheta_2, \ldots, \vartheta_K)^{\text{T}}$ is a K-dimensional real vector (for clarity, we use ϑ to denote the component of θ). Then the input layer has K neurons. Suppose the hidden layer has L neurons. The output of the hidden layer is an L-dimensional real vector $h = \text{ReLU}(w\theta + b)$. The output layer has only two neurons corresponding to the magnitude and phase of the guiding function α. Their outputs are two real numbers $A = w_A h + b_A$ and $B = w_B h + b_B$. The guiding function is parametrized as

$$\alpha(\theta; \lambda) = W(A)e^{iB}F(\theta), \tag{4}$$

Input layer **Hidden layer** **Output layer**

θ h

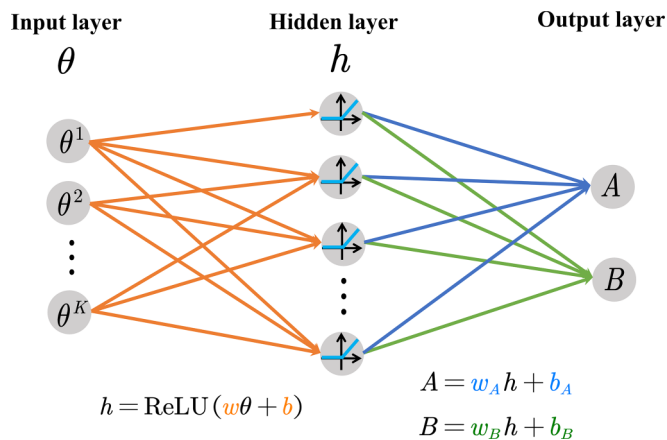$h = \text{ReLU}(w\theta + b)$

$A = w_A h + b_A$
$B = w_B h + b_B$

FIG. 3. The rectifier neural network with one hidden layer for parametrizing the guiding function $\alpha(\theta; \lambda)$, where $\lambda = (w, b, w_A, w_B, b_A, b_B)$.

where $\lambda = (w, b, w_A, b_A, w_B, b_B)$ is the parameter vector of the network, $W(A)$ is a nonnegative function [for instance, $W(A) = |A|$, $W(A) = e^{-A}$, and $W(A) = \exp(-\frac{1}{2}A^2)$], and $F(\theta)$ is a prior guiding function (a complex-valued function in general).

We introduce the prior guiding function to incorporate knowledge/intuition about the target state, such as which states $|\phi(\theta)\rangle$ may contribute more to the target state. We can also use the prior guiding function to confine θ to a subset of Θ, i.e., taking $F(\theta) = 0$ for θ outside the subset. Without any intuition or confinement, we can take $F(\theta) = 1$.

B. General formalism

The core idea of VQCMC is expressing a quantum state as a weighted average of states in the sample space. Given the expression, we can evaluate the expected value of operators with the Monte Carlo method and then the loss function used to find optimal parameters.

In VQAs, we solve problems by optimizing a variational wave function. In the deterministic-circuit approach, if $U(\theta)$ is the *Ansatz* circuit, the state $|\phi(\theta)\rangle$ is the variational wave function, i.e., we aim to find the solution in the state subset Ω.

In the random-circuit approach, we aim to express the quantum state as an average of states in Ω. Specifically, the variational wave function reads

$$|\psi(\lambda)\rangle = \int d\theta P(\theta; \lambda) e^{i\gamma(\theta; \lambda)} |\phi(\theta)\rangle. \quad (5)$$

Here, both the distribution $P(\theta; \lambda)$ and the phase $\gamma(\theta; \lambda)$ are determined by the guiding function $\alpha(\theta; \lambda)$. In Sec. IV, we theoretically prove that $|\psi(\lambda)\rangle$ in VQCMC has an expressive power not lower than $|\phi(\theta)\rangle$ in the conventional VQA methods with the same time cost. Further, by increasing the time cost, $|\psi(\lambda)\rangle$ can have a higher expressive power. Notice that this variational wave function is a function of network parameters λ. We optimize λ to solve problems, in contrast with optimizing circuit parameters θ in the deterministic-circuit approach.

In many VQAs, we determine the parameters of the variational wave function by minimizing a loss function. In this

paper, we focus on VQAs in this category, including VQE, quantum approximate optimization algorithm, and quantum neural networks. However, we would like to note that the random-circuit approach can also be employed in variational quantum simulators that update parameters following certain differential equations.

We take VQE as an example, in which the loss function is the expected value of a Hamiltonian operator H. Because the variational wave function $|\psi(\lambda)\rangle$ is unnormalized, the expected value of H reads

$$L(\lambda) = \frac{\langle\psi(\lambda)|H|\psi(\lambda)\rangle}{\langle\psi(\lambda)|\psi(\lambda)\rangle}. \quad (6)$$

By minimizing the loss function, we can find the optimal $|\psi(\lambda)\rangle$ to approximate the ground state of H.

In general, the loss function is a function of such expected values. We define the expected value of an operator O in the state $|\psi(\lambda)\rangle$ as

$$\langle O\rangle(\lambda) = \langle\psi(\lambda)|O|\psi(\lambda)\rangle, \quad (7)$$

and the expected value in the normalized state as

$$\widetilde{\langle O\rangle}(\lambda) = \frac{\langle\psi(\lambda)|O|\psi(\lambda)\rangle}{\langle\psi(\lambda)|\psi(\lambda)\rangle} = \frac{\langle O\rangle(\lambda)}{\langle\mathbb{1}\rangle(\lambda)}, \quad (8)$$

where $\mathbb{1}$ is the identity operator. With the notations, the loss function in VQE can be rewritten as $L(\lambda) = \widetilde{\langle H\rangle}(\lambda)$. Let $O_1, O_2, \ldots$ be a set of operators. The general loss function reads

$$L(\lambda) = L(\widetilde{\langle O_1\rangle}(\lambda), \widetilde{\langle O_2\rangle}(\lambda), \ldots). \quad (9)$$

We need to evaluate such loss functions to carry out VQAs.

Next, we present the method of evaluating an expected value $\langle O\rangle(\lambda)$ by sampling states in Ω, and we also give a detailed pseudocode.

The operator estimator

We only demonstrate how to evaluate $\langle O\rangle(\lambda)$. Then it is straightforward to evaluate expected values in the normalized state $\widetilde{\langle O\rangle}(\lambda)$ and the loss function $L(\lambda)$.

Given the expression of the variational wave function in Eq. (5), we can rewrite the expected value as

$$\langle O\rangle(\lambda) = \int d\theta d\theta' P(\theta; \lambda) P(\theta'; \lambda) \exp\{i[\gamma(\theta; \lambda) - \gamma(\theta'; \lambda)]\}$$
$$\times \langle\phi(\theta')|O|\phi(\theta)\rangle. \quad (10)$$

Notice that Eq. (10) is a double integral, where θ and θ' are independent of each other. In this way, we have transformed the expected value in a quantum state into the expected value of the quantity:

$$X_{\theta,\theta'}(\lambda) = \exp\{i[\gamma(\theta; \lambda) - \gamma(\theta'; \lambda)]\}\langle\phi(\theta')|O|\phi(\theta)\rangle. \quad (11)$$

According to the Monte Carlo method, we can evaluate $\langle O\rangle(\lambda)$ by computing the sample mean in randomly drawn (θ, θ').

We measure the quantity $X_{\theta,\theta'}$ on a quantum computer. We can use the Hadamard test [81] to measure such a quantity, including methods with and without ancilla qubit. See Appendix A for a brief review of these methods. Regardless of which method we use, quantum computing outputs an

---

ALGORITHM 1. Operator estimator.

1: Input an operator O, the neural network parameter λ, a sample
space Ω, and the sample size M.
2: **for** $l = 1$ to M
3:     Draw circuit parameters (θ, θ') according to the distribution
$P(\theta; \lambda)P(\theta'; \lambda)$. ▷ Use Markov chain Monte Carlo.
4:     $\hat{X}_{\theta,\theta'} \leftarrow \text{QUANTUMCOMPUTING}\lambda, \theta, \theta', O$ ▷ Via Hadamard
test shown in Appendix A.
5:     $\mu_l \leftarrow \hat{X}_{\theta,\theta'}$
6: Output $\langle \hat{O} \rangle \leftarrow \frac{1}{M} \sum_{l=1}^{M} \mu_l$.

---

estimate of $X_{\theta,\theta'}$, denoted by $\hat{X}_{\theta,\theta'}$. Due to the randomness
in quantum measurement, $\hat{X}_{\theta,\theta'}$ is usually inexact. In what
follows, we suppose that $\hat{X}_{\theta,\theta'}$ is unbiased and has a finite
variance, which is true in the Hadamard test.

The pseudocode for evaluating $\langle O \rangle(\lambda)$ is shown in
Algorithm 1. Here, $\langle \hat{O} \rangle$ is the output value. With
Algorithm 1 as the subroutine, the procedure of VQCMC is
like conventional VQAs; see Algorithm 2.

C. Statistical errors

There are two sources of statistical errors: quantum mea-
surement and randomized circuits. We will show that both
of them can be suppressed by taking a large sample size M
(see Algorithm 1). We will also show that statistical errors
are amplified when $\langle \mathbb{1} \rangle(\lambda)$ is small, which is known as the
sign problem. Therefore, we have to take a larger sample size
(i.e., time cost) when $\langle \mathbb{1} \rangle(\lambda)$ is smaller. Through $\langle \mathbb{1} \rangle(\lambda)$, we
can control the time cost.

For the first error source, when we use a quantum circuit to
measure the quantity $X_{\theta,\theta'}$, the measurement outcome in each
run of the circuit is random. To reduce this statistical error due
to quantum measurement, we can repeat the measurement. For
example, in the Hadamard test, we take $\hat{X}_{\theta,\theta'}$ as the mean of
outcomes in the repeated measurements. Then the variance
of $\hat{X}_{\theta,\theta'}$ decreases with the number of measurements, denoted
by M_Q. In the ancilla-qubit Hadamard test, supposing O is a
unitary Hermitian operator, the variance of $\hat{X}_{\theta,\theta'}$ has the upper
bound:

$$\text{Var}(\hat{X}_{\theta,\theta'}) \leqslant \sigma_O^2 = \frac{1}{M_Q}. \tag{12}$$

The result is similar for a general operator: For a general
operator, σ_O^2 depends on properties of O and details of the
measurement protocol, see Appendix A.

ALGORITHM 2. Variational quantum-circuit Monte Carlo.

1: Input a set of operator $\{O_1, O_2, \ldots\}$, initialized neural network
parameter λ, a sample space Ω, the sample size M, and
the training steps N.
2: **for** $t = 1$ to N
3:     Estimate the loss function $L(\lambda)$ [Eq. (9)] using Algorithm 1.
4:     Optimize λ to minimize $L(\lambda)$
5: Output the solution $L(\lambda)$.

For the second error source, $\langle O \rangle$ is the mean of the random
variable $X_{\theta,\theta'}$, which has fluctuation.

Overall, the variance of $\langle \hat{O} \rangle$ has two terms and is propor-
tional to $1/M$:

$$\text{Var}(\langle \hat{O} \rangle) = \frac{1}{M}(A + B - \langle O \rangle^2), \tag{13}$$

where

$$A = \int d\theta d\theta' P(\theta; \lambda)P(\theta'; \lambda) \, \text{Var}(\hat{X}_{\theta,\theta'}), \tag{14}$$

$$B = \int d\theta d\theta' P(\theta; \lambda)P(\theta'; \lambda)|X_{\theta,\theta'}|^2. \tag{15}$$

Here, A and B correspond to the first and second error sources,
respectively.

Let $\|O\|$ be the spectral norm of O. Then $B \leqslant \|O\|^2$, and
the overall variance has the upper bound:

$$\text{Var}\left(\langle \hat{O} \rangle\right) \leqslant \frac{1}{M}\left(\|O\|^2 + \sigma_O^2\right). \tag{16}$$

When $\sigma_O^2 \propto 1/M_Q$, it is optimal to take $M_Q = 1$ to minimize
the total number of measurements MM_Q required for achiev-
ing a certain value of the variance.

To complete the discussion on statistical properties of $\langle \hat{O} \rangle$,
we note that, when $\hat{X}_{\theta,\theta'}$ is unbiased, $\langle \hat{O} \rangle$ is also unbiased.

1. Sign problem and time cost

In classical computing, quantum Monte Carlo suffers from
the sign problem. It arises when the probability distribution
sampled in the Monte Carlo simulation has a complex phase,
resulting in a cancellation of positive and negative contribu-
tions. VQCMC has a similar problem: When the phase of
$X_{\theta,\theta'}$ is oscillatory, the absolute value of $\langle O \rangle$ is small; then
the relative error is large. The large relative error can lead to
a large absolute error when evaluating expected values in the
normalized state $\widetilde{\langle O \rangle}$.

Let $\delta_O = \langle \hat{O} \rangle - \langle O \rangle$ be the error in $\langle \hat{O} \rangle$. The error in eval-
uating $\widetilde{\langle O \rangle}$ is

$$\frac{\langle \hat{O} \rangle}{\langle \hat{\mathbb{1}} \rangle} - \widetilde{\langle O \rangle} \simeq \frac{\delta_O - \widetilde{\langle O \rangle}\delta_{\mathbb{1}}}{\langle \mathbb{1} \rangle}. \tag{17}$$

We can find that the error is amplified by a factor of $1/\langle \mathbb{1} \rangle$.
To suppress the error, we have to take a sufficiently large M,
such that δ_O and $\delta_{\mathbb{1}}$ are sufficiently small compared with $\langle \mathbb{1} \rangle$.
Therefore, when $\langle \mathbb{1} \rangle$ is small, the sample size is large. This
observation is summarized in the following theorem, and the
proof is in Appendix B.

Theorem 1. Let ε and κ be any positive numbers. When the
sample size satisfies

$$M \geqslant \frac{\chi^2}{\kappa \varepsilon^2 \langle \mathbb{1} \rangle^2}, \tag{18}$$

where

$$\chi = \sqrt{\|O\|^2 + \sigma_O^2} + \sqrt{1 + \sigma_{\mathbb{1}}^2}\|O\| + \varepsilon \sqrt{1 + \sigma_{\mathbb{1}}^2}, \tag{19}$$

the statistical error is smaller than ε with the probability:

$$\text{Pr}\left(\left|\frac{\langle \hat{O} \rangle}{\langle \hat{\mathbb{1}} \rangle} - \widetilde{\langle O \rangle}\right| \leqslant \varepsilon\right) \geqslant 1 - 2\kappa. \tag{20}$$
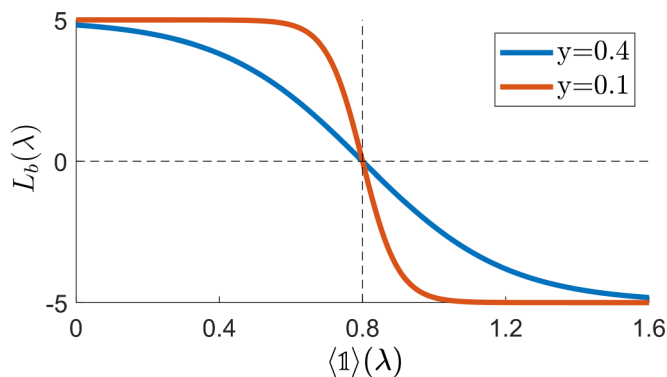
FIG. 4. The barrier term $L_b(\lambda) = -z \tanh[\frac{\langle\mathbb{1}\rangle(\lambda)-x}{y}]$ in the modified loss function, where the parameter x specifies the position of the barrier. We illustrate this with two examples: for $y = 0.4$ and 0.1, while holding $z = 5$ and $x = 0.8$ constant.

According to Theorem 1, when we take $M = \chi^2/(\kappa \varepsilon^2 \langle\mathbb{1}\rangle^2)$, the sample size (i.e., time cost) is sufficient for achieving the accuracy ε with a probability not lower than $1 - 2\kappa$.

2. Modified loss functions and time-cost control

According to Theorem 1, we can control the time cost (i.e., the sample size) by confining the variational wave function in the region where $\langle\mathbb{1}\rangle$ is large. Here, we propose two methods for this purpose. The first method is generic for all VQAs that work through minimizing a loss function. The second method is specific to VQE. We will give a rigorous theoretical justification of the second method.

In the first method, we modify the loss function and take

$$L'(\lambda) = L(\lambda) - z \tanh\left[\frac{\langle\mathbb{1}\rangle(\lambda) - x}{y}\right], \quad (21)$$

where the tanh function is added to the raw loss function. The tanh function plays the role of a barrier at $\langle\mathbb{1}\rangle(\lambda) = x$ (see Fig. 4): In the vicinity of the barrier, the loss function increases rapidly (y determines how rapidly) when $\langle\mathbb{1}\rangle(\lambda)$ decreases. The height of the barrier is determined by z. With the tanh function, λ prefers to stay where $\langle\mathbb{1}\rangle > x$, thus controlling the time cost. Some other functions, e.g., the rectifier and sigmoid functions, can be used in a similar way (after a preliminary numerical test, we find that the tanh may perform better than the other two).

In the second method, we regularize the loss function of VQE by adding positive constants to the denominator and numerator, i.e.,

$$L''(\lambda) = \frac{\langle H\rangle(\lambda) + \eta_H}{\langle\mathbb{1}\rangle(\lambda) + \eta_{\mathbb{1}}}. \quad (22)$$

Now we explain why such a loss function can prevent $\langle\mathbb{1}\rangle$ from vanishing. The exact value of $\langle\mathbb{1}\rangle$ is always positive. Adding $\eta_{\mathbb{1}}$ to the denominator makes sure that the denominator is positive (with a certain probability) even with the presence of statistical error, noticing that $\langle\mathbb{1}\rangle$ may be negative due to the statistical error. When the denominator is positive, adding η_H to the numerator increases the value of the loss function. The increment is larger when the denominator is smaller.

Therefore, minimizing the modified loss function can prevent the denominator from vanishing.

By taking proper values of $\eta_{\mathbb{1}}$ and η_H, the modified loss function $L''(\lambda)$ has two good properties. First, it is variational, i.e., $L''(\lambda) \geqslant E_g$, where E_g is the ground-state energy of H. Therefore, minimizing $L''(\lambda)$ always leads to a better result of the ground-state energy. The estimator of the modified loss function inherits this property. Second, we have an analytical upper bound of the statistical error in the energy [the raw loss function $L(\lambda) = \widetilde{\langle H\rangle}(\lambda)$ is the expected value of the energy in the normalized state]. These two properties are summarized in the following theorem, and the proof is in Appendix C.

Theorem 2. Let κ be any positive number. Take

$$\eta_H = \sqrt{\frac{\|H\|^2 + \sigma_H^2}{\kappa M}} \text{ and } \eta_{\mathbb{1}} = \sqrt{\frac{1 + \sigma_{\mathbb{1}}^2}{\kappa M}}.$$

An estimate of the modified loss function

$$\hat{L}''(\lambda) = \frac{\langle \hat{H}\rangle(\lambda) + \eta_H}{\langle \hat{\mathbb{1}}\rangle(\lambda) + \eta_{\mathbb{1}}} \quad (23)$$

is in the interval:

$$\min\{\widetilde{\langle H\rangle}(\lambda), 0\} \leqslant \hat{L}'' \leqslant \widetilde{\langle H\rangle}(\lambda) + 2\frac{\eta_H + \|H\|\eta_{\mathbb{1}}}{\langle\mathbb{1}\rangle(\lambda)}, \quad (24)$$

with a probability of at least $1 - 2\kappa$.

We can find that the estimator of the modified loss function is variational when the true ground-state energy E_g is negative. The negativity condition can always be satisfied by subtracting a sufficiently large positive constant from the Hamiltonian, i.e., taking $H \leftarrow H - E_{\text{const}}$. When E_g is negative, $\hat{L}'' \geqslant E_g$ holds in the entire parameter space up to a controllable probability (notice that $\widetilde{\langle H\rangle}(\lambda) \geqslant E_g$).

III. NUMERICAL DEMONSTRATION

In this section, we demonstrate the random-circuit approach by numerical simulations. We take VQE as an example of VQAs and solve the ground state of an antiferromagnetic Heisenberg model. We choose the barrier method to control the time cost, and we illustrate the trade-off between expressive power and time cost. The numerical simulations are performed on a high-performance server with an Intel Xeon Gold 6230 CPU @ 2.10GHz CPU. They can also run on ordinary laptops.

The antiferromagnetic Heisenberg model has six qubits on a randomly generated graph, as shown in Fig. 5. On the graph, each vertex represents a spin-$\frac{1}{2}$ particle, and each edge represents the coupling between the two spins. The Hamiltonian reads

$$H = J \sum_{\langle i,j\rangle} (X_i X_j + Y_i Y_j + Z_i Z_j), \quad (25)$$

where X_i, Y_i, and Z_i denote Pauli operators of the ith qubit, J is the coupling strength, and $\langle i, j\rangle$ denotes two coupled spins on the graph. We take the coupling strength J such that the Hamiltonian is normalized by the spectral norm, i.e., $\|H\| = 1$.

We parametrize the circuit with a simplified version of the Hamiltonian *Ansatz*. The circuit has only two parameters θ_{XY}
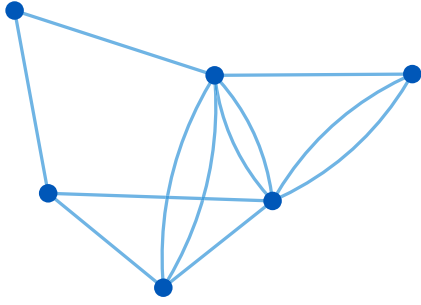
FIG. 5. A six-qubit Heisenberg model. Qubits are denoted by vertices. Two qubits are coupled if there is an edge incident on corresponding vertices.

and θ_Z:

$$U(\theta) = [R(\theta)]^{N_T} R_0, \qquad (26)$$

where

$$R(\theta) = \prod_{\langle i,j \rangle} \exp\left(-\frac{iZ_iZ_j\theta_Z}{2}\right) \exp\left(-\frac{iY_iY_j\theta_{XY}}{2}\right)$$
$$\times \exp\left(-\frac{iX_iX_j\theta_{XY}}{2}\right), \qquad (27)$$

N_T denotes the number of gate layers, R_0 prepares the pairwise singlet state:

$$|\Psi_0\rangle = R_0|0\rangle^{\otimes n} = |\Phi\rangle_{1,2} \otimes |\Phi\rangle_{3,4} \otimes \cdots, \qquad (28)$$

and

$$|\Phi\rangle_{i,j} = \frac{1}{\sqrt{2}}(|0\rangle_i \otimes |1\rangle_j - |1\rangle_i \otimes |0\rangle_j) \qquad (29)$$

is the singlet state of spins i and j.

With this simplified Hamiltonian *Ansatz* circuit, we can plot the energy landscape as shown in Fig. 6. This energy landscape has undesired properties. First, it has a rugged surface. When using gradient descent and randomly choosing the initial point, we find that the point easily falls into a local minimum. Second, even for the global minimum, its error in the energy is ~ 0.05 (the exact ground-state energy is -1, and the global minimum is about -0.95). In other words,
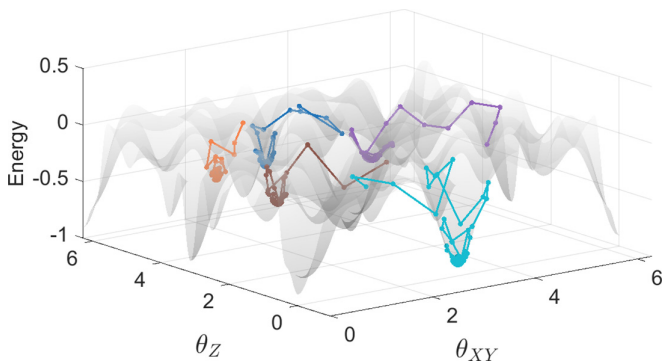

FIG. 6. Energy landscape of the simplified Hamiltonian *Ansatz* circuit with $N_T = 2$. The colored lines represent the training process of finding the minimum value using the variational quantum algorithm (VQA) method from different initial points.

the expressive power of the *Ansatz* circuit is insufficient for approximating the ground state to achieve an error <0.05. Therefore, the performance of the *Ansatz* circuit is poor in the deterministic-circuit approach.

In the random-circuit approach, we can solve the ground state to a satisfactory accuracy even with the poor-performance *Ansatz* circuit. By controlling the time cost, we can systematically increase the expressive power and approach the exact solution with a sufficiently large time cost.

In the random-circuit approach, we parametrize the guiding function using a rectifier ANN with two hidden layers, each consisting of 200 neurons. We take $\theta = (\theta_{XY}, \theta_Z)$ as the input to ANN. For the output function, we take $W(A) = e^{-A}$. We choose the prior guiding function $F(\theta) = \sum_{i=1}^{N_\theta} \delta(\theta - \theta_i)$, where θ_i are $N_\theta = 100$ uniformly generated points in the parameter space. With this prior guiding function, we effectively utilize a finite sample space to simplify the numerical simulation.

Implementing the barrier method requires a proper initial value of the parameter λ. Suppose the initial λ is randomly chosen in the parameter space, the corresponding value of $\langle \mathbb{1} \rangle(\lambda)$ may be small and violate the $\langle \mathbb{1} \rangle(\lambda) > x$ restriction. In this case, the statistical error can be large and cause problems in the initial stage of training. To avoid this issue, we must choose the initial value satisfying $\langle \mathbb{1} \rangle(\lambda) > x$. We give two methods of doing this. First, we can take λ such that $P(\theta; \lambda) \approx \delta(\theta - \theta_0)$ is a δ-function distribution. For such a distribution, $\langle \mathbb{1} \rangle(\lambda) \approx 1$. In Sec. IV, we will show how to take λ to approximate the δ-function distribution by the ANN. Second, we can employ two-stage training, as shown in Fig. 7. In the first stage, we can take $L_b(\lambda) = -z \tanh[\frac{\langle \mathbb{1} \rangle(\lambda) - x}{y}]$ as the loss function to find a value of λ satisfying $\langle \mathbb{1} \rangle(\lambda) > x$. This tanh loss function is robust to the statistical error in $\langle \mathbb{1} \rangle(\lambda)$ even when $\langle \mathbb{1} \rangle(\lambda)$ is small [in contrast with $L(\lambda)$, in which $\langle \mathbb{1} \rangle(\lambda)$ is the denominator]. In the second stage, we take the value λ determined in the first stage as the initial value and minimize $L'(\lambda)$. We obtain our numerical results with the two-stage method.

We set the barrier at $x = 0.1, 0.2, \ldots, 0.9$. Then the denominator $\langle \mathbb{1} \rangle(\lambda)$ is confined in the $\langle \mathbb{1} \rangle(\lambda) > x$ regime, respectively. For each value of x, we repeat the numerical experiment 10 times. The results are shown in Fig. 8. We can find that the error decreases when we reduce the value of x, which illustrates the power-cost trade-off. In Figs. 7 and 8, the VQCMC method achieves a loss value $L(\lambda)$ very close to -1, in contrast to the conventional VQE's result of ~ 0.95 in Fig. 6. This numerical result demonstrates the advantage of the proposed VQCMC method.

IV. THEORETICAL RESULTS ON THE EXPRESSIVE POWER

In this section, we present a few theoretical results on the expressive power of the random-circuit approach. First of all, let us introduce the ways of characterizing *expressive power* and *time cost*.

We characterize the expressive power in two straightforward ways: the subset of variational wave functions and the error in the ground-state energy. Discussions based on the
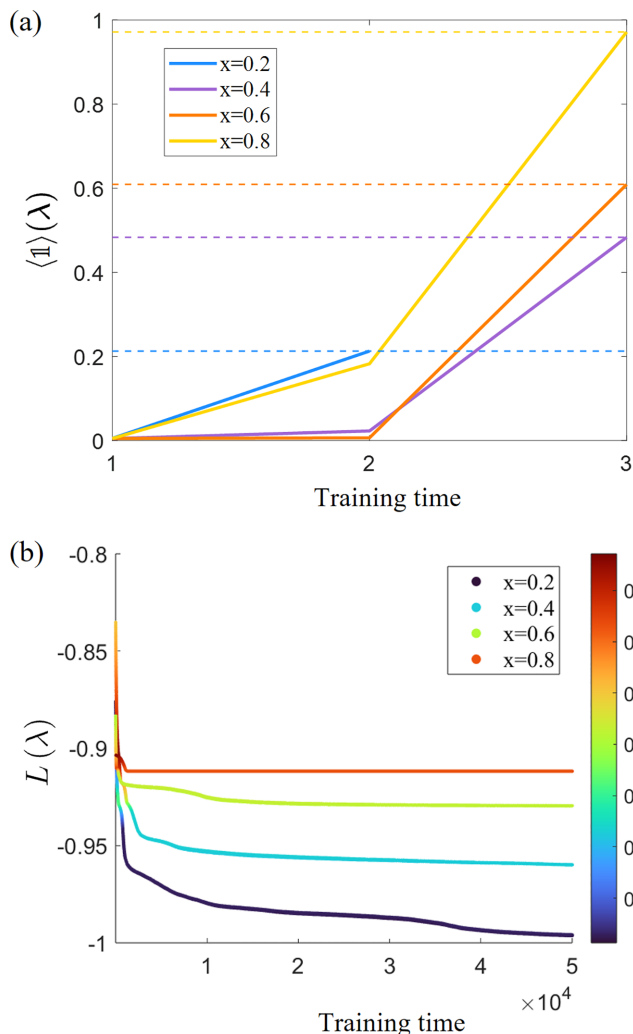
(a)

(b)

FIG. 7. The two-stage training in the barrier method. (a) In the first stage, we minimize the loss function L_b, and it ends when $\langle \mathbb{1} \rangle(\lambda) > x$ is satisfied. Here, we take $x = 0.2$, 0.4, and 0.8 as examples. (b) In the second stage, we take the resulting value of λ trained in the first stage as the initial value, then we minimize the loss function $L'(\lambda)$. $L(\lambda)$ decreases with the training times. The color bar represents the value of $\langle \mathbb{1} \rangle(\lambda)$. In both stages, we take $z = 5$, $y = 1$, and we use a gradient descent algorithm Adam to minimize loss functions.

subset are general for various VQA tasks. When focusing on VQE, we employ the energy error.

We define the variational–wave function subset of the $|\psi(\lambda)\rangle$ as

$$\mathcal{V} = \left\{ |\widetilde{\psi}(\lambda)\rangle = \frac{|\psi(\lambda)\rangle}{\sqrt{\langle \mathbb{1} \rangle(\lambda)}} \,\middle|\, \lambda \in \Lambda \right\}, \qquad (30)$$

where Λ denotes the space of parameters [$\lambda = (w, b, w_A, b_A, w_B, b_B)$ has $N = KL + 3L + 2$ real parameters; $w \in \mathbb{R}^{L \times K}$; $b, w_A, w_B \in \mathbb{R}^L$; and $b_A, b_B \in \mathbb{R}$; therefore, $\Lambda = \mathbb{R}^N$]. Notice that states in this subset are normalized. The normalization factor $\langle \mathbb{1} \rangle(\lambda)$ is related to the sample size M (see Theorem 1). Given a finite permitted run time, the sample size is finite, and we can only utilize a subset of $\mathcal{V}$ in VQAs: We cannot evaluate a state to adequate accuracy when
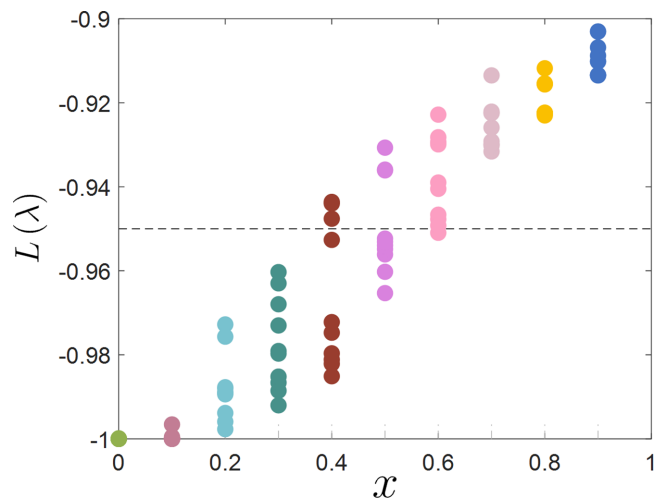

FIG. 8. The ground-state energy estimated in the random-circuit approach with different barriers. Each point represents the final value of $L(\lambda)$ in a numerical experiment. For each value of x, the experiment is repeated 10 times. The dashed line indicates the global minimum energy in the deterministic-circuit approach.

$\langle \mathbb{1} \rangle(\lambda)$ is too small. To reflect this time cost constraint, we define

$$\mathcal{V}_x = \{ |\widetilde{\psi}(\lambda)\rangle \,|\, \langle \mathbb{1} \rangle(\lambda) \geqslant x \}. \qquad (31)$$

For a larger subset $\mathcal{V}_x$, its expressive power is higher. As shown in Appendix D, there is a simple relation between the variational–wave function subset and the covering number of the hypothesis space (a measure of expressive power).

For the variational–wave function subset $\mathcal{V}_x$, we define the minimum error in the ground-state energy as

$$\epsilon_g(\mathcal{V}_x) = \min_{|\widetilde{\psi}(\lambda)\rangle \in \mathcal{V}_x} \langle \widetilde{\psi}(\lambda)|H|\widetilde{\psi}(\lambda)\rangle - E_g. \qquad (32)$$

The aim of VQE is to find the ground state of a Hamiltonian. The error ϵ_g describes how well the variational wave function can approximate the ground state.

We characterize the time cost with x. According to Theorem 1, the sample size M (i.e., time cost) is proportional to $1/\langle \mathbb{1} \rangle(\lambda)^2$. When a larger M is taken, we can search for the solution in a subset $\mathcal{V}_x$ [Eq. (31)] with a smaller x. Specifically, according to Theorem 1, given the sample size M, an upper bound of the statistical error ε, and failure probability 2κ, we can search for the solution in $\mathcal{V}_x$ with

$$x = \frac{\chi}{\varepsilon\sqrt{\kappa M}}. \qquad (33)$$

Therefore, the number of measurements M, quantifying the time cost, is directly proportional to $1/x^2$. In short, when the time cost is larger, x is smaller, and vice versa.

A. Low-cost limit

The time cost of evaluating a state in the random-circuit approach is determined by the normalization factor $\langle \mathbb{1} \rangle(\lambda)$. The largest value that $\langle \mathbb{1} \rangle(\lambda)$ can take is one, i.e., $\langle \mathbb{1} \rangle(\lambda) = 1$ corresponds to the low-cost limit.

Our first theoretical result is that Ω is a subset of the closure of $\mathcal{V}_x$ (for all $x < 1$). In other words, we can express all states
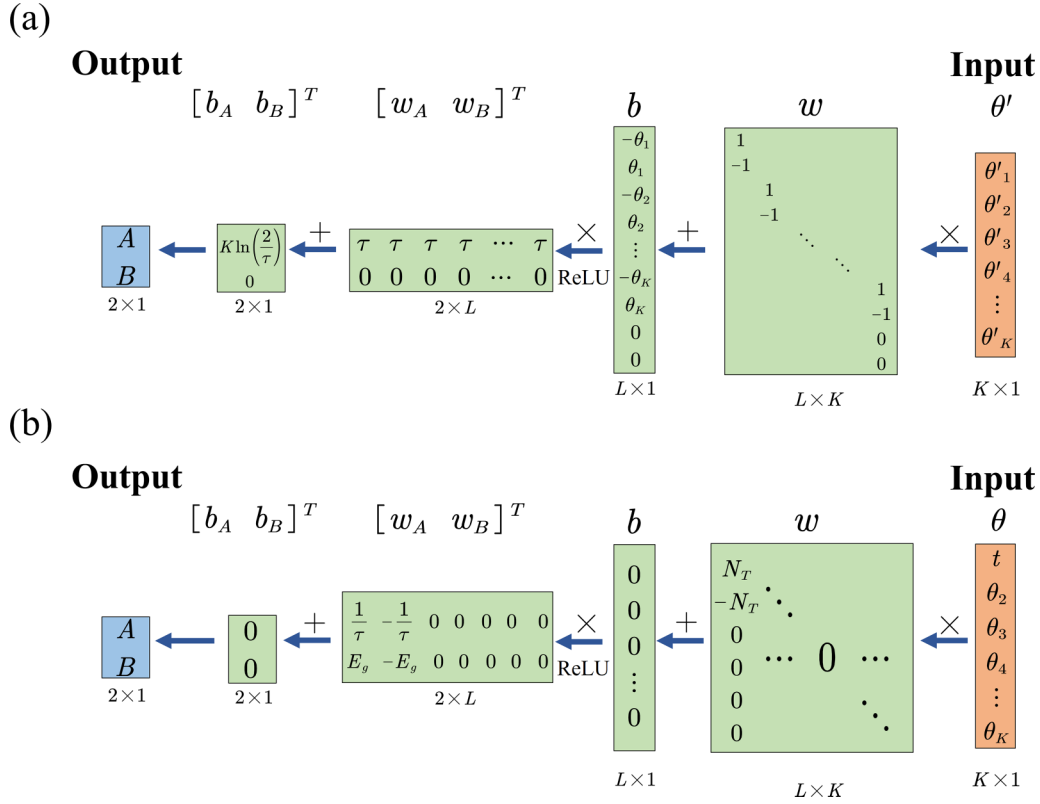
(a)

(b)

FIG. 9. (a) The values of artificial neural network (ANN) parameters when approximating $|\phi(\theta)\rangle$ with the random-circuit variational wave function. (b) The values of ANN parameters when considering Hamiltonian *Ansatz* to approximate the ground state.

$|\phi(\theta)\rangle \in \Omega$ to arbitrary accuracy with the random-circuit variational wave function $|\psi(\lambda)\rangle$, and the normalization factor $\langle \mathbb{1} \rangle(\lambda)$ approaches one. Therefore, the expressive power of $|\psi(\lambda)\rangle$ in the low-cost limit is not lower than $|\phi(\theta)\rangle$.

We assume that $|\phi(\theta)\rangle$ has a finite gradient with respect to θ. Let $\mathbf{n}$ be a unit vector in the space of θ. We assume there exists a positive number ξ such that, for all $\mathbf{n}$ and θ,

$$\|\mathbf{n} \cdot \nabla|\phi(\theta)\rangle\|_2 \leqslant \xi. \tag{34}$$

Here, $\| \bullet \|_2$ denotes the ℓ^2 norm. This assumption is true for all *Ansatz* circuits to the best of our knowledge. For example, if each parameter component ϑ_i is the angle of a rotation gate $\exp(-i\sigma_i\vartheta_i)$, $\|\frac{\partial}{\partial \vartheta_i}|\phi(\theta)\rangle\|_2 = 1$; then $\|\mathbf{n} \cdot \nabla|\phi(\theta)\rangle\|_2 \leqslant \sqrt{K}$ (K is the dimension of θ). Under this assumption, we have the following theorem.

Theorem 3. Suppose the guiding function α is parametrized as a rectifier ANN, the hidden layer has $L \geqslant 2K$ neurons, $W(A) = e^{-A}$, and $F(\theta) = 1$. Then for all $|\phi(\theta)\rangle \in \Omega$ and $x < 1$,

$$\inf_{|\widetilde{\psi}(\lambda)\rangle \in \mathcal{V}_x} \||\widetilde{\psi}(\lambda)\rangle - |\phi(\theta)\rangle\|_2 = 0. \tag{35}$$

The proof is given in Appendix E. To approximate $|\phi(\theta)\rangle$ with the random-circuit variational wave function, we consider values of ANN parameters as follows [see Fig. 9(a)]. We use θ' to denote the input to the neural network for clarity): (i) For the jth hidden neuron, we take $w_{j,i} = \delta_{j,2i-1} - \delta_{j,2i}$ and $b_j = -\sum_i (\delta_{j,2i-1} - \delta_{j,2i})\vartheta_i$. (ii) For the output neuron A, we take $w_{A,j} = \tau$ and $b_A = K \ln(2/\tau)$. (iii) For the output neuron B, we take $w_{B,j} = b_B = 0$. With these parameters, hidden

neurons realize the modulus calculation, i.e., $x_{2i-1} + x_{2i} = |\vartheta_i' - \vartheta_i|$. Outputs of the neural network are $A = \tau \sum_i |\vartheta_i' - \vartheta_i| + K \ln(2/\tau)$ and $B = 0$. Here, τ is a positive number. Then the corresponding random-circuit variational wave function is

$$|\psi(\lambda)\rangle = \left(\frac{\tau}{2}\right)^K \int d\theta' \exp(-\tau\||\theta' - \theta\|_1)|\phi(\theta')\rangle. \tag{36}$$

In the limit $\tau \to +\infty$, $|\psi(\lambda)\rangle$ [as well as the normalized state $|\widetilde{\psi}(\lambda)\rangle$] approaches $|\phi(\theta)\rangle$, and the distance between the two states vanishes.

We remark that, when $|\psi(\lambda)\rangle$ approaches $|\phi(\theta)\rangle$, the corresponding value of $\langle \mathbb{1} \rangle(\lambda)$ approaches one. Therefore, we can approximate $|\phi(\theta)\rangle$ with $|\psi(\lambda)\rangle$ in the low-cost limit.

B. Power-cost trade-off

Our second theoretical result is a simple observation that justifies the trade-off between expressive power and time cost. The following proposition is straightforward, and we present it without proof. It shows that the variational–wave function subset $\mathcal{V}_x$ enlarges monotonically when x decreases. Therefore, the expressive power increases monotonically with the time cost.

Proposition 1. For all $x_1 \leqslant x_2$, $\mathcal{V}_{x_1} \supseteq \mathcal{V}_{x_2}$.

C. High-cost limit: Universal approximation theorem

Now we discuss the extreme case of expressive power in the random-circuit approach. We ask whether all states in the Hilbert space can be approximated with the random-circuit

variational wave function when an arbitrarily large time cost is allowed.

Suppose Ω is a spanning set of the entire Hilbert space. All states can be expressed in the form:

$$|\varphi\rangle = \int d\theta \beta(\theta)|\phi(\theta)\rangle. \tag{37}$$

If we can approximate $\beta(\theta)$ with the guiding function $\alpha(\theta; \lambda)$, we can approximate $|\varphi\rangle$ with the variational wave function $|\psi(\lambda)\rangle$. Owing to the universal approximation theorem of ANNs, we can approximate an arbitrary continuous function to arbitrary accuracy [82]. Therefore, we have the following theorem. The proof is in Appendix F.

Theorem 4. Suppose the guiding function α is parametrized as a rectifier neural network, the hidden layer has L neurons, $W(A) = e^{-A}$, and $F(\theta) = 1$. Under conditions (i) Ω is a spanning set of the Hilbert space $\mathcal{H}$, (ii) the circuit parameter space Θ is compact, and (iii) $|\phi(\theta)\rangle$ has a finite gradient with respect to θ, for all normalized states $|\varphi\rangle \in \mathcal{H}$ and $\epsilon > 0$, there exist $L \in \mathbb{N}$ and $\lambda \in \Lambda$ such that

$$\|\,|\widetilde{\psi}(\lambda)\rangle - |\varphi\rangle\,\|_2 < \epsilon. \tag{38}$$

The compact condition holds when each component is a rotation angle of a single-qubit gate, i.e., its value is in the interval $[0, 2\pi]$. The spanning-set condition holds for many *Ansatz* circuits, for example, a circuit with one layer of single-qubit rotation gates on the initial state $|0\rangle^{\otimes n}$.

D. Hamiltonian Ansatz: Energy error vs time cost

In what follows, we discuss the trade-off feature taking the ground-state problem and the Hamiltonian *Ansatz* as an example. We take the error in the ground-state energy $\epsilon_g(\mathcal{V}_x)$ as the measure of the expressive power. We show that (1) an upper bound of $\epsilon_g(\mathcal{V}_x)$ decreases with $1/x$, and (2) the error upper bound approaches zero in the limit $1/x \to \infty$. It is noteworthy that the error upper bound approaches zero for Hamiltonian *Ansätze* with any circuit depth.

1. Hamiltonian Ansatz

The Hamiltonian *Ansatz* is a parametrized Trotterization circuit [58]. Suppose that $H = \sum_{j=1}^{N_H} h_j \sigma_j$, where σ_j are Pauli operators, h_j are real coefficients, and N_H is the number of terms. The Hamiltonian *Ansatz* reads

$$U(\theta) = R(\omega_{N_T}) \cdots R(\omega_2) R(\omega_1) R_0, \tag{39}$$

where R_0 is a unitary operator that prepares the initial state $|\Psi_0\rangle = R_0 |0\rangle^{\otimes n}$,

$$R(\omega_i) = \exp(-i\sigma_{N_H}\omega_{i,N_H}) \cdots \exp(-i\sigma_2\omega_{i,2})\exp(-i\sigma_1\omega_{i,1}) \tag{40}$$

is the unitary operator of one Trotter step, $\omega_{i,j}$ is the angle of the jth rotation gate in the ith Trotter step, and N_T is the number of Trotter steps. We take the circuit parameter vector as $\theta = (t, \omega_1, \omega_2, \ldots, \omega_{N_T})$, in which t is a redundant real parameter. The usage of the redundant parameter will be shown later.

2. ANN configuration and prior guiding function

We use a rectifier ANN with $L \geqslant 2$ hidden-layer neurons to parametrize the guiding function, and we take $W(A) = \exp(-\frac{1}{2}A^2)$. We choose a nontrivial prior guiding function $F(\theta)$ inspired by the Pauli-operator-expansion formula of real-time evolution [83]; see Appendix G. The prior guiding function has the following property:

$$\int d\omega_1 \cdots d\omega_{N_T} F(\theta) R(\omega_{N_T}) \cdots R(\omega_1) = \exp(-iHN_T t). \tag{41}$$

To approximate the ground state, we consider values of ANN parameters as follows [see Fig. 9(b)]: (i) We suppose that $\vartheta_1 = t$ is the first input neuron. (ii) For the jth hidden neuron, we take $w_{j,i} = (\delta_{j,1} - \delta_{j,2})\delta_{i,1}N_T$ and $b_j = 0$, such that only the first two hidden neurons are nontrivial. (iii) For the output neuron A, we take $w_{A,j} = (\delta_{j,1} - \delta_{j,2})\tau^{-1}$ and $b_A = 0$. (iv) For the output neuron B, we take $w_{B,j} = (\delta_{j,1} - \delta_{j,2})E_g$ and $b_B = 0$. Here, τ is a positive number.

With the above configuration of parameters, outputs are $A = \tau^{-1}N_T t$ and $B = E_g N_T t$. Then the corresponding variational wave function is

$$\begin{aligned}
|\psi(\lambda)\rangle &= \frac{1}{C(\lambda)} \int d\theta \exp\left(-\frac{N_T^2 t^2}{2\tau^2}\right) \\
&\quad \times \exp(iE_g N_T t) F(\theta) U(\theta)|0\rangle^{\otimes n} \\
&= \frac{\tau\sqrt{2\pi}}{C(\lambda)N_T} \exp\left[-\frac{1}{2}(H - E_g)^2\tau^2\right]|\Psi_0\rangle. \quad (42)
\end{aligned}$$

The operator $\exp[-\frac{1}{2}(H - E_g)^2\tau^2]$ is a result of the integral over t [84], and it projects the initial state onto the ground state in the limit $\tau \to +\infty$.

3. Projection onto the ground state

The operator $\exp[-\frac{1}{2}(H - E_g)^2\tau^2]$ partially projects the initial state onto the ground state when τ is finite. Let Δ be the energy gap between the ground state $|\Psi_g\rangle$ and the first excited state. When $\Delta \geqslant (\sqrt{2}\tau)^{-1}$, the error in the energy decreases exponentially with τ; otherwise, the error is inversely proportional to τ.

Lemma 1. Let $E(\lambda)$ be the energy of the state in Eq. (42). Then the error in the energy has the upper bound:

$$E(\lambda) - E_g \leqslant \frac{1 - p_g}{p_g}\exp(-\tilde{\Delta}^2\tau^2)\tilde{\Delta}, \tag{43}$$

where $p_g = |\langle\Psi_g|\Psi_0\rangle|^2$ and $\tilde{\Delta} = \max\{\Delta, (\sqrt{2}\tau)^{-1}\}$.

The proof is given in Appendix H.

4. Monotonic dependence on the time cost

First, we give a lower bound of the normalization factor $\langle\mathbb{1}\rangle(\lambda)$, which determines the time cost. See the following lemma, and the proof is in Appendix I.

Lemma 2. Let $\langle\mathbb{1}\rangle(\lambda)$ be the normalization factor of the state in Eq. (42). Then the normalization factor has the lower bound:

$$\langle\mathbb{1}\rangle(\lambda) \geqslant \frac{2\pi p_g}{c(\tau)^2}, \tag{44}$$

where

$$c(\tau) = \int du \exp\left(-\frac{u^2}{2}\right)$$

$$\times \left[2\exp\left(\frac{h_{\text{tot}}\tau}{N_T}|u|\right) - 1 - 2\frac{h_{\text{tot}}\tau}{N_T}|u|\right]^{N_T}, \quad (45)$$

and $h_{\text{tot}} = \sum_{j=1}^{N_H} |h_j|$. When $N_T > 4h_{\text{tot}}^2\tau^2$,

$$c(\tau) \leqslant \sqrt{\frac{2\pi}{1 - 4h_{\text{tot}}^2\tau^2/N_T}}. \quad (46)$$

Then with the lower bound in Eq. (44), we can conclude that the state in Eq. (42) with $\tau = c^{-1}(\sqrt{2\pi p_g/x})$ is in the subset $\mathcal{V}_x$. Here, we have used that $c(\tau)$ is strictly monotonic, and $c^{-1}(\bullet)$ denotes the inverse function. Then we can apply the error upper bound in Eq. (43) to $\mathcal{V}_x$. We have the following result (which holds for all N_T), and the proof is straightforward.

Theorem 5. Take the Hamiltonian *Ansatz* circuit. Suppose the guiding function α is parametrized as a rectifier neural network, the hidden layer has $L \geqslant 2$ neurons, and $W(A) = \exp(-\frac{1}{2}A^2)$. With a proper prior guiding function $F(\theta)$, the error in the energy satisfies

$$\epsilon_g(\mathcal{V}_x) \leqslant \frac{1 - p_g}{p_g}\exp(-\tilde{\Delta}^2\tau^2)\tilde{\Delta}, \quad (47)$$

where $\tau = c^{-1}(\sqrt{2\pi p_g/x})$. The error upper bound decreases monotonically with $1/x$ and approaches zero (τ approaches ∞) in the limit $1/x \to \infty$.

According to Eq. (33), $1/x \propto \sqrt{M}$. Therefore, the error upper bound decreases monotonically with the time cost. We remark that a condition of the above result is $p_g > 0$, which is a requirement on the initial state $|\Psi_0\rangle$.

E. Hamiltonian Ansatz: Scaling with the circuit depth

In Theorem 5, the monotonic relation between the error upper bound and the time-cost factor $1/x$ holds for all circuit depths N_T. It raises the question: What is the advantage of using a more powerful quantum computer that can realize quantum circuits with more gates? We answer the question in this section. We show that, for all $x < p_g$, the error upper bound decreases with the circuit depth and vanishes in the limit of a large depth.

The upper bound of $c(\tau)$ in Eq. (46) implies a lower bound of $\tau = c^{-1}(\sqrt{2\pi p_g/x})$. When $x < p_g$,

$$c^{-1}\left(\sqrt{\frac{2\pi p_g}{x}}\right) \geqslant \frac{1}{2h_{\text{tot}}}\sqrt{N_T\left(1 - \frac{x}{p_g}\right)}. \quad (48)$$

Therefore, the upper bound of $\epsilon_g(\mathcal{V}_x)$ in Eq. (47) decreases with N_T. In the limit $N_T \to \infty$, the error upper bound approaches zero.

Let ϵ be the permissible error in the ground-state energy. According to Lemma 1, we can approximate the ground state

and satisfy the permissible error ϵ by taking

$$\tau = \begin{cases} \frac{1-p_g}{\sqrt{2e}p_g\epsilon}, & \epsilon > \frac{1-p_g}{\sqrt{e}p_g}\Delta, \\ \frac{1}{\Delta}\sqrt{\ln\frac{(1-p_g)\Delta}{p_g\epsilon}}, & \epsilon \leqslant \frac{1-p_g}{\sqrt{e}p_g}\Delta. \end{cases} \quad (49)$$

This value of τ determines the required circuit depth. Substituting τ into $\tau = c^{-1}(\sqrt{2\pi p_g/x})$ and considering Eq. (48), we can work out N_T.

Theorem 6. Take the same setup as in Theorem 5. For all $\epsilon > 0$ and all $x < p_g$, $\epsilon_g(\mathcal{V}_x) \leqslant \epsilon$ holds when

$$N_T = \left\lceil \frac{4h_{\text{tot}}^2\tau^2}{1 - x/p_g} \right\rceil = O\left(\frac{1}{\epsilon^2}\right). \quad (50)$$

Here, τ is given by Eq. (49).

Because the gate number is proportional to N_T, the gate number scales as $O(1/\epsilon^2)$.

V. CONCLUSIONS AND OUTLOOK

In this paper, we have presented a method of realizing a variational wave function with randomized quantum circuits in VQAs. This method can systematically increase the expressive power without changing the gate number. The cost is an enlarged time for evaluating a quantity in the Monte Carlo calculation. The trade-off between the expressive power and time cost is analyzed theoretically and illustrated numerically. Especially in VQE, we have shown that the energy error that is due to the finite expressive power decreases with the time cost and eventually vanishes. These results demonstrate the viability and potential advantages of the random-circuit approach in VQAs, providing a pathway for improving the performance of quantum computing with the constraints of gate numbers and associated errors.

Further research directions include applying the random-circuit approach to other VQAs, not just VQE. In addition to sampling from different gate parameters, it is interesting to construct the sample space Ω with varying circuit structure [85]. Additionally, we foresee the possibility of refining our algorithm by incorporating more advanced neural networks and optimization methods.

The source codes for the numerical simulation are available at Ref. [86].

APPENDIX A: HADAMARD TEST

In this section, we review two methods for measuring $X_{\theta,\theta'}$ on a quantum computer: the Hadamard test with and without ancilla qubit. In the methods, we analyze the variance of $\hat{X}_{\theta,\theta'}$ in the ancilla-qubit Hadamard test in detail.
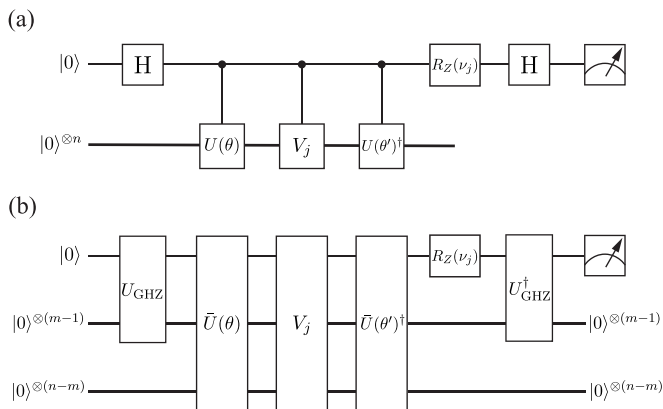
(a)

(b)

FIG. 10. (a) Hadamard test with an ancilla qubit. The rotation gate $R_Z(\nu_j) = \exp(-iZ\nu_j/2)$. (b) Hadamard test without ancilla qubit. $U_{\text{GHZ}} = \prod_{i=1}^{m-1} \Lambda_{i,i+1} H_1$, where $\Lambda_{i,j}$ is a controlled-NOT gate with the ith qubit as the control qubit and the jth qubit as the target qubit, and H_1 is a Hadamard gate on the first qubit. The gate U_{GHZ} prepares a Greenberger-Horne-Zeilinger (GHZ) state $|\text{GHZ}\rangle = \frac{1}{\sqrt{2}}(|0\rangle^{\otimes m} + |1\rangle^{\otimes m})$.

1. Ancilla-qubit Hadamard test

The conventional Hadamard test circuit requires an ancilla qubit [81], as shown in Fig. 10(a). In the final state of the circuit, the expected value of the ancilla-qubit Z Pauli operator is

$$\langle Z_A \rangle = \text{Re}[\exp(i\nu_j)\langle\phi(\theta')|V_j|\phi(\theta)\rangle]. \quad (A1)$$

Therefore, at the end of the circuit, we measure the ancilla qubit in the computational (i.e., Z) basis. For each circuit shot, we obtain a measurement outcome $\mu = \pm 1$, and $\langle Z_A \rangle = \text{E}[\mu]$.

The measurement protocol depends on the property of operator O. The simplest case is that O is unitary and Hermitian. In this case, we have

$$X_{\theta,\theta'} = \text{Re}[e^{i\nu}\langle\phi(\theta')|O|\phi(\theta)\rangle], \quad (A2)$$

where $\nu = \gamma(\lambda, \theta) - \gamma(\lambda, \theta')$. This quantity can be measured by taking $V_j = O$ and $\nu_j = \nu$. If the number of circuit shots is M_Q, the variance of $\hat{X}_{\theta,\theta'}$ has the upper bound $\sigma_O^2 = 1/M_Q$.

When O is Hermitian, e.g., the Hamiltonian H in VQE, we can evaluate $X_{\theta,\theta'}$ with the Hadamard test incorporating Monte Carlo. First, we express the operator O as a linear combination of unitary operators, $O = \sum_j a_j V_j$. Here, a_j are complex coefficients, and V_j are unitary operators. Then we can express $X_{\theta,\theta'}$ as

$$X_{\theta,\theta'} = C_O \sum_j P_j \text{Re}[\exp(i\nu_j)\langle\phi(\theta')|V_j|\phi(\theta)\rangle], \quad (A3)$$

where $P_j = |a_j|/C_O$ is the probability, $C_O = \sum_j |a_j|$ is the normalization factor, and $\nu_j = \gamma(\lambda, \theta) - \gamma(\lambda, \theta') + \arg a_j$. Finally, in each circuit shot, we randomly choose V_j and ν_j according to the probability P_j. With the measurement outcomes, we evaluate $X_{\theta,\theta'}$ according to $X_{\theta,\theta'} = C_O \text{E}[\mu]$. Therefore, the variance of $\hat{X}_{\theta,\theta'}$ has the upper bound $\sigma_O^2 = C_O^2/M_Q$.

A similar approach applies to a general operator O. In general, $X_{\theta,\theta'}$ has real and imaginary parts, i.e.,

$$X_{\theta,\theta'} = C_O \sum_j P_j \bigg(\text{Re}[\exp(i\nu_j)\langle\phi(\theta')|V_j|\phi(\theta)\rangle]$$
$$+ i\,\text{Re}\left\{\exp\left[i\left(\nu_j - \frac{\pi}{2}\right)\right]\langle\phi(\theta')|V_j|\phi(\theta)\rangle\right\}\bigg). \quad (A4)$$

We can measure the two parts separately. If we measure each part with M_Q circuit shots, the variance of $\hat{X}_{\theta,\theta'}$ has the upper bound $\sigma_O^2 = 2C_O^2/M_Q$.

2. Ancilla-free Hadamard test

For certain models, the Hadamard test can be implemented without the ancilla qubit, e.g., fermion models with particle number conservation [62,88,89]. The ancilla-free circuit with particle number conservation is illustrated in Fig. 10(b). Suppose that the fermion system in states $|\phi(\theta)\rangle$ contains m particle. If we take the Jordan-Wigner transformation for encoding fermions into qubits, qubit states $|0\rangle$ and $|1\rangle$ denote unoccupied and occupied fermion modes, respectively. Then the number of ones is the same as the number of particles. Let us introduce an m-particle state as the initial state, $|m\rangle = |1\rangle^{\otimes m} \otimes |0\rangle^{\otimes(n-m)}$. Given this m-particle initial state, we can prepare $|\phi(\theta)\rangle$ via a particle-number-preserving transformation on $|m\rangle$. Let $\bar{U}(\theta) = U(\theta)\prod_{i=1}^{m} X_i$, where X_i is the X Pauli operator on the ith qubit. It can be verified that $|\phi(\theta)\rangle = \bar{U}(\theta)|m\rangle$. The ancilla-free Hadamard test works under the condition that all $\bar{U}(\theta)$ and V_j are particle-number-preserving operators. In the final state, we have

$$\langle Z \otimes |0\rangle\langle 0|^{\otimes(n-1)}\rangle = \text{Re}[\exp(i\nu_j)\langle\phi(\theta')|V_j|\phi(\theta)\rangle]. \quad (A5)$$

Accordingly, we measure all qubits in the computational basis. The measurement outcome takes three values $\mu = 0, \pm 1$: The outcome is the eigenvalue of the first-qubit Z Pauli operator subjected to the condition that all other qubits are in the state $|0\rangle$; otherwise, the outcome is zero.

APPENDIX B: PROOF OF THEOREM 1

Proof. Suppose the sample size is M for both $\langle\hat{O}\rangle$ and $\langle\hat{\mathbb{1}}\rangle$. Let $\varepsilon_O = \sqrt{\frac{1}{M\kappa}(\|O\|^2 + \sigma_O^2)}$ and $\varepsilon_{\mathbb{1}} = \sqrt{\frac{1}{M\kappa}(1 + \sigma_{\mathbb{1}}^2)}$. Then using Eq. (16), we have $\text{Var}(\langle\hat{O}\rangle) \leqslant \kappa\varepsilon_O^2$ and $\text{Var}(\langle\hat{\mathbb{1}}\rangle) \leqslant \kappa\varepsilon_{\mathbb{1}}^2$. According to Chebyshev's inequality,

$$\Pr(|\delta_O| \leqslant \varepsilon_O) \geqslant 1 - \kappa, \quad (B1)$$

$$\Pr(|\delta_{\mathbb{1}}| \leqslant \varepsilon_{\mathbb{1}}) \geqslant 1 - \kappa. \quad (B2)$$

Therefore,

$$\Pr(|\delta_O| \leqslant \varepsilon_O \text{ and } |\delta_{\mathbb{1}}| \leqslant \varepsilon_{\mathbb{1}}) \geqslant 1 - 2\kappa. \quad (B3)$$

When M satisfies the condition in Eq. (18), the following two inequalities hold:

$$\varepsilon_{\mathbb{1}} \leqslant \langle\mathbb{1}\rangle, \quad (B4)$$

$$\frac{\langle\mathbb{1}\rangle\varepsilon_O + |\langle O\rangle|\varepsilon_{\mathbb{1}}}{\langle\mathbb{1}\rangle(\langle\mathbb{1}\rangle - \varepsilon_{\mathbb{1}})} \leqslant \varepsilon. \quad (B5)$$

Then when $|\delta_O| \leqslant \varepsilon_O$ and $|\delta_{\mathbb{1}}| \leqslant \varepsilon_{\mathbb{1}}$,

$$
\left| \frac{\langle \hat{O} \rangle}{\langle \hat{\mathbb{1}} \rangle} - \widetilde{\langle O \rangle} \right| = \left| \frac{\langle O \rangle + \delta_O}{\langle \mathbb{1} \rangle + \delta_{\mathbb{1}}} - \frac{\langle O \rangle}{\langle \mathbb{1} \rangle} \right|
$$

$$
\leqslant \frac{\langle \mathbb{1} \rangle |\delta_O| + |\langle O \rangle||\delta_{\mathbb{1}}|}{\langle \mathbb{1} \rangle (\langle \mathbb{1} \rangle - |\delta_{\mathbb{1}}|)} \leqslant \varepsilon. \tag{B6}
$$

∎

APPENDIX C: PROOF OF THEOREM 2

Proof. According to Chebyshev's inequality,

$$
\Pr[|\langle \hat{\mathbb{1}} \rangle - \langle \mathbb{1} \rangle| \geqslant \kappa^{-1/2} \sqrt{\mathrm{Var}(\langle \hat{\mathbb{1}} \rangle)}] \leqslant \kappa. \tag{C1}
$$

Because $\kappa^{-1/2} \sqrt{\mathrm{Var}(\langle \hat{\mathbb{1}} \rangle)} \leqslant \eta_{\mathbb{1}}$,

$$
\Pr(|\langle \hat{\mathbb{1}} \rangle - \langle \mathbb{1} \rangle| \geqslant \eta_{\mathbb{1}}) \leqslant \kappa. \tag{C2}
$$

Similarly,

$$
\Pr(|\langle \hat{H} \rangle - \langle H \rangle| \geqslant \eta_H) \leqslant \kappa. \tag{C3}
$$

Therefore,

$$
\Pr(|\langle \hat{\mathbb{1}} \rangle - \langle \mathbb{1} \rangle| \leqslant \eta_{\mathbb{1}} \text{ and } |\langle \hat{H} \rangle - \langle H \rangle| \leqslant \eta_H) \geqslant 1 - 2\kappa. \tag{C4}
$$

In what follows, we focus on the case that $|\langle \hat{\mathbb{1}} \rangle - \langle \mathbb{1} \rangle| \leqslant \eta_{\mathbb{1}}$ and $|\langle \hat{H} \rangle - \langle H \rangle| \leqslant \eta_H$ are true.

Because $\langle \mathbb{1} \rangle > 0$, $\langle \mathbb{1} \rangle + 2\eta_{\mathbb{1}} \geqslant \langle \hat{\mathbb{1}} \rangle + \eta_{\mathbb{1}} > 0$. Therefore,

$$
\frac{\langle H \rangle}{\langle \hat{\mathbb{1}} \rangle + \eta_{\mathbb{1}}} \leqslant \hat{L}'' \leqslant \frac{\langle H \rangle + 2\eta_H}{\langle \hat{\mathbb{1}} \rangle + \eta_{\mathbb{1}}}. \tag{C5}
$$

When $\langle H \rangle \leqslant 0$,

$$
\frac{\langle H \rangle}{\langle \hat{\mathbb{1}} \rangle + \eta_{\mathbb{1}}} \geqslant \widetilde{\langle H \rangle}. \tag{C6}
$$

When $\langle H \rangle > 0$,

$$
\frac{\langle H \rangle}{\langle \hat{\mathbb{1}} \rangle + \eta_{\mathbb{1}}} > 0. \tag{C7}
$$

Therefore,

$$
\hat{L}'' \geqslant \min\{\widetilde{\langle H \rangle}, 0\}. \tag{C8}
$$

When $\langle H \rangle \leqslant -2\eta_H$,

$$
\frac{\langle H \rangle + 2\eta_H}{\langle \hat{\mathbb{1}} \rangle + \eta_{\mathbb{1}}} \leqslant \frac{\langle H \rangle + 2\eta_H}{\langle \mathbb{1} \rangle + 2\eta_{\mathbb{1}}}
$$

$$
\leqslant \widetilde{\langle H \rangle} + 2 \frac{\eta_H + \|H\|\eta_{\mathbb{1}}}{\langle \mathbb{1} \rangle}. \tag{C9}
$$

Here, we have used that $|\widetilde{\langle H \rangle}| \leqslant \|H\|$. When $\langle H \rangle > -2\eta_H$,

$$
\frac{\langle H \rangle + 2\eta_H}{\langle \hat{\mathbb{1}} \rangle + \eta_{\mathbb{1}}} \leqslant \widetilde{\langle H \rangle} + \frac{2\eta_H}{\langle \mathbb{1} \rangle}. \tag{C10}
$$

Therefore,

$$
\hat{L}'' \leqslant \widetilde{\langle H \rangle} + 2 \frac{\eta_H + \|H\|\eta_{\mathbb{1}}}{\langle \mathbb{1} \rangle}. \tag{C11}
$$

∎

APPENDIX D: COVERING NUMBER

The covering number has been introduced as an efficient measure for the expressivity of VQAs [40]. Let $\mathcal{A}$ be a set of variational wave functions. Following the reference, we define the hypothesis space:

$$
\mathcal{H}(\mathcal{A}) = \{\langle \widetilde{\psi}|O|\widetilde{\psi} \rangle || \widetilde{\psi} \rangle \in \mathcal{A}\}. \tag{D1}
$$

We use $\mathcal{N}(\mathcal{H}, \epsilon, \|\cdot\|)$ to denote the covering number of $\mathcal{H}$ with the scale ϵ and norm $\|\cdot\|$. Then the larger $\mathcal{N}(\mathcal{H}, \epsilon, \|\cdot\|)$ is, the more expressive $\mathcal{H}$ is.

For two sets of variational wave functions $\mathcal{A}_1$ and $\mathcal{A}_2$, assuming $\mathcal{A}_1 \subseteq \mathcal{A}_2$, it is obvious that $\mathcal{H}(\mathcal{A}_1) \subseteq \mathcal{H}(\mathcal{A}_2)$. Therefore, $\mathcal{N}(\mathcal{H}(\mathcal{A}_1), \epsilon, \|\cdot\|) \leqslant \mathcal{N}(\mathcal{H}(\mathcal{A}_2), \epsilon, \|\cdot\|)$.

APPENDIX E: PROOF OF THEOREM 3

Proof. For all $|\phi(\theta')\rangle, |\phi(\theta)\rangle \in \Omega$,

$$
\||\phi(\theta')\rangle - |\phi(\theta)\rangle\|_2 \leqslant \xi \|\theta' - \theta\|_2. \tag{E1}
$$

Then for the random-circuit variational wave function in Eq. (36):

$$
\||\psi(\lambda)\rangle - |\phi(\theta)\rangle\|_2
$$

$$
\leqslant \left(\frac{\tau}{2}\right)^K \int d\theta' \exp\left(-\tau \|\theta' - \theta\|_1\right) \xi \|\theta' - \theta\|_2
$$

$$
\leqslant \left(\frac{\tau}{2}\right)^K \int d\theta' \exp\left(-\tau \|\theta' - \theta\|_1\right) \xi \|\theta' - \theta\|_1
$$

$$
= \frac{K\xi}{\tau}. \tag{E2}
$$

Because $\||\phi(\theta)\rangle\|_2 = 1$,

$$
\langle \mathbb{1} \rangle(\lambda) \geqslant \left(1 - \frac{K\xi}{\tau}\right)^2, \tag{E3}
$$

i.e., $|\widetilde{\psi}(\lambda)\rangle \in \mathcal{V}_x$, where $x = (1 - \frac{K\xi}{\tau})^2$.

Taking the limit $\tau \to +\infty$, we have $|\widetilde{\psi}(\lambda)\rangle \in \mathcal{V}_x$ for all $x < 1$, and

$$
\lim_{\tau \to +\infty} \||\psi(\lambda)\rangle - |\phi(\theta)\rangle\|_2
$$

$$
= \lim_{\tau \to +\infty} \||\widetilde{\psi}(\lambda)\rangle - |\phi(\theta)\rangle\|_2 = 0. \tag{E4}
$$

∎

APPENDIX F: PROOF OF THEOREM 4

Proof. Let $\{|\phi(\theta_1)\rangle, |\phi(\theta_2)\rangle, \ldots, |\phi(\theta_d)\rangle\} \subset \Omega$ be a basis of $\mathcal{H}$. Here, d is the dimension of $\mathcal{H}$. Then there exist coefficients $\beta_1, \beta_2, \ldots, \beta_d \in \mathbb{C}$ such that

$$
|\varphi\rangle = \sum_{i=1}^d \beta_i |\phi(\theta_i)\rangle. \tag{F1}
$$

Accordingly, we can express $|\varphi\rangle$ as an integral, as given in Eq. (37), in which

$$
\beta(\theta) = \sum_{i=1}^d \beta_i \delta(\theta - \theta_i). \tag{F2}
$$

Now inspired by the wave function in Eq. (36), we approximate $\beta(\theta)$ with a continuous function:

$$\beta_\tau(\theta) = \left(\frac{\tau}{2}\right)^K \sum_{i=1}^d \beta_i \exp\left(-\tau\|\theta - \theta_i\|_1\right). \quad \text{(F3)}$$

The corresponding state is

$$|\varphi_\tau\rangle = \int d\theta \beta_\tau(\theta)|\phi(\theta)\rangle = \sum_{i=1}^d \beta_i|\phi_\tau(\theta_i)\rangle, \quad \text{(F4)}$$

where

$$|\phi_\tau(\theta_i)\rangle = \left(\frac{\tau}{2}\right)^K \int d\theta \exp\left(-\tau\|\theta - \theta_i\|_1\right)|\phi(\theta)\rangle. \quad \text{(F5)}$$

This approximation introduces an error:

$$\||\varphi_\tau\rangle - |\varphi\rangle\|_2 \leqslant \frac{K\xi}{\tau} \sum_{i=1}^d |\beta_i|. \quad \text{(F6)}$$

Here, we have used that

$$\||\phi_\tau(\theta_i)\rangle - |\phi(\theta_i)\rangle\|_2 \leqslant \frac{K\xi}{\tau}, \quad \text{(F7)}$$

which can be proved the same as in Appendix E.

To apply the universal approximation theorem [82], we map $\beta_\tau(\theta)$ to two real functions $\ln|\beta_\tau(\theta)|$ and $\arg\beta_\tau(\theta)$. When $\beta_\tau(\theta)$ is continuous, $\ln|\beta_\tau(\theta)|$ and $\arg\beta_\tau(\theta)$ are continuous functions. According to the universal approximation theorem, for all $\epsilon_{AB} > 0$, there exist $L \in \mathbb{N}$ and $\lambda \in \Lambda$ such that

$$\|A(\bullet) - \ln|\beta_\tau(\bullet)|\|_\infty \leqslant \epsilon_{AB}, \quad \text{(F8)}$$

and

$$\|B(\bullet) - \arg\beta_\tau(\bullet)\|_\infty \leqslant \epsilon_{AB}. \quad \text{(F9)}$$

Here, $\|f(\bullet) - g(\bullet)\|_\infty = \sup_x |f(x) - g(x)|$. With L and λ, we have

$$\|\alpha(\theta; \lambda) - \beta_\tau(\bullet)\|_\infty \leqslant \epsilon_\beta, \quad \text{(F10)}$$

where

$$\epsilon_\beta = [\exp(\epsilon_{AB}) - 1 + |\exp(i\epsilon_{AB}) - 1|] \sum_{i=1}^d |\beta_i|, \quad \text{(F11)}$$

This error in the function leads to an error in the state:

$$\||\psi(\lambda)\rangle - |\varphi_\tau\rangle\|_2 \leqslant \epsilon_\beta \int d\theta. \quad \text{(F12)}$$

Notice that $|\beta_i|$ and $\int d\theta$ ($\theta \in \Theta$, and Θ is compact) are finite. Therefore, there exists τ such that $\||\varphi_\tau\rangle - |\varphi\rangle\|_2 \leqslant \epsilon/2$, and there exist $L \in \mathbb{N}$ and $\lambda \in \Lambda$ such that $\||\psi(\lambda)\rangle - |\varphi_\tau\rangle\|_2 \leqslant \epsilon/2$. When $|\psi(\lambda)\rangle$ approaches $|\varphi\rangle$, it is also normalized. ∎

APPENDIX G: PRIOR GUIDING FUNCTION

We consider a prior guiding function related to the real-time evolution operator $\exp(-iHtN_T)$. Before giving the prior guiding function, we introduce some relevant notations: (i) $\mathcal{L}_k = \{\vec{l} = (l_1, l_2, \ldots, l_k) \,|\, l_q = 1, 2, \ldots, N_H\}$ is a set of k-dimensional vectors; (ii) for a vector $\vec{l} \in \mathcal{L}_k$,

$$m(\vec{l}, j) = \sum_{q=1}^k \delta_{l_q, j} \quad \text{(G1)}$$

is the number of j in components of $\vec{l}$; (iii)

$$S(\vec{l}) = \sigma_{l_k} \cdots \sigma_{l_2} \sigma_{l_1} \quad \text{(G2)}$$

and

$$T(\vec{l}) = \sigma_{N_H}^{m(\vec{l}, N_H)} \cdots \sigma_2^{m(\vec{l}, 2)} \sigma_1^{m(\vec{l}, 1)} \quad \text{(G3)}$$

are Pauli operators depending on the vector; and (iv)

$$a(\vec{l}) = \frac{\text{Tr}[S(\vec{l})T(\vec{l})^\dagger]}{\text{Tr}(\mathbb{1})}. \quad \text{(G4)}$$

Because two Pauli operators are either commutative or anticommutative, $S(\vec{l}) = a(\vec{l})T(\vec{l})$ and $a(\vec{l}) = \pm 1$. We remark that computing $m(\vec{l}, j)$ and $a(\vec{l})$ on a classical computer is efficient, and the resource cost increases polynomially with the vector dimension k.

The prior guiding function reads

$$F(\theta) = \prod_{i=1}^{N_T} f(t, \omega_i), \quad \text{(G5)}$$

where

$$f(t, \omega_i) = \prod_{j=1}^{N_H} \delta(\omega_{i,j} - h_j t) + \sum_{k=2}^\infty \sum_{\vec{l} \in \mathcal{L}_k} \frac{\prod_{q=1}^k (-ih_{l_q} t)}{k!} [a(\vec{l}) - 1] \prod_{j=1}^{N_H} i^{m(\vec{l}, j)} \delta\left[\omega_{i,j} - m(\vec{l}, j)\frac{\pi}{2}\right]. \quad \text{(G6)}$$

The property of the above prior guiding function is given by Eq. (41). We can prove it as follows. We only have to consider the integral over one ω_i:

$$\int d\omega_i f(t, \omega_i)R(\omega_i) = \exp(-i\sigma_{N_H} h_{N_H} t) \cdots \exp(-i\sigma_2 h_2 t)\exp(-i\sigma_1 h_1 t) + \sum_{k=2}^\infty \sum_{\vec{l} \in \mathcal{L}_k} \frac{\prod_{q=1}^k (-ih_{l_q} t)}{k!} [a(\vec{l}) - 1]T(\vec{l}). \quad \text{(G7)}$$

The second line in the above equation corresponds to the first term of $f(t, \omega_i)$ in Eq. (G6), i.e., taking $\omega_{i,j} = h_j t$ in the Hamiltonian *Ansatz*. Here, $T(\vec{l})$ corresponds to taking $\omega_{i,j} = m(\vec{l}, j)\frac{\pi}{2}$: Taking this value of $\omega_{i,j}$, we have $R(\omega_i) = \prod_{j=1}^{N_H} i^{-m(\vec{l}, j)} T(\vec{l})$,

and the phase is canceled by $\prod_{j=1}^{N_H} i^{m(\vec{l},j)}$ in $f(t, \omega_i)$. Using $S(\vec{l}) = a(\vec{l})T(\vec{l})$, we have

$$\int d\omega_i f(t, \omega_i) R(\omega_i) = \exp\left(-i\sigma_{N_H} h_{N_H} t\right) \cdots \exp\left(-i\sigma_2 h_2 t\right) \exp\left(-i\sigma_1 h_1 t\right) + \sum_{k=2}^{\infty} \sum_{\vec{l} \in \mathcal{L}_k} \frac{\prod_{q=1}^{k}(-ih_{l_q} t)}{k!}[S(\vec{l}) - T(\vec{l})]. \quad \text{(G8)}$$

Then applying the Taylor expansion, we have

$$\exp\left(-i\sigma_{N_H} h_{N_H} t\right) \cdots \exp\left(-i\sigma_2 h_2 t\right) \exp\left(-i\sigma_1 h_1 t\right)$$

$$= \mathbb{1} + \sum_{k=1}^{\infty} \sum_{\vec{l} \in \mathcal{L}_k} \frac{\prod_{q=1}^{k}(-ih_{l_q} t)}{k!} T(\vec{l}) \quad \text{(G9)}$$

and

$$e^{-iHt} = \mathbb{1} + \sum_{k=1}^{\infty} \sum_{\vec{l} \in \mathcal{L}_k} \frac{\prod_{q=1}^{k}(-ih_{l_q} t)}{k!} S(\vec{l}). \quad \text{(G10)}$$

Notice that $S(\vec{l}) = T(\vec{l})$ for all $\vec{l} \in \mathcal{L}_1$. Therefore,

$$\int d\omega_i f(t, \omega_i) R(\omega_i) = e^{-iHt}. \quad \text{(G11)}$$

APPENDIX H: PROOF OF LEMMA 1

Proof. We consider the spectral decomposition of the Hamiltonian $H = \sum_{l=1}^{2^n} E_l |\varphi_l\rangle\langle\varphi_l|$. Here, $E_1 \leqslant E_2 \leqslant \cdots \leqslant E_{2^n}$ are eigenenergies in ascending order, and $|\varphi_l\rangle$ are eigenstates. Also, $E_1 = E_g$ is the ground-state energy, $|\Phi_1\rangle = |\Psi_g\rangle$ is the ground state, and $\Delta = E_2 - E_1$ is the energy gap. Without loss of generality, we suppose $|\Psi_0\rangle = \sum_{l=1}^{2^n} \sqrt{p_l}|\varphi_l\rangle$, where p_l is the probability of the eigenstate $|\varphi_l\rangle$, and $p_1 = p_g$. Then

$$E(\lambda) = \frac{\sum_{l=1}^{2^n} \exp[-(E_l - E_g)^2 \tau^2] p_l E_l}{\sum_{l=1}^{2^n} \exp[-(E_l - E_g)^2 \tau^2] p_l}$$

$$= E_g + \frac{\sum_{l=2}^{2^n} \exp[-(E_l - E_g)^2 \tau^2] p_l (E_l - E_g)}{p_g + \sum_{l=2}^{2^n} \exp[-(E_l - E_g)^2 \tau^2] p_l}$$

$$\leqslant E_g + \frac{1 - p_g}{p_g} Q, \quad \text{(H1)}$$

where

$$Q = \max_{2 \leqslant l \leqslant 2^n} \exp[-(E_l - E_g)^2 \tau^2](E_l - E_g). \quad \text{(H2)}$$

We derive two upper bounds of Q. First, notice that the maximum value of $\exp(-x^2\tau^2)x$ is $\exp(-\frac{1}{2})(\sqrt{2}\tau)^{-1}$ at $x = (\sqrt{2}\tau)^{-1}$. Then

$$\exp[-(E_l - E_g)^2 \tau^2](E_l - E_g) \leqslant \exp\left(-\frac{1}{2}\right)(\sqrt{2}\tau)^{-1}, \quad \text{(H3)}$$

for all l. Therefore, $Q \leqslant \exp(-\frac{1}{2})(\sqrt{2}\tau)^{-1}$. Second, we suppose that $\Delta \geqslant (\sqrt{2}\tau)^{-1}$. Under this condition,

$$Q = \exp[-(E_2 - E_g)^2 \tau^2](E_2 - E_g)$$

$$= \exp(-\Delta^2 \tau^2)\Delta. \quad \text{(H4)}$$

∎

APPENDIX I: PROOF OF LEMMA 2

Proof. The normalization factor of the state in Eq. (42) is

$$\langle\mathbb{1}\rangle(\lambda) = \left[\frac{\tau\sqrt{2\pi}}{C(\lambda)N_T}\right]^2 \langle\Psi_0|\exp[-(H - E_g)^2 \tau^2]|\Psi_0\rangle$$

$$\geqslant \left[\frac{\tau\sqrt{2\pi}}{C(\lambda)N_T}\right]^2 p_g. \quad \text{(I1)}$$

Here,

$$C(\lambda) = \int d\theta \exp\left(-\frac{N_T^2 t^2}{2\tau^2}\right)|F(\theta)|$$

$$= \int dt \exp\left[-\frac{N_T^2 t^2}{2\tau^2}\right]\left[\int d\omega_i |f(t, \omega_i)|\right]^{N_T}. \quad \text{(I2)}$$

The contribution of each Trotter step is

$$\int d\omega_i |f(t, \omega_i)| \leqslant 1 + 2\sum_{k=2}^{\infty} \sum_{\vec{l} \in \mathcal{L}_k} \frac{\prod_{q=1}^{k}(|h_{l_q} t|)}{k!}$$

$$= 1 + 2[\exp(h_{\text{tot}}|t|) - 1 - h_{\text{tot}}|t|]$$

$$= 2\exp(h_{\text{tot}}|t|) - 1 - 2h_{\text{tot}}|t|$$

$$\leqslant \exp\left(2h_{\text{tot}}^2 t^2\right). \quad \text{(I3)}$$

Therefore,

$$C(\lambda) \leqslant \int dt \exp\left(-\frac{N_T^2 t^2}{2\tau^2}\right)[2\exp(h_{\text{tot}}|t|) - 1 - 2h_{\text{tot}}|t|]^{N_T}$$

$$= \frac{\tau}{N_T} c(\tau). \quad \text{(I4)}$$

When $N_T > 4h_{\text{tot}}^2 \tau^2$,

$$\frac{\tau}{N_T} c(\tau) \leqslant \int dt \exp\left(-\frac{N_T^2 t^2}{2\tau^2}\right)\exp\left(2N_T h_{\text{tot}}^2 t^2\right)$$

$$= \frac{\tau\sqrt{2\pi}}{N_T} \frac{1}{\sqrt{1 - \frac{4h_{\text{tot}}^2 \tau^2}{N_T}}}. \quad \text{(I5)}$$

Substituting the upper bound of $C(\lambda)$, we prove the lower bounds of $\langle\mathbb{1}\rangle(\lambda)$. ∎

[1] J. W. Z. Lau, K. H. Lim, H. Shrotriya, and L. C. Kwek, NISQ computing: Where are we and where do we go? AAPPS Bull. **32**, 27 (2022).

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

[3] 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).

[4] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, Nature (London) **549**, 242 (2017).

[5] J. R. McClean, M. E. Kimchi-Schwartz, J. Carter, and W. A. de Jong, Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states, Phys. Rev. A **95**, 042308 (2017).

[6] O. Higgott, D. Wang, and S. Brierley, Variational quantum computation of excited states, Quantum **3**, 156 (2019).

[7] K. M. Nakanishi, K. Mitarai, and K. Fujii, Subspace-search variational quantum eigensolver for excited states, Phys. Rev. Res. **1**, 033062 (2019).

[8] R. M. Parrish, E. G. Hohenstein, P. L. McMahon, and T. J. Martínez, Quantum computation of electronic transitions using a variational quantum eigensolver, Phys. Rev. Lett. **122**, 230401 (2019).

[9] W. J. Huggins, J. Lee, U. Baek, B. O'Gorman, and K. B. Whaley, A non-orthogonal variational quantum eigensolver, New J. Phys. **22**, 073009 (2020).

[10] J. Tilly, H. Chen, S. Cao, D. Picozzi, K. Setia, Y. Li, E. Grant, L. Wossnig, I. Rungger, G. H. Booth *et al.*, The variational quantum eigensolver: A review of methods and best practices, Phys. Rep. **986**, 1 (2022).

[11] Y. Li and S. C. Benjamin, Efficient variational quantum simulator incorporating active error minimization, Phys. Rev. X **7**, 021050 (2017).

[12] 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).

[13] X. Yuan, S. Endo, Q. Zhao, Y. Li, and S. C. Benjamin, Theory of variational quantum simulation, Quantum **3**, 191 (2019).

[14] S. Endo, J. Sun, Y. Li, S. C. Benjamin, and X. Yuan, Variational quantum simulation of general processes, Phys. Rev. Lett. **125**, 010501 (2020).

[15] P. J. Ollitrault, S. Jandura, A. Miessen, I. Burghardt, R. Martinazzo, F. Tacchino, and I. Tavernelli, Quantum algorithms for grid-based variational time evolution, Quantum **7**, 1139 (2023).

[16] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm, arXiv:1411.4028.

[17] Z. Wang, S. Hadfield, Z. Jiang, and E. G. Rieffel, Quantum approximate optimization algorithm for MaxCut: A fermionic view, Phys. Rev. A **97**, 022304 (2018).

[18] L. Zhou, S.-T. Wang, S. Choi, H. Pichler, and M. D. Lukin, Quantum approximate optimization algorithm: Performance, mechanism, and implementation on near-term devices, Phys. Rev. X **10**, 021067 (2020).

[19] L. Li, M. Fan, M. Coram, P. Riley, and S. Leichenauer, Quantum optimization with a novel Gibbs objective function and ansatz architecture search, Phys. Rev. Res. **2**, 023074 (2020).

[20] M. M. Wauters, E. Panizon, G. B. Mbeng, and G. E. Santoro, Reinforcement-learning-assisted quantum optimization, Phys. Rev. Res. **2**, 033446 (2020).

[21] S. H. Sack and M. Serbyn, Quantum annealing initialization of the quantum approximate optimization algorithm, Quantum **5**, 491 (2021).

[22] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, Nature (London) **549**, 195 (2017).

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

[24] I. Cong, S. Choi, and M. D. Lukin, Quantum convolutional neural networks, Nat. Phys. **15**, 1273 (2019).

[25] K. Beer, D. Bondarenko, T. Farrelly, T. J. Osborne, R. Salzmann, D. Scheiermann, and R. Wolf, Training deep quantum neural networks, Nat. Commun. **11**, 808 (2020).

[26] A. Abbas, D. Sutter, C. Zoufal, A. Lucchi, A. Figalli, and S. Woerner, The power of quantum neural networks, Nat. Comput. Sci. **1**, 403 (2021).

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

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

[29] M. Schuld and N. Killoran, Quantum machine learning in feature Hilbert spaces, Phys. Rev. Lett. **122**, 040504 (2019).

[30] A. Pérez-Salinas, A. Cervera-Lierta, E. Gil-Fuster, and J. I. Latorre, Data re-uploading for a universal quantum classifier, Quantum **4**, 226 (2020).

[31] M. Schuld, A. Bocharov, K. M. Svore, and N. Wiebe, Circuit-centric quantum classifiers, Phys. Rev. A **101**, 032308 (2020).

[32] X. Xu, J. Sun, S. Endo, Y. Li, S. C. Benjamin, and X. Yuan, Variational algorithms for linear algebra, Sci. Bull. **66**, 2181 (2021).

[33] P. J. P. O'Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding *et al.*, Scalable quantum simulation of molecular energies, Phys. Rev. X **6**, 031007 (2016).

[34] C. Cîrstoiu, Z. Holmes, J. Iosue, L. Cincio, P. J. Coles, and A. Sornborger, Variational fast forwarding for quantum simulation beyond the coherence time, npj Quantum Inf. **6**, 82 (2020).

[35] Y. Kim, A. Eddins, S. Anand, K. X. Wei, E. Van Den Berg, S. Rosenblatt, H. Nayfeh, Y. Wu, M. Zaletel, K. Temme *et al.*, Evidence for the utility of quantum computing before fault tolerance, Nature (London) **618**, 500 (2023).

[36] S. Sim, P. D. Johnson, and A. Aspuru-Guzik, Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms, Adv. Quantum Technol. **2**, 1900070 (2019).

[37] J. Tangpanitanon, S. Thanasilp, N. Dangniam, M.-A. Lemonde, and D. G. Angelakis, Expressibility and trainability of parametrized analog quantum systems for machine learning applications, Phys. Rev. Res. **2**, 043364 (2020).

[38] Y. Du, M.-H. Hsieh, T. Liu, and D. Tao, Expressive power of parametrized quantum circuits, Phys. Rev. Res. **2**, 033125 (2020).

[39] K. Nakaji and N. Yamamoto, Expressibility of the alternating layered ansatz for quantum computation, Quantum **5**, 434 (2021).

[40] Y. Du, Z. Tu, X. Yuan, and D. Tao, Efficient measure for the expressivity of variational quantum algorithms, Phys. Rev. Lett. **128**, 080506 (2022).

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

[42] C. Ortiz Marrero, M. Kieferová, and N. Wiebe, Entanglement-induced barren plateaus, PRX Quantum **2**, 040316 (2021).

[43] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Cost function dependent barren plateaus in shallow parametrized quantum circuits, Nat. Commun. **12**, 1791 (2021).

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

[45] A. Pesah, M. Cerezo, S. Wang, T. Volkoff, A. T. Sornborger, and P. J. Coles, Absence of barren plateaus in quantum convolutional neural networks, Phys. Rev. X **11**, 041011 (2021).

[46] T. Volkoff and P. J. Coles, Large gradients via correlation in random parameterized quantum circuits, Quantum Sci. Technol. **6**, 025008 (2021).

[47] M. Ostaszewski, E. Grant, and M. Benedetti, Structure optimization for parameterized quantum circuits, Quantum **5**, 391 (2021).

[48] A. Choquette, A. Di Paolo, P. K. Barkoutsos, D. Sénéchal, I. Tavernelli, and A. Blais, Quantum-optimal-control-inspired ansatz for variational quantum algorithms, Phys. Rev. Res. **3**, 023092 (2021).

[49] T. L. Patti, K. Najafi, X. Gao, and S. F. Yelin, Entanglement devised barren plateau mitigation, Phys. Rev. Res. **3**, 033090 (2021).

[50] K. Bharti and T. Haug, Quantum-assisted simulator, Phys. Rev. A **104**, 042418 (2021).

[51] H. L. Tang, V. O. Shkolnikov, G. S. Barron, H. R. Grimsley, N. J. Mayhall, E. Barnes, and S. E. Economou, Qubit-ADAPT-VQE: An adaptive algorithm for constructing hardware-efficient *Ansätze* on a quantum processor, PRX Quantum **2**, 020310 (2021).

[52] R. Haghshenas, J. Gray, A. C. Potter, and G. K.-L. Chan, Variational power of quantum circuit tensor networks, Phys. Rev. X **12**, 011047 (2022).

[53] A. W. Harrow and S. Mehraban, Approximate unitary *t*-designs by short random quantum circuits using nearest-neighbor and long-range gates, Commun. Math. Phys. **401**, 1531 (2023).

[54] A. K. Jain, J. Mao, and K. M. Mohiuddin, Artificial neural networks: A tutorial, Computer **29**, 31 (1996).

[55] Y. Bengio, Learning deep architectures for AI, Found. Trends Mach. Learn. **2**, 1 (2009).

[56] X. Glorot, A. Bordes, and Y. Bengio, Deep sparse rectifier neural networks, in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics* (JMLR Workshop and Conference Proceedings, Fort Lauderdale, FL, 2011), pp. 315–323.

[57] A. L. Maas, A. Y. Hannun, and A. Y. Ng, Rectifier nonlinearities improve neural network acoustic models, in *Proceedings of the 30th International Conference on Machine Learning* (Atlanta, GA, 2013), Vol. 28, p. 3.

[58] D. Wecker, M. B. Hastings, and M. Troyer, Progress towards practical quantum variational algorithms, Phys. Rev. A **92**, 042303 (2015).

[59] W. Foulkes, L. Mitas, R. Needs, and G. Rajagopal, Quantum Monte Carlo simulations of solids, Rev. Mod. Phys. **73**, 33 (2001).

[60] B. M. Austin, D. Y. Zubarev, and W. A. Lester, Jr., Quantum Monte Carlo and related approaches, Chem. Rev. **112**, 263 (2012).

[61] W. J. Huggins, B. A. O'Gorman, N. C. Rubin, D. R. Reichman, R. Babbush, and J. Lee, Unbiasing fermionic quantum Monte Carlo with a quantum computer, Nature (London) **603**, 416 (2022).

[62] X. Xu and Y. Li, Quantum-assisted Monte Carlo algorithms for fermions, Quantum **7**, 1072 (2023).

[63] G. Mazzola and G. Carleo, Exponential challenges in unbiasing quantum Monte Carlo algorithms with quantum computers, arXiv:2205.09203.

[64] A. Montanaro and S. Stanisic, Accelerating variational quantum Monte Carlo using the variational quantum eigensolver, arXiv:2307.07719.

[65] M. S. Moss, S. Ebadi, T. T. Wang, G. Semeghini, A. Bohrdt, M. D. Lukin, and R. G. Melko, Enhancing variational Monte Carlo using a programmable quantum simulator, Phys. Rev. A **109**, 032410 (2024).

[66] Y. Zhang, Y. Huang, J. Sun, D. Lv, and X. Yuan, Quantum computing quantum Monte Carlo, arXiv:2206.10431.

[67] S. Kanno, H. Nakamura, T. Kobayashi, S. Gocho, M. Hatanaka, N. Yamamoto, and Q. Gao, Quantum computing quantum Monte Carlo with hybrid tensor network toward electronic structure calculations of large-scale molecular and solid systems, arXiv:2303.18095.

[68] K. C. Tan, D. Bhowmick, and P. Sengupta, Sign-problem free quantum stochastic series expansion algorithm on a quantum computer, npj Quantum Inf. **8**, 44 (2022).

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

[70] G. Cybenko, Approximation by superpositions of a sigmoidal function, Math. Control Signal Systems **2**, 303 (1989).

[71] K. Hornik, M. Stinchcombe, and H. White, Multilayer feedforward networks are universal approximators, Neural Networks **2**, 359 (1989).

[72] K. Hornik, Approximation capabilities of multilayer feedforward networks, Neural Networks **4**, 251 (1991).

[73] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, Science **355**, 602 (2017).

[74] D. Luo and B. K. Clark, Backflow transformations via neural networks for quantum many-body wave functions, Phys. Rev. Lett. **122**, 226401 (2019).

[75] K. T. Schütt, M. Gastegger, A. Tkatchenko, K.-R. Müller, and R. J. Maurer, Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions, Nat. Commun. **10**, 5024 (2019).

[76] J. Hermann, Z. Schätzle, and F. Noé, Deep-neural-network solution of the electronic Schrödinger equation, Nat. Chem. **12**, 891 (2020).

[77] D. Pfau, J. S. Spencer, A. G. D. G. Matthews, and W. M. C. Foulkes, *Ab initio* solution of the many-electron Schrödinger equation with deep neural networks, Phys. Rev. Res. **2**, 033429 (2020).

[78] W. Ren, W. Fu, X. Wu, and J. Chen, Towards the ground state of molecules via diffusion Monte Carlo on neural networks, Nat. Commun. **14**, 1860 (2023).

[79] W. K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, Biometrika **57**, 97 (1970).

[80] J. Brownlee, A gentle introduction to the rectified linear unit (ReLU), Machine Learning Mastery (2020), https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/.

[81] A. K. Ekert, C. M. Alves, D. K. L. Oi, M. Horodecki, P. Horodecki, and L. C. Kwek, Direct estimations of linear and nonlinear functionals of a quantum state, Phys. Rev. Lett. **88**, 217901 (2002).

[82] B. C. Csáji, Approximation with artificial neural networks, Master's thesis, EötvösLoránd University, Hungary, 2001.

[83] M. Suzuki, Fractal decomposition of exponential operators with applications to many-body theories and monte carlo simulations, Phys. Lett. A **146**, 319 (1990).

[84] M. R. Wall and D. Neuhauser, Extraction, through filter-diagonalization, of general quantum eigenvalues or classical normal mode frequencies from a small number of residues or a short-time segment of a signal. I. Theory and application to a quantum-dynamics model, J. Chem. Phys. **102**, 8011 (1995).

[85] K. Nakaji, L. B. Kristensen, J. A. Campos-Gonzalez-Angulo, M. G. Vakili, H. Huang, M. Bagherimehrab, C. Gorgulla, F. Wong, A. McCaskey, J.-S. Kim *et al.*, The generative quantum eigensolver (GQE) and its application for ground state search, arXiv:2401.09253.

[86] https://github.com/yyl19/variational_quantum_circuit_monte_carlo.

[87] https://github.com/guochu/VQC.jl.

[88] T. E. O'Brien, S. Polla, N. C. Rubin, W. J. Huggins, S. McArdle, S. Boixo, J. R. McClean, and R. Babbush, Error mitigation via verified phase estimation, PRX Quantum **2**, 020317 (2021).

[89] S. Lu, M. C. Bañuls, and J. I. Cirac, Algorithms for quantum simulation at finite energies, PRX Quantum **2**, 020321 (2021).