Imaginary Hamiltonian variational *Ansatz* for combinatorial optimization problems

Xiaoyang Wang [1,2,3,4] Yahui Chai,[5] Xu Feng [1,2,3] Yibin Guo,[5] Karl Jansen,[5,6] and Cenk Tüysüz [5,7]

[1]*School of Physics, Peking University, Beijing 100871, China*
[2]*Collaborative Innovation Center of Quantum Matter, Beijing 100871, China*
[3]*Center for High Energy Physics, Peking University, Beijing 100871, China*
[4]*Interdisciplinary Theoretical and Mathematical Sciences Program, RIKEN, Wako 351-0198, Japan*
[5]*Deutsches Elektronen-Synchrotron DESY, Platanenallee 6, 15738 Zeuthen, Germany*
[6]*Computation-Based Science and Technology Research Center, Cyprus Institute, 20 Kavafi Street, 2121 Nicosia, Cyprus*
[7]*Institut für Physik, Humboldt-Universität zu Berlin, Newtonstraße 15, 12489 Berlin, Germany*

(Received 28 August 2024; accepted 3 March 2025; published 18 March 2025)

Obtaining exact solutions to combinatorial optimization problems using classical computing is computationally expensive. The current tenet in the field is that quantum computers can address these problems more efficiently. While promising algorithms require fault-tolerant quantum hardware, variational algorithms have emerged as viable candidates for near-term devices. The success of these algorithms hinges on multiple factors, with the design of the *Ansatz* being of the utmost importance. It is known that popular approaches such as the quantum approximate optimization algorithm (QAOA) and quantum annealing suffer from adiabatic bottlenecks, which lead to either larger circuit depth or evolution time. On the other hand, the evolution time of imaginary-time evolution is bounded by the inverse energy gap of the Hamiltonian, which is constant for most noncritical physical systems. In this work we propose an imaginary Hamiltonian variational *Ansatz* (*i*HVA) inspired by quantum imaginary-time evolution to solve the MaxCut problem. We introduce a tree arrangement of the parametrized quantum gates, enabling the exact solution of arbitrary tree graphs using the one-round *i*HVA. For randomly generated D-regular graphs, we numerically demonstrate that the *i*HVA solves the MaxCut problem with a small constant number of rounds and sublinear depth, outperforming the QAOA, which requires rounds increasing with the graph size. Furthermore, our *Ansatz* solves the MaxCut problem exactly for graphs with up to 24 nodes and $D \leqslant 5$, whereas only approximate solutions can be derived by the classical near-optimal Goemans-Williamson algorithm. We validate our simulated results with hardware demonstrations on a graph with 67 nodes.

DOI: 10.1103/PhysRevA.111.032612

I. INTRODUCTION

Many applications of quantum computers involve the preparation of the ground state of a Hamiltonian system in fields such as chemistry [1], drug design [2,3], particle physics [4,5], combinatorial optimization [6,7], and quantum machine learning [8]. The variational quantum eigensolver (VQE) [9,10] is an algorithm designed for ground-state preparation on quantum computers. It combines classical optimization techniques with expectation values evaluated on quantum computers. Although the VQE has been explored for use on noisy intermediate-scale quantum (NISQ) [11] devices due to its relatively shallow circuit depth compared to other quantum algorithms, its practical suitability and effectiveness on these devices remain open questions.

The success of the VQE highly relies on the efficient parametrization of the quantum circuits. The parametrized quantum circuit is a variational *Ansatz* determining what quantum states can be prepared. There have been many efforts to construct the variational *Ansatz* to guarantee that the ground state of a quantum system can be prepared with high accuracy [1,10,12,13]. Among them, the quantum approximate optimization algorithm (QAOA) *Ansatz* is designed to solve combinatorial optimization problems, inspired by the adiabatic evolution [14]. Its performance has been extensively studied both analytically and numerically [12,15–21]. For many-body quantum systems, a widely used *Ansatz* following the same spirit of the QAOA is the Hamiltonian variational *Ansatz* [13].

Many challenges exist for the QAOA *Ansatz*. It has been shown that the number of QAOA *Ansatz* rounds should grow linearly with the system size even in some classically solvable tasks to find the solution with high accuracy [22] and there exists a fundamental limitation if the rounds do not increase faster than a logarithmic function of the system size [17]. This requirement leads to other caveats related to the variational optimization of the QAOA *Ansatz*. For example, the *Ansatz* with many rounds is susceptible to noise in NISQ devices [23,24] and its energy landscape has many local minima [25]. More importantly, generic variational *Ansätze* with linearly increasing rounds suffer from the barren plateau (BP) phenomenon [26–28], as demonstrated in Ref. [29], so that the gradient of the QAOA *Ansatz* cannot be measured efficiently if its number of rounds grows linearly with the system size.

The linear behavior originates from the real-time adiabatic evolution that inspires the QAOA *Ansatz*. The adiabatic evolution should be slow enough to avoid diabatic excitation [30], leading to the requirement of many rounds of the QAOA

Ansatz [14]. There are many efforts to enhance the QAOA *Ansatz* using ideas of, e.g., shortcuts to adiabaticity [31]. However, the variational *Ansatz* with high-order counterdiabatic terms contains many unitary gates and is difficult to implement on NISQ devices [32–35].

The imaginary Hamiltonian variational *Ansatz* (*i*HVA), inspired by works of quantum imaginary-time evolution (QITE) [36–38], is distinguished from the QAOA *Ansatz*. Imaginary-time evolution has no problem of diabatic excitation and is widely used in state preparation algorithms such as tensor networks [39,40] and Monte Carlo methods [41]. The *i*HVA has been applied to the Gibbs state preparation in previous studies [42]. In this work we propose to tackle the ground-state problems using the *i*HVA. The *Ansatz* uses unitary gates constrained by system symmetries as building blocks, which can be realized on gate-based quantum devices. In this work we apply the *i*HVA to the combinatorial optimization MaxCut problem.

For the MaxCut problem, the arrangement of the parametrized quantum gates in the *i*HVA impacts the solution accuracy. We propose a tree arrangement of gates in the *i*HVA for arbitrary graphs, and the corresponding *Ansatz* is called the *i*HVA tree. We provide a theorem which states that arbitrary tree graphs can be maximally cut exactly using the *i*HVA tree with one round and sublinear depth, which cannot be achieved using the constant-round QAOA *Ansatz* [22]. For more complicated random D-regular graphs, we perform numerical simulations using noiseless quantum simulators. The results show that the *i*HVA tree can solve the MaxCut problem of 3-regular graphs exactly up to 14 graph nodes using constant rounds and sublinear depth, while the QAOA *Ansatz* requires rounds growing with the graph nodes. For D-regular graphs with the number of nodes up to 24 and $D \leqslant 5$, the two-round *i*HVA tree can exactly solve the MaxCut problem, whereas only an approximate solution can be derived by the classical polynomial-time Goemans-Williamson (GW) algorithm [15,43]. Furthermore, we validate our results on a real quantum device by running an instance of a graph with 67 nodes.

We show that the constant-round *i*HVA on D-regular graphs does not exhibit BPs. It is known that circuits of constant depth are free from BPs and can be trained efficiently for local Hamiltonians [29]. For D-regular graphs, the constant-round *i*HVA has linear or sublinear depth, where the previous results cannot be applied directly. By exploring the feature that the number of noncommuting gates acting on each qubit has no dependence on the system size, we prove that the variance of the constant-round *i*HVA does not decay exponentially with the graph nodes. Therefore, we prove that the constant-round *i*HVA of D-regular graphs is free from BPs.

The remainder of this paper is structured as follows. In Sec. II we present how to choose parametrized quantum gates in the *i*HVA by leveraging system symmetries and an introduction to the MaxCut problem. In Sec. III we explicitly construct the *i*HVA for the MaxCut problem following the tree arrangement. In Sec. IV numerical simulations are performed to compare the performance of the QAOA, the *i*HVA, and the GW algorithm. In Sec. V we demonstrate that the constant-round *i*HVA is free from BPs. In Sec. VI we summarize our results and propose some open questions to be explored in future works.

II. FRAMEWORK

In this section we review the construction of the *i*HVA proposed in Ref. [42]. Then we introduce the combinatorial optimization MaxCut problem and basic concepts of graphs that are used in the following sections.

A. Imaginary Hamiltonian variational *Ansatz*

The imaginary Hamiltonian variational *Ansatz* is inspired by the QITE algorithm [36–38]. The QITE algorithm performs imaginary-time evolution on quantum computers with no need for ancillary qubits. Consider a k-local Hamiltonian

$$H = \sum_{\mu} H_{\mu}, \tag{1}$$

where H_μ is a local interaction term acting on at most k qubits. The imaginary-time propagator of H can be decomposed by a Trotterized-type formula

$$\begin{aligned} e^{-\tau H}|\psi\rangle &= (e^{-\Delta\tau H})^L|\psi\rangle \\ &= \left(\prod_{\mu} e^{-\Delta\tau H_\mu}\right)^L |\psi\rangle + O\left(\frac{\tau^2}{L}\right), \end{aligned} \tag{2}$$

where $\Delta\tau = \tau/L$ is the Trotter step. The resulting state approaches the ground state of H when τ is larger than the inverse energy gap of H [41]. The inverse of the energy gap typically remains constant with system size for many noncritical physical systems, such as the classical Ising chain [14]. Combinatorial optimization problems, including the MaxCut problem, are often modeled using the classical Ising model. Thus one can expect that the imaginary-time evolution converges fast to the ground state in these cases.

For each local interaction term H_μ supported on a set of qubits $\mathbb{S}_\mu$, the QITE algorithm [37] shows that the imaginary-time propagator of each local interaction term can be approximated by unitary gates

$$e^{-\Delta\tau H_\mu}|\psi\rangle \propto \prod_{m\in\mathbf{P}_{\mathbb{S}_i}} e^{-i\theta_\mu^{(m)}\sigma_\mu^{(m)}/2}|\psi\rangle, \tag{3}$$

where $\mathbf{P}_{\mathbb{S}_\mu}$ includes linear combinations of Pauli strings on the support $\mathbb{S}_\mu$ except for identity. For example, if $\mathbb{S}_\mu$ includes two qubits, then

$$\sigma_\mu^{(m)} \in \mathrm{span}(\{IX, IY, IZ, \ldots, ZY, ZZ\}), \tag{4}$$

with real spanning coefficients. We call $\sigma_\mu^{(m)}$ a Pauli series on $\mathbb{S}_\mu$. Equation (3) is approximately valid in the case that the correlation length of the initial state $|\psi\rangle$ and the imaginary time $\Delta\tau$ are finite and not very large [37].

Since the imaginary-time propagator preserves symmetries of the Hamiltonian system, Pauli series $\sigma_\mu^{(m)}$ should also preserve symmetries and thus can be determined. Specifically, assuming $\mathcal{G}$ is a unitary symmetry group, H_μ and $|\psi\rangle$ are invariant under transformations of the symmetry group, i.e.,

$$[U_g, H_\mu] = 0, \quad U_g|\psi\rangle = e^{if(g)}|\psi\rangle \forall\, g \in \mathcal{G}, \tag{5}$$

where U_g is a unitary representation of the symmetry group element $g \in \mathcal{G}$ and $f(g)$ is a real scalar function. Then $\sigma_\mu^{(m)}$

should be invariant by the conjugation of U_g [42,44–48],

$$U_g \sigma_\mu^{(m)} U_g^\dagger = \sigma_\mu^{(m)} \, \forall \, g \in \mathcal{G}. \tag{6}$$

This equation can be solved either by constructing linear systems of equations [42] or by implementing twirling operations on $\sigma_\mu^{(m)}$ [48]. Then the *i*HVA is constructed by applying unitary gates in Eq. (3) for each local interaction term

$$\left|\phi_I^{(L)}(\boldsymbol{\theta})\right\rangle = \prod_{l=1}^{L} \prod_{m} \prod_{\mu} e^{-i\theta_{l,\mu}^{(m)} \sigma_\mu^{(m)}/2}, \tag{7}$$

where L is the number of *Ansatz* layers and μ and m index local interaction terms and symmetry preserving Pauli series, respectively.

In this work we compare the *i*HVA and the QAOA *Ansatz* [12,13]. The QAOA *Ansatz* encodes the real-time evolution e^{-itH} of the Hamiltonian H. Since the real-time propagator also preserves unitary symmetries of the Hamiltonian system, the Pauli series utilized in the QAOA *Ansatz* also satisfy Eq. (6). Thus, the Pauli series obtained by Eq. (6) include the ones in the QAOA if only unitary symmetry groups are considered. These two kinds of *Ansätze* are distinguished if the Hamiltonian system possesses antiunitary time-reversal symmetry, which means that the Hamiltonian and the initial state have only purely real entries. Time-reversal symmetry is preserved by many chemical, quantum field, and combinatorial optimization Hamiltonians, such as the MaxCut problem studied in this work. For these Hamiltonian systems, the Pauli series used in the *i*HVA and the QAOA *Ansatz* are distinguished as follows:

$$\sigma_\mu^{(m)} : \begin{cases} \text{contains odd } Y \text{ letters} & \text{for the } i\text{HVA} \\ \text{contains even } Y \text{ letters} & \text{for the QAOA.} \end{cases} \tag{8}$$

For example, in the two-qubit case, $\sigma_\mu^{(m)}$ of the *i*HVA is spanned by

$$\sigma_\mu^{(m)} \in \text{span}(\{IY, XY, YI, YX, YZ, ZY\}) \tag{9}$$

and of the QAOA *Ansatz* by

$$\sigma_\mu^{(m)} \in \text{span}(\{IX, IZ, XI, XX, XZ, YY, ZI, ZX, ZZ\}).$$

The discriminative criterion (8) follows intuition. For the *i*HVA, $\sigma_\mu^{(m)}$ with odd Y letters is purely imaginary. Since H_μ is purely real, $e^{-i\theta\sigma_\mu^{(m)}/2}$ can be regarded as performing the real-time dynamics of the imaginary Hamiltonian $-iH_\mu$, which corresponds to the imaginary-time propagator $e^{-\Delta\tau H_i}$. This is the reason we refer to this *Ansatz* as the imaginary Hamiltonian variational *Ansatz*. For the QAOA, on the other hand, $\sigma_\mu^{(m)}$ with even Y letters is purely real, which is consistent with the realness of H_i. We refer to the Pauli series used in the *i*HVA as the relevant series.

To highlight the difference between the *i*HVA and the QAOA *Ansatz*, we present two toy examples. The variational *Ansatz* state of the *i*HVA and QAOA are denoted by $|\phi_I\rangle$ and $|\phi_R\rangle$, respectively. Consider a one-qubit and a two-qubit Hamiltonian

$$H_1 = -Z, \quad H_2 = ZZ, \tag{10}$$

whose *i*HVA and QAOA *Ansatz* are

$$H_1 : \begin{cases} |\phi_I(\theta)\rangle = R_Y(\theta)|+\rangle \\ |\phi_R(\theta)\rangle = R_X(\theta)R_Z(\theta)|+\rangle, \end{cases}$$
$$H_2 : \begin{cases} |\phi_I(\theta_1,\theta_2)\rangle = R_{YZ}(\theta_2)R_{ZY}(\theta_1)|{+}{+}\rangle \\ |\phi_R(\theta_1,\theta_2)\rangle = R_{XI}(\theta_2)R_{IX}(\theta_2)R_{ZZ}(\theta_1)|{+}{+}\rangle, \end{cases} \tag{11}$$

where $R_\sigma(\theta) = e^{-i\sigma\theta/2}$ is the Pauli exponential of Pauli string σ. The Pauli strings in the *i*HVA and QAOA *Ansatz* contain odd and even Y letters, respectively, and all satisfy the symmetry constraint in Eq. (6) ($U_g = XX$ for the two-qubit case). The *i*HVAs for these two toy examples are closely related to the imaginary-time evolution of H_1 and H_2, as one can check that

$$e^{\tau Z}|+\rangle \propto e^{-i\theta(\tau)Y/2}|+\rangle,$$
$$e^{-\tau ZZ}|{+}{+}\rangle \propto e^{-i\theta_1(\tau)ZY/2}|{+}{+}\rangle = e^{-i\theta_2(\tau)YZ/2}|{+}{+}\rangle,$$

where $\theta(\tau)$, $\theta_1(\tau)$, and $\theta_2(\tau)$ are functions of the imaginary time τ. In these formulas, all the imaginary-time propagators and the unitary gates can be represented by real matrices in the computational basis, as a result of the time-reversal symmetry kept by H_1 and H_2.

We study the ground-state preparation trajectories in these two *Ansätze* of the one- and two-qubit examples, as presented in Fig. 1. Figure 1(a) shows two trajectories in the Bloch sphere as we perform gradient descent using the *i*HVA and QAOA of H_1 starting at $\theta = 0$. We see that the *i*HVA trajectory (red) is geodesic between the initial state $|+\rangle$ and the ground state of H_1 ($|0\rangle$) on the Bloch sphere, while the trajectory of the QAOA (blue) is nongeodesic and would require more iteration steps during the gradient descent. Figure 1(b) shows the energy landscapes of the *i*HVA and QAOA for H_2. We see that a saddle point in the QAOA landscape appears at $(\theta_1, \theta_2) = (0, 0)$, which complicates the optimization process. In contrast, the *i*HVA landscape does not have this problem and is thus more favorable for optimization.

B. Graph and oriented spanning tree

We review the concepts of the tree graph and D-regular graph, which are the types of graphs mainly studied in this work. A graph $G = (\mathcal{V}, \mathcal{E})$ consists of a set of N nodes $i \in \mathcal{V}$, labeled by integers $i = 0, \ldots, N-1$ and undirected edges $(i, j) \in \mathcal{E}$. A tree graph is defined as a graph without a cycle. A graph is called as a D-regular graph if each node in the graph has D edges connected with the other nodes. Figure 2(a) presents an example of the tree graph with six nodes, five edges and the 3-regular graph with six nodes and nine edges.

A key concept used in this work is the oriented spanning tree, as shown in Fig. 2(b). A spanning tree of an undirected graph G is a tree subgraph that includes all of the vertices of G. An oriented spanning tree is obtained by choosing a tree node as the root node such that the tree hierarchy is subsequently constructed. Each edge of the oriented tree connects a parent node and a child node. For a tree graph, its oriented spanning tree is not unique, which is determined by the chosen root. Oriented spanning trees of a tree graph and a 3-regular graph are shown in Fig. 2(b), as well as other useful concepts of the oriented tree.

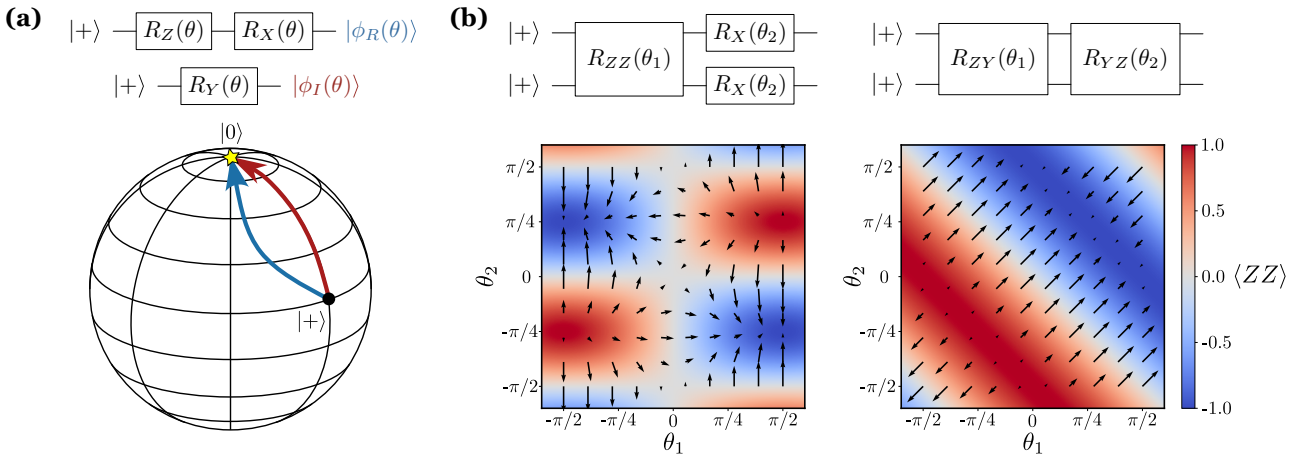

FIG. 1. Comparison of the *i*HVA and the QAOA *Ansatz* finding the ground state of (a) $H_1 = -Z$ and (b) $H_2 = ZZ$. (a) Gradient-descent trajectories of the *i*HVA (red) and QAOA (blue) on the Bloch sphere starting at $\theta = 0$. (b) Energy landscape of the QAOA (left) and *i*HVA (right). The arrows indicate trajectories of the gradient descent.

C. MaxCut problem

The MaxCut problem is a paradigmatic test for various *Ansätze* used in the VQE [12]. Given that a graph consists of edges and nodes, MaxCut aims to partition the graph's nodes into two complementary sets such that the number of edges between these two sets is as large as possible. This problem can be formulated as follows. Suppose $G = (\mathcal{V}, \mathcal{E})$ is a graph with N nodes. Given an N-bit string $x = x_{N-1}, \ldots, x_0$, $x_i \in \{0, 1\}$, assume cut(x) is the set of edges (i, j) such that $x_i \neq x_j$. The object of the MaxCut problem is to maximize the cut size $C(x) = |\text{cut}(x)|$, i.e., the number of edges in cut(x). For example, Fig. 2(c) gives solutions of the MaxCut problem of the two graphs in Fig. 2(a), where the black (white) node denotes $x_i = 1$ (0). The MaxCut solution of the tree graph is $x_{\text{tree}} = 101011$ with cut size $C(x_{\text{tree}}) = 5$ and that of the 3-regular graph is $x_{\text{regular}} = 011001$ with cut size $C(x_{\text{regular}}) = 7$. The MaxCut solution of a graph is not unique. Specifically, the bitwise inverse $\bar{x}_{\text{tree}} = 010100$ and $\bar{x}_{\text{regular}} = 100110$ are also MaxCut solutions of their corresponding graphs.

For arbitrary graphs, the MaxCut solution is equal to the maximum eigenvalue of an N-qubit Hamiltonian

$$\hat{C} = \frac{1}{2} \sum_{(i,j)\in\mathcal{E}} (I - Z_i Z_j), \tag{12}$$

where Z_i is the Pauli-Z operator on the ith qubit and I is the identity operator. Finding the maximum of $\hat{C}$ is equivalent to finding the minimum of $-\hat{C}$. Thus, we aim to find the minimum eigenvalue of the MaxCut Hamiltonian

$$H_{\text{MC}} = \sum_{(i,j)\in\mathcal{E}} Z_i Z_j, \tag{13}$$

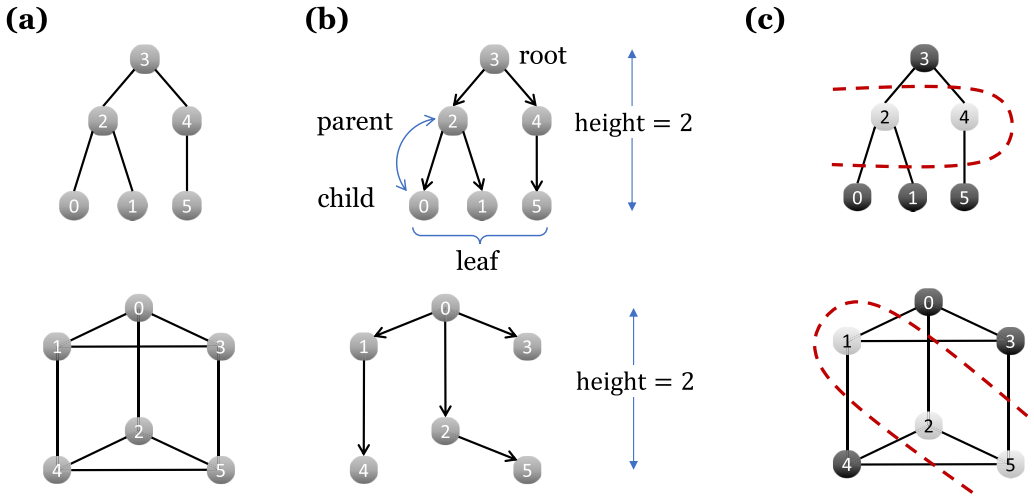

FIG. 2. (a) Examples of a tree graph (top) and a 3-regular graph (bottom). (b) Oriented spanning trees of the tree graph and the 3-regular graph. Each oriented edge connects a child node at its tip and a parent node at its tail. A root node is the topmost node in an oriented tree that has no parent node, and the leaf node does not have child nodes. The height of an oriented tree is the length of the longest downward path from the unique root node to one of the leaf nodes. (c) MaxCut solution of the tree graph and the 3-regular graph.

where an irrelevant constant is discarded. Thus, the MaxCut problem is mapped to a ground-state problem and can be solved using the VQE [12,18].

Finding the MaxCut solution for arbitrary graphs is known to be NP-complete [49]. For this reason, we intend to find an approximate solution for the MaxCut problem. The performance of an algorithm approximately solving the MaxCut problem can be estimated by the approximation ratio, which is defined as

$$\alpha \equiv \frac{C(x)}{C_{\max}}, \tag{14}$$

where $C_{\max}$ is the exact maximum cut size of the graph, $C(x)$ is the cut size provided by a given algorithm, and $\alpha \to 1$ indicates that the algorithm could solve the MaxCut problem with high accuracy. The Goemans-Williamson algorithm is a classical polynomial-time algorithm that guarantees an approximation ratio of 0.8785 [43], which is optimal under the unique game conjecture [50]. In the quantum scenario, $C(x)$ is transformed to a functional of a given quantum state $|\phi\rangle$, which is a superposition of bit strings $|x\rangle$, and the cut size is evaluated as

$$\begin{aligned} C(x) \to C[\phi] &\equiv \langle\phi|\hat{C}|\phi\rangle \\ &= \tfrac{1}{2}(|\mathcal{E}| - \langle\phi|H_{\mathrm{MC}}|\phi\rangle), \end{aligned} \tag{15}$$

where $\hat{C}$ is defined in Eq. (12), N is the number of nodes, $|\mathcal{E}|$ is the total number of edges of the graph, and $|\phi\rangle$ is the variational *Ansatz* state, which can be either the *i*HVA state $|\phi_I\rangle$ or the QAOA *Ansatz* state $|\phi_R\rangle$ in this work.

III. *i*HVA FOR THE MAXCUT PROBLEM

This section explicitly constructs the *i*HVA for the MaxCut problem. Apart from providing relevant series of the MaxCut Hamiltonian, we focus on choosing an appropriate arrangement of the parametrized quantum gates, where the tree arrangement and the *i*HVA tree are introduced for arbitrary graphs.

A. Relevant series of the MaxCut problem

The relevant series of the MaxCut problem and the general structure of its *i*HVA are given as follows. The MaxCut Hamiltonian H_{MC} commutes with the symmetry transformation $U_g = \prod_{i\in\mathcal{V}} X_i$, which corresponds to the global bit-flip symmetry [16]. Additionally, H_{MC} is purely real such that the time-reversal symmetry is preserved. The relevant series corresponding to the local interaction term Z_iZ_j are

$$\sigma^{(1)}_{(i,j)} = Z_iY_j, \quad \sigma^{(2)}_{(i,j)} = Y_iZ_j. \tag{16}$$

Then we define parametrized subcircuits of these two relevant series

$$U^{(l)}_{ZY} \equiv \prod_{(i,j)\in\mathcal{E}} e^{-i\theta_{l,ij}Z_iY_j/2}, \quad U^{(l)}_{YZ} \equiv \prod_{(i,j)\in\mathcal{E}} e^{-i\theta_{l,ij}Y_iZ_j/2}, \tag{17}$$

where $\theta_{l,ij}$ are variational parameters. The variational *Ansatz* is constructed by alternating these subcircuits in order and applying to the initial state

$$\left|\phi^{(p)}_I(\boldsymbol{\theta})\right\rangle \equiv U^{(p)}_{ZY\,(YZ)}\cdots U^{(2)}_{YZ}U^{(1)}_{ZY}|+\rangle^{\otimes N}, \tag{18}$$

where $|+\rangle^{\otimes N}$ is the tensor product of N single-qubit states $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ and $U^{(p)}_{ZY\,(YZ)}$ denotes that the last round is $U^{(p)}_{ZY\,(YZ)}$ if p is odd (even). Similar to L used in Eq. (7), here we define p as the number of rounds of the *i*HVA. In the first round of the *i*HVA, we apply one ZY gate on each edge of the graph. In the second round, we reverse the qubits of Z and Y such that both $e^{-i\theta_{l,ij}Z_iY_j/2}$ and $e^{-i\theta_{l,ij}Y_iZ_j/2}$ are applied on one edge, as required by the imaginary-time evolution of Z_iZ_j. In one round of the *i*HVA, only one ZY gate is applied for each edge of the graph. So in this way, the number of two-qubit gates in one round of the *i*HVA can be compared with that in one round of the QAOA *Ansatz*, as will be detailed later.

The QAOA *Ansatz* for the MaxCut problem is distinguished from the *i*HVA, which reads

$$\left|\phi^{(p)}_R(\boldsymbol{\beta},\boldsymbol{\gamma})\right\rangle = \prod_{l=1}^{p}\left(\prod_{i\in\mathcal{V}} e^{-i\beta_{l,i}X_i/2}\prod_{(i,j)\in\mathcal{E}} e^{-i\gamma_{l,ij}Z_iZ_j/2}\right)|+\rangle^{\otimes N},$$

where $\beta_{l,i}$ and $\gamma_{l,ij}$ are variational parameters. This *Ansatz* is the multiangle QAOA (MQAOA) *Ansatz* [51], which has better expressibility than the original QAOA *Ansatz* [52]. The number of two-qubit gates in one round of the QAOA *Ansatz* is the same as in one round of the *i*HVA, since the two-qubit Pauli exponentials in these two *Ansätze* can be converted by single-qubit gates

$$e^{-i\theta_{l,ij}Z_iY_j/2} = e^{i(\pi/4)X_j}e^{-i\theta_{l,ij}Z_iZ_j/2}e^{-i(\pi/4)X_j}. \tag{19}$$

One round of the *i*HVA has ZY gates $e^{-i\theta_{l,ij}Z_iY_j/2}$ on different edges that do not commute with each other. This allows different arrangements. The arrangement of the ZY gates impacts the solution accuracy. Additionally, depending on the arrangement of the ZY gates, the circuit depth of $U^{(l)}_{ZY}$ could be a constant or grow logarithmically or even linearly to the graph nodes N. We discuss the arrangement of the *i*HVA and its depth in detail in the following sections.

B. *i*HVA on trees and tree arrangement

In this section we demonstrate how to choose an appropriate arrangement of Pauli exponentials $e^{-i\theta_{l,ij}Z_iY_j/2}$ in one subcircuit $U^{(l)}_{ZY}$ defined in Eq. (17). Our choice of the tree arrangement is based on an observation that the MaxCut of tree graphs can be exactly achieved by one round of the *i*HVA by choosing the tree arrangement, as demonstrated below.

We use the tree graph in Fig. 2(a) as an example, whose MaxCut solution is shown in the top panel of Fig. 2(c). The two solutions $x_{\mathrm{tree}} = 101011$ and $\bar{x}_{\mathrm{tree}} = 010100$ can be obtained by preparing the ground state of its MaxCut Hamiltonian H_{MC},

$$|\mathrm{g.s.}\rangle = \frac{1}{\sqrt{2}}(|010100\rangle + |101011\rangle). \tag{20}$$

This ground state is locally equivalent to the six-qubit Greenberger-Horne-Zeilinger (GHZ) state $(|0\rangle^{\otimes 6} + |1\rangle^{\otimes 6})/\sqrt{2}$. The GHZ state has long-range entanglement and cannot be prepared by a constant-depth quantum circuit [53]. This provides the intuition to arrange the ZY gates following an orientation of the tree to increase the circuit depth. Figure 3(a) shows the oriented tree and the corresponding tree

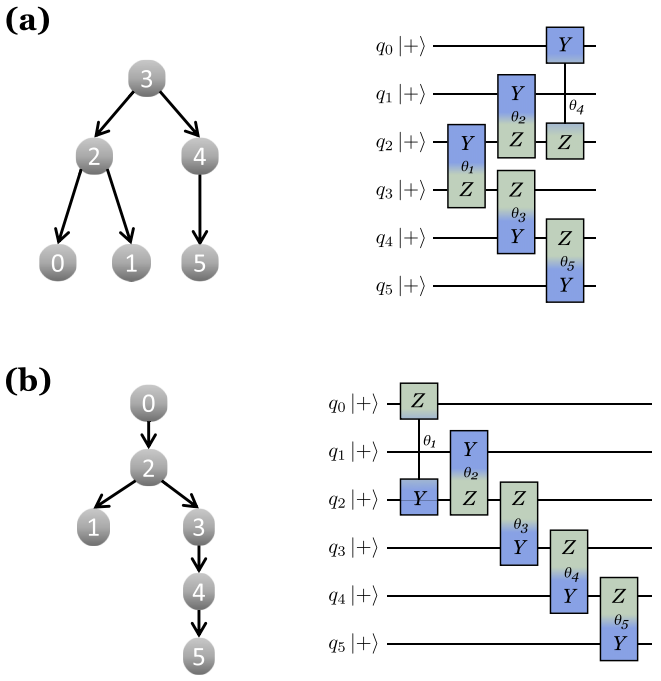

FIG. 3. (a) Oriented tree for the tree graph in Fig. 2(a) and the corresponding tree arrangement of the ZY gates. Each colored rectangle is a ZY exponential $e^{-i\theta_{l,ij}Z_iY_j/2}$ with the value of $\theta_{l,ij}$ shown at the center of the rectangle. This oriented tree has node 3 as the root, which is the lowest among all orientations of the tree. (b) The highest-oriented tree of the tree graph has node 0 as the root node. Its corresponding tree arrangement of ZY gates is illustrated.

arrangement of ZY gates in $U_{ZY}^{(l)}$. This circuit with variational parameters θ_i, $i = 1, \ldots, 5$, is a one-round *i*HVA. One can check that the ground state (20) can be prepared by setting

$$\theta_1 = \theta_2 = \cdots = \theta_5 = \frac{\pi}{2}. \tag{21}$$

Thus, the MaxCut problem of the tree is exactly solved. This example can be generalized to arbitrary trees, and we have the following theorem.

Theorem 1. The MaxCut of arbitrary trees can be achieved by the one-round *i*HVA following the tree arrangement.

The proof of this theorem is presented in Appendix A.

In the tree arrangement, we construct an oriented tree whose root can be chosen arbitrarily. We can use this arbitrariness to reduce the depth of the tree arrangement *Ansatz*. The reduction in depth reduces the runtime of the algorithm and improves robustness against noise on real hardware. Roughly speaking, the tree arrangement circuit is shallower if the corresponding oriented tree is lower in height. For example, the tree arrangement in Fig. 3(a) has depth 3 in the unit of the depth of the ZY exponential, and the height of the oriented tree is 2. On the other hand, if the node 0 is chosen as the root, as shown in Fig. 3(b), the corresponding tree arrangement has depth 5 and the height of the oriented tree is 4. Thus, to construct a tree arrangement circuit with relatively small depth, we choose the root node corresponding to the lowest oriented tree among all nodes. This can be done on a classical computer by first enumerating all the N nodes as the root and then calculating the corresponding height of the oriented tree. As calculating the height of an oriented tree recursively requires the time complexity of $O(N)$ [54], the above procedure can be accomplished with the time complexity of $O(N^2)$.

ALGORITHM I. Arrangement of gates in one round of the *i*HVA tree.

Require: A connected graph $G = (\mathcal{V}, \mathcal{E})$
procedure ARRANGE(G)
1. Randomly pick a root of G. Construct a breadth-first spanning tree T.
2. Find a root r for the tree with minimum height. Arrange ZY gates following oriented spanning tree T with root r.
3. Define the remaining graph $G' \leftarrow G - T$.
Delete isolated nodes in G'.
if G' has no nodes **then**
return
else
for each connected subgraph g' of G' **do**
ARRANGE(g').
end for
end if
end procedure

C. *i*HVA on arbitrary graphs

In the preceding section we saw that the one-round *i*HVA following the tree arrangement can achieve the MaxCut of arbitrary trees. In this section we generalize the tree arrangement of tree graphs to arbitrary connected graphs.

The generalization proceeds by decomposing a connected graph into several breadth-first spanning (BFS) trees [55]. We choose the BFS tree because the BFS tree is usually lower in height than the other spanning trees, so the corresponding quantum circuit has a smaller depth. One way of decomposing a connected graph into BFS trees is shown in Fig. 4(a). In the first step, we randomly pick a root and construct a BFS tree, which means that a parent node connects all the adjacent nodes as child nodes if the tree has never visited the nodes. Constructing a BFS tree with a given root is efficient for all connected graphs. In the second step, we find a root of the spanning tree leading to the minimum height and arrange the ZY gates at the rightmost end of the circuit, as described in the preceding section. Third, apart from the spanning tree, the remaining graph is obtained by subtracting edges in the spanning tree from the original graph and deleting isolated nodes that possibly appear. We delete the isolated nodes to provide an explicit judgment on when the procedure should be stopped, as will be detailed later. The remaining graph can be connected or disconnected. For every connected part of the remaining graph, we return to the first step and repeatedly construct its BFS tree, as shown in the second line of Fig. 4(a). This repetition is stopped in the third step if no nodes are left after deleting isolated nodes. These steps are summarized in the following Algorithm 1.

This algorithm can be performed efficiently on classical computers. The root-finding procedure in the second step is the most time-consuming part of the algorithm. Assume that the number of edges in each spanning tree derived by the above procedure is $M_{\mathcal{E}_\alpha}$, where $\mathcal{E} = \bigcup_\alpha \mathcal{E}_\alpha$ is the whole edge

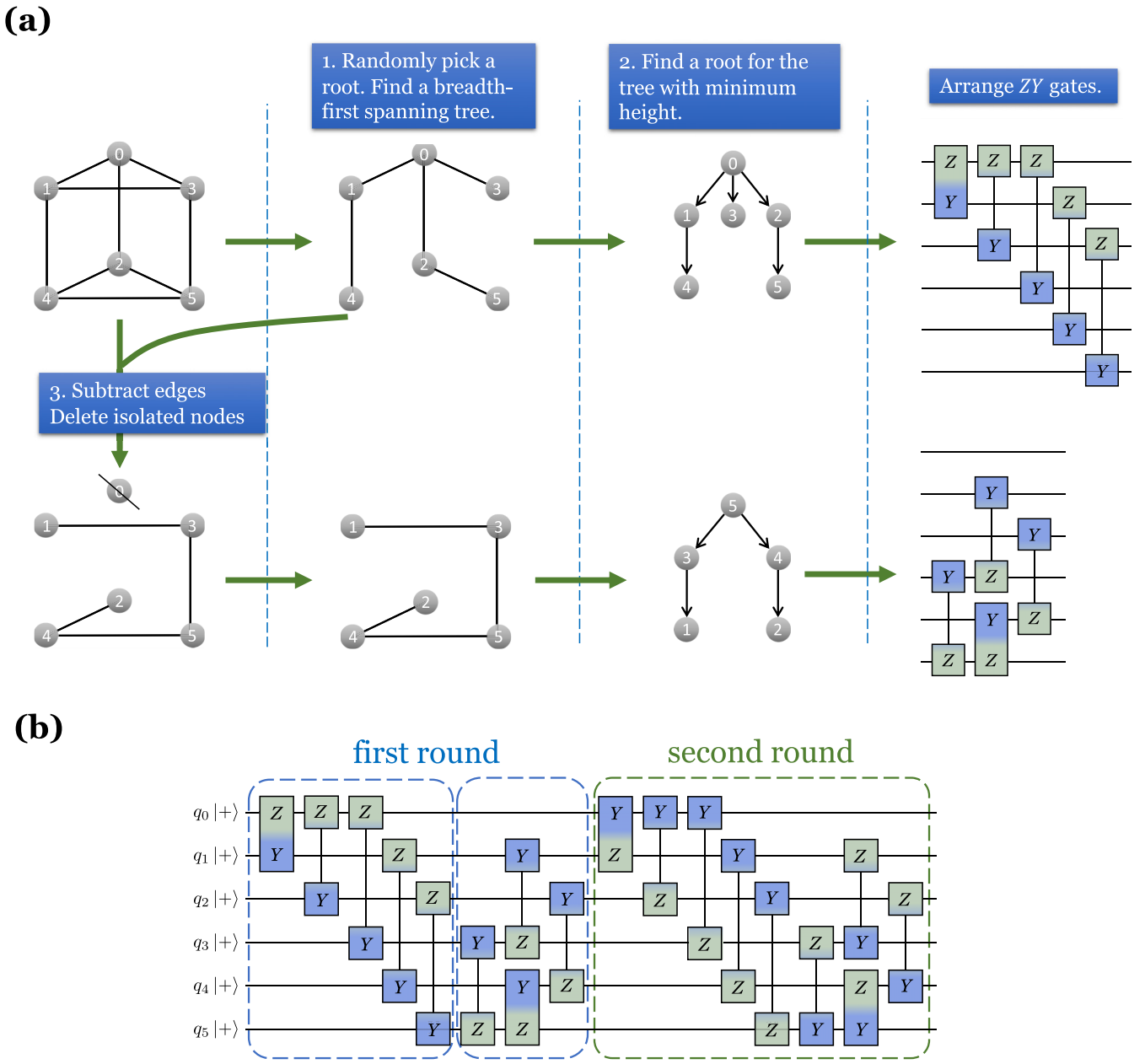

FIG. 4. (a) Construction of the *i*HVA tree solving the MaxCut problem of arbitrary graphs. These three steps give one round $U_{ZY}^{(l)}$ of the *i*HVA tree. Each two-qubit gate represents one ZY exponential $e^{-i\theta_{l,ij}Z_iY_j/2}$. (b) The *i*HVA tree with two rounds. The first round has two parts that arrange ZY gates by the procedure in (a). The second round is constructed by reversing the orientation of ZY gates to YZ.

set of G and $M=\sum_\alpha M_{\mathcal{E}_\alpha}$ is the total number of edges of G. The number of nodes in each tree is $M_{\mathcal{E}_\alpha}+1$. Thus, using the time complexity $O(N^2)$ of one tree derived in the preceding section, the total time complexity of the algorithm is upper bounded by

$$\sum_\alpha \left(M_{\mathcal{E}_\alpha}+1\right)^2 \sim O(M^2), \tag{22}$$

which grows polynomially with the system size.

Using the above procedure, an explicit product order of ZY gates in the $U_{ZY}^{(l)}$ is obtained. The same order can be defined for another subcircuit $U_{YZ}^{(l)}$ by reversing the orientation of ZY gates in $U_{ZY}^{(l)}$ to YZ, as shown in Fig. 4(b). We call an *i*HVA with each round given by the above procedure an *i*HVA tree. In Fig. 4(b) we show an example of an *i*HVA tree with two rounds.

For tree graphs, the one-round *i*HVA tree is reduced to the tree arrangement introduced in Sec. III B. Thus, the one-round *i*HVA tree can exactly cut arbitrary tree graphs, as shown in Theorem 1. In contrast, the QAOA requires linearly growing rounds to exactly cut lines [22], and the MQAOA with one round can exactly cut only star graphs [51]. Both lines and star graphs are particular tree graphs and can be cut exactly using the one-round *i*HVA tree.

The advantage of the *i*HVA tree over the QAOA on tree graphs is in exchange for the larger depth of the quantum circuit. Assuming an all-to-all qubit connectivity of the quantum

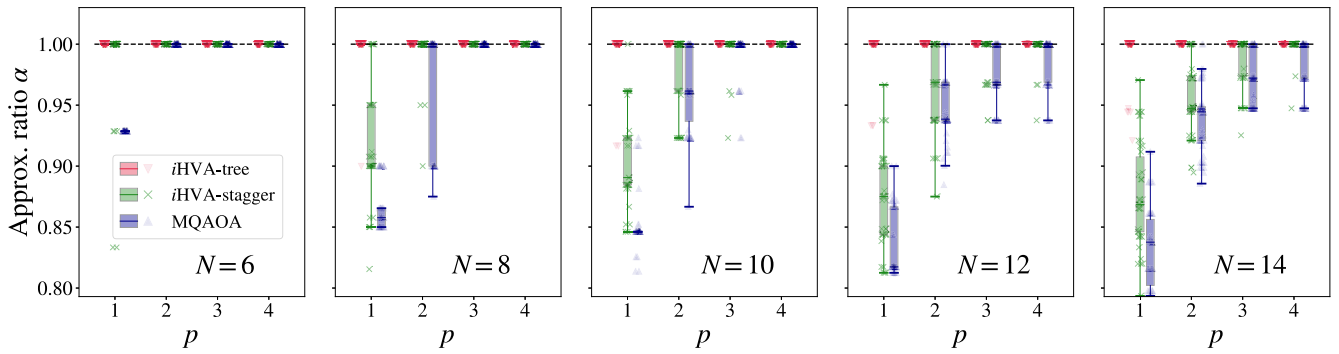

FIG. 5. Simulated results for the approximation ratio α of 3-regular graphs as a function of p circuit rounds. The *Ansätze* considered here include the *i*HVA tree, *i*HVA stagger, and MQAOA, with results marked by red down triangles, green crosses, and blue up triangles, respectively. Each plot corresponds to a fixed number of nodes N, with 50 randomly generated 3-regular graphs. The box plot is used to reflect the statistical properties of the 50 ratios. For $N \leqslant 14$, approximation ratios achieved by the *i*HVA tree are all close to 1 as $p \geqslant 2$.

chip, the depth of a p-round QAOA for a D-regular graph is $O(p)$, which has no dependence on the number of nodes N [16]. In contrast, the depth of the p-round *i*HVA tree is lower bounded by

$$d_p = \Omega(p \ln N) \tag{23}$$

and upper bounded by

$$d_p = O(pN), \tag{24}$$

which is distinguished from the QAOA. These bounds are derived in Appendix B. This depth scaling could bring up fundamental differences between the accuracy of solving the MaxCut problem using the *i*HVA tree and QAOA. Although deeper quantum circuits suffer more from errors on NISQ devices, there exist error suppression methods such as dynamical decoupling [56–58] that are suitable for the *i*HVA-tree-type circuits.

The arrangement of the ZY gates in the *i*HVA impacts its ability to find the MaxCut solution. To manifest this, we introduce another arrangement of the *i*HVA in Appendix C, constructed as shallow as possible among all the arrangements by a staggered layout of the ZY gates. The *i*HVA following this arrangement is called the *i*HVA stagger. In the following section we numerically compare the *i*HVA tree, *i*HVA stagger, and QAOA ansatz by testing their performance of finding the MaxCut solution of random regular graphs.

IV. NUMERICAL RESULTS

We have seen that the *i*HVA tree performs better than the QAOA *Ansatz* in solving the MaxCut problem of tree graphs. In this section we numerically demonstrate that the outperformance of the *i*HVA tree can be observed in solving the MaxCut problem of more complicated D-regular graphs. We compare the performance of the *i*HVA tree with the MQAOA *Ansatz* and further with the classical, polynomial-time Goemans-Williamson (GW) algorithm. The numerical simulations are performed using the noiseless simulator of Qiskit [59] and using the `ibm_brisbane` superconducting quantum computer.

A. Simulated results

We perform numerical simulations using the Qiskit noiseless state-vector quantum simulator [59]. In Fig. 5 we plot the approximation ratio as a function of *Ansatz* rounds, with the number of nodes $N \in \{6, 8, 10, 12, 14\}$. For each plot with a fixed N, we generate 50 random 3-regular graphs [60] with the corresponding number of nodes. The ground state of the Hamiltonian H_{MC} is prepared using the VQE algorithm, with the classical optimizer `SLSQP` [61]. To avoid local minima of the energy landscape as much as possible, for each optimization trajectory, we adopt small constant initialization [62] for the variational parameter $\boldsymbol{\theta}$, where each variational parameter $\theta_{l,ij}$ is chosen independently and uniformly from $[0, 0.001]$. The optimization is performed five times with different initializations for each graph, and the largest approximation ratio α among the five repetitions is selected and plotted. The *i*HVA tree, *i*HVA stagger, and MQAOA *Ansatz* results are marked by red down triangles, green crosses, and blue up triangles, respectively. The statistical properties of the results are reflected using the box plot [63], where the middle line of the box denotes the median of the data. For all the number of nodes and the circuit rounds, the performance of the *i*HVA tree is better than that of the *i*HVA stagger, and both are better than the MQAOA. For $N \leqslant 14$, approximation ratios achieved by the *i*HVA tree are all close to one as $p \geqslant 2$. To achieve the same accuracy, the MQAOA requires more rounds as N increases, i.e., $p \geqslant 2, 3, 4$ for $N = 6, 8, 10$, respectively. This behavior of the MQAOA is consistent with previous studies [22].

For larger graph sizes and regular graphs beyond $D = 3$, we perform numerical simulations using the *i*HVA tree with fixed $p = 2$. During the optimization, we use a conditional value at risk (`CVaR`) with a confidence level 0.1 as the objective function, which has been shown to accelerate optimization for combinatorial optimization problems [64]. Figure 6 plots the approximation ratios for 50 graphs and their box plots as a function of graph nodes N, where 50 random D-regular graphs are generated for each $N \in \{8, 12, \ldots, 24\}$ and $D \in \{3, 4, 5\}$. The performance of the classical, polynomial-time GW algorithm on the same test graphs is plotted with gray circles. For each graph, both the GW algorithm and VQE

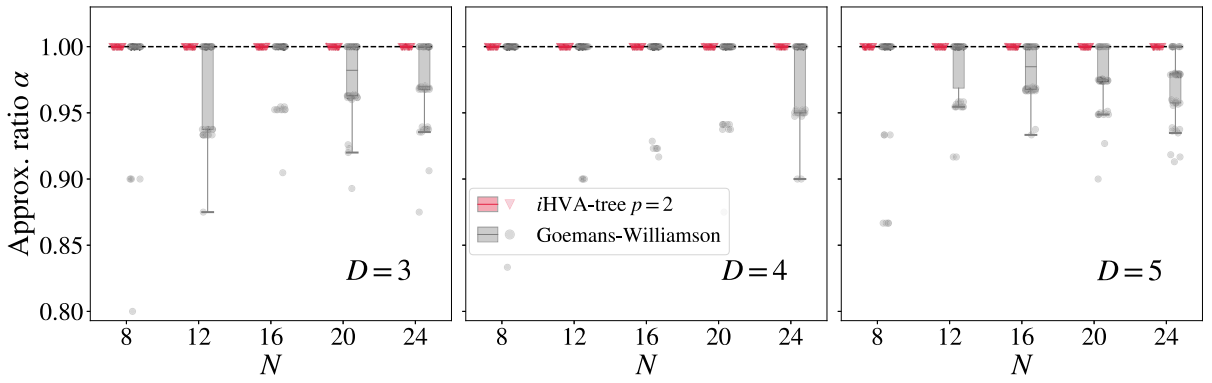

FIG. 6. Approximation ratio α of D-regular graphs as a function of graph nodes N, reached by the two-round *i*HVA tree and GW algorithm, where $D = 3, 4, 5$ for each plot. For each D and N, approximation ratios of 50 random D-regular graphs derived by these two algorithms are plotted and marked by red down triangles and gray circles, respectively. Compared with the GW results, the *i*HVA tree can exactly solve the MaxCut problem for all the randomly generated regular graphs with N up to 24 and $D \leqslant 5$.

with small constant initialization are repeated five times, and the largest approximation ratio is selected and plotted. We see that the *i*HVA tree can exactly solve the MaxCut problem for all the randomly generated regular graphs with N up to 24 and $D \leqslant 5$, while the GW algorithm struggles to achieve the exact solutions for some particular graphs.

The advantage of the *i*HVA tree compared with the *i*HVA stagger and QAOA *Ansatz* can be partially explained by looking at the backward light cone of an observable Z_iZ_j in these *Ansätze*, as shown in Fig. 7. The light cone of the one-round *i*HVA stagger and QAOA covers only a constant number of qubits, while the light cone of the *i*HVA tree covers the whole graph. This means that the one-round *i*HVA tree is accessible to the global information of the graph, while the *i*HVA stagger and QAOA are not. The global information is important for the accurate solution of the MaxCut problem [17], and this is also a part of the reason why the MaxCut problem is hard to solve using classical computers. Additionally, the global backward light cone indicates that the expectation of Z_iZ_j cannot be calculated directly in the Heisenberg picture. On the other hand, for the one-round *i*HVA stagger and QAOA, the expectation of Z_iZ_j can be calculated efficiently on classical computers by involving a constant number of qubits. From this perspective, the *i*HVA tree could be harder to simulate classically and have more quantum effects involved than the *i*HVA stagger and QAOA ansatz, thus providing higher accuracy than the latter two.

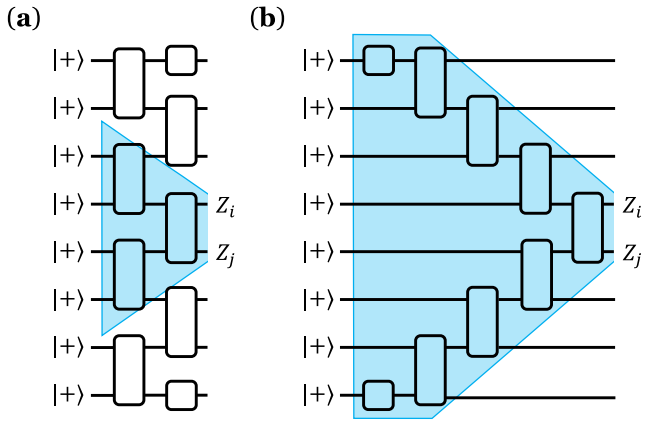

FIG. 7. Illustration of the backward light cone of the observable Z_iZ_j in one round of (a) the *i*HVA stagger or QAOA and (b) the *i*HVA tree on a cycle graph. The two-qubit gate denotes $e^{-i\gamma ZZ/2}$ for the QAOA and $e^{-i\theta ZY/2}$ for the *i*HVA. Single-qubit gates $e^{-i\beta X/2}$ in the QAOA have no impact on the backward light cone and are not explicitly presented.

B. Hardware results

We compare the *i*HVA tree and QAOA ansatz using IBM's quantum hardware to solve the maximum eigenvalue problem. Consider a random weighted Hamiltonian

$$H_w = \frac{1}{2}\sum_{(i,j)\in\mathcal{E}} (I - w_{ij}Z_iZ_j), \tag{25}$$

where $\mathcal{E}$ is a set of edges of a heavy-hex connectivity graph $G = (\mathcal{V}, \mathcal{E})$ with the number of nodes $N = 67$ and w_{ij} are randomly chosen to be ± 1. Here G is tailored to the connectivity of the IBM Eagle-class heavy-hex devices [59,65], as shown in Appendix D. The performance of the variational *Ansatz* is evaluated using an approximation ratio defined by

$$\alpha(x) = \frac{\langle x|H_w|x\rangle}{\max_{x_0}\langle x_0|H_w|x_0\rangle}, \tag{26}$$

given a classical bit string x by measuring a quantum state in the Pauli-Z basis. This definition is analogous to the approximation ratio of the MaxCut problem in Eq. (14). The exact maximum eigenvalue $\max_{x_0}\langle x_0|H_w|x_0\rangle$ for the 67-node heavy-hex graph G can be obtained by the greedy algorithm. The greedy algorithm can provide a good approximation to $\max_{x_0}\langle x_0|H_w|x_0\rangle$ since G has a small number of cycles (see Appendix D). However, the greedy algorithm does not work for general graphs.

For the hardware demonstrations, we use the equal-angle version of the *i*HVA tree and the QAOA *Ansatz* with one

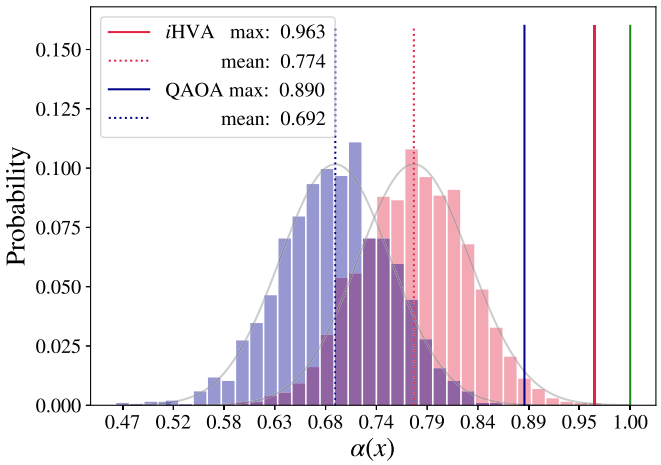

FIG. 8. Hardware results for the probability distributions of $\alpha(x)$ by 2048 measurements using the *i*HVA tree (red) and the QAOA *Ansatz* (blue). The solid and dotted lines label the positions of the maximum and mean values of the distributions, respectively. The green solid line denotes the position of $\alpha(x) = 1$. The demonstration is performed using 67 qubits of `ibm_brisbane`.

round, defined by

$$|\phi_I(\theta)\rangle = \prod_{(i,j)\in\mathcal{E}} e^{-i\theta w_{ij} Z_i Y_j/2} |+\rangle^{\otimes N},$$
$$|\phi_R(\beta,\gamma)\rangle = \prod_{i\in\mathcal{V}} e^{-i\beta X_i/2} \prod_{(i,j)\in\mathcal{E}} e^{-i\gamma w_{ij} Z_i Z_j/2} |+\rangle^{\otimes N}, \quad (27)$$

which have one and two variational parameters to be optimized, respectively. For the *i*HVA tree, the ZY gates are arranged according to the steps provided in Fig. 4 and the QAOA gates are arranged to be as shallow as possible [16].

We perform the VQE using the above two *Ansätze* on the quantum hardware `ibm_brisbane` [59]. The *i*HVA-tree circuit performed on the hardware and its layout on the `ibm_brisbane` coupling map are shown in Appendix D. The optimization is performed using the classical optimizer `COBYLA` [61] and `CVaR` as the objective function. Each `COBYLA` optimization trajectory starts using small constant initialization and iterates 20 steps, and 1024 measurement shots are used to evaluate one `CVaR` expectation. After the optimization, we take 2048 measurement shots using the optimized variational parameters and calculate their $\alpha(x)$. The probability distributions of $\alpha(x)$ are plotted in Fig. 8(b). We see that both the maximum and mean α of the *i*HVA tree are correspondingly larger than the ones of the QAOA *Ansatz*, and the maximum α of the *i*HVA tree reaches an approximation ratio of 0.963. These results show the benefit of using the *i*HVA tree over the QAOA in solving large-scale combinatorial optimization problems.

When the above two *Ansätze* are executed on the hardware, dynamic decoupling with a super-Hahn sequence [66] is used to suppress the decoherence error during the idle periods of the qubits. This technique significantly improves the behavior of the *i*HVA tree since its qubits have long idle periods due to its treelike structure.

V. ABSENCE OF A BARREN PLATEAU FOR THE CONSTANT-ROUND *i*HVA

A variational *Ansatz* with a constant number of rounds is efficient in the number of quantum gates and could be more resilient to decoherence in quantum devices, compared to the *Ansätze* with rounds growing with the system size. More importantly, such *Ansätze* could be free from barren plateaus (BPs) [26]. A variational *Ansatz* with BPs cannot be optimized efficiently due to the exponentially vanishing gradients of its energy landscape. Barren plateaus can be diagnosed by calculating the variance of the energy expectation over the variational parameters

$$\mathrm{Var}(\langle\overline{H_{\mathrm{MC}}}\rangle) \equiv \mathrm{Var}_{\boldsymbol{\theta}}\left(\langle\phi(\boldsymbol{\theta})|\frac{H_{\mathrm{MC}}}{E_0}|\phi(\boldsymbol{\theta})\rangle\right), \quad (28)$$

where E_0 is the minimum eigenvalue of the Hamiltonian H_{MC}, as a normalization factor. If $\mathrm{Var}(\langle\overline{H_{\mathrm{MC}}}\rangle)$ vanishes exponentially with the number of nodes N, then the energy landscape of the *Ansatz* is said to exhibit BPs [67].

For the constant-round *i*HVA with arbitrary arrangements of the ZY gates, the following theorem holds.

Theorem 2. For the p-round *i*HVA in Eq. (18) solving the MaxCut problem on D-regular graph with N nodes, if p is even, the variance of the energy expectation is lower bounded by

$$\mathrm{Var}(\langle\overline{H_{\mathrm{MC}}}\rangle) \geqslant \frac{DN}{E_0^2 2^{D(p+1)-1}}. \quad (29)$$

This theorem is proved by explicitly calculating $\mathrm{Var}(\langle\overline{H_{\mathrm{MC}}}\rangle)$ in the Heisenberg picture, as shown in Appendix E.

According to this theorem, if the degree D is a constant and the ground-state energy E_0 is of $O(\mathrm{poly}(N))$, because the exponent in Eq. (29) does not depend on graph nodes N, the variance decays at most polynomially in N. Therefore, the constant-round *i*HVA does not exhibit BPs to solve the MaxCut problem of D-regular graphs. This theorem guarantees that the gradient calculated in the previous numerical simulations can be measured efficiently using real quantum devices.

Figure 9 depicts the variance of the normalized energy expectation as a function of the graph nodes N, where the two-round *i*HVA tree is used. We randomly generate 50 graphs, uniformly sample variational parameters 1024 times from $[0, 4\pi]$, and calculate the variance of the energy expectations. Each data point is averaged over the 50 graphs. It is shown that the variance of D-regular graphs (where $D = 3, 4, 5$) decays polynomially with N, which is consistent with the theoretical lower bound given in Eq. (29).

The above theorem can be applied to graphs with more edges than regular graphs with a constant D. For example, the all-to-all connected complete graph is an $(N-1)$-regular graph, and the Erdős-Rényi graph connecting each pair of nodes with probability q is effectively a $q(N-1)$-regular graph. For these graphs, the lower bound of the variance decays exponentially with N, so the gradient of their constant-round *i*HVA vanishes exponentially with N. As numerically demonstrated in Fig. 9, for the two-round *i*HVA tree of the

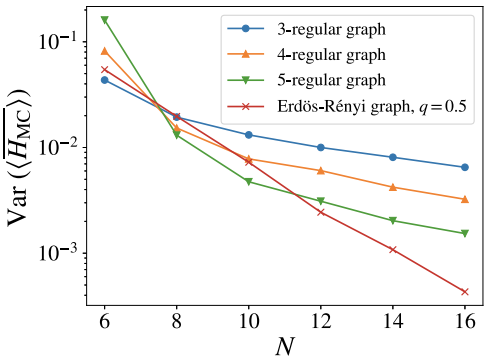

FIG. 9. Variance of the normalized energy expectation as a function of the graph nodes N. The *Ansatz* used is the two-round *i*HVA tree. We test D-regular graphs (where $D = 3, 4, 5$) and Erdős-Rényi graphs with $q = 0.5$, with results marked by blue circles, orange up triangles, green down triangles, and red crosses, respectively. Within the plot range, the variance decays polynomially with N for regular graphs and exponentially for Erdős-Rényi graphs. The y axis is the logarithmic scale.

Erdős-Rényi graph ($q = 0.5$), the variance decays exponentially with the graph nodes up to $N = 16$.

This result shows that the constant round *i*HVA tree is no panacea. Therefore, special attention has to be paid when using the *i*HVA tree for arbitrary graphs. Although both the *i*HVA tree and QAOA suffer from BPs for certain problems, the *i*HVA tree would still need fewer rounds compared to the QAOA.

VI. CONCLUSION AND OUTLOOK

Common quantum computing approaches to solve the MaxCut problem such as the QAOA or quantum annealing suffer from an adiabatic bottleneck that leads to either larger circuit depth or longer evolution time [24,30,68]. On the other hand, the evolution time of imaginary-time evolution is bounded by the inverse energy gap of the problem Hamiltonian [41]. The inverse of the energy gap typically remains constant with system size for many noncritical physical systems, such as the classical Ising chain [14]. Combinatorial optimization problems, including the MaxCut problem, are often modeled using the classical Ising model. This constitutes the motivation to build a variational *Ansatz* based on imaginary-time evolution.

In this work we introduced a variational *Ansatz*, the imaginary Hamiltonian variational *Ansatz*, to solve the combinatorial optimization MaxCut problem. The construction of the *i*HVA leverages the bit-flip and time-reversal symmetries of the MaxCut Hamiltonian and the imaginary-time evolution. Although the *i*HVA is built on the principles of imaginary-time evolution, we do not perform imaginary-time evolution on quantum circuits, which is distinguished from previous works of the QITE algorithm [36–38,42].

The *i*HVA for the MaxCut problem arranges ZY gates utilizing notions of graph theory. We proposed the tree arrangement of ZY gates based on the graph's spanning tree and showed that the MaxCut of arbitrary tree graphs can be achieved exactly by the tree arrangement. Generalizing the tree arrangement of tree graphs onto arbitrary graphs leads to the *i*HVA tree. We numerically showed the advantage of the *i*HVA tree over the QAOA in solving the MaxCut problem. The performance of the constant-round *i*HVA tree is better than the classical Goemans-Williamson algorithm in solving the MaxCut problem on D-regular graphs with up to 24 nodes. Additionally, we performed demonstrations using real quantum hardware on a 67-node graph with heavy-hex connectivity, further demonstrating the advantage of the *i*HVA tree over the QAOA on the large-scale problem. These results validate our ideas of constructing variational *Ansätze* according to imaginary-time evolution and the oriented spanning tree or, more generally, a directed graph [55]. The idea of arranging quantum gates using directed graphs can be adapted to other variational *Ansätze* such as the Hamiltonian variational *Ansatz* (HVA) and the unitary coupled-cluster *Ansatz* [9].

Theoretically, the performance guarantees of the *i*HVA can be derived similarly to the ones given in the QAOA [12,17–19]. The performance guarantees in the QAOA are derived based on the locality of the *Ansatz*. However, the *i*HVA tree proposed in this work is highly nonlocal, so the method used in the QAOA cannot be adapted to the *i*HVA tree directly. There is no performance guarantee proposed in the literature for nonlocal variational *Ansätze*. While our numerical results demonstrate the benefits of the proposed *i*HVA regarding local minima, saddle points, and the necessary number of rounds, we leave the study of theoretical performance guarantees to future work.

We showed that the constant-round *i*HVA tree of regular graphs does not exhibit BPs. This result has many implications. First, this allows the *i*HVA tree to outperform the QAOA on MaxCut problems with regular graphs. Second, this opens up the question of classical simulability of the constant-round *i*HVA tree based on the recent conjecture by Cerezo *et al.* [69]. Based on this conjecture, it may be possible that the constant-round *i*HVA tree is classically simulable. For instance, the recently introduced $\mathfrak{g}$-sim method [70] could be a potential method to simulate constant-round *i*HVA-tree circuits. Such a result would imply the classical easiness of solving the MaxCut problem on D-regular graphs. Moreover, this would make the study of the *i*HVA tree on Erdős-Rényi graphs more valuable, since we demonstrate that the constant-round *i*HVA tree exhibits BPs. In this case, warm start methods such as the one proposed by Chai *et al.* [71] could be used to support the optimization of the *i*HVA.

Finally, the *i*HVA can be constructed for a broader range of quantum systems, such as particle-number-preserving chemical and condensed-matter models, and lattice gauge theories preserving local gauge symmetries. These systems also preserve time-reversal symmetries. Thus, their *i*HVAs are distinguished from the commonly employed HVA. One may observe the advantage of the *i*HVA over the HVA in the ground-state preparation of these quantum systems. Similarly, Pelofske *et al.* studied the performance of the QAOA on higher-order Ising models [24]. In their results, it can be seen that even short-depth QAOA circuits lead to saddle points and local minima. Employing the *i*HVA on higher-order problems has the potential to simplify the optimization processes.

ACKNOWLEDGMENTS

We thank Maria Demidik, Manuel S. Rudolph, and Xiao Yuan for helpful discussions. X.W. and X.F. were supported in part by NSFC of China under Grants No. 12125501, No. 12070131001, and No. 12141501 and National Key Research and Development Program of China under Grant No. 2020YFA0406400. C.T. was supported in part by the Helmholtz Association through Innopool Project Variational Quantum Computer Simulations. This work was supported by the RIKEN TRIP initiative (RIKEN Quantum) and funds from the Ministry of Science, Research and Culture of the State of Brandenburg within the Centre for Quantum Technologies and Applications. This work was funded by the European Union's Horizon Europe Framework Programme under the ERA Chair scheme with Grant Agreement No. 101087126. The views expressed here are those of the authors and do not reflect the official policy or position of IBM or the IBM Quantum team.

APPENDIX A: PROOF OF THEOREM 1

The MaxCut problem of tree graphs can be exactly solved by one round of the *i*HVA tree, as guaranteed by Theorem 1. Now we present the proof.

Proof. Theorem 1 can be proved by mathematical induction. Our goal is to prepare the ground state of H_{MC} from the initial state $|+\rangle^{\otimes N}$. In the ground state, for each pair of parent and child nodes of the tree, as defined in Fig. 2(a), their 0 and 1 states are opposite. The first step of preparing this ground state is choosing an arbitrary node $p_0 \in \mathcal{V}$ as the tree's root. Then we choose an arbitrary child node c_0 of the root to implement a ZY exponential with the parameter $\theta = \pi/2$,

$$e^{-i(\pi/4)Z_{p_0}Y_{c_0}}|{+}{+}\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle), \tag{A1}$$

which leads to opposite 0 and 1 states between p_0 and c_0. For each pair of a parent node p and a child node c, the induction hypothesis is that the parent node has been rotated into a component of the GHZ-type state by the upstream ZY gates of the parent node p, and the child node remains as the initial state

$$|\phi\rangle \equiv \frac{1}{\sqrt{2}}(|s\rangle|0\rangle_p + |\bar{s}\rangle|1\rangle_p)|+\rangle_c, \tag{A2}$$

where s is a 0 or 1 bit string of the upstream qubits of p and $\bar{s}$ is its bitwise inverse. The downstream qubits of c are in the $|+\rangle$ state and are omitted. Implementing a ZY exponential on this state leads to

$$e^{-i(\pi/4)Z_pY_c}|\phi\rangle = \frac{1}{\sqrt{2}}(|s\rangle|0\rangle_p|1\rangle_c + |\bar{s}\rangle|1\rangle_p|0\rangle_c). \tag{A3}$$

An example of this implementation is illustrated in Fig. 10, where the ZY exponential $e^{-i(\pi/4)Z_pY_c}$ rotates the state $|\phi\rangle = \frac{1}{\sqrt{2}}(|10\rangle_{23}|1\rangle_p + |01\rangle_{23}|0\rangle_p)|+\rangle_c|{+}{+}\rangle_{01}$ to $\frac{1}{\sqrt{2}}(|10\rangle_{23}|1\rangle_p|0\rangle_c + |01\rangle_{23}|0\rangle_p|1\rangle_c)|{+}{+}\rangle_{01}$. Since the resulting state in Eq. (A3) is still a GHZ-type state and the 0 and 1 states between p and c are opposite, by induction, ZY gates following the tree arrangement can generate the ground state of H_{MC} of the tree. Thus, we prove that the targeting ground state can be prepared using the one-round *i*HVA following the tree arrangement. ■

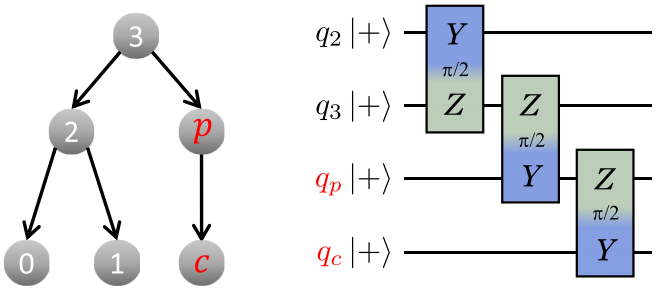

FIG. 10. Illustration for the proof of Theorem 1. The rightmost ZY exponential $e^{-i(\pi/4)Z_pY_c}$ on the parent node p and the child node c rotates the state $|\phi\rangle = \frac{1}{\sqrt{2}}(|10\rangle_{23}|1\rangle_p + |01\rangle_{23}|0\rangle_p)|+\rangle_c|{+}{+}\rangle_{01}$ to $\frac{1}{\sqrt{2}}(|10\rangle_{23}|1\rangle_p|0\rangle_c + |01\rangle_{23}|0\rangle_p|1\rangle_c)|{+}{+}\rangle_{01}$. This circuit is a subcircuit of the one shown in Fig. 3(a).

A natural corollary of Theorem 1 is that arbitrary bipartite graphs can be cut exactly using the one-round *i*HVA tree. The MaxCut of any bipartite graphs can be obtained by cutting its arbitrary spanning trees, as shown in Fig. 11. To cut the spanning tree, we set the parameters in the one-round *i*HVA tree $\theta_{1,ij} = \pi/2$ for edges (i, j) in the spanning tree (black line) and $\theta_{1,i'j'} = 0$ for edges (i', j') in the rest of the graph (gray line). In this example, we see that setting all parameters free in the *i*HVA tree helps to improve the solution accuracy of the *i*HVA tree. On the other hand, if we sets all parameters equal, the bipartite graph cannot be cut exactly using the one-round *i*HVA tree. This is one of the reasons we choose all free parameters in the construction of the *i*HVA.

APPENDIX B: CIRCUIT DEPTH OF THE *i*HVA TREE

The arrangement of ZY gates in the *i*HVA tree leads to the depth of one round $U^{(l)}_{ZY\,(YZ)}$ growing with the number of graph nodes. In this Appendix we provide bounds on the depth of the *i*HVA tree of D-regular graphs.

We first consider 2-regular graphs, which are rings; their *i*HVA tree are ladder arranged in one round and the depth grows linearly with the graph nodes N. For a D-regular graph G with $D > 2$, the number of edges is $ND/2$. The depth of one round of the *i*HVA tree cannot be larger than the number of edges. Thus we derive an upper bound on the depth of the p-round *i*HVA tree

$$d_p = O(pN). \tag{B1}$$

On the other hand, the depth of the one-round *i*HVA tree is lower bounded by the height of the graph's spanning tree [see the definition of tree height in Fig. 2(b)]. Because one descendant edge of a node in the spanning tree increases the circuit depth by one ZY gate depth, the spanning tree of a D-regular graph is a $(D-1)$-ary tree, which means that each node has at most $D-1$ child nodes. Fixing the tree height h, a $(D-1)$-ary tree has the maximum number of nodes if every node has $D-1$ child nodes. Thus, we have

$$N \leqslant \sum_{i=0}^{h}(D-1)^i = \frac{(D-1)^{h+1}-1}{D-2}. \tag{B2}$$

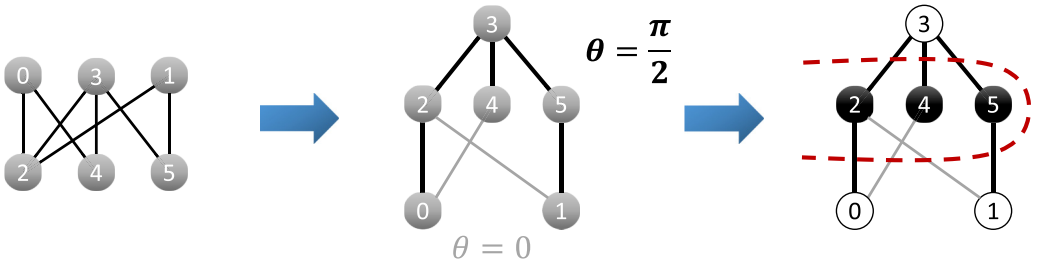

FIG. 11. Illustration showing that the one-round *i*HVA tree can exactly cut all bipartite graphs. The bipartite graph in the left panel can be exactly cut by cutting its spanning tree (black lines in the middle panel) and the other edges are cut automatically due to the bipartition of the graph (see the right panel). The spanning tree can be cut exactly using the one-round *i*HVA tree, where the tree edges (black lines in the middle panel) have the parameter $\theta = \pi/2$ and the rest edges (gray lines in the middle panel) have the parameter $\theta = 0$.

As $D > 2$, the height of the spanning tree is lower bounded by

$$h \geqslant \frac{\ln[N(D-2)+1]}{\ln(D-1)} - 1. \tag{B3}$$

Thus the depth of the p-round *i*HVA tree is lower bounded by

$$d_p = \Omega(p \ln N). \tag{B4}$$

To verify the derived bounds, we plot the depth of the one-round *i*HVA tree as a function of N, shown in Fig. 12. Here we generate 200 random D-regular graphs and count their *i*HVA-tree depth by the depth of the ZY gate. The figure plots the average depth, and the colored bands denote the maximum and minimum depths among the 200 graphs. We see that for $D \in \{3, 4, 5\}$, the depth grows sublinearly with the number of nodes for the randomly generated graphs. This behavior is consistent with the upper and lower bounds given by Eqs. (B1) and (B4), respectively.

APPENDIX C: *i*HVA STAGGER *ANSATZ*

In the main text we introduced the *i*HVA tree for the Max-Cut problem, where the arrangement of ZY gates in the *i*HVA is provided explicitly. In this Appendix we provide another arrangement of the *i*HVA. This arrangement is inspired by the shallowest arrangement proposed for the QAOA *Ansatz*

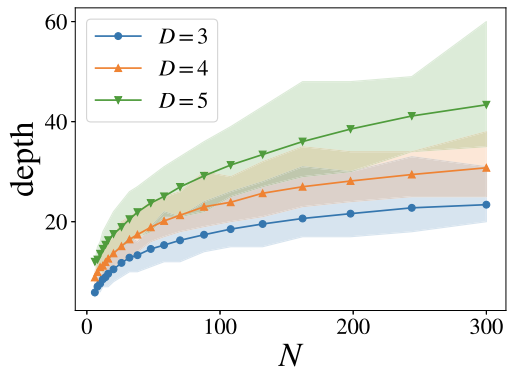

FIG. 12. Depth of one round of the *i*HVA tree on D-regular graphs as a function of the graph nodes. Results for $D \in \{3, 4, 5\}$ are marked by blue circles, orange up triangles, and green down triangles, respectively. Each data point is the average depth and the colored band denotes the maximum and minimum depth among 200 randomly generated D-regular graphs.

[16], which makes the *i*HVA as shallow as possible. The corresponding *i*HVA with this arrangement is called the *i*HVA stagger.

Consider one round of the *i*HVA

$$U_{ZY}^{(l)} \equiv \prod_{(i,j)\in\mathcal{E}} e^{-i\theta_{l,ij} Z_i Y_j/2}, \tag{C1}$$

defined on a graph $G = (\mathcal{V}, \mathcal{E})$ with N nodes and a maximum degree D. Its *i*HVA stagger is constructed as follows. First, we make an edge coloring of G, which means each edge is assigned a color so that no two incident edges have the same color. For example, Fig. 13(a) gives an edge coloring for a 3-regular graph with six nodes. According to Vizing's theorem [72], there exists an edge coloring utilizing at most $D+1$ colors. Assume $\mathcal{E} = \mathcal{E}_1 \cup \cdots \cup \mathcal{E}_{D+1}$ is such an edge coloring. For each color $c \in \{1, \ldots, D+1\}$ we define the unitary gate

$$U_c^{(l)} = \prod_{(i,j)\in\mathcal{E}_c} e^{-i\theta_{l,ij} Z_i Y_j/2}. \tag{C2}$$

As all the edges in $\mathcal{E}_c$ are not adjacent, the order of ZY gates in this product is well defined, and all ZY gates can be realized on quantum devices in parallel. Then the one-round *i*HVA stagger is arranged by $U_{ZY}^{(l)} = U_{D+1}^{(l)} \cdots U_2^{(l)} U_1^{(l)}$. The $U_{ZY}^{(l)}$ of the 3-regular graph in Fig. 13(a) is shown in Fig. 13(b). This subcircuit has a depth of at most $D+1$. Another subcircuit $U_{YZ}^{(l)}$ is constructed by reversing the qubits of Z and Y in $U_{ZY}^{(l)}$, which also has a depth of at most $D+1$. Therefore, the *i*HVA stagger with p rounds has a depth of at most $p(D+1)$, which has no dependence on the number of nodes.

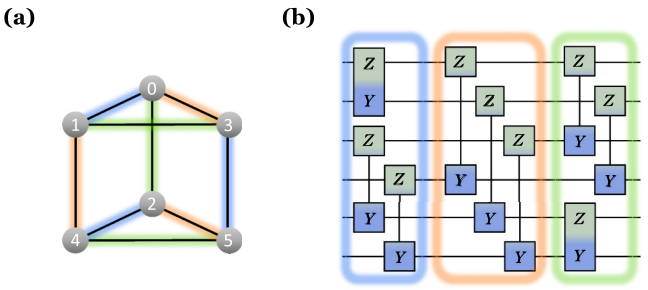

FIG. 13. (a) Edge coloring of a 3-regular graph. (b) Subcircuit $U_{ZY}^{(l)}$ following the edge coloring of the 3-regular graph. Two-qubit gates $e^{-i\theta_{l,ij} Z_i Y_j/2}$ on edges with the same color can be applied simultaneously.

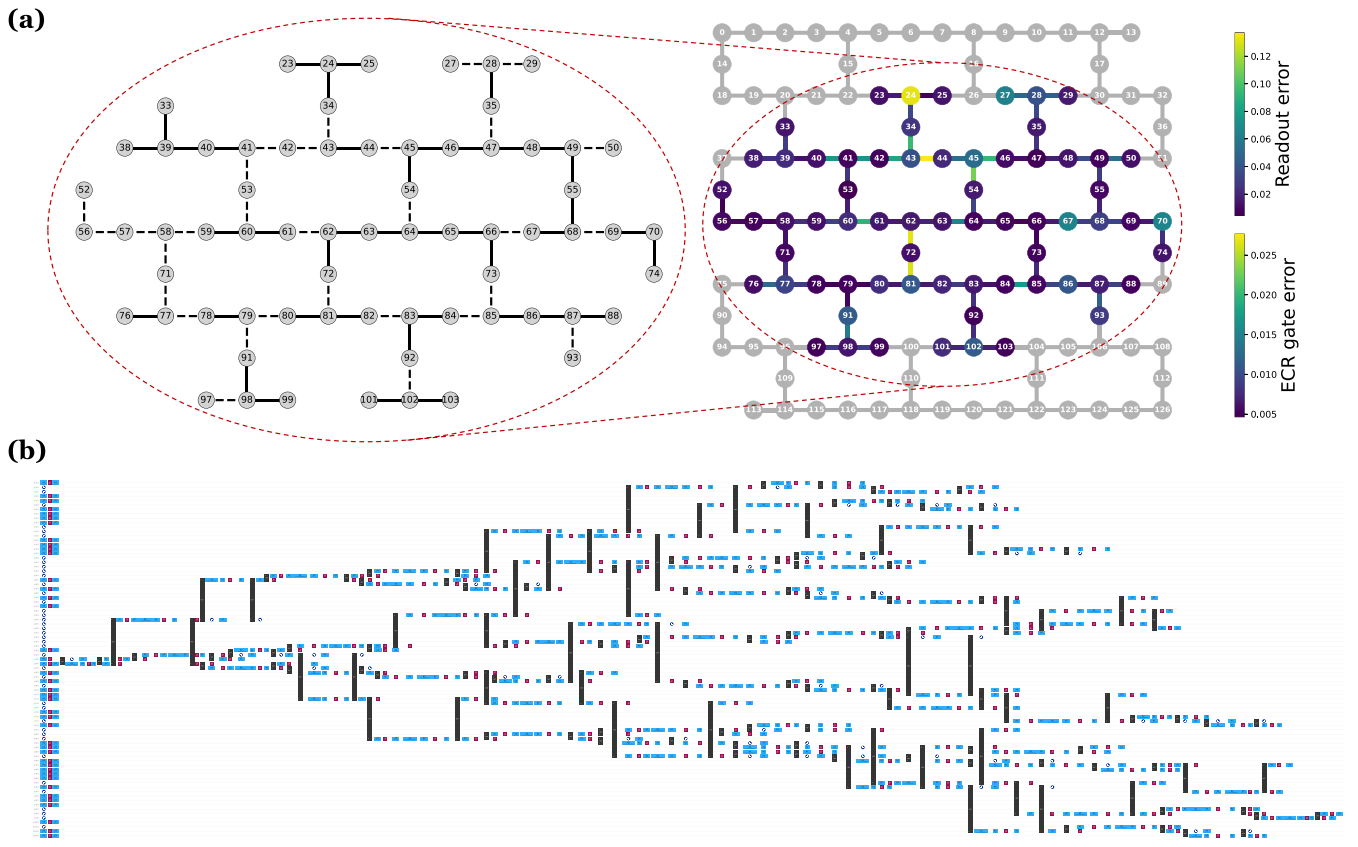

FIG. 14. (a) Heavy-hex connectivity graph with 67 nodes used in the hardware demonstration (left panel) and its layout on `ibm_brisbane`'s coupling map (right panel). A solid (dashed) line in the graph denotes a weight $w_{ij} = +1$ (-1). The coupling map of `ibm_brisbane` is colored to represent the readout error for each qubit and the two-qubit echoed cross-resonant (ECR) gate error for each qubit connection. (b) One-round *i*HVA-tree circuit executed on 67 qubits of `ibm_brisbane`. The definition of the quantum gates can be found in the IBM quantum platform [59].

For an arbitrary graph, to construct the *i*HVA stagger as shallow as possible, we need to find an edge coloring using as few colors as possible. The general problem of finding an optimal edge coloring is NP-hard. In practice, we use the greedy coloring algorithm [73] to derive an edge coloring with a few colors.

APPENDIX D: SETUP OF THE HARDWARE DEMONSTRATION

In the hardware demonstration shown in Sec. IV B, we found the maximum eigenvalue of the random weighted Hamiltonian (25), i.e., $H_w = \frac{1}{2}\sum_{(i,j)\in\mathcal{E}}(I - w_{ij}Z_iZ_j)$, where $\mathcal{E}$ is a set of edges of a heavy-hex connectivity graph $G = (\mathcal{V}, \mathcal{E})$ with 67 nodes and w_{ij} are edge weights randomly chosen as ± 1, as shown by the solid and dashed lines, respectively, in Fig. 14(a). The graph is tailored to the coupling map of the IBM Eagle-class heavy-hex devices `ibm_brisbane` [65] shown in the right panel of Fig. 14(a). The coupling map is colored to represent the readout error for each qubit and the two-qubit echoed cross-resonant gate error for each qubit connection. Other single-qubit properties of `ibm_brisbane` are summarized in Table I. All hardware data are obtained from the IBM cloud quantum platform [59] and more details are available in [74].

The one-round *i*HVA-tree circuit of the heavy-hex connectivity graph executed on `ibm_brisbane` is illustrated in Fig. 14(b). This figure is generated by the IBM cloud quantum platform [59].

APPENDIX E: PROOF OF THEOREM 2

In this Appendix we prove that the constant-round *i*HVA for the MaxCut problem is free from the barren plateau phenomenon. As explained in the main text, we aim to provide a lower bound on the variance of the Hamiltonian expectation $\mathrm{Var}(\langle \overline{H_{\mathrm{MC}}} \rangle)$. For convenience, we first consider the 2-round

TABLE I. Summary of single-qubit properties of `ibm_brisbane` on the same day the hardware demonstrations were performed.

Parameter	Median	Mean	Min	Max
Frequency (GHz)	4.91	4.90 ± 0.11	4.61	5.12
Anharmonicity (MHz)	308.42	308.66 ± 5.38	289.81	359.05
T_1 (μs)	222.18	218.62 ± 71.64	43.43	380.24
T_2 (μs)	142.66	151.25 ± 87.70	13.17	459.64

*i*HVA. Its variance is given by

$$\mathrm{Var}(\overline{\langle H_{\mathrm{MC}}\rangle}) = \mathrm{Var}_{\boldsymbol{\theta}}\left(\langle\phi_I^{(2)}(\boldsymbol{\theta})|\frac{H_{\mathrm{MC}}}{E_0}|\phi_I^{(2)}(\boldsymbol{\theta})\rangle\right) = \frac{1}{E_0^2}\mathrm{Var}_{\boldsymbol{\theta}}\left(\sum_{(i,j)\in\mathcal{E}}\langle\phi_I^{(2)}(\boldsymbol{\theta})|Z_iZ_j|\phi_I^{(2)}(\boldsymbol{\theta})\rangle\right), \tag{E1}$$

where

$$|\phi_I^{(2)}(\boldsymbol{\theta})\rangle = U_{YZ}^{(2)}U_{ZY}^{(1)}|+\rangle^{\otimes N}, \quad U_{YZ}^{(2)} = \prod_{(i,j)\in\mathcal{E}} e^{-i\theta_{2,ij}Y_iZ_j/2}, \quad U_{ZY}^{(1)} = \prod_{(i,j)\in\mathcal{E}} e^{-i\theta_{1,ij}Z_iY_j/2}. \tag{E2}$$

Then we have the following lemma.

Lemma 1. For the 2-round *i*HVA in Eq. (E2) solving the MaxCut problem on the D-regular graph with N nodes, the variance of the energy expectation is lower bounded by

$$\mathrm{Var}(\overline{\langle H_{\mathrm{MC}}\rangle_2}) \geqslant \frac{DN}{E_0^2 2^{3D-1}}. \tag{E3}$$

Proof. First, we show that the mean of the Hamiltonian expectation is zero. The mean of the Hamiltonian expectation reads

$$\mathbb{E}_{\boldsymbol{\theta}}\left(\sum_{(i,j)\in\mathcal{E}}\langle\phi_I^{(2)}(\boldsymbol{\theta})|Z_iZ_j|\phi_I^{(2)}(\boldsymbol{\theta})\rangle\right) = \sum_{(i,j)\in\mathcal{E}}\int \mathcal{D}\boldsymbol{\theta}\langle\phi_I^{(2)}(\boldsymbol{\theta})|Z_iZ_j|\phi_I^{(2)}(\boldsymbol{\theta})\rangle, \tag{E4}$$

where $\mathcal{D}\boldsymbol{\theta} \equiv \prod_{(i,j)\in\mathcal{E}}(\frac{\mathrm{d}\theta_{1,ij}}{2\pi})(\frac{\mathrm{d}\theta_{2,ij}}{2\pi})$ is the measure over the circuit parameters. Viewing from the Heisenberg picture, the expectation $\langle\phi_I^{(2)}(\boldsymbol{\theta})|Z_iZ_j|\phi_I^{(2)}(\boldsymbol{\theta})\rangle$ can be derived by sequentially conjugating ZY gates on Z_iZ_j, for example,

$$e^{i\theta Y_iZ_k/2}Z_iZ_je^{-i\theta Y_iZ_k/2} = \cos\theta Z_iZ_j - \sin\theta X_iZ_jZ_k. \tag{E5}$$

Thus, new Pauli strings such as $X_iZ_jZ_k$ appear in the expression. After conjugating all the ZY gates, it remains a linear combination of Pauli strings. The only Pauli strings contributing to the expectation are those consisting only of Pauli-X letters because

$$\langle+|^{\otimes N}\sigma|+\rangle^{\otimes N} = \begin{cases}1 & \text{for } \sigma \in \{I, X_0, X_1, X_0X_1, \ldots, X_0X_1\ldots X_{N-1}\}\\ 0 & \text{otherwise.}\end{cases} \tag{E6}$$

Meanwhile, as each of the ZY gates in $|\phi_I^{(2)}(\boldsymbol{\theta})\rangle$ have a free angle, it can be seen that the coefficient of each Pauli string in the linear combination must be a product of cos or sin functions and each $\cos\theta_{1\,(2),ij}$ or $\sin\theta_{1\,(2),ij}$ appears at most once. After the integration $\int\mathcal{D}\boldsymbol{\theta}$, since

$$\int_0^{2\pi}\frac{\mathrm{d}\theta}{2\pi}\cos\theta = \int_0^{2\pi}\frac{\mathrm{d}\theta}{2\pi}\sin\theta = 0, \tag{E7}$$

each coefficient in the linear combination vanishes. Thus, the integration $\int\mathcal{D}\boldsymbol{\theta}\langle\phi_I^{(2)}(\boldsymbol{\theta})|Z_iZ_j|\phi_I^{(2)}(\boldsymbol{\theta})\rangle$ and the mean of the Hamiltonian expectation in Eq. (E4) vanish consequently.

To simplify the calculation of the variance, we assume that all the expectations of Z_iZ_j are mutually independent, i.e.,

$$\begin{aligned}&\mathbb{E}_{\boldsymbol{\theta}}\left(\langle\phi_I^{(2)}(\boldsymbol{\theta})|Z_iZ_j|\phi_I^{(2)}(\boldsymbol{\theta})\rangle\langle\phi_I^{(2)}(\boldsymbol{\theta})|Z_kZ_l|\phi_I^{(2)}(\boldsymbol{\theta})\rangle\right)\\ &= \mathbb{E}_{\boldsymbol{\theta}}\left(\langle\phi_I^{(2)}(\boldsymbol{\theta})|Z_iZ_j|\phi_I^{(2)}(\boldsymbol{\theta})\rangle\right)\mathbb{E}_{\boldsymbol{\theta}}\left(\langle\phi_I^{(2)}(\boldsymbol{\theta})|Z_kZ_l|\phi_I^{(2)}(\boldsymbol{\theta})\rangle\right)\end{aligned} \tag{E8}$$

in the case $(i,j)\neq(k,l)$. This assumption is examined numerically as shown in Fig. 15. With this assumption, the variance and summation in Eq. (E1) can be exchanged, i.e.,

$$\mathrm{Var}(\overline{\langle H_{\mathrm{MC}}\rangle}) = \frac{1}{E_0^2}\sum_{(i,j)\in\mathcal{E}}\mathrm{Var}_{\boldsymbol{\theta}}\left[\langle\phi_I^{(2)}(\boldsymbol{\theta})|Z_iZ_j|\phi_I^{(2)}(\boldsymbol{\theta})\rangle\right]. \tag{E9}$$

Next we provide a rigorous lower bound on the variance $\mathrm{Var}_{\boldsymbol{\theta}}[\langle\phi_I^{(2)}(\boldsymbol{\theta})|Z_iZ_j|\phi_I^{(2)}(\boldsymbol{\theta})\rangle]$. When we calculate the expectation $\langle\phi_I^{(2)}(\boldsymbol{\theta})|Z_iZ_j|\phi_I^{(2)}(\boldsymbol{\theta})\rangle$, a general structure of the two-round *i*HVA is shown in Fig. 16(a). In the middle of the circuit, there exists an $e^{-i\theta_1Y_iZ_j/2}$ and an $e^{-i\theta_2Z_iY_j/2}$, shown as the colored gates in the figure. As our conclusion has no concern about which qubit is i or j, the order of these two gates is irrelevant. Other ZY gates do not connect qubit i and

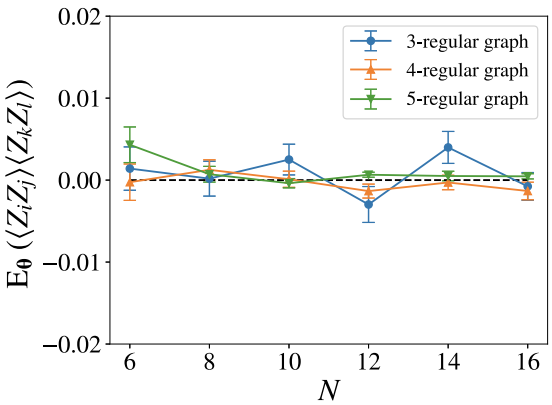

FIG. 15. Numerical verification of the assumption in Eq. (E8) that the expectations of observables Z_iZ_j and Z_kZ_l with the *i*HVA are mutually independent. Here a D-regular graph with N nodes and the edges $(i,j),(k,l)\in\mathcal{E}$ are randomly chosen and the variational parameters $\boldsymbol{\theta}$ are uniformly sampled 2048 times for each data point. The covariance $\mathbb{E}_{\boldsymbol{\theta}}(\langle\phi_I^{(2)}(\boldsymbol{\theta})|Z_iZ_j|\phi_I^{(2)}(\boldsymbol{\theta})\rangle\langle\phi_I^{(2)}(\boldsymbol{\theta})|Z_kZ_l|\phi_I^{(2)}(\boldsymbol{\theta})\rangle)$ vanishes for arbitrary D-regular graphs with N nodes within the error of statistics. So the left-hand side of Eq. (E8) is zero with high probability, equal to the theoretical value of the right-hand side of Eq. (E8).

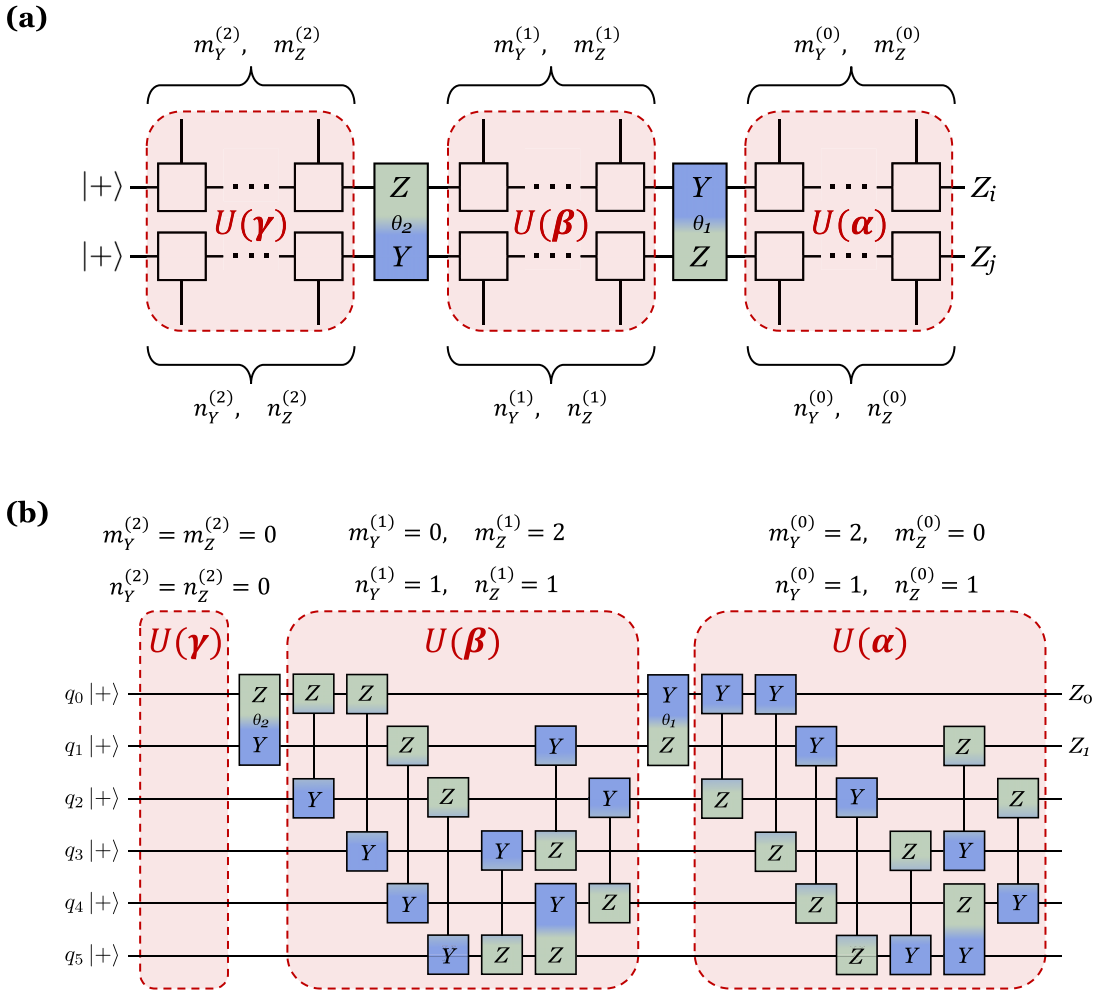

FIG. 16. (a) General structure of ZY gates in the two-round iHVA. Here $U(\boldsymbol{\alpha})$, $U(\boldsymbol{\beta})$, and $U(\boldsymbol{\gamma})$ are gate blocks of ZY gates and $m_Y^{(i)}$, $m_Z^{(i)}$, $n_Y^{(i)}$, and $n_Z^{(i)}$ ($i \in \{0, 1, 2\}$) are the numbers of gates. (b) Concrete example of the gate blocks $U(\boldsymbol{\alpha})$, $U(\boldsymbol{\beta})$, and $U(\boldsymbol{\gamma})$ and the numbers of gates.

j and thus are oriented outside the figure. The expectation of Z_iZ_j reads

$$\begin{aligned}&\langle\phi_I^{(2)}(\boldsymbol{\theta})|Z_iZ_j|\phi_I^{(2)}(\boldsymbol{\theta})\rangle\\&\quad=\langle+|^{\otimes N}U^\dagger(\boldsymbol{\gamma})e^{i\theta_2 Z_iY_j/2}U^\dagger(\boldsymbol{\beta})e^{i\theta_1 Y_iZ_j/2}U^\dagger(\boldsymbol{\alpha})Z_iZ_jU(\boldsymbol{\alpha})\\&\qquad\times e^{-i\theta_1 Y_iZ_j/2}U(\boldsymbol{\beta})e^{-i\theta_2 Z_iY_j/2}U(\boldsymbol{\gamma})|+\rangle^{\otimes N},\end{aligned}\tag{E10}$$

where the ZY gate blocks $U(\boldsymbol{\alpha})$, $U(\boldsymbol{\beta})$, and $U(\boldsymbol{\gamma})$ indicating ZY gates with the parameters $\{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}\} = \{\boldsymbol{\theta}\}$, as shown in Fig. 16(a). In the figure $m_Y^{(i)}$, $m_Z^{(i)}$, $n_Y^{(i)}$, and $n_Z^{(i)}$ ($i \in \{0, 1, 2\}$) denote the number of gates. For example, $m_Z^{(0)}$ denotes there are $m_Z^{(0)}$ gates in $U(\boldsymbol{\alpha})$ connecting qubit i and other qubits like $e^{-i\theta Z_iY_k/2}$ with $k \in \mathcal{V}/\{i, j\}$. Figure 16(b) provides an explicit example of the concrete components of $U(\boldsymbol{\alpha})$, $U(\boldsymbol{\beta})$, and $U(\boldsymbol{\gamma})$. The corresponding values of $m_Y^{(i)}$, $m_Z^{(i)}$, $n_Y^{(i)}$, and $n_Z^{(i)}$ are presented explicitly. These 12 integers are not independent. Since the iHVA is constructed for the D-regular graph, the total number of gates connecting qubit i and other qubits is $2D$ (for two rounds), where D gates are like $e^{-i\theta Z_iY_k/2}$ and the other D gates are like $e^{-i\theta Y_iZ_k/2}$. Thus, the 12 gate numbers satisfy the relations

$$\sum_{i=0}^{2} m_Y^{(i)} = \sum_{i=0}^{2} m_Z^{(i)} = \sum_{i=0}^{2} n_Y^{(i)} = \sum_{i=0}^{2} n_Z^{(i)} = D - 1. \tag{E11}$$

With the above notation, we first calculate how the rightmost gate block $U(\boldsymbol{\alpha})$ in Fig. 16(a) conjugates on the observable Z_iZ_j under Heisenberg picture. An illustration of two gates conjugating on Z_iZ_j is shown in Fig. 17, where we utilize the relation

$$e^{i\alpha Y_iZ_k/2}Z_ie^{-i\alpha Y_iZ_k/2} = \cos\alpha Z_i - \sin\alpha X_iZ_k. \tag{E12}$$

It results in a series of Pauli strings with coefficients $\cos\alpha_1\cos\alpha_2$, $-\cos\alpha_1\sin\alpha_2$, $-\sin\alpha_1\cos\alpha_2$, and $\sin\alpha_1\sin\alpha_2$. This result can be generalized for $U^\dagger(\boldsymbol{\alpha})Z_iZ_jU(\boldsymbol{\alpha})$. We define $l_Y = m_Y^{(0)} + n_Y^{(0)}$, which is the number of gates $e^{-i\alpha Y_iZ_k/2}$ and $e^{-i\alpha Y_jZ_k/2}$ not commuted with Z_iZ_j in $U(\boldsymbol{\alpha})$; the corresponding angles on these l_Y gates are denoted by $\alpha_1, \ldots, \alpha_{l_Y}$. The observable conjugated by $U(\boldsymbol{\alpha})$ is a periodic function of $\boldsymbol{\alpha}$, which can be expanded as a high-dimensional Fourier series. The result is a linear

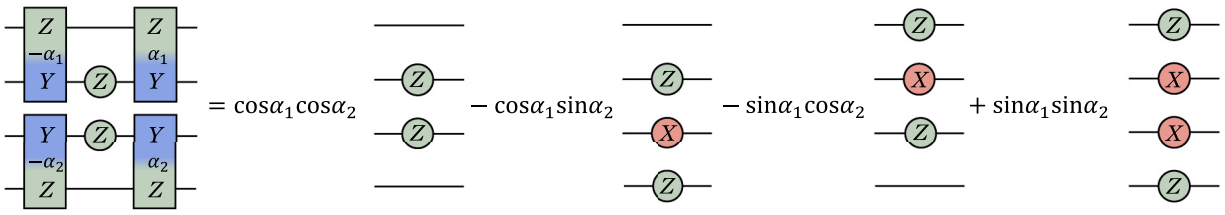

FIG. 17. Schematic calculation of the observable Z_iZ_j conjugated by two ZY gates. Each circle denotes a single-qubit Pauli operator and each rectangle denotes a matrix $e^{-i\alpha ZY/2}$ with the value of α shown at the center of the rectangle.

combination of 3^{l_Y} subterms [75]

$$U^\dagger(\boldsymbol{\alpha})Z_iZ_jU(\boldsymbol{\alpha}) = \sum_{\boldsymbol{\xi}\in\{0,1,2\}^{l_Y}} \Phi_{\boldsymbol{\xi}}(Z_iZ_j) \prod_{l:\xi_l=0} \cos\alpha_l \prod_{l:\xi_l=1} \sin\alpha_l \times \prod_{l:\xi_l=2} 1, \tag{E13}$$

where ξ_l is the lth component of the vector $\boldsymbol{\xi}\in\{0,1,2\}^{l_Y}$ and

$$\Phi_{\boldsymbol{\xi}}(Z_iZ_j) \equiv \sum_{\sigma_{\mathbf{i}}\in\mathbf{P}^N} d_{\mathbf{i}}^{(\boldsymbol{\xi})}\sigma_{\mathbf{i}} \tag{E14}$$

is a linear combination of all possible Pauli strings $\sigma_{\mathbf{i}}\in\mathbf{P}^N = \{I,X,Y,Z\}^N$ on N qubits. Here $d_{\mathbf{i}}^{(\boldsymbol{\xi})}$ are some real coefficients that have no dependence on $\alpha_1,\alpha_2,\ldots,\alpha_{l_Y}$. The $\boldsymbol{\xi}=\mathbf{0}$ term with coefficients $\prod_{l=1}^{l_Y}\cos\alpha_l$ is explicitly known, i.e.,

$$\Phi_{\mathbf{0}}(Z_iZ_j) = Z_iZ_j, \tag{E15}$$

This is similar to the results in Fig. 17. Then the whole expectation can be derived due to its linearity

$$\langle\phi_I(\boldsymbol{\theta})|Z_iZ_j|\phi_I(\boldsymbol{\theta})\rangle = \sum_{\boldsymbol{\xi}\in\{0,1,2\}^{l_Y}} a_{\boldsymbol{\xi}} \prod_{l:\xi_l=0}\cos\alpha_l \prod_{l:\xi_l=1}\sin\alpha_l \prod_{l:\xi_l=2} 1, \tag{E16}$$

where the real coefficients $a_{\boldsymbol{\xi}}$ is given by

$$a_{\boldsymbol{\xi}} \equiv \langle+|^{\otimes N}U^\dagger(\boldsymbol{\gamma})e^{i\theta_2Z_iY_j/2}U^\dagger(\boldsymbol{\beta})e^{i\theta_1Y_iZ_j/2}\Phi_{\boldsymbol{\xi}}(Z_iZ_j) \times e^{-i\theta_1Y_iZ_j/2}U(\boldsymbol{\beta})e^{-i\theta_2Z_iY_j/2}U(\boldsymbol{\gamma})|+\rangle^{\otimes N}. \tag{E17}$$

As the mean of the expectation vanishes, as proved previously, the variance is the integration of the square of the expectation, which reads

$$\mathrm{Var}_{\boldsymbol{\theta}}\left[\left|\phi_I^{(2)}(\boldsymbol{\theta})\middle|Z_iZ_j\middle|\phi_I^{(2)}(\boldsymbol{\theta})\right|\right] = \int \mathcal{D}\boldsymbol{\theta}\left|\phi_I^{(2)}(\boldsymbol{\theta})\middle|Z_iZ_j\middle|\phi_I^{(2)}(\boldsymbol{\theta})\right|^2. \tag{E18}$$

Its integrand given in Eq. (E16) is a Fourier series of $\boldsymbol{\alpha}$, so the variance can be evaluated using Parseval's equation, which states that if a periodic function can be expanded using the Fourier series

$$f(\alpha) = \frac{a_0}{2} + \sum_{n=1}^{\infty}(a_n\cos n\alpha + b_n\sin n\alpha), \tag{E19}$$

then the integration of the square of $f(\alpha)$ can be derived by

$$\int_0^{2\pi}\frac{\mathrm{d}\alpha}{2\pi}f(\alpha)^2 = \left(\frac{a_0}{2}\right)^2 + \frac{1}{2}\sum_{n=1}^{\infty}\left(a_n^2+b_n^2\right), \tag{E20}$$

due to the orthogonality of the Fourier basis. Generalizing this equation to higher dimensions and integrating out $\boldsymbol{\alpha}$ in Eq. (E18), the variance reads

$$\begin{aligned}&\mathrm{Var}_{\boldsymbol{\theta}}[\langle\phi_I(\boldsymbol{\theta})|Z_iZ_j|\phi_I(\boldsymbol{\theta})\rangle]\\ &= \int\mathcal{D}\boldsymbol{\theta}/\{\boldsymbol{\alpha}\}\sum_{\boldsymbol{\xi}\in\{0,1,2\}^{l_Y}}\frac{a_{\boldsymbol{\xi}}^2}{2^{(\text{No. of 0's and 1's in }\boldsymbol{\xi})}}\\ &\geqslant \frac{1}{2^{l_Y}}\int\mathcal{D}\boldsymbol{\theta}/\{\boldsymbol{\alpha}\}a_{\mathbf{0}}^2.\end{aligned} \tag{E21}$$

In the first line, the denominator appears because $\int_0^{2\pi}\frac{\mathrm{d}\theta}{2\pi}\cos^2\theta = \int_0^{2\pi}\frac{\mathrm{d}\theta}{2\pi}\sin^2\theta = \frac{1}{2}$. Each 0 and 1 in $\boldsymbol{\xi}$ corresponds to a cos and a sin function and leads to one $\frac{1}{2}$. The integration measure $\mathcal{D}\boldsymbol{\theta}/\{\boldsymbol{\alpha}\}$ means that the parameters in $U(\boldsymbol{\alpha})$ are integrated out. In the second line, we only retain the term $\boldsymbol{\xi}=\mathbf{0}$, whose number of 0's and 1's is l_Y. So the factor $1/2^{l_Y}$ appears.

Then we provide a lower bound of $\int\mathcal{D}\boldsymbol{\theta}/\{\boldsymbol{\alpha}\}a_{\mathbf{0}}^2$ using Parseval's equation again. The expression of $a_{\mathbf{0}}$ can be calculated explicitly,

$$\begin{aligned}a_{\mathbf{0}} &= \langle+|^{\otimes N}U^\dagger(\boldsymbol{\gamma})e^{i\theta_2Z_iY_j/2}U^\dagger(\boldsymbol{\beta})e^{i\theta_1Y_iZ_j/2}Z_iZ_j\\ &\quad\times e^{-i\theta_1Y_iZ_j/2}U(\boldsymbol{\beta})e^{-i\theta_2Z_iY_j/2}U(\boldsymbol{\gamma})|+\rangle^{\otimes N}\\ &= \cos\theta_1\langle+|^{\otimes N}U^\dagger(\boldsymbol{\gamma})e^{i\theta_2Z_iY_j/2}U^\dagger(\boldsymbol{\beta})Z_iZ_jU(\boldsymbol{\beta})\\ &\quad\times e^{-i\theta_2Z_iY_j/2}U(\boldsymbol{\gamma})|+\rangle^{\otimes N}\\ &\quad-\sin\theta_1\langle+|^{\otimes N}U^\dagger(\boldsymbol{\gamma})e^{i\theta_2Z_iY_j/2}U^\dagger(\boldsymbol{\beta})X_iU(\boldsymbol{\beta})\\ &\quad\times e^{-i\theta_2Z_iY_j/2}U(\boldsymbol{\gamma})|+\rangle^{\otimes N}.\end{aligned} \tag{E22}$$

These two terms have a similar structure, as we have seen in Eq. (E10), and we can repeat the procedure from Eqs. (E13)–(E21) for both terms.

Things can be simplified by observing how the factor $1/2^{l_Y}$ appears in Eq. (E21). The exponent l_Y is the number of ZY gates not commuted with the observable Z_iZ_j. In gate block $U(\boldsymbol{\beta})$, the number of gates not commuted with Z_iZ_j and X_i is $m_Y^{(1)}+n_Y^{(1)}$ and $m_Y^{(1)}+m_Z^{(1)}$, respectively. Thus, we arrive at the lower bound

$$\int\mathcal{D}\boldsymbol{\theta}/\{\boldsymbol{\alpha}\}a_{\mathbf{0}}^2 \geqslant \frac{1}{2^{m_Y^{(1)}+n_Y^{(1)}+1}}\int\mathcal{D}\boldsymbol{\theta}/\{\boldsymbol{\alpha},\boldsymbol{\beta}\}b_{\mathbf{0}}^2 + \frac{1}{2^{m_Y^{(1)}+m_Z^{(1)}+1}}\int\mathcal{D}\boldsymbol{\theta}/\{\boldsymbol{\alpha},\boldsymbol{\beta}\}b_{\mathbf{0}}'^2, \tag{E23}$$

where

$$\begin{aligned}b_{\mathbf{0}} &= \langle+|^{\otimes N}U^\dagger(\boldsymbol{\gamma})e^{i\theta_2Z_iY_j/2}Z_iZ_je^{-i\theta_2Z_iY_j/2}U(\boldsymbol{\gamma})|+\rangle^{\otimes N},\\ b_{\mathbf{0}}' &= \langle+|^{\otimes N}U^\dagger(\boldsymbol{\gamma})e^{i\theta_2Z_iY_j/2}X_ie^{-i\theta_2Z_iY_j/2}U(\boldsymbol{\gamma})|+\rangle^{\otimes N}.\end{aligned} \tag{E24}$$

Repeating the above procedure and integrating the parameters in $U(\boldsymbol{\gamma})$, we derive the lower bound for each term in Eq. (E23),

$$\int \mathcal{D}\boldsymbol{\theta}/\{\boldsymbol{\alpha},\boldsymbol{\beta}\} b_{\mathbf{0}}^2 \geqslant \frac{1}{2^{n_Y^{(2)}+n_Z^{(2)}+1}},$$
$$\int \mathcal{D}\boldsymbol{\theta}/\{\boldsymbol{\alpha},\boldsymbol{\beta}\} b_{\mathbf{0}}'^2 \geqslant \frac{1}{2^{m_Y^{(2)}+m_Z^{(2)}+1}}, \tag{E25}$$

where we have used the fact that only Pauli-X strings contribute to the expectation [see Eq. (E6)].

Combining Eqs. (E18), (E23), and (E25), the variance of the expectation is lower bounded by

$$\begin{aligned}\mathrm{Var}_{\boldsymbol{\theta}}\left[\left\langle\phi_I^{(2)}(\boldsymbol{\theta})\middle|Z_iZ_j\middle|\phi_I^{(2)}(\boldsymbol{\theta})\right\rangle\right] &\geqslant \frac{1}{2^{m_Y^{(0)}+n_Y^{(0)}}}\left(\frac{1}{2^{m_Y^{(1)}+n_Y^{(1)}+1}}\frac{1}{2^{n_Y^{(2)}+n_Z^{(2)}+1}}+\frac{1}{2^{m_Y^{(1)}+m_Z^{(1)}+1}}\frac{1}{2^{m_Y^{(2)}+m_Z^{(2)}+1}}\right)\\ &= \frac{1}{2^{D+1+m_Y^{(0)}+m_Y^{(1)}+n_Z^{(2)}}}+\frac{1}{2^{D+1+n_Y^{(0)}+m_Z^{(1)}+m_Z^{(2)}}}\\ &\geqslant \frac{1}{2^D}\sqrt{\frac{1}{2^{m_Y^{(0)}+m_Y^{(1)}+n_Z^{(2)}+n_Y^{(0)}+m_Z^{(1)}+m_Z^{(2)}}}}\\ &\geqslant \frac{1}{2^{3D-2}}.\end{aligned} \tag{E26}$$

In the second line we used the constraints in Eq. (E11). The third line utilizes the basic inequality $a+b\geqslant 2\sqrt{ab}$. The fourth line is derived by maximizing the exponent $m_Y^{(0)}+m_Y^{(1)}+n_Z^{(2)}+n_Y^{(0)}+m_Z^{(1)}+m_Z^{(2)}$, i.e., choosing $m_Y^{(0)}+m_Y^{(1)}=m_Z^{(1)}+m_Z^{(2)}=n_Z^{(2)}=n_Y^{(0)}=D-1$. Substituting the last inequality in Eq. (E9), we arrive at the conclusion

$$\mathrm{Var}(\langle\overline{H_{\mathrm{MC}}}\rangle) \geqslant \frac{1}{E_0^2}\frac{DN}{2^{3D-1}}, \tag{E27}$$

where we have used the fact that the number of edges in the D-regular graph is $DN/2$. ■

Now we prove Theorem 2 from the main text. Lemma 1 can be generalized to the iHVA with even p rounds

$$\left|\phi_I^{(p)}(\boldsymbol{\theta}_p)\right\rangle \equiv U_{YZ}^{(p)}U_{ZY}^{(p-1)}\cdots U_{YZ}^{(4)}U_{ZY}^{(3)}U_{YZ}^{(2)}U_{ZY}^{(1)}|+\rangle^{\otimes N} = \mathcal{U}^{(p-2)}\left|\phi_I^{(2)}(\boldsymbol{\theta}_2)\right\rangle, \tag{E28}$$

where $\{\boldsymbol{\theta}_p\}\equiv\{\theta_{1,ij},\ldots,\theta_{p,ij}\}$ and $\{\boldsymbol{\theta}_2\}\equiv\{\theta_{1,ij},\theta_{2,ij}\}$, with $(i,j)\in\mathcal{E}$, and $\mathcal{U}^{(p-2)}\equiv U_{YZ}^{(p)}U_{ZY}^{(p-1)}\cdots U_{YZ}^{(4)}U_{ZY}^{(3)}$. The variance of the Hamiltonian expectation reads

$$\mathrm{Var}(\langle\overline{H_{\mathrm{MC}}}\rangle) = \frac{1}{E_0^2}\sum_{(i,j)\in\mathcal{E}}\mathrm{Var}_{\boldsymbol{\theta}_p}\left[\left\langle\phi_I^{(p)}(\boldsymbol{\theta}_p)\middle|Z_iZ_j\middle|\phi_I^{(p)}(\boldsymbol{\theta}_p)\right\rangle\right] = \frac{1}{E_0^2}\sum_{(i,j)\in\mathcal{E}}\mathrm{Var}_{\boldsymbol{\theta}_p}\left[\left\langle\phi_I^{(2)}(\boldsymbol{\theta}_2)\middle|\mathcal{U}^{(p-2)\dagger}Z_iZ_j\mathcal{U}^{(p-2)}\middle|\phi_I^{(2)}(\boldsymbol{\theta}_2)\right\rangle\right].$$

Consider $\mathcal{U}^{(p-2)\dagger}Z_iZ_j\mathcal{U}^{(p-2)}$ with even p. For each two-round structure $U_{YZ}^{(k+1)}U_{ZY}^{(k)}$ in $\mathcal{U}^{(p-2)}$, as shown in Fig. 16(b), there are $2D$ ZY gates not commuted with Z_iZ_j. These $2D$ ZY gates lead to a production of $2D$ cosine functions as a precoefficient of Z_iZ_j. Since $\mathcal{U}^{(p-2)}$ has $(p-2)/2$ two-round structures $U_{YZ}^{(k+1)}U_{ZY}^{(k)}$, $\mathcal{U}^{(p-2)\dagger}Z_iZ_j\mathcal{U}^{(p-2)}$ generates $(p-2)D$ cosine functions as a precoefficient of Z_iZ_j. Similar to the procedure we used in the proof of Lemma 1, integrating out the free parameters $\{\boldsymbol{\theta}_p\}/\{\boldsymbol{\theta}_2\}$ in the variance leads to

$$\mathrm{Var}_{\boldsymbol{\theta}_p}\left[\left\langle\phi_I^{(2)}(\boldsymbol{\theta}_2)\middle|\mathcal{U}^{(p-2)\dagger}Z_iZ_j\mathcal{U}^{(p-2)}\middle|\phi_I^{(2)}(\boldsymbol{\theta}_2)\right\rangle\right] \geqslant \frac{1}{2^{(p-2)D}}\mathrm{Var}_{\boldsymbol{\theta}_2}\left[\left\langle\phi_I^{(2)}(\boldsymbol{\theta}_2)\middle|Z_iZ_j\middle|\phi_I^{(2)}(\boldsymbol{\theta}_2)\right\rangle\right]. \tag{E29}$$

Combining the lower bound of $\mathrm{Var}_{\boldsymbol{\theta}_2}[\langle\phi_I^{(2)}(\boldsymbol{\theta}_2)|Z_iZ_j|\phi_I^{(2)}(\boldsymbol{\theta}_2)\rangle]$ in Lemma 1, $\mathrm{Var}(\langle\overline{H_{\mathrm{MC}}}\rangle)$ is lower bounded by

$$\mathrm{Var}(\langle\overline{H_{\mathrm{MC}}}\rangle) \geqslant \frac{1}{E_0^2}\frac{DN}{2^{3D-1}}\frac{1}{2^{(p-2)D}} = \frac{1}{E_0^2}\frac{DN}{2^{D(p+1)-1}}. \tag{E30}$$

Therefore, we prove Theorem 2, and the BP is absent for the constant-round iHVA of D-regular graphs.

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

[2] J. J. M. Kirsopp, C. Di Paola, D. Z. Manrique, M. Krompiec, G. Greene-Diniz, W. Guba, A. Meyder, D. Wolf, M. Strahm, and D. Muñoz Ramo, Quantum computational quantification of protein–ligand interactions, Int. J. Quantum Chem. **122**, e26975 (2022).

[3] Y. Cao, J. Romero, and A. Aspuru-Guzik, Potential of quantum computing for drug discovery, IBM J. Res. Dev. **62**, 6:1 (2018).

[4] N. Klco, E. F. Dumitrescu, A. J. McCaskey, T. D. Morris, R. C. Pooser, M. Sanz, E. Solano, P. Lougovski, and M. J. Savage, Quantum-classical computation of Schwinger model dynamics using quantum computers, Phys. Rev. A **98**, 032331 (2018).

[5] N. Klco, M. J. Savage, and J. R. Stryker, SU(2) non-Abelian gauge field theory in one dimension on digital quantum computers, Phys. Rev. D **101**, 074512 (2020).

[6] E. Farhi, J. Goldstone, S. Gutmann, J. Lapan, A. Lundgren, and D. Preda, A quantum adiabatic evolution algorithm applied to random instances of an NP-complete problem, Science **292**, 472 (2001).

[7] F. Gemeinhardt, A. Garmendia, M. Wimmer, B. Weder, and F. Leymann, Quantum combinatorial optimization in the NISQ era: A systematic mapping study, ACM Comput. Surv. **56**, 1 (2023).

[8] M. Cerezo, G. Verdon, H.-Y. Huang, L. Cincio, and P. J. Coles, Challenges and opportunities in quantum machine learning, Nat. Comput. Sci. **2**, 567 (2022).

[9] A. Peruzzo, J. McClean, P. Shadbolt, M. Yung, X. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, A variational eigenvalue solver on a photonic quantum processor, Nat. Commun. **5**, 4213 (2014).

[10] 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).

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

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

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

[14] E. Farhi, J. Goldstone, S. Gutmann, and M. Sipser, Quantum computation by adiabatic evolution, arXiv:quant-ph/0001106.

[15] G. E. Crooks, Performance of the quantum approximate optimization algorithm on the maximum cut problem, arXiv:1811.08419.

[16] S. Bravyi, A. Kliesch, R. Koenig, and E. Tang, Obstacles to variational quantum optimization from symmetry protection, Phys. Rev. Lett. **125**, 260505 (2020).

[17] E. Farhi, E. Gamarnik, and S. Gutmann, The quantum approximate optimization algorithm needs to see the whole graph: Worst case examples, arXiv:2005.08747.

[18] J. Wurtz and P. Love, Maxcut quantum approximate optimization algorithm performance guarantees for $p > 1$, Phys. Rev. A **103**, 042612 (2021).

[19] J. Basso, E. Farhi, K. Marwaha, B. Villalonga, and L. Zhou, in *Proceedings of the 17th Conference on the Theory of Quantum Computation, Communication and Cryptography, Urbana-Champaign*, edited by F. Le Gall and T. Morimae (Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Saarbrücken, 2022).

[20] V. Vijendran, A. Das, D. E. Koh, S. M Assad, and P. K. Lam, An expressive ansatz for low-depth quantum approximate optimisation, Quantum Sci. Technol. **9**, 025010 (2024).

[21] L. Zhu, H. L. Tang, G. S. Barron, F. A. Calderon-Vargas, N. J. Mayhall, E. Barnes, and S. E. Economou, Adaptive quantum approximate optimization algorithm for solving combinatorial problems on a quantum computer, Phys. Rev. Res. **4**, 033029 (2022).

[22] G. B. Mbeng, R. Fazio, and G. Santoro, Quantum annealing: A journey through digitalization, control, and hybrid quantum variational schemes, arXiv:1906.08948.

[23] S. Ebadi, A. Keesling, M. Cain, T. T. Wang, H. Levine, D. Bluvstein, G. Semeghini, A. Omran, J.-G. Liu, R. Samajdar, X.-Z. Luo, B. Nash, X. Gao, B. Barak, E. Farhi, S. Sachdev, N. Gemelke, L. Zhou, S. Choi, H. Pichler *et al.*, Quantum optimization of maximum independent set using Rydberg atom arrays, Science **376**, 1209 (2022).

[24] E. Pelofske, A. Bärtschi, and S. Eidenbenz, Short-depth QAOA circuits and quantum annealing on higher-order ising models, npj Quantum Inf. **10**, 30 (2024).

[25] Z. Wang, P. L. Zheng, B. Wu, and Y. Zhang, Quantum dropout: On and over the hardness of quantum approximate optimization algorithm, Phys. Rev. Res. **5**, 023171 (2023).

[26] 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).

[27] M. Ragone, B. N. Bakalov, F. Sauvage, A. F. Kemper, C. O. Marrero, M. Larocca, and M. Cerezo, A Lie algebraic theory of barren plateaus for deep parametrized quantum circuits, Nat. Commun. **15**, 7172 (2024).

[28] E. Fontana, D. Herman, S. Chakrabarti, N. Kumar, R. Yalovetzky, J. Heredge, S. H. Sureshbabu, and M. Pistoia, Characterizing barren plateaus in quantum ansätze with the adjoint representation, Nat. Commun. **15**, 7171 (2024).

[29] 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).

[30] T. Bode and F. K. Wilhelm, Adiabatic bottlenecks in quantum annealing and nonequilibrium dynamics of paramagnons, Phys. Rev. A **110**, 012611 (2024).

[31] D. Guéry-Odelin, A. Ruschhaupt, A. Kiely, E. Torrontegui, S. Martínez-Garaot, and J. G. Muga, Shortcuts to adiabaticity: Concepts, methods, and applications, Rev. Mod. Phys. **91**, 045001 (2019).

[32] J. Wurtz and P. J. Love, Counterdiabaticity and the quantum approximate optimization algorithm, Quantum **6**, 635 (2022).

[33] Y. Chai, Y. J. Han, Y. C. Wu, Y. Li, M. Dou, and G. P. Guo, Shortcuts to the quantum approximate optimization algorithm, Phys. Rev. A **105**, 042415 (2022).

[34] H. Guan, F. Zhou, F. Albarrán-Arriagada, X. Chen, E. Solano, N. N. Hegade, and H. Huang, Single-layer digitized-counterdiabatic quantum optimization for p-spin models, arXiv:2311.06682.

[35] T. E. Morris, A. Kaushik, M. Roetteler, and E. C. Lotshaw, Performant near-term quantum combinatorial optimization, arXiv:2404.16135.

[36] 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).

[37] M. Motta, C. Sun, A. T. K. Tan, M. J. O'Rourke, E. Ye, A. J. Minnich, F. G. S. L. Brandão, and G. K.-L. Chan, Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution, Nat. Phys. **16**, 205 (2020).

[38] S. N. Sun, M. Motta, R. N. Tazhigulov, A. T. K. Tan, G. K.-L. Chan, and A. J. Minnich, Quantum computation of finite-temperature static and dynamical properties of spin systems using quantum imaginary time evolution, PRX Quantum **2**, 010317 (2021).

[39] G. Vidal, Efficient classical simulation of slightly entangled quantum computations, Phys. Rev. Lett. **91**, 147902 (2003).

[40] G. Vidal, Efficient simulation of one-dimensional quantum many-body systems, Phys. Rev. Lett. **93**, 040502 (2004).
[41] C. Gattringer and C. Lang, *Quantum Chromodynamics on the Lattice: An Introductory Presentation* (Springer, Berlin, 2009).
[42] X. Wang, Y. Chai, M. Demidik, X. Feng, K. Jansen, and C. Tüysüz, Symmetry enhanced variational quantum imaginary time evolution, arXiv:2307.13598.
[43] M. X. Goemans and D. P. Williamson, Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming, J. ACM **42**, 1115 (1995).
[44] M. Larocca, F. Sauvage, F. M. Sbahi, G. Verdon, P. J. Coles, and M. Cerezo, Group-invariant quantum machine learning, PRX Quantum **3**, 030341 (2022).
[45] J. J. Meyer, M. Mularski, E. Gil-Fuster, A. A. Mele, F. Arzani, A. Wilms, and J. Eisert, Exploiting symmetry in variational quantum machine learning, PRX Quantum **4**, 010328 (2023).
[46] F. Sauvage, M. Larocca, P. J. Coles, and M. Cerezo, Building spatial symmetries into parameterized quantum circuits for faster training, Quantum Sci. Technol. **9**, 015029 (2024).
[47] Q. T. Nguyen, L. Schatzki, P. Braccia, M. Ragone, P. J. Coles, F. Sauvage, M. Larocca, and M. Cerezo, Theory for equivariant quantum neural networks, PRX Quantum **5**, 020328 (2024).
[48] M. Ragone, P. Braccia, Q. T. Nguyen, L. Schatzki, P. J. Coles, F. Sauvage, M. Larocca, and M. Cerezo, Representation theory for geometric quantum machine learning, arXiv:2210.07980.
[49] R. M. Karp, *Complexity of Computer Computations*, edited by R. E. Miller and J. W. Thatcher (Springer, Boston, 1972), pp. 85–103.
[50] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell, Optimal inapproximability results for MAX-CUT and other 2-variable CSPs? SIAM J. Comput. **37**, 319 (2007).
[51] R. Herrman, P. C. Lotshaw, J. Ostrowski, T. S. Humble, and G. Siopsis, Multi-angle quantum approximate optimization algorithm, Sci. Rep. **12**, 6781 (2022).
[52] K. Blekos, D. Brand, A. Ceschini, C.-H. Chou, R.-H. Li, K. Pandya, and A. Summer, A review on quantum approximate optimization algorithm and its variants, Phys. Rep. **1068**, 1 (2024).
[53] S. Bravyi, M. B. Hastings, and F. Verstraete, Lieb-Robinson bounds and the generation of correlations and topological quantum order, Phys. Rev. Lett. **97**, 050401 (2006).
[54] S. Gaurav, Height of binary tree, https://www.scaler.com/topics/height-of-binary-tree/ (Scaler, Bangalore, 2023).
[55] J. Kleinberg and E. Tardos, *Algorithm Design* (Addison Wesley Longman, Boston, 2005).
[56] L. Viola and S. Lloyd, Dynamical suppression of decoherence in two-state quantum systems, Phys. Rev. A **58**, 2733 (1998).
[57] L. Duan and G. Guo, Suppressing environmental noise in quantum computation through pulse control, Phys. Lett. A **261**, 139 (1999).
[58] L. Viola, E. Knill, and S. Lloyd, Dynamical decoupling of open quantum systems, Phys. Rev. Lett. **82**, 2417 (1999).
[59] H. Abraham *et al.*, *Qiskit: An Open-Source Framework for Quantum Computing* (Zenodo, Geneva, 2019).
[60] A. Steger and N. C. Wormald, Generating random regular graphs quickly, Comb. Probab. Comput. **8**, 377 (1999).
[61] P. Virtanen *et al.*, SciPy 1.0: Fundamental algorithms for scientific computing in Python, Nat. Methods **17**, 261 (2020).
[62] J. Rivera-Dean, P. Huembeli, A. Acín, and J. Bowles, Avoiding local minima in variational quantum algorithms with neural networks, arXiv:2104.02955.
[63] Wikipedia, *Box plot*, https://en.wikipedia.org/wiki/Box_plot (Wikipedia, San Francisco, 2024).
[64] P. K. Barkoutsos, G. Nannicini, A. Robert, I. Tavernelli, and S. Woerner, Improving variational quantum optimization using CVaR, Quantum **4**, 256 (2020).
[65] Y. Kim, A. Eddins, S. Anand, K. X. Wei, E. van den Berg, S. Rosenblatt, H. Nayfeh, Y. Wu, M. Zaletel, K. Temme, and A. Kandala, Evidence for the utility of quantum computing before fault tolerance, Nature (London) **618**, 500 (2023).
[66] N. Ezzell, B. Pokharel, L. Tewala, G. Quiroz, and D. A. Lidar, Dynamical decoupling for superconducting qubits: A performance survey, Phys. Rev. Appl. **20**, 064027 (2023).
[67] A. Arrasmith, Z. Holmes, M. Cerezo, and P. J. Coles, Equivalence of quantum barren plateaus to cost concentration and narrow gorges, Quantum Sci. Technol. **7**, 045015 (2022).
[68] E. Pelofske, A. Bärtschi, and S. Eidenbenz, in *High Performance Computing*, edited by A. Bhatele, J. Hammond, M. Baboulin, and C. Kruse, Lecture Notes in Computer Science Vol. 13948 (Springer, Cham, 2023), pp. 240–258.
[69] M. Cerezo, M. Larocca, D. García-Martín, N. L. Diaz, P. Braccia, E. Fontana, M. S. Rudolph, P. Bermejo, A. Ijaz, S. Thanasilp, E. R. Anschuetz, and Z. Holmes, Does provable absence of barren plateaus imply classical simulability? Or, why we need to rethink variational quantum computing, arXiv:2312.09121.
[70] M. L. Goh, M. Larocca, L. Cincio, M. Cerezo, and F. Sauvage, Lie-algebraic classical simulations for variational quantum computing, arXiv:2308.01432.
[71] Y. Chai, K. Jansen, S. Kühn, T. Schwägerl, and T. Stollenwerk, Structure-inspired ansatz and warm start of variational quantum algorithms for quadratic unconstrained binary optimization problems, arXiv:2407.02569.
[72] M. Stiebitz, D. Scheide, B. Toft, and L. M. Favrholdt, *Graph Edge Coloring: Vizing's Theorem and Goldberg's Conjecture* (Wiley, New York, 2012).
[73] A. Kosowski and K. Manuszewski, in *Graph Colorings*, edited by M. Kubale, Contemporary Mathematics Vol. 352 (American Mathematical Society, Providence, 2004), Chap. 1, pp. 1–19.
[74] X. Wang, *Demonstration Data and Properties of the Quantum Device ibm_Brisbane* (Figshare, London, 2024).
[75] X. You and X. Wu, in *Proceedings of the 38th International Conference on Machine Learning*, edited by M. Meila and T. Zhang (PMLR, Cambridge, 2021), Vol. 139.