Combinatorial optimization with quantum imaginary time evolution

Nora M. Bauer,[*] Rizwanul Alam,[†] and George Siopsis[‡]
Department of Physics and Astronomy, University of Tennessee, Knoxville, Tennessee 37996-1200, USA

James Ostrowski[§]
Department of Industrial and Systems Engineering, University of Tennessee, Knoxville, Tennessee 37996-2315, USA

(Received 8 January 2024; accepted 3 May 2024; published 21 May 2024)

We use quantum imaginary-time evolution (QITE) to solve polynomial unconstrained binary optimization (PUBO) problems. We show that a linear ansatz yields good results for a wide range of PUBO problems, often outperforming standard classical methods, such as the Goemans-Williamson (GW) algorithm. We obtain numerical results for the low-autocorrelation binary sequences (LABS) and weighted MaxCut combinatorial optimization problems, thus extending an earlier demonstration of successful application of QITE on MaxCut for unweighted graphs. We find the performance of QITE on the LABS problem with a separable ansatz comparable to QAOA at level 10 for up to 18 vertices and do not see a significant advantage with an entangling ansatz. On weighted MaxCut, QITE with a separable ansatz often outperforms the GW algorithm on graphs with up to 150 vertices.

DOI: 10.1103/PhysRevA.109.052430

I. INTRODUCTION

Many important combinatorial optimization problems can be mapped onto an Ising-type Hamiltonian. Solving a generic Ising model is an NP-hard problem [1]. It would be interesting to witness quantum algorithms outperform their classical counterparts in such problems, but demonstrating their superiority with noisy intermediate-scale quantum (NISQ) devices has proved to be challenging. One of the quantum algorithms extensively studied on NISQ hardware is the quantum approximate optimization algorithm (QAOA) [2]. Another interesting approach was recently proposed that performs better on NISQ devices based on a quantum-enhanced classical optimization algorithm [3].

After mapping onto a Hamiltonian, the combinatorial optimization problem reduces to finding the ground-state energy. Various methods have been employed to compute the ground state of a system, such as adiabatic evolution, quantum annealing, and quantum imaginary-time evolution (QITE). QITE has been widely employed in the analysis of quantum many-body systems, serving as a valuable technique for diverse purposes, including computation of energy levels of multiparticle systems and generation of states at finite temperatures [4,5]. Given that evolution in imaginary time effectively reduces the system to zero temperature [6], the ground state can be precisely prepared using QITE without the need for variational optimization. Nevertheless, in practice, approximations are necessary due to limited computational resources, prompting the need for an approach involving variational calculus.

The progression in imaginary time τ is executed through the nonunitary operator $U(\tau) = e^{-\tau\mathcal{H}}$, where $\mathcal{H}$ denotes the Hamiltonian of the system of interest. Starting with an initial state featuring a nonzero overlap with the system's ground state, the evolved state converges towards the ground state as τ approaches infinity. Access to excited states is also attainable by selecting an appropriate initial state, one that is orthogonal to the ground state. Simulating QITE on a quantum computer is not straightforward because $U(\tau)$ is a nonunitary operator. Motta *et al.* [7] proposed a QITE algorithm that dispensed with the need for classical optimization, distinguishing it from QAOA and the variational quantum eigensolver (VQE) [2]. The approach found practical application in the quantum computation of chemical energy levels on NISQ hardware [8–11] and in the simulation of open quantum systems [12]. The impact of noise on QITE in NISQ hardware was addressed in [13] using error mitigation and randomized compiling. Error mitigation was also addressed with a different method based on deep reinforcement learning [14]. The reduction of quantum circuit depth for QITE through a nonlocal approximation was discussed in [15]. An efficient algorithm to reduce the hardware requirements for QITE was presented in [16] using *causal cones* that facilitated the simulation of finite systems that surpassed the dimensions of the underlying quantum hardware. The implementation of real- and imaginary-time evolution using compressed quantum circuits on NISQ hardware was demonstrated in [17].

The QITE algorithm starts by representing the Hamiltonian in local terms and employs Trotterization to implement $U(\tau)$. Subsequently, the nonunitary evolution for a short imaginary-time interval is approximately implemented with a unitary operator. This unitary operator is expressed in terms of Pauli

[*]nbauer1@vols.utk.edu
[†]ralam4@vols.utk.edu
[‡]siopsis@tennessee.edu
[§]jostrows@tennessee.edu

spin operators, with coefficients determined from measurements on quantum hardware. This approach was employed in Ref. [18] to solve combinatorial optimization problems. For the approximate unitary operator, a separable ansatz was used that was linear in the Pauli operators, and therefore, its implementation required no entanglement of qubits. The method was applied to the MaxCut problem on thousands of randomly selected unweighted graphs with up to 50 vertices. Results compared favorably with the performance of classical algorithms, such as the greedy [19–22] and Goemans-Williamson (GW) [23] algorithms. The overlap of the final state of the QITE algorithm with the ground state was also discussed as a performance metric, which is a quantum feature not shared by other classical algorithms. These results indicate that the linear QITE method is efficient and quantum advantage due to entanglement is likely to be found only for larger graphs ($N \gtrsim 100$) requiring deep quantum circuits which cannot currently be handled by NISQ hardware.

Given the demonstrated success of QITE based on a linear ansatz for MaxCut problems on unweighted graphs [18], it is crucial to assess its efficacy in more general polynomial unconstrained binary optimization (PUBO) problems, which are a more general class of problems containing the more popular quadratic unconstrained binary optimization problems. This analysis is important for assessing the utility of NISQ devices.

Here we extend the results of Ref. [18] to solve PUBO problems using QITE. We concentrate on two problems: (1) the MaxCut problem on weighted graphs and (2) the low-autocorrelation binary sequences (LABS) problem. In both cases we map the problem onto an Ising-type Hamiltonian.

In the MaxCut problem on weighted graphs, we tested QITE with a separable linear ansatz on graphs with up to 150 vertices (qubits) and found that QITE often outperformed the GW algorithm, attaining an average approximation ratio (AR) of ∼0.98 for $N > 100$ vertices. Even as the energy gaps between the ground and first excited states became small, QITE was able to find a state with an energy equal to or lower than that of GW. We point out that a variation for MaxCut, performing better than VQE and QAOA, was developed in Ref. [24] based on filtering operator and causal cones and was demonstrated on three regular weighted graphs with up to 24 vertices. It should be noted that QITE with a separable ansatz can be simulated efficiently classically. Our results indicate that quantum advantage in combinatorial optimization problems will be hard to witness on NISQ devices.

In the LABS problem, complexity grows quickly for large N. Optimal solutions are known only for $N \leqslant 66$, so LABS is a promising candidate for quantum advantage because no classical solutions currently exist [25]. The relevant regime where classical heuristics produce poor solutions is $N \approx 200$, so the number of qubits required for a solution to a classically intractable problem is on the order of hundreds of qubits. In Ref. [25], the authors obtained a scaling advantage over classical algorithms for a problem size of up to 40 qubits using QAOA up to level $p = 40$, obtaining exact solutions. We applied linear ansatz QITE and obtained a probability of measuring the ground state $P(\mathrm{GS})$ comparable to level $p = 10$ QAOA results for relatively low circuit depth and hardware requirements. We also investigated the performance of an entangling quadratic ansatz with QITE but found no significant advantage over the linear ansatz QITE for $N < 10$.

Our discussion is organized as follows. In Sec. II, we discuss the QITE procedure and its hardware requirements. In Sec. III, we define the weighted MaxCut problem and give results for graphs of up to 150 vertices (qubits). In Sec. IV, we define the LABS problem, discuss how we solve it with QITE, and present results for a problem size of up to 28 qubits. Finally, in Sec. V, we summarize our results and discuss further research directions.

II. METHOD

In this section, we review the QITE procedure introduced in Ref. [18] and discuss the hardware requirements of the algorithm.

To implement QITE, we perform evolution in small imaginary-time intervals τ. In the zero-temperature limit, the ground state of the Hamiltonian $\mathcal{H}$ for any state $|\Psi\rangle$ is given by

$$|\Omega\rangle = \lim_{\beta\to\infty} \frac{e^{-\beta\mathcal{H}}|\Psi\rangle}{\|e^{-\beta\mathcal{H}}|\Psi\rangle\|} \tag{1}$$

as long as $\langle\Omega|\Psi\rangle \neq 0$. The Hamiltonian is defined on a graph $G \equiv (V, E)$, where V (E) is the set of vertices (edges). The system consists of qubits lying on the vertices of the graph G.

We initialize the system in the state $|\Psi[0]\rangle$, which can be arbitrarily chosen, as long as it has finite overlap with the ground state of the system. Suppose that after $s-1$ steps we arrive at the state $|\Psi[s-1]\rangle$. In the sth step, we approximate the evolution of $|\Psi[s-1]\rangle$ in (small) imaginary time τ by the action of the unitary $e^{-i\tau A[s]}$, where $A[s]$ is a Hermitian operator. Thus, after s steps, we arrive at the state

$$|\Psi[s]\rangle = e^{-i\tau A[s]}|\Psi[s-1]\rangle = \prod_{s'=1}^{s} e^{-i\tau A[s']}|\Psi[0]\rangle. \tag{2}$$

For the unitary updates, we adopt the *linear ansatz*

$$A[s] = \sum_{j\in V} a_j[s] Y_j. \tag{3}$$

This unitary update is optimal when the coefficients $a_j[s]$ obey the linear system of equations [18]

$$\boldsymbol{S}\cdot\boldsymbol{a} = \boldsymbol{b}, \quad S_{ij}[s] = \langle Y_i Y_j\rangle, \quad b_j[s] = -\frac{i}{2}\langle[\mathcal{H}, Y_j]\rangle, \tag{4}$$

where all expectation values are evaluated with respect to the state $|\Psi[s-1]\rangle$ obtained in the previous step.

We choose the initial state to be the tensor product of eigenstates of X and Z,

$$|\Psi[0]\rangle = \bigotimes_{j=1}^{|V|} |s_j\rangle, \quad |s_j\rangle \in \{|0\rangle, |1\rangle, |+\rangle, |-\rangle\}, \tag{5}$$

where $|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$, introducing no entanglement. Consequently, we have $\boldsymbol{S} = \mathbb{I}$, and therefore, $\boldsymbol{a} = \boldsymbol{b}$. Since all unitary updates commute with each other, we may write the

state after s steps as

$$|\Psi[s]\rangle = e^{-i\tau\mathcal{A}[s]}|\Psi[0]\rangle, \quad \mathcal{A}[s] = \sum_{s'=1}^{s} A[s']. \tag{6}$$

We quantify the quality of the solution by measuring the expectation value of the Hamiltonian and the probability of measuring the ground state $P(\mathrm{GS})$, given as

$$P(\mathrm{GS}) = \sum_{\boldsymbol{\sigma}\in\mathrm{GS}} P(\boldsymbol{\sigma}), \quad P(\boldsymbol{\sigma}) = |\langle\boldsymbol{\sigma}|\Psi[s]\rangle|^2 . \tag{7}$$

In terms of hardware requirements, at each QITE step we need to measure only $\boldsymbol{b}$, which requires N measurements due to basis rotations. In the results shown in the following sections, the value of the small imaginary-time parameter τ is chosen by increasing by steps of $\Delta\beta = \frac{\tau}{T}$, where $T \ll \tau$, starting from zero. At each increase, the cost function $\langle\boldsymbol{\mathcal{H}}\rangle$ is measured, and the process is continued until the cost function (energy) starts increasing. At that point, the previous τ value is selected. We choose a maximum $\beta_{\max}$ and perform fewer than $\beta_{\max}/\Delta\beta$ measurements. Since the applications of $e^{-ia_j[s]Y_j}$ commute, the circuit depth does not scale with the number of steps. The number of measurements scales linearly with the number of steps.

In the following results, we simulate the QITE algorithm exactly using state vectors without noise. For linear QITE, all unitary operations applied are single-qubit operations, so we do not consider the entire Hilbert space, only separable states of size $2N$. For quadratic QITE, we consider the full Hilbert space of size 2^N because the unitary operations are entangling. We perform simulations using PYTHON with standard NUMPY and SCIPY libraries.

III. WEIGHTED MAXCUT

Here we define the weighted MaxCut problem and the technique of imaginary-time-dependent (ITD) edges that we use to improve convergence. We present results of graphs with up to 150 vertices (qubits) and find that QITE sometimes outperforms the classical GW algorithm [23].

Given a graph $G = (V, E)$ consisting of a set of vertices V and edges E joining the vertices, the unweighted MaxCut problem on G is the combinatorial optimization problem of partitioning V into two disjoint sets such that the number of edges with end points in each set C is maximized ($C = C_{\max}$). It can be formulated as a Hamiltonian ground-state problem by associating a qubit with every vertex in V and defining the Hamiltonian

$$\boldsymbol{\mathcal{H}} = \sum_{(i,j)\in E} Z_i Z_j, \tag{8}$$

where Z_i is the Pauli Z matrix acting on the ith qubit. The ground state energy $\mathcal{E}_0$ of $\boldsymbol{\mathcal{H}}$ is related to $C_{\max}$ by

$$C_{\max} = \frac{|E| - \mathcal{E}_0}{2}. \tag{9}$$

For the weighted MaxCut problem, the edges $(i, j) \in E$ of the graph G^w have associated weights w_{ij}, and the Hamiltonian is modified as

$$\boldsymbol{\mathcal{H}}^w = \sum_{(i,j)\in E} w_{ij} Z_i Z_j. \tag{10}$$

The ground-state energy $\mathcal{E}_0^w$ is related to the maximum cut $C_{\max}^w$ as

$$C_{\max}^w = \frac{\sum_{(i,j)\in E} w_{ij} - \mathcal{E}_0^w}{2}. \tag{11}$$

After s steps, QITE produces an approximation to $C_{\max}^w$, given by

$$C_{\mathrm{QITE}}^w = \frac{1}{2}\left(\sum_{(i,j)\in E} w_{ij} - \langle\Psi[s]|\mathcal{H}^w|\Psi[s]\rangle\right). \tag{12}$$

Evidently, $C_{\mathrm{QITE}}^w \leqslant C_{\max}^w$. We compare the cost function obtained from QITE to the one obtained with the classical GW algorithm [23]. Since for large graph sizes we cannot guarantee that GW produces $C_{\max}^w$, we define the AR R_{A} as the cost function obtained by QITE divided by the cost function produced by GW,

$$R_{\mathrm{A}} = \frac{C_{\mathrm{QITE}}^w}{C_{\mathrm{GW}}^w}. \tag{13}$$

As in [18], we use the method of ITD edges to improve convergence. This is done by interpolating between a Hamiltonian that corresponds to a subgraph of G^w and $\boldsymbol{\mathcal{H}}$ using an ITD Hamiltonian $\mathcal{H}^w[s]$ given by

$$\boldsymbol{\mathcal{H}}^w[s] = \sum_{(i,j)\in E} w_{ij}[s] Z_i Z_j. \tag{14}$$

A subgraph is selected for which the corresponding weights vanish initially and then are increased with each step s until $w_{ij}[s] \to w_{ij}$ for sufficiently many steps and $\mathcal{H}^w[s] \to \mathcal{H}^w$.

We restrict our focus to N-vertex weighted Newman-Watts-Strogatz (NWS) graphs with $k = 4$ and $p = 0.5$, where the weights are selected from the uniform random distribution (0,1]. We apply the QITE algorithm with a linear ansatz with one ITD edge and measure the AR as the measured energy from QITE divided by the lowest energy obtained from GW. We also measure the probability of obtaining this lowest-energy value as the probability of obtaining the ground state $P(\mathrm{GS})$. First, we perform QITE on graphs with $6 \leqslant N \leqslant 30$ vertices and measure the average number of random subgraphs to obtain convergence [$P(\mathrm{GS}) > 0.995$ overlap] to the ground state. The algorithm is performed repeatedly with different selections for the ITD edge and different random initial states until all attempted graphs have converged to the ground state. This procedure is shown in Fig. 1 for an example eight-vertex graph. We compute this result for 25 and 50 QITE steps on weighted graphs and also 25 QITE steps on unweighted graphs. We compute this as a function of the number of graph vertices N, as shown in Fig. 2.

We observe similar results for 25 and 50 steps on weighted graphs for $N < 18$, while the unweighted graphs require fewer resources. For $N > 18$, the results for 25 and 50 QITE steps begin to diverge, with almost twice the resources required at $N = 30$. At this point, the energy gap between the ground state and first excited state is on the order of 0.01.

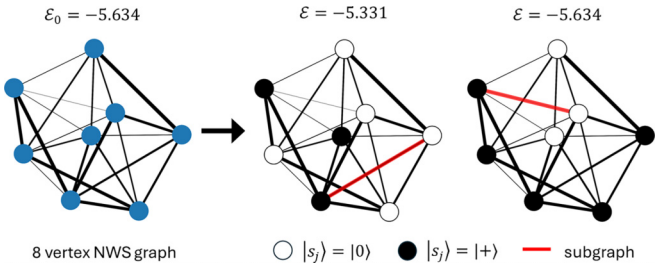

FIG. 1. Example eight-vertex NWS graph where we randomly choose a random ITD edge and initial states $|s_j\rangle$ for each qubit, then apply the QITE algorithm with 25 steps. The first choice results in convergence to an excited state where $\mathcal{E} = -5.331$, while the second choice results in convergence to the ground state where $\mathcal{E} = -5.634$.

For larger graphs with $N \geqslant 20$, the state produced from QITE sometimes converges to a state with lower energy than the lowest energy obtained from GW. To account for this, we compute the approximation ratio, the QITE energy divided by the GW energy, for graphs with $6 < N < 150$ vertices for single selections of random subgraphs and initial states, as shown in Fig. 3.

For 10 QITE steps, the average AR is less than 0.97 for $N > 10$, indicating that 10 steps are not sufficient to solve larger MaxCut problems. For 25 and 50 steps, we observe that the AR initially decreases with graph size to a minimum $\sim$0.97 for $N = 30$ vertices, then increases for larger values of N to a maximum value of $\sim$0.98 (0.975) for $N = 120$ vertices at 50 (25) steps. We attribute the decrease to QITE performing better on smaller N when GW produces the exact solution and the increase to GW finding suboptimal solutions when QITE produces a state with a finite overlap with a lower-energy solution. To this end, we also plot the best AR for each number of steps in Fig. 4, showing that certain instances of QITE produce solutions superior to those of GW. For $N > 10$, 25, and 50 vertices, QITE remains above 1, indicating that QITE

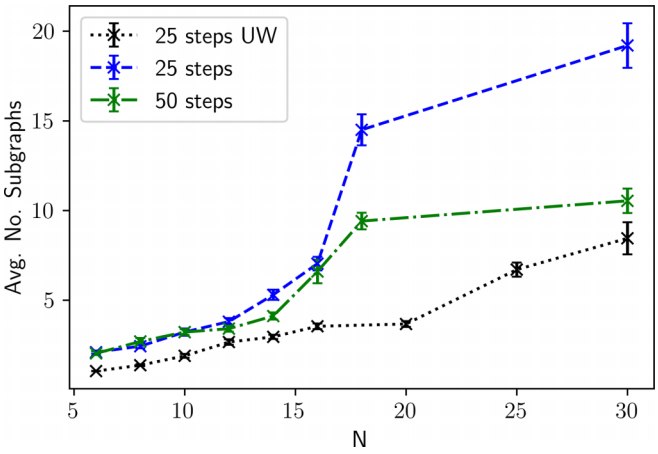

FIG. 2. QITE results for the weighted MaxCut problem showing the average number of random subgraph choices required for N-qubit graphs to obtain $P(\mathrm{GS}) > 0.995$. Results are shown for 25 QITE steps on unweighted (black dotted line) and weighted (blue dashed line) graphs, as well as 50 QITE steps on weighted graphs (green dash-dotted line). The bars denote one standard error over 250 trials.

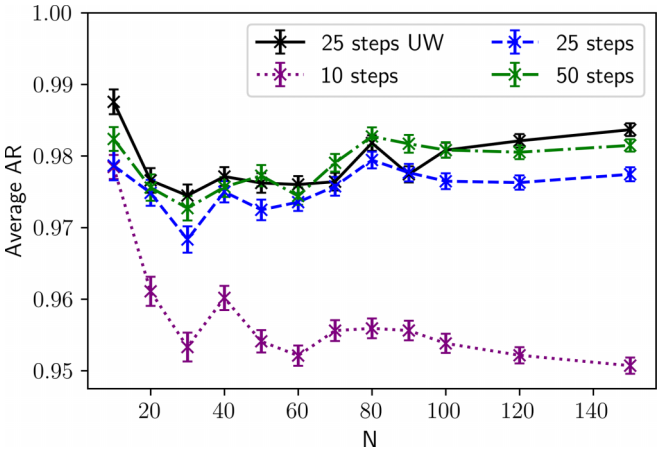

FIG. 3. QITE results for the weighted MaxCut problem showing the average approximation ratio, computed by dividing the energy from QITE by the energy from GW, for single subgraph trials for N-qubit weighted graphs. Results are shown for 10 (dotted purple line), 25 (dashed blue line), and 50 (dash-dotted green line) QITE steps, in addition to 25 steps for the unweighted (UW) case (solid black line).

produces solutions that are better than those found by the GW classical algorithm. For unweighted graphs, the average AR remains in the range of 0.975–0.99 for all N, and the best AR is $>$1.02 for $N > 70$.

An example of a 20-vertex weighted graph where 50-step QITE outperforms the GW classical algorithm is shown in Fig. 5. The cuts chosen by both algorithms are indicated, in addition to the cuts unique to QITE and GW. QITE chooses a solution which is superior to the one produced by GW, containing 3 cuts which differ from those in GW's solution. The difference in energy between QITE and GW is $\Delta E = -0.0404$.

QITE with a linear ansatz can be simulated efficiently classically, so the computation can be done for larger graph sizes

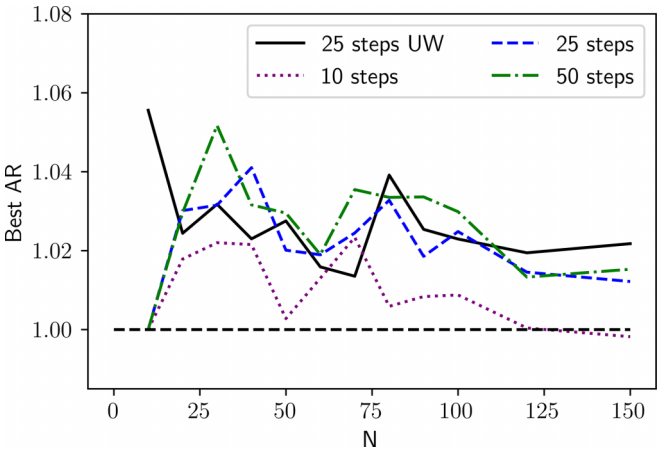

FIG. 4. QITE results for the weighted MaxCut problem showing the best approximation ratio, computed by dividing the energy from QITE by the energy from GW, for N-qubit weighted graphs. Results are shown for 10 (dotted purple line), 25 (dashed blue line), and 50 QITE steps (dash-dotted green line), in addition to 25 steps for the unweighted (UW) case (solid black line).

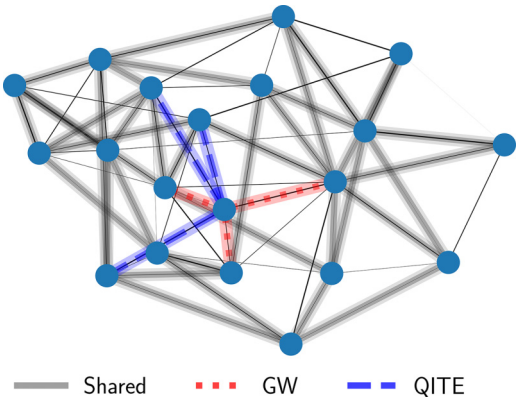

FIG. 5. Example 20-vertex graph where 50-step QITE outperforms GW for the weighted MaxCut problem. The cuts common to GW and QITE are solid lines highlighted in gray. The cuts unique to QITE are dashed lines highlighted in blue, while the cuts unique to GW are dotted lines highlighted in red.

than considered here without issue. Since here we chose one ITD edge, it would be interesting to see if choosing more ITD edges or choosing a different scheme would produce better results on larger graphs. We are currently investigating this in addition to the scaling of QITE on graphs with 500+ vertices.

IV. LABS

Here we discuss how we solve the LABS problem with QITE using both a linear ansatz and a quadratic ansatz, present results for problem sizes of up to 28 qubits, and compare them with recent QAOA results [25].

The goal of the LABS problem is to minimize the "sidelobe energy" for a system of N spins $\sigma_i \in \{+1, -1\}$ with autocorrelation $\mathcal{A}_k(\boldsymbol{\sigma})$:

$$\mathcal{E}_{\text{sidelobe}}(\boldsymbol{\sigma}) = \sum_{k=1}^{N-1} \mathcal{A}_k^2(\boldsymbol{\sigma}), \quad \mathcal{A}_k(\boldsymbol{\sigma}) = \sum_{k=1}^{N-k} \sigma_i \sigma_{i+k}. \tag{15}$$

This can be mapped to the ground-state problem of the quantum Hamiltonian

$$\mathcal{H}^{\text{LABS}} = 2 \sum_{i=1}^{N-3} \sum_{t=1}^{\frac{N-i-1}{2}} \sum_{k=1}^{N-i-t} Z_i Z_{i+t} Z_{i+k} Z_{i+t+k} + \sum_{i=1}^{N-2} \sum_{k=1}^{\frac{N-i}{2}} Z_i Z_{i+2k}. \tag{16}$$

The quality of a solution can be quantified by the overlap with the ground state (7) and the ratio of the merit factor of the measured state over the merit factor of the ground-state solution,

$$R_{\text{A}} = \frac{\mathcal{F}_{\text{QITE}}}{\mathcal{F}_{\text{GS}}}, \tag{17}$$

where the merit factor $\mathcal{F}$ is given as

$$\mathcal{F}(|\psi\rangle) = \frac{N^2}{2\langle \mathcal{H} \rangle} = \sum_{\boldsymbol{\sigma}} P(\boldsymbol{\sigma}) \mathcal{F}(\boldsymbol{\sigma}),$$
$$\mathcal{F}(\boldsymbol{\sigma}) = \frac{N^2}{2\mathcal{E}_{\text{sidelobe}}(\boldsymbol{\sigma})} \tag{18}$$

and the overlap probabilities $P(\boldsymbol{\sigma})$ are defined in (7). As with weighted MaxCut, we introduce imaginary-time dependence to improve convergence to the ground state instead of an excited state. Since QITE with a linear ansatz performs well with a Hamiltonian with quadratic terms, we introduce an ITD coefficient $\alpha[s]$ to the quartic terms of the Hamiltonian and define the ITD Hamiltonian

$$\mathcal{H}^{\text{LABS}}[s] = 2\alpha[s] \sum_{i=1}^{N-3} \sum_{t=1}^{\frac{N-i-1}{2}} \sum_{k=1}^{N-i-t} Z_i Z_{i+t} Z_{i+k} Z_{i+t+k} + \sum_{i=1}^{N-2} \sum_{k=1}^{\frac{N-i}{2}} Z_i Z_{i+2k}, \quad \alpha[s] = \frac{a\left\lfloor \frac{s}{a} \right\rfloor}{n_{\text{steps}}}, \tag{19}$$

where s are the integer time indices ($s = 0, \ldots, n_{\text{steps}}$) and a determines whether $\alpha[s]$ is piecewise or linear ($a = 1$) with respect to s.

To improve the results further, we can add another time-dependent parameter, $\beta[s]$, that depends on the range of the quartic interaction terms, slowly adding increasingly nonlocal Hamiltonian parameters,

$$\mathcal{H}^{\text{LABS}}[s] = 2\alpha[s] \sum_{i,t,k} \beta_{itk}[s] Z_i Z_{i+t} Z_{i+k} Z_{i+t+k} + \sum_{i,k} Z_i Z_{i+2k}, \tag{20}$$

where

$$\beta_{itk}[s] = \begin{cases} \frac{b\lfloor \frac{s}{b} \rfloor}{n_{\text{steps}}}, & \max(t, k, t+k, |t-k|) > R_{\max}, \\ 1, & \text{otherwise.} \end{cases} \tag{21}$$

We reference a LABS solution bank [26] when computing the approximation ratio and the overlap with the ground state $P(\text{GS})$, which is a quantum metric. Using Eq. (20), we perform the QITE algorithm with 40 steps on 50 random initial states composed of $|0\rangle$ and $|+\rangle$ for $6 \leqslant N \leqslant 21$. We first compute the approximation ratio by dividing the resulting QITE energy by the energy obtained from the solution bank. The average and maximum ARs for the 50 initial states are plotted in Fig. 6. The probability of measuring the ground state $P(\text{GS})$ is given in Fig. 7. We plot the QITE results alongside the $P(\text{GS})$ values for $p = 10$ QAOA obtained in [25]. We find the average $P(\text{GS})$ of the linear ansatz QITE comparable to the $p = 10$ QAOA, while the best-case $P(\text{GS})$ consistently exceeds both values. In Ref. [25], LABS was solved exactly with simulated noiseless QAOA for problem sizes up to $N = 40$. However, the authors used QAOA levels up to $p = 40$, which results in a gate depth on the order of 10^3 for problem size $N = 18$. This is challenging to execute on NISQ devices or classical simulators. For comparison, linear ansatz QITE requires shallow quantum circuits and readily available classical computing resources.

Since the linear ansatz QITE does not give $P(\text{GS}) = 1$ as in weighted MaxCut, we considered a higher-order quadratic ansatz. The details of the formulation are as follows. For the quadratic ansatz, we used one containing the linear terms in addition to terms quadratic in Y:

$$A_{\text{quad}}[s] = \sum_{j \in V} a_{0,j}[s] Y_j + \sum_{i,j \in V} a_{i,j}[s] Y_i Y_j. \tag{22}$$

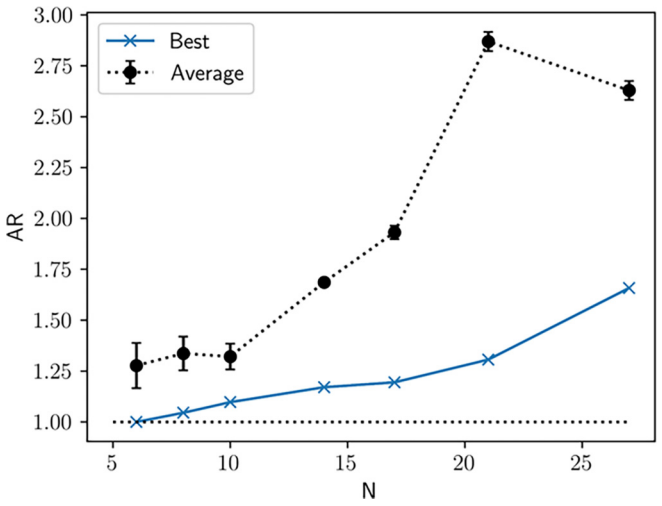

FIG. 6. QITE results for the LABS problem showing the average approximation ratio (black circles) and best approximation ratio (blue crosses) for 40 QITE steps using the linear ansatz for problem size $6 \leqslant N \leqslant 28$ for 100 different random initial states. The bars denote one standard error.

It is convenient to define the vector consisting of all operators appearing in Eq. (22), $\boldsymbol{\mathcal{Y}} \equiv (Y_1, Y_2, \ldots, Y_N, Y_1Y_2, Y_1Y_3, \ldots, Y_1Y_N, Y_2Y_3, \ldots)$, where the quadratic terms contain all unique choices of pairs of vertices. Evidently, this vector is of length $N(N+1)/2$.

As in the linear ansatz case, the coefficients $a_I[s]$, where $I = \{i, j\}$, obey the linear system of equations

$$\boldsymbol{S} \cdot \boldsymbol{a} = \boldsymbol{b}, \tag{23}$$

where we defined

$$S_{IJ}[s] = \langle \mathcal{Y}_I \mathcal{Y}_J \rangle, \quad b_J[s] = -\frac{i}{2} \langle [\boldsymbol{\mathcal{H}}^{\mathrm{LABS}}, \mathcal{Y}_J] \rangle, \tag{24}$$

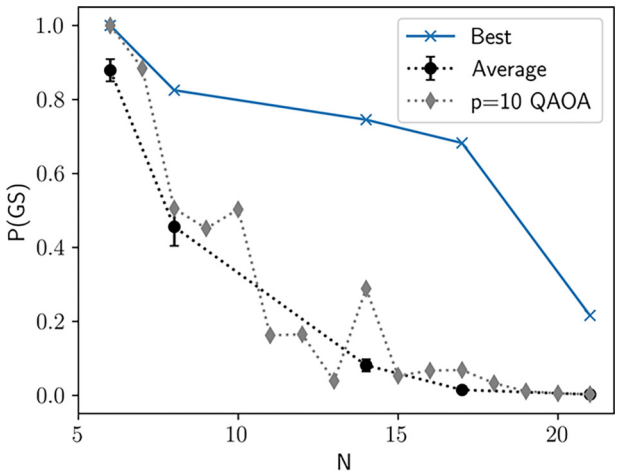

FIG. 7. QITE results for the LABS problem showing the average (black circles) and best (blue crosses) probabilities of measuring the ground state $P(\mathrm{GS})$ for problem size $6 \leqslant N \leqslant 18$ for 40 QITE steps using the linear ansatz. The results from [25] for $p = 10$ QAOA are also shown (gray diamonds).

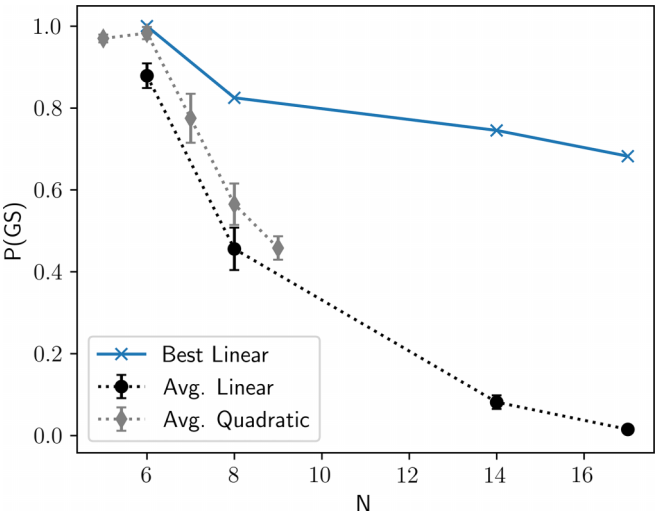

FIG. 8. Probability of measuring the ground state $P(\mathrm{GS})$ after 40 QITE steps, shown for problem sizes $6 \leqslant N \leqslant 18$ for the linear ansatz and $5 \leqslant N \leqslant 9$ for the quadratic ansatz. For the linear ansatz, the average $P(\mathrm{GS})$ is shown by black circles, and the best $P(\mathrm{GS})$ is shown by blue crosses. For the quadratic ansatz, the average $P(\mathrm{GS})$ is shown by gray diamonds.

where all expectation values are evaluated with respect to the state $|\Psi[s-1]\rangle$ obtained in the previous step. If the initial state is chosen to be the tensor product of eigenstates of X and Z, then we have $\boldsymbol{S} = \mathbb{I}$, and therefore, $\boldsymbol{a} = \boldsymbol{b}$. The algorithm then proceeds as in the linear ansatz case with updates of the form

$$|\Psi[s]\rangle = e^{-i\tau A_{\mathrm{quad}}[s]} |\Psi[s-1]\rangle. \tag{25}$$

The quadratic ansatz gives a minor improvement in the $P(\mathrm{GS})$ over the linear ansatz, as shown in Fig. 8. Therefore, we do not see a benefit to including higher-order terms in the ansatz in the range of $5 < N < 10$. It is conceivable that there may be benefits at larger problems sizes beyond our current resources to test. Additionally, the higher-order terms might not provide a significant benefit because the LABS problem has a separable ground-state solution.

In [25], the scaling advantage of QAOA over classical solvers was found in the range of $28 \leqslant N \leqslant 40$, which is larger than the problem sizes we tested. Therefore, we need to investigate the performance of linear and quadratic ansatz QITE compared to QAOA and classical algorithms in this range. The quadratic ansatz simulation requires significantly more computational resources and precision than the linear ansatz, but we are currently working on extending our results to larger problem sizes.

In this regime, introducing entanglement into the algorithm does not provide a significant result, as expected, as the results from linear ansatz QITE are comparable to both quadratic ansatz QITE and $p = 10$ QAOA. We are currently investigating larger problem sizes and higher-order PUBO problems to search for a case where introducing entanglement at a NISQ-era circuit depth provides a substantial benefit.

V. CONCLUSION

The QITE algorithm effectively cools a system to zero temperature, at which the system settles to its ground state, reaching the lowest-energy level of its Hamiltonian. By encoding combinatorial optimization problems in terms of a quantum Hamiltonian, one can solve these problems by finding the ground state of the Hamiltonian corresponding to the optimal solution. QITE allows for a flexible ansatz which can be chosen to be separable or entangling. Although solutions of combinatorial optimization problems involve separable states, quantum algorithms, such as QAOA, introduce entangling operations to solve these problems.

In this work, we investigated the performance of QITE on PUBO problems, concentrating on the weighted MaxCut and LABS combinatorial optimization problems. Our method was a generalization of the approach introduced in Ref. [18] that was successfully applied to unweighted MaxCut. In general, we expected an increased difficulty in PUBO cases due to smaller gaps between the ground- and first-excited-state energies. We compared the performance of QITE on NWS graphs with up to 150 vertices with the commonly used classical algorithm GW, which is widely believed to offer the best performance guarantee. We found that QITE with a linear ansatz and ITD edges often outperforms the classical GW algorithm on weighted MaxCut instances after a sufficient number of steps, yielding an average AR of $\sim$0.98 for $N > 100$ vertices. Our separable ansatz can be simulated efficiently classically. In Ref. [27], the performance of noiseless QAOA was assessed on MaxCut problems with up to 17 vertices, and it was observed that level $p = 5$ was required for results to be comparable with GW. Especially for larger graphs, this is outside of the range of NISQ devices. In Ref. [28], it was estimated that graph sizes of several hundreds to thousands of vertices are required for quantum advantage over classical solvers on the MaxCut problem. Our results indicate that much larger graphs need to be considered in order to observe quantum advantage. Given the attendant increase in the depth of the quantum circuit, a quantum computation may not be suitable on NISQ devices. It is important to analyze the applicability of our method further to better assess the utility of NISQ devices.

We also analyzed the performance of QITE on the LABS problem. This is a PUBO problem, as the quantum Hamiltonian encoding the LABS problem includes quartic terms in the Pauli Z matrix. We tested the performance of QITE with a separable linear ansatz and a randomly chosen separable initial state. We compared the results to known solutions of the problem from a solution bank and calculated the AR and probability of measuring the ground state P(GS). Although, in general, we did not obtain convergence to the ground state, we found that the average P(GS) of the linear ansatz QITE with 40 steps was comparable to $p = 10$ QAOA [25]. More importantly, the best-case P(GS) for each problem size was consistently higher than the QAOA results, indicating that an appropriate choice of initial state gives a high probability of solving the LABS problem.

Expecting improvement and possibly quantum advantage to be found using an entangling ansatz containing higher-order terms, we considered an entangling ansatz with both linear and quadratic terms. Comparing it to the separable linear ansatz, we found that the entangling ansatz gave a P(GS) which was no more than 10% higher than the result from the linear ansatz for problem sizes up to $N = 9$. Thus, the entangling quadratic ansatz did not perform significantly better than the separable linear ansatz in the problem-size range we tested, indicating that we might need to explore larger problem sizes to see a significant advantage due to quantum entanglement.

Obtaining quantum advantage in combinatorial optimization problems appears to require larger problem sizes than those that can be handled by NISQ devices or, perhaps, more complicated PUBO problems than the ones studied here. The PUBO problems are of particular interest for QITE since these problems become classically intractable at smaller problem sizes, which are better suited for NISQ devices. Both research directions, i.e., searching for a combinatorial optimization problem where an entangling ansatz outperforms the linear ansatz significantly even at sufficiently small problem sizes implementable on NISQ devices and simulating a large enough problem among those considered here where an entangling ansatz outperforms the linear ansatz, are currently being pursued.

ACKNOWLEDGMENTS

We thank P. C. Lotshaw for useful discussions. This work was supported by the DARPA ONISQ program under Award No. W911NF-20-2-0051 and NSF Award No. DGE-2152168. A portion of the computation for this work was performed using the University of Tennessee Infrastructure for Scientific Applications and Advanced Computing (ISAAC) computational resources.

[1] F. Barahona, On the computational complexity of Ising spin glass models, J. Phys. A **15**, 3241 (1982).

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

[3] M. Dupont, B. Evert, M. J. Hodson, B. Sundar, S. Jeffrey, Y. Yamaguchi, D. Feng, F. B. Maciejewski, S. Hadfield, M. S. Alam, Z. Wang, S. Grabbe, P. A. Lott, E. G. Rieffel, D. Venturelli, and M. J. Reagor, Quantum-enhanced greedy combinatorial optimization solver, Sci. Adv. **9**, eadi0487 (2023).

[4] 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).

[5] M. J. S. Beach, R. G. Melko, T. Grover, and T. H. Hsieh, Making trotters sprint: A variational imaginary time ansatz for quantum many-body systems, Phys. Rev. B **100**, 094434 (2019).

[6] P. J. Love, Cooling with imaginary time, Nat. Phys. **16**, 130 (2020).

[7] M. Motta, C. Sun, A. T. Tan, M. J. O'Rourke, E. Ye, A. J. Minnich, F. G. Brandao, and G. K.-L. Chan, Determining

eigenstates and thermal states on a quantum computer using quantum imaginary time evolution, Nat. Phys. **16**, 205 (2020).

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

[9] K. Yeter-Aydeniz, R. C. Pooser, and G. Siopsis, Practical quantum computation of chemical and nuclear energy levels using quantum imaginary time evolution and Lanczos algorithms, npj Quantum Inf. **6**, 63 (2020).

[10] K. Yeter-Aydeniz, B. T. Gard, J. Jakowski, S. Majumder, G. S. Barron, G. Siopsis, T. S. Humble, and R. C. Pooser, Benchmarking quantum chemistry computations with variational, imaginary time evolution, and Krylov space solver algorithms, Adv. Quantum Technol. **4**, 2100012 (2021).

[11] S. Barison, D. E. Galli, and M. Motta, Quantum simulations of molecular systems with intrinsic atomic orbitals, Phys. Rev. A **106**, 022404 (2022).

[12] H. Kamakari, S.-N. Sun, M. Motta, and A. J. Minnich, Digital quantum simulation of open quantum systems using quantum imaginary time evolution, PRX Quantum **3**, 010320 (2022).

[13] J.-L. Ville *et al.*, Leveraging randomized compiling for the quantum imaginary-time-evolution algorithm, Phys. Rev. Res. **4**, 033140 (2022).

[14] C. Cao, Z. An, S.-Y. Hou, D. L. Zhou, and B. Zeng, Quantum imaginary time evolution steered by reinforcement learning, Commun. Phys. **5**, 57 (2022).

[15] H. Nishi, T. Kosugi, and Y.-i. Matsushita, Implementation of quantum imaginary-time evolution method on NISQ devices by introducing nonlocal approximation, npj Quantum Inf. **7**, 85 (2021).

[16] M. Benedetti, M. Fiorentini, and M. Lubasch, Hardware-efficient variational quantum algorithms for time evolution, Phys. Rev. Res. **3**, 033083 (2021).

[17] S.-H. Lin, R. Dilip, A. G. Green, A. Smith, and F. Pollmann, Real- and imaginary-time evolution with compressed quantum circuits, PRX Quantum **2**, 010342 (2021).

[18] R. Alam, G. Siopsis, R. Herrman, J. Ostrowski, P. C. Lotshaw, and T. S. Humble, Solving MaxCut with quantum imaginary time evolution, Quantum Inf. Process. **22**, 281 (2023).

[19] H. Schröder, A. May, I. Vrt'o, and O. Sýkora, Approximation algorithms for the vertex bipartization problem, in *SOFSEM'97: Theory and Practice of Informatics*, edited by F. Plášil and K. G. Jeffery (Springer, Berlin, 1997), pp. 547–554.

[20] S. Kahruman, E. Kolotoglu, S. Butenko, and I. Hicks, On greedy construction heuristics for the max-cut problem, Int. J. Comput. Sci. Eng. **3**, 211 (2007).

[21] C. Mathieu and W. Schudy, Yet another algorithm for dense max cut: Go greedy, in *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '08* (Society for Industrial and Applied Mathematics, Philadelphia, 2008), pp. 176–182.

[22] Y. Bian, A. Gronskiy, and J. M. Buhmann, Greedy maxcut algorithms and their information content, in *Proceedings of 2015 IEEE Information Theory Workshop (ITW)* (IEEE, Piscataway, NJ, 2015), pp. 1–5.

[23] M. X. Goemans and D. P. Williamson, .878-approximation algorithms for MAX CUT and MAX 2SAT, in *Proceedings of the Twenty-Sixth Annual ACM Symposium on Theory of Computing* (ACM, New York, 1994), pp. 422–431.

[24] D. Amaro, C. Modica, M. Rosenkranz, M. Fiorentini, M. Benedetti, and M. Lubasch, Filtering variational quantum algorithms for combinatorial optimization, Quantum Sci. Technol. **7**, 015021 (2022).

[25] R. Shaydulin *et al.*, Evidence of scaling advantage for the quantum approximate optimization algorithm on a classically intractable problem, arXiv:2308.02342.

[26] B. Bošković, F. Brglez, and J. Brest, A GitHub archive for solvers and solutions of the `labs` problem, 2016, https://github.com/borkob/git_labs.

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

[28] G. G. Guerreschi and A. Y. Matsuura, QAOA for Max-Cut requires hundreds of qubits for quantum speed-up, Sci. Rep. **9**, 6903 (2019).