

Lawrence Berkeley National Laboratory

LBL Publications

Title

Classical optimization with imaginary-time block encoding on quantum computers: The MaxCut problem

Permalink

<https://escholarship.org/uc/item/4cf185mr>

Journal

Physical Review A, 112(4)

ISSN

2469-9926

Authors

Zhong, Dawei

Francis, Akhil

Rrapaj, Ermal

Publication Date

2025-10-01

DOI

10.1103/gtq3-j37b

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Classical optimization with imaginary time block encoding on quantum computers: The MaxCut problem

Dawei Zhong^{*}

Department of Physics & Astronomy, University of Southern California, Los Angeles, CA 90089, USA

Akhil Francis[†]

*Applied Mathematics and Computational Research Division,
Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*

Ermal Rrapaj[‡]

*Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA
Department of Physics, University of California, Berkeley, CA 94720, USA and
RIKEN iTHEMS, Wako, Saitama 351-0198, Japan*

Optimization problems in finance, physics and computer science are typically very hard to tackle in classical computing and quantum computing could help speed up computations and provide efficient methods for tackling large problems. Typically, to treat the problem with a quantum computer, the optimal solution is cast as the ground state of a diagonal Hamiltonian. We develop a new method, called ITE-BE, based on a recent imaginary time algorithm, which requires no variational parameter optimization as all parameters can be derived analytically from the target Hamiltonian. We also demonstrate that our method can be successfully combined with other quantum algorithms such as quantum approximate optimization algorithm (QAOA). For illustration, here we study the MaxCut problem. We find that the QAOA ansatz increases the post-selection success of ITE-BE, and shallow QAOA circuits, when boosted with ITE-BE, achieve better performance than deeper QAOA circuits. For the special case of the transverse initial state, we adapt our block encoding scheme to allow for a deterministic application of the first layer of the circuit.

I. INTRODUCTION

The advent of quantum computers has spurred a wide range of research into quantum advantage: problems that can be solved much faster, and/or at a much larger scale through quantum computing. Some of the most well known cases are integer prime factorization [1], and unstructured search [2], to name a few. An interesting venue of research is discrete optimization problems [3]. These problems can be expressed as the minimization of a polynomial function of binary variables, and the cost function can be written as a diagonal Hamiltonian which encodes the optimal solution in the ground state [4, 5]. For instance, Quadratic Unconstrained Binary Optimization (QUBO) models several problems such as MaxCut, minimum vertex cover, and graph coloring [6, 7]. These problems have found applications in industry [8, 9], condensed matter physics [10], and beyond.

Apart from special cases, finding the exact ground state of such diagonal Hamiltonians is usually difficult, and generally NP-hard [4, 5]. Often classical heuristic and approximate methods, such as simulated annealing [11] or semi-definite programming [12], are used to approximate solutions. Simulated annealing motivated by the annealing process in statistical mechanics, is a

probabilistic method where an appropriate temperature and annealing schedule is defined to ‘cool’ the system, while the new configuration is accepted with an appropriate acceptance probability.

Semi-definite programming (SDP) minimizes a linear cost function subject to a semi-definite constraint, resulting in convex optimization problem which can be solved efficiently [13]. SDP could be applied by relaxing the original optimization problems (e.g., the MaxCut problem), followed by an appropriate rounding scheme to obtain approximate solutions [12].

Quantum computers could potentially find solutions beyond the realm of classical computation [3] as quantum hardware moves beyond the noisy intermediate-scale quantum (NISQ) era to fault-tolerant computing [14]. Many quantum algorithms such as quantum approximate optimization algorithm (QAOA) [15–19], adiabatic evolution [20], or imaginary time evolution methods [21–24] have been proposed to help determine the ground state. All these methods have their own drawbacks too, such as barren plateaus [25, 26] and other optimization related issues for QAOA [27], longer depth circuits for implementing adiabatic evolution [28] and larger number of measurements for successful post-selection of block encoded implementation of imaginary time evolution [29].

QAOA has been widely explored recently in large sparse graphs, and spin glasses [30], in particular in Sherrington-Kirkpatrick (SK) model of spin glass [31], on MaxCut problems [32] in both software and hardware [33–35]. In addition to two-body interactions, QAOA has also been applied to four-body interactions,

^{*} dawei@usc.edu

[†] afrancis@lbl.gov

[‡] ermalrrapaj@lbl.gov

such as the low autocorrelation binary sequences (LABS) problem [9].

Imaginary time evolution, otherwise known as Wick rotation [36], connects Euclidean and Minkowski space by performing an analytical continuation from real to imaginary time (see figure 6.1 from [37]), and has been used extensively in auxiliary field diffusion Monte Carlo studies [38–40]. This approach exponentially improves the overlap towards the ground state when applied to an initial state with non-zero overlap with the ground state. Since it is a non-unitary operation, implementing it on a quantum computer is not trivial, and there has been several approaches in the literature. One approach is to find equivalent unitary evolutions, with appropriate quantum measurements and classical computations, such as in variational imaginary time evolution [21, 41, 42] where a fixed ansatz is updated to follow McLachlan’s variational principle. Another approach discretizes the imaginary time evolution and approximates each step by a unitary constructed through tomography [22, 43, 44]. These methods require classical processing and multiple measurements during the time evolution which scale with both system and length of time. Other methods such as block-encoding schemes [45–47], encode the desired non-unitary operation in a bigger unitary matrix, which can then be implemented with ancilla qubits and post-selection. Recently an exact block-encoding scheme for discretized imaginary time evolution, based on unitary Restricted Boltzmann Machines (RBM), has been proposed [29] (for similar recent implementations see [24]).

Here we utilize this imaginary time evolution block encoding (ITE-BE) scheme of the first order Trotter decomposition, to implement imaginary time evolution in solving classical optimization problems such as MaxCut. We also implement a new parametrization that allows to correct for failed post-selection when the initial state is in the traverse state. We find that ITE-BE is beneficial in providing solutions to MaxCut, and also as a boosting technique to improve other methods such as QAOA. In Sec. II we define the MaxCut problem, QAOA and the block encoding method ITE-BE. Then, we proceed to perform numerical evaluations of these methods in Sec. III. In Sec. IV we summarize our findings and discuss their implications.

II. PRELIMINARY

A. The MaxCut Problem Definition

In this work, we focus on the MaxCut problem for a given undirected graph $G = (V, \tilde{E})$, where $V = \{1, \dots, N\}$ represents the set of vertices, and $\tilde{E} = \{(\langle j, k \rangle, w_{jk})\}$ denotes the set of edges with non-negative weights $w_{jk} \geq 0$ associated with the edge connecting vertices j and k . The objective of this problem is to partition vertices of graph G into two disjoint subsets such that the sum of weights of the edges linking the two subsets

is maximized.

To solve the MaxCut problem, one approach is to formulate it as an integer quadratic programming problem, where the goal is to find a solution $\vec{s} \in \{-1, +1\}^N$ that maximizes the cost function

$$C(\vec{s}) = \frac{1}{2} \sum_{\langle j, k \rangle \in \tilde{E}} w_{jk} (1 - s_j s_k). \quad (1)$$

Here, the solution $\vec{s}$ defines a partition of the graph G , with each node j is labeled by $s_j \in \{-1, +1\}$ indicating which of the two subsets it belongs to. Approximation algorithms, such as the Goemans-Williamson algorithm [12], are employed to find near-optimal solutions efficiently. The Goemans-Williamson algorithm gives the highest known approximation ratio of $r \approx 0.87856$ among all other classical algorithms for generic graphs [12], while the bound becomes $r \approx 0.9326$ for u3R graphs [48], where $r = C(\vec{s})/C_{\max}$ and $C_{\max}$ is the maximum total cost of the cut.

Quantum computing-based methods can also solve the MaxCut problem, which are designed to search for a quantum state that maximize the quantum cost function $-\langle H_C \rangle$. The diagonal Hamiltonian H_C for the MaxCut problem on graph G is given by

$$H_C = - \sum_{\langle j, k \rangle \in \tilde{E}} \frac{w_{jk}}{2} (I - Z_j Z_k), \quad (2)$$

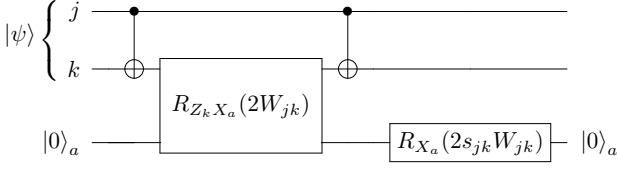
where I is identity operator for N qubit Hilbert space, Z_j and Z_k are Pauli-Z operators acting on qubits corresponding to vertices j and k . Ideally, we expect the quantum algorithm to return a quantum state $|\psi\rangle$ as a superposition of computational basis $|\vec{z}\rangle$, where $\vec{z} \in \{0, 1\}^N$ is a binary string representing a solution of the problem. In this context, each state $|\vec{z}\rangle$ is an eigenvector of $Z_j Z_k$ with eigenvalue ± 1 for any edge $\langle j, k \rangle$, and the quantum cost function $C(\vec{z}) = -\langle \vec{z} | H_C | \vec{z} \rangle$ has a similar structure to the classical cost function $C(\vec{s})$ in the integer quadratic programming problem.

B. Quantum Approximate Optimization Algorithm

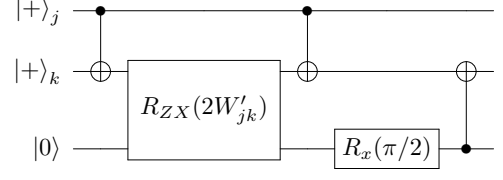
The Quantum Approximate Optimization Algorithm (QAOA) is a parameterized quantum algorithm designed to tackle general combinatorial optimization problems, including MaxCut. A p -level QAOA utilizes two parameterized unitary operators, $U(\gamma) = e^{-i\gamma H_C}$ and $U(\beta) = e^{-i\beta H_B}$, to construct a wavefunction for the graph $G = (V, E)$,

$$|\psi(\vec{\beta}, \vec{\gamma})\rangle = e^{-i\beta_p H_B} e^{-i\gamma_p H_C} \dots e^{-i\beta_1 H_B} e^{-i\gamma_1 H_C} |+\rangle^{\otimes N}. \quad (3)$$

Here, $H_B = \sum_{j=1}^N X_j$ is called the mixing Hamiltonian, where X_j are Pauli-X operators acting on qubit j . The expectation value of H_C can be obtained by repeated



(a) RBM inspired block encoding requires post-selection of the ancillary qubit on $|0\rangle$.



(b) Conditional operation based on the ancilla qubit for the physical initial state $|+\rangle_j \otimes |+\rangle_k$ removes the need for post-selection.

FIG. 1: Block-encoding-based circuit implementations for the imaginary-time propagator $e^{-\tau w_{jk} Z_j Z_k / 2}$ with (a) a generic input state $|\psi\rangle$ and (b) the input state $|+\rangle_j \otimes |+\rangle_k$. Circuit parameters are given in Eq. (8) and Eq. (A4).

These block encodings can be applied for a generic Pauli string propagator $e^{-K \prod_i \sigma_i}$.

measurements of the quantum system in the computational basis,

$$\langle H_C(\vec{\beta}, \vec{\gamma}) \rangle = \langle \psi(\vec{\beta}, \vec{\gamma}) | H_C | \psi(\vec{\beta}, \vec{\gamma}) \rangle = \sum_{\vec{z}} \text{Pr}(\vec{z}) \langle \vec{z} | H_C | \vec{z} \rangle, \quad (4)$$

where $\text{Pr}(\vec{z})$ is the probability of measuring the computational basis state $|\vec{z}\rangle$.

To find the optimal parameters $(\vec{\beta}, \vec{\gamma})$, one typically starts from an initial guess and then performs numerical optimization to maximize $\langle H_C(\vec{\beta}, \vec{\gamma}) \rangle$ using feedback from the quantum computer. This iterative process continues until the maximum cost function value is found. A good choice of initial parameters can significantly reduce the computational cost of optimization. Heuristic strategies have been proposed to optimize $\langle H_C(\vec{\beta}, \vec{\gamma}) \rangle$ with respect to parameters $(\vec{\beta}, \vec{\gamma})$ using a good initial guess [34]. Moreover, instead of searching on an instance-by-instance basis, the optimal parameters can be determined in advance for some special cases, including $p \leq 8$ QAOA for Sherrington-Kirkpatrick (SK) model [31] and $p \leq 17$ QAOA for MaxCut on large-girth D -regular graphs [49]. These values have been adopted in some software packages for QAOA simulation such as QOKit from JPMorgan Chase [50].

C. ITE-BE Method

Imaginary-time evolution (ITE) is a widely used approach to determine the ground state of a Hamiltonian H , by evolving the quantum system through the imaginary-time propagator

$$|\psi(\tau)\rangle = e^{-\tau H} |\psi(0)\rangle, \tau > 0. \quad (5)$$

As long as the initial state $|\psi(0)\rangle$ is not orthogonal to the ground state and τ is sufficiently large.

The ITE method can also be applied to classical optimization problems if the classical cost function can be expressed as the expectation value of a problem Hamiltonian. For example, the ITE method can solve the

MaxCut problem by finding ground states of Hamiltonian H_C , which maximize the MaxCut cost function. In this case, the imaginary-time propagator is given by

$$e^{-\tau H_C} = e^{\tau \sum_{\langle j,k \rangle} \frac{w_{jk}}{2} I} e^{-\tau \sum_{\langle j,k \rangle} \frac{w_{jk}}{2} Z_j Z_k}. \quad (6)$$

Here, $e^{\tau \sum_{\langle j,k \rangle} \frac{w_{jk}}{2} I}$ is a global phase factor that can be ignored, and the second term can be exactly decomposed as

$$e^{-\tau \sum_{\langle j,k \rangle} \frac{w_{jk}}{2} Z_j Z_k} = \prod_{\langle j,k \rangle} e^{-\tau \frac{w_{jk}}{2} Z_j Z_k}, \quad (7)$$

which does not incur Trotter error because all $Z_j Z_k$ operators in the sum on the left-hand side commute with each other. Therefore, we can choose a sufficiently large time step τ to ensure convergence. Since each term on the right-hand side is non-unitary, we implement them using a block-encoding-based method proposed in Ref. [29]. The circuit block for $e^{-\tau \frac{w_{jk}}{2} Z_j Z_k}$ is shown in Fig. 1a, and the parameters in the circuit block can be determined analytically by

$$W_{jk} = \frac{1}{2} \cos^{-1} [\exp(-|\tau w_{jk}|)], \quad (8)$$

$$s_{jk} = \text{sign}(W_{jk}).$$

In this approach, data will be discarded unless the ancilla qubit is measured to be $|0\rangle$. If the input state to a certain circuit block is $|+\rangle_j \otimes |+\rangle_k$, we can modify the block encoding parameters to deterministically apply the imaginary time propagator, see Fig. 1b. In particular, the last CNOT gate could “fix” the data qubit if ancilla is measured to be $|1\rangle$, so no post-selection is needed for this circuit block. We refer readers to Appendix A for further details. For a given undirected graph $G = (V, E)$, the ITE-BE method requires $|V| + 1$ qubits if reusing ancilla qubit for each block in Fig. 1 and a circuit with $O(|E|)$ depth according to Eq. (7).

III. PROTOCOLS AND EVALUATION

In this section, we design two protocols built upon the imaginary-time evolution to solve the MaxCut problem. The first protocol is a purely ITE-BE method starting with initial state $|+\rangle^{\otimes N}$. In this case, we adopt the circuit shown in Fig. 1b for the first layer, which improves the post-selection rate, and apply the circuit in Fig. 1a for the subsequent layers. The second protocol starts with the QAOA circuit, followed by ITE-BE, see Fig. 2. In essence, this method uses a QAOA circuit to prepare an initial state with a non-minimal overlap with the ground state. Then, the state is evolved through ITE-BE to significantly enhance the overlap.

Both methods can be applied to general settings. As proof of concept, we numerically evaluate the performance of these protocols on unweighted 3-regular (u3R) graphs with vertex number $6 \leq N \leq 12$. We finally report the 99.7% uncertainty intervals for approximation ratio r , the probability of obtaining the optimal solution $p_{\text{opt}} = N_{\text{opt}}/N_{\text{tot}}$ and the post-selection rate. Here, N_{tot} is the number of measurements made on data qubits after the post-selection of the ITE-BE method, and N_{opt} is the number of times one of the optimal solutions is obtained. Both metrics estimate the average quality of an algorithm, but p_{opt} is more conservative because it directly reflects the algorithm's ability to find exact solutions. The post-selection rate denotes the success rate of the block encoding procedure, which is the probability in getting all $|0\rangle$ state for the ancilla qubits.

A. ITE-BE Only

Our first method is purely based on the imaginary time evolution. Since all $Z_j Z_k$ terms commute in the sum on the left-hand side of Eq. (7), we can rearrange the order of multiplication on the right-hand side of the same equation and separate them into two parts V_1 and V_2 . For the first part, we define $V_1 = \prod_{\langle j,k \rangle} e^{-\tau \frac{w_{jk}}{2} Z_j Z_k}$ using edges $\langle j,k \rangle$ that form a maximal matching in the graph, i.e., a set of disjoint edges without common vertices that includes as many nodes as possible. All imaginary propagators in V_1 can be implemented by the circuit with correction shown in Fig. 1b when the protocol starts with initial state $|+\rangle^{\otimes N}$. The second part consists of all remaining edges in the graph, and they can be only implemented using the circuit in Fig. 1a as the input state of each block is no longer $|+\rangle_j \otimes |+\rangle_k$. We have verified that this choice enhances post-selection success. To reduce the overhead of this approach, ancilla qubits for each block can be reused by performing mid-circuit measurements and resetting them to $|0\rangle$. In this case, the current shot terminates and a new shot begins if any of the ancilla qubits in the second part is measured to be $|1\rangle$. All data qubits are measured only after all circuit blocks for the imaginary time propagator $e^{-\tau H_C}$ have been executed.

We then apply this method to solve the MaxCut problem on 100 randomly u3R graph instances generated by `networkx` Python package with fixed N , where $w_{jk} = 1$ for any j, k in edges E and each node is connected exactly to three other nodes. The imaginary time τ is varied from 0 to 2 which allows us to observe how the results evolve until they are sufficiently close to the ideal solutions. For each τ , we run the ITE-BE circuit with 10^5 shots for each individual graph and calculate the approximation ratio, the probability of obtaining the optimal solution, and the post-selection success rate using data from successful executions. We repeat this process 10 times to accumulate enough samples and reduce statistical fluctuations. Finally, we exclude graphs with at least one overall failure (i.e., all 10^5 shots discarded) in 10 repetitions, and report 99.7% uncertainty intervals of above three metrics in Fig. 3. All quantum circuits were simulated with the `AerSimulator` backend provided by the `Qiskit` Python package [52].

Our simulation shows that, as expected, both the approximation ratio r and the probability of obtaining optimal solutions p_{opt} increase monotonically with imaginary time τ across all graph sizes and eventually saturate near the optimal value of 1.0 as τ approaches 2. Meanwhile, the post-selection success rate decreases exponentially with τ , which is consistent with the prediction in Ref. [29]. Such behavior is typical of repeated applications of block encoding due to the post selection probability being less than 1. If one chooses to employ a separate auxiliary qubit per Pauli string in the Hamiltonian, and post select at the end of all propagator applications, then one can resort to amplitude amplification [53] to improve the post-selection rate. In addition to the increased cost in terms of the number auxiliary qubits, this additional step would also increase the overall circuit depth. We note that smaller graphs converge more quickly to the ground state and exhibit higher post-selection success rates compared to larger graphs due to fewer terms in the Hamiltonian.

Further analysis shows that for bipartite graph, the ITE-BE method shows a faster convergence to the ground state and higher success rate compared to non-bipartite graphs of the same size. As a result, we do not consider bipartite graphs in Fig. 3 and provide a detailed investigation of the ITE-BE method on MaxCut problem for them in Appendix B.

B. QAOA + ITE-BE

Our second method combines the QAOA with the imaginary time evolution (see Fig. 2). In the first part of the protocol, we directly use near-optimal values $(\vec{\beta}, \vec{\gamma})$ presented in Ref. [49] without any further training of QAOA to prepare an input state for the ITE-BE part. The QAOA output state has a higher overlap with the ground state than the standard equal superposition state $|+\rangle^{\otimes n}$. As such we expect this protocol to improve the

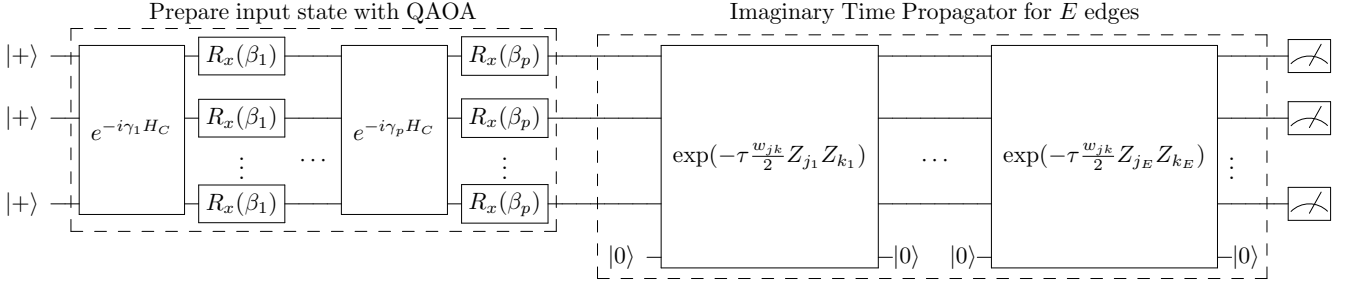


FIG. 2: The protocol of QAOA+ITE-BE method with a p -level QAOA circuit state preparation for ITE-BE rather than using the equal superposition state $|+\rangle^{\otimes n}$.

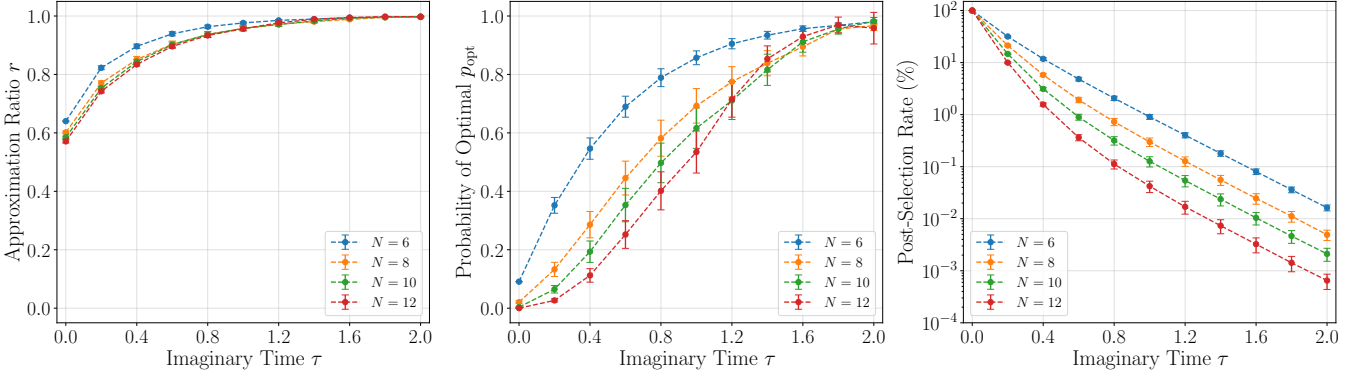


FIG. 3: The averaged approximation ratio r , probability of obtaining optimal solution p_{opt} , and the post-selection success rate for randomly generated unweighted 3-regular graphs within three standard deviations of the mean for ITE-BE method. Here, the approximation ratio starts from 0.5 at $\tau = 0$, which is expected according to Section 6.2.1 of Ref. [51]. This method guarantees convergence to the optimal solution, while the post-selection success rate decays exponentially. The convergence becomes slower and the success rate decreases when the problem size N increases.

post-selection success rate in the subsequent ITE-BE part. In the second part, we apply the ITE-BE method to further search for the optimal solution, where each imaginary time propagator for edge $\langle j, k \rangle$ is implemented by the circuit block in Fig. 1a. To implement the overall protocol for a given undirected graph $G = (V, E)$, we could reuse ancilla qubits for each circuit block in the ITE-BE part such that only $|V| + 1$ qubits is required. The data qubits are measured only after all circuit blocks have been executed. Also, the circuit depth of $e^{-i\gamma_k H_C}$ in state preparation is $O(|E|)$, so the circuit for overall protocol has $O(p|E| + |E|)$ depth.

In order to evaluate its performance, we apply this hybrid approach to solve the MaxCut problem on the same set of randomly generated unweighted 3-regular(u3R) graph instances as in Sec. III A, following the same settings. We report the 99.7% averaged uncertainty interval for the averaged approximation ratio r , probability of optimal p_{opt} and post-selection success rate as a function of imaginary time τ for graphs with $N = 12$ and different QAOA levels in Fig. 4.

Similar to the pure ITE-BE results, both r and p_{opt} increase monotonically with time τ , and the final results

always converge to the ground state at $\tau \sim 2$ regardless of the QAOA level. Also, the post-selection success rate decays exponentially with τ , but this trend is independent on the QAOA level. These observations suggest that a shallow QAOA+ITE-BE circuit with a lower p and a large τ is sufficient to obtain the optimal solution. Additional results for graphs with $N = 6, 8, 10$ lead to the same conclusion, and the data can be found in Appendix C. For bipartite graphs, this method demonstrates rapid convergence and high success rate which eventually reach a plateau. We do not consider bipartite graphs in Fig. 4, but study the performance on Appendix B.

It is worth noting that at $\tau = 0$, the imaginary time propagator does not act on the input state. Thus, the values of both r and p_{opt} at $\tau = 0$ reflect the performance of p -level QAOA with near-optimal parameter values from Ref. [49]. As shown in Fig. 4, the QAOA output state achieves a higher approximation ratio but a lower probability of obtaining optimal solution. This behavior indicates that the QAOA output state is a superposition of the ground state and multiple excited state, $|\psi\rangle = \sum_j c_j |\tilde{z}_j\rangle$. The presence of excited states lead to a

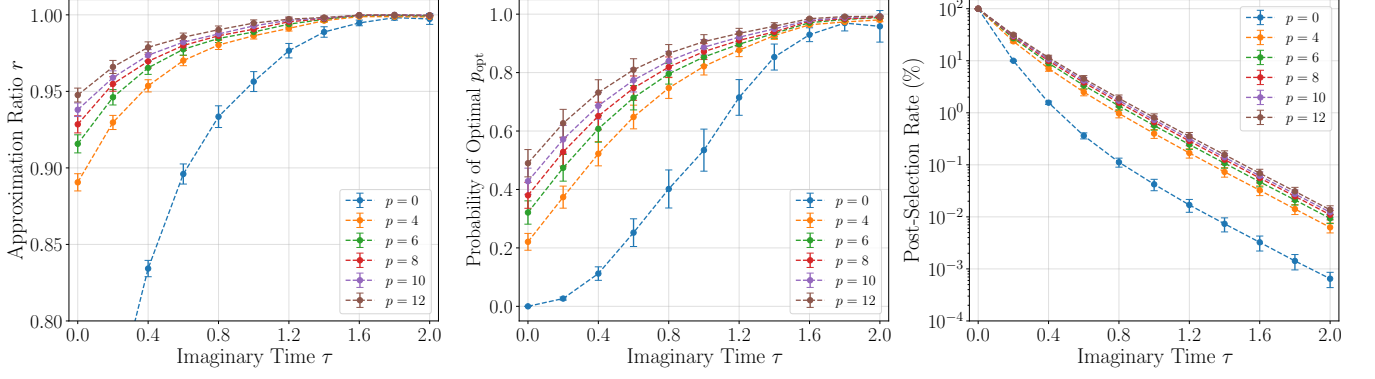


FIG. 4: The averaged approximation ratio r , probability of obtaining optimal solution p_{opt} , and the post-selection success rate within three standard deviations of the mean for the QAOA+ITE-BE method. Here we use p -level QAOA circuit for state preparation, where $p = 4, 6, 8, 10, 12$. The QAOA input state leads to faster convergence of ITE-BE and a post-selection rate that is one order of magnitude higher than using an equal superposition state to solve same graph for $p = 12$. As comparison, we show the results from purely ITE-BE method labeled as $p = 0$ in above figures, which indicates that the ITE-BE has faster convergence to the optimal solution and nearly 10 times higher post-selection rate when using p -level QAOA input state.

lower p_{opt} , but their total contribution to the cost function $\langle H_C \rangle$ is significantly smaller than that of the ground state, which results in a high approximation ratio. We can conclude that parameter values in Ref. [49] are tuned to maximize r rather than p_{opt} , and the ITE-BE layer significantly enhance performance of QAOA to achieve a higher p_{opt} while also improving r .

Another interesting comparison can be made on the post-selection rate as function of the initial state. Not only does the ITE-BE method improve the results of QAOA, but also the initial state provided by QAOA increases the post-selection rate with respect to the $|+\rangle^{\otimes n}$. Indeed, for $\tau = 2$ and $N = 12$ the post-selection rate increases by about an order of magnitude.

IV. CONCLUSION AND OUTLOOK

In this paper, we propose two quantum algorithms based on the block-encoding implementation of the imaginary time evolution propagator (ITE-BE) and evaluate their performance on the MaxCut problem. We conduct numerical simulations on randomly generated unweighted 3-regular graphs with vertex number from $N = 6$ to $N = 12$. Using circuits with N qubits and $O(|E|)$ depth, our results show that both ITE-BE based methods converge to the ground state when the imaginary time τ is sufficiently large. This performance surpasses that of the QAOA algorithm. In comparison to classical MaxCut solvers, current semi-definite programming approaches can solve dense graph instances (density up to 0.9) for up to $N = 100$ nodes (BiqMac [54]) or $N = 180$ nodes (BiqBin [55]). Moreover, modern linear programming solver can handle sparse graphs with hundreds or thousands of nodes [56], and even more than $N = 10,000$ nodes (McSparse [57]). Thus, demonstrating

quantum advantage would require access to more than 100 qubits. However, the guaranteed convergence of the ITE-BE based methods still indicates their promising potential for solving MaxCut problems on large graphs.

The success of QAOA+ITE-BE method emphasizes the importance of concatenating quantum algorithms in order to boost overall performance and overcome individual limitations. When preparing the initial state with a p -level QAOA for the ITE-BE layer, our results show a faster convergence of ITE-BE to the optimal solution and a post-selection rate that is an order of magnitude higher than using $|+\rangle^{\otimes n}$ as initial state. The result of QAOA+ITE-BE also reveals a trade-off between the QAOA level p and the imaginary time τ , namely, one can choose a lower QAOA level with a large τ to ensure convergence using a shallow circuit, or choose higher QAOA level to have quicker convergence with higher post-selection rate. Moreover, integrating QAOA with the ITE-BE method also benefits QAOA, as it avoids the need for additional training of QAOA with near-optimal parameters to achieve better accuracy. Instead, the subsequent ITE-BE layer enhances the result with only $O(|E|)$ circuit depth, where $|E|$ is the number of edges in the graph. This improvement is particularly significant when solving MaxCut on large graphs.

Our work focuses mainly on the MaxCut problem, but the principle of both the ITE-BE and QAOA+ITE-BE methods can be applied to other combinatorial optimization problems such as the Traveling Salesman problem or the Knapsack problem. The practical application of these methods is limited by the exponential decay in post-selection success rates with increasing imaginary time τ and graph size N . Therefore, developing methods to improve the post-selection success rate or enhance the convergence speed is a crucial research direction we leave for exploration in future work. By solving this challenge with

advanced algorithm designs, we will be able to realize the full potential of imaginary time evolution on optimization problems and achieve practical quantum advantage.

V. ACKNOWLEDGMENT

This research was supported by the U.S. Department of Energy (DOE) under Contract No. DE-AC02-05CH11231, through the National Energy Research Sci-

entific Computing Center (NERSC), an Office of Science User Facility located at Lawrence Berkeley National Laboratory. A.F. acknowledges funding by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research Quantum Testbed Program under contract DE-AC02-05CH11231. The authors acknowledge the Center for Advanced Research Computing (CARC) at the University of Southern California for providing computing resources that have contributed to the research results reported within this publication. URL: <https://carc.usc.edu>.

-
- [1] P. Shor, Algorithms for quantum computation: discrete logarithms and factoring, in *Proceedings 35th Annual Symposium on Foundations of Computer Science* (1994) pp. 124–134.
 - [2] L. K. Grover, Quantum mechanics helps in searching for a needle in a haystack, *Phys. Rev. Lett.* **79**, 325 (1997).
 - [3] A. Abbas *et al.*, Challenges and opportunities in quantum optimization, *Nature Reviews Physics* **10.1038/s42254-024-00770-9** (2024).
 - [4] F. Barahona, On the computational complexity of ising spin glass models, *Journal of Physics A: Mathematical and General* **15**, 3241 (1982).
 - [5] A. Lucas, Ising formulations of many np problems, *Frontiers in Physics* **2**, 10.3389/fphy.2014.00005 (2014).
 - [6] G. Kochenberger, J.-K. Hao, F. Glover, M. Lewis, Z. Lü, H. Wang, and Y. Wang, The unconstrained binary quadratic programming problem: a survey, *Journal of Combinatorial Optimization* **28**, 58 (2014).
 - [7] A. Mazumder and S. Tayur, Five starter problems: Solving quadratic unconstrained binary optimization models on quantum computers (2024), [arXiv:2401.08989 \[quant-ph\]](https://arxiv.org/abs/2401.08989).
 - [8] M. Streif, S. Yarkoni, A. Skolik, F. Neukart, and M. Leib, Beating classical heuristics for the binary paint shop problem with the quantum approximate optimization algorithm, *Phys. Rev. A* **104**, 012403 (2021).
 - [9] R. Shaydulin, C. Li, *et al.*, Evidence of scaling advantage for the quantum approximate optimization algorithm on a classically intractable problem, *Science Advances* **10**, eadm6761 (2024).
 - [10] M. Mezard, G. Parisi, and M. Virasoro, *Spin Glass Theory and Beyond* (WORLD SCIENTIFIC, 1986).
 - [11] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, Optimization by simulated annealing, *Science* **220**, 671 (1983).
 - [12] M. X. Goemans and D. P. Williamson, Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming, *J. ACM* **42**, 1115–1145 (1995).
 - [13] L. Vandenberghe and S. Boyd, Semidefinite programming, *SIAM Review* **38**, 49 (1996).
 - [14] P. C. Lothaw *et al.*, Scaling quantum approximate optimization on near-term hardware, *Scientific Reports* **12**, 12388 (2022).
 - [15] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm (2014), [arXiv:1411.4028 \[quant-ph\]](https://arxiv.org/abs/1411.4028).
 - [16] E. Farhi and A. W. Harrow, Quantum supremacy through the quantum approximate optimization algorithm (2019), [arXiv:1602.07674 \[quant-ph\]](https://arxiv.org/abs/1602.07674).
 - [17] S. Lloyd, Quantum approximate optimization is computationally universal (2018), [arXiv:1812.11075 \[quant-ph\]](https://arxiv.org/abs/1812.11075).
 - [18] E. Farhi, J. Goldstone, S. Gutmann, and H. Neven, Quantum algorithms for fixed qubit architectures (2017), [arXiv:1703.06199 \[quant-ph\]](https://arxiv.org/abs/1703.06199).
 - [19] D. Wecker, M. B. Hastings, and M. Troyer, Training a quantum optimizer, *Phys. Rev. A* **94**, 022309 (2016).
 - [20] 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).
 - [21] S. McArdle *et al.*, Variational ansatz-based quantum simulation of imaginary time evolution, *npj Quantum Information* **5**, 75 (2019).
 - [22] 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, *Nature Physics* **16**, 205 (2020).
 - [23] Y. Huang, Y. Shao, W. Ren, J. Sun, and D. Lv, Efficient quantum imaginary time evolution by drifting real-time evolution: An approach with low gate and measurement complexity, *Journal of Chemical Theory and Computation* **19**, 3868 (2023).
 - [24] C. Leadbeater, N. Fitzpatrick, D. M. Ramo, and A. J. W. Thom, Non-unitary trotter circuits for imaginary time evolution, *Quantum Science and Technology* **9**, 045007 (2024).
 - [25] S. Wang *et al.*, Noise-induced barren plateaus in variational quantum algorithms, *Nature Communications* **12**, 6961 (2021).
 - [26] Z. Holmes, K. Sharma, M. Cerezo, and P. J. Coles, Connecting ansatz expressibility to gradient magnitudes and barren plateaus, *PRX Quantum* **3**, 010313 (2022).
 - [27] L. Bittel and M. Kliesch, Training variational quantum algorithms is np-hard, *Phys. Rev. Lett.* **127**, 120502 (2021).
 - [28] W. van Dam, M. Mosca, and U. Vazirani, How powerful is adiabatic quantum computation?, in *Proceedings 42nd IEEE Symposium on Foundations of Computer Science* (2001) pp. 279–287.
 - [29] E. Rrapaj and E. Rule, Exact block encoding of imaginary time evolution with universal quantum neural networks, *Phys. Rev. Res.* **7**, 013306 (2025).
 - [30] J. Basso, D. Gamarnik, S. Mei, and L. Zhou, Performance and limitations of the qaoa at constant levels on large sparse hypergraphs and spin glass models, in *2022*

- IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)* (2022) pp. 335–343.
- [31] E. Farhi, J. Goldstone, S. Gutmann, and L. Zhou, The Quantum Approximate Optimization Algorithm and the Sherrington-Kirkpatrick Model at Infinite Size, *Quantum* **6**, 759 (2022).
- [32] G. E. Crooks, Performance of the quantum approximate optimization algorithm on the maximum cut problem (2018), [arXiv:1811.08419 \[quant-ph\]](#).
- [33] R. Shaydulin and M. Pistoia, Qaoawith $n \cdot p \geq 200$, in *2023 IEEE International Conference on Quantum Computing and Engineering (QCE)*, Vol. 01 (2023) pp. 1074–1077.
- [34] L. Zhou, S.-T. Wang, S. Choi, H. Pichler, and M. D. Lukin, Quantum approximate optimization algorithm: Performance, mechanism, and implementation on near-term devices, *Phys. Rev. X* **10**, 021067 (2020).
- [35] M. P. Harrigan *et al.*, Quantum approximate optimization of non-planar graph problems on a planar superconducting processor, *Nature Physics* **17**, 332 (2021).
- [36] G. C. Wick, Properties of bethe-salpeter wave functions, *Phys. Rev.* **96**, 1124 (1954).
- [37] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory* (Westview Press, 1995) reading, USA: Addison-Wesley (1995) 842 p.
- [38] F. Pederiva, A. Sarsa, K. Schmidt, and S. Fantoni, Auxiliary field diffusion monte carlo calculation of ground state properties of neutron drops, *Nuclear Physics A* **742**, 255 (2004).
- [39] D. Lonardonì, S. Gandolfi, J. E. Lynn, C. Petrie, J. Carlson, K. E. Schmidt, and A. Schwenk, Auxiliary field diffusion monte carlo calculations of light and medium-mass nuclei with local chiral interactions, *Phys. Rev. C* **97**, 044318 (2018).
- [40] J. Lee, H. Q. Pham, and D. R. Reichman, Twenty years of auxiliary-field quantum monte carlo in quantum chemistry: An overview and assessment on main group chemistry and bond-breaking, *Journal of Chemical Theory and Computation* **18**, 7024 (2022).
- [41] J. Gacon *et al.*, Variational quantum time evolution without the quantum geometric tensor, *Phys. Rev. Res.* **6**, 013143 (2024).
- [42] N. M. Bauer, R. Alam, J. Ostrowski, and G. Siopsis, Combinatorial optimization with quantum imaginary time evolution (2023), [arXiv:2312.16664 \[quant-ph\]](#).
- [43] Y. Huang *et al.*, Efficient quantum imaginary time evolution by drifting real-time evolution: An approach with low gate and measurement complexity, *Journal of Chemical Theory and Computation* **19**, 3868 (2023).
- [44] F. Turro, A. Roggero, V. Amitrano, P. Luchi, K. A. Wendt, J. L. Dubois, S. Quaglioni, and F. Pederiva, Imaginary-time propagation on a quantum chip, *Phys. Rev. A* **105**, 022440 (2022).
- [45] A. M. Childs and N. Wiebe, Hamiltonian simulation using linear combinations of unitary operations, *Quantum Info. Comput.* **12**, 901–924 (2012).
- [46] G. H. Low and I. L. Chuang, Optimal hamiltonian simulation by quantum signal processing, *Phys. Rev. Lett.* **118**, 010501 (2017).
- [47] A. Gilyén, Y. Su, G. H. Low, and N. Wiebe, Quantum singular value transformation and beyond: exponential improvements for quantum matrix arithmetics, in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing* (2019) p. 193–204.
- [48] E. Halperin, D. Livnat, and U. Zwick, Max cut in cubic graphs, *Journal of Algorithms* **53**, 169 (2004).
- [49] J. Basso *et al.*, The quantum approximate optimization algorithm at high depth for maxcut on large-girth regular graphs and the sherrington-kirkpatrick model, in *17th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2022)* (2022).
- [50] D. Lykov, R. Shaydulin, Y. Sun, Y. Alexeev, and M. Pistoia, Fast simulation of high-depth qaoa circuits, in *Proceedings of the SC '23 Workshops of The International Conference on High Performance Computing, Network, Storage, and Analysis*, SC-W 2023 (ACM, 2023).
- [51] M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data Analysis*, 2nd ed. (Cambridge University Press, USA, 2017).
- [52] A. Javadi-Abhari *et al.*, Quantum computing with qiskit (2024), [arXiv:2405.08810 \[quant-ph\]](#).
- [53] G. Brassard and P. Hoyer, An exact quantum polynomial-time algorithm for simon’s problem, in *Proceedings of the Fifth Israeli Symposium on Theory of Computing and Systems* (1997) pp. 12–23.
- [54] F. Rendl, G. Rinaldi, and A. Wiegeler, Solving max-cut to optimality by intersecting semidefinite and polyhedral relaxations, *Mathematical Programming* **121**, 307 (2010).
- [55] N. Gusmeroli *et al.*, Biqbin: A parallel branch-and-bound solver for binary quadratic problems with linear constraints, *ACM Trans. Math. Softw.* **48**, 10.1145/3514039 (2022).
- [56] D. Rehfeldt, T. Koch, and Y. Shinano, Faster exact solution of sparse maxcut and qubo problems, *Mathematical Programming Computation* **15**, 445 (2023).
- [57] J. Charfreitag, M. Jünger, S. Mallach, and P. Mutzel, Mcsparse: Exact solutions of sparse maximum cut and sparse unconstrained binary quadratic optimization problems, in *2022 Proceedings of the Symposium on Algorithm Engineering and Experiments (ALENEX)*, pp. 54–66, <https://epubs.siam.org/doi/pdf/10.1137/1.9781611977042.5>.

Appendix A: ITE Parameters

We provide here the general form of the ansatz for a single qubit (the same analysis applies to n-qubit Pauli string),

$$e^{-\tau K \sigma_r} = \mathcal{N} \langle 0_a | e^{-i(c\sigma_a^x + W\sigma_r\sigma_a^x)} | 0_a \rangle, \quad (\text{A1})$$

where W is the coupling between ancilla qubit a and the target qubit r . For a generic initial state, $|\Psi\rangle = \alpha|0\rangle + \beta|1\rangle$ with $|\beta|^2 = 1 - |\alpha|^2$, the ratio of the probability to measure $|0_a\rangle$ versus $|1_a\rangle$ is,

$$\frac{P(0_a)}{P(1_a)} = \left| \frac{\alpha \cos(c+W) + \beta \cos(c-W)}{\alpha \sin(c+W) + \beta \sin(c-W)} \right|^2. \quad (\text{A2})$$

In [29], the values of the parameters are,

$$\begin{aligned} \mathcal{N} &= \exp(|\tau K|)/2, \\ W &= \frac{1}{2} \cos^{-1}[\exp(-2|\tau K|)], \\ c &= Ws, \quad s = \text{sign}(K). \end{aligned} \quad (\text{A3})$$

When post-selection fails, one must repeat the operation. However, for special cases, one can correct instead of re-starting computations. For instance, instead of the parameters in the (A3), we can use the following parameters,

$$\begin{aligned} \mathcal{N} &= \frac{\sqrt{e^{4\tau K} + 1}}{2\sqrt{e^{2\tau K}}}, \\ W &= \tan^{-1}(e^{2\tau K}) - \pi/4, \\ c &= \pi/4. \end{aligned} \quad (\text{A4})$$

The state post operation is,

$$\begin{aligned} U(|\Psi\rangle \otimes |0_a\rangle) &= \frac{1}{\sqrt{e^{4K\tau} + 1}} (\alpha|0\rangle + \beta e^{2K\tau}|1\rangle) \otimes |0_a\rangle \\ &\quad - \frac{ie^{2K\tau}}{\sqrt{e^{4K\tau} + 1}} (\alpha|0\rangle + \beta e^{-2K\tau}|1\rangle) \otimes |1_a\rangle. \end{aligned} \quad (\text{A5})$$

Thus, when the ancilla is measure in the $|0_a\rangle$ state, the desired operation is performed in the system qubit, and when the ancilla is measure in the $|1_a\rangle$ state, the opposite operation is performed. If $\alpha = \beta$, we can flip the amplitudes with a CX gate, effectively correcting post-selection, and no measurement is required. The ancilla can be reset for re-use.

Appendix B: The MaxCut Solution of Bipartite Graphs

During the process of data analysis, we observe that both ITE-BE methods converge more quickly for some graphs than for others in our sample. Moreover, these graphs have higher post-selection success rate compared to other graph with same size, and the success rate reaches a plateau after a certain imaginary time τ . To understand the conditions under which the ITE-BE method converges more quickly, we compute the spectra of both the adjacency matrices and Hamiltonians of these special graphs. We found that the minimum eigenvalue of the adjacency matrices of these graphs is -3 , which is a characteristic feature of unweighted bipartite 3-regular graphs. In addition, the minimum eigenvalue of their Hamiltonian H_C is given by $(\frac{3N}{2} - 1)/2$, where $\frac{3N}{2}$ is the total number of edges in a 3-regular graph. This implies that the MaxCut solution is a partition where all edges in the graph connect the two parts of the partition. These observations conclusively identify the special graphs as bipartite 3-regular graphs.

To verify whether the ITE-BE method converges quickly on other bipartite 3-regular graphs, we randomly generate 40 instances of bipartite graphs with $6 \leq N \leq 12$ nodes, and solve the MaxCut problem using both ITE-BE and QAOA + ITE-BE methods. The numerical simulation setting and data analysis process are consistent with those described in the main text. The results are shown in Fig. 5 and Fig. 6.

It is worth noting that any bipartite graphs with non-negative weights and only one connected component, the solution to the MaxCut problem is unique up to a “flipping” of all vertices and corresponds to the bipartition of the graph. This is because, in a bipartite graph, all edges connect vertices from one part to the other, so the total weight of the edges crossing the bipartition is maximized and cannot be increased further. In this case, the MaxCut solution can be found using classical algorithms such as breadth-first search (BFS) or depth-first search (DFS) with linear time complexity $O(V + E)$.

Appendix C: QAOA+ITE-BE Results for Graphs with 6, 8, 10 Nodes

In this appendix, we show the results of the QAOA+ITE-BE method on u3R graph instances with $N = 6, 8$ and 10 in Fig. 7. Due to the optimal parameters for QAOA, at $\tau = 0$, the approximation ratio is already above 0.9 and further improves with ITE-BE. The p_{opt} at $\tau = 0$, on the other hand, is smaller in value and decreases significantly as the system become larger, compared to approximation ratio, for the optimal QAOA. The application of the ITE-BE layer increases the overlap with the ground state, as expected.

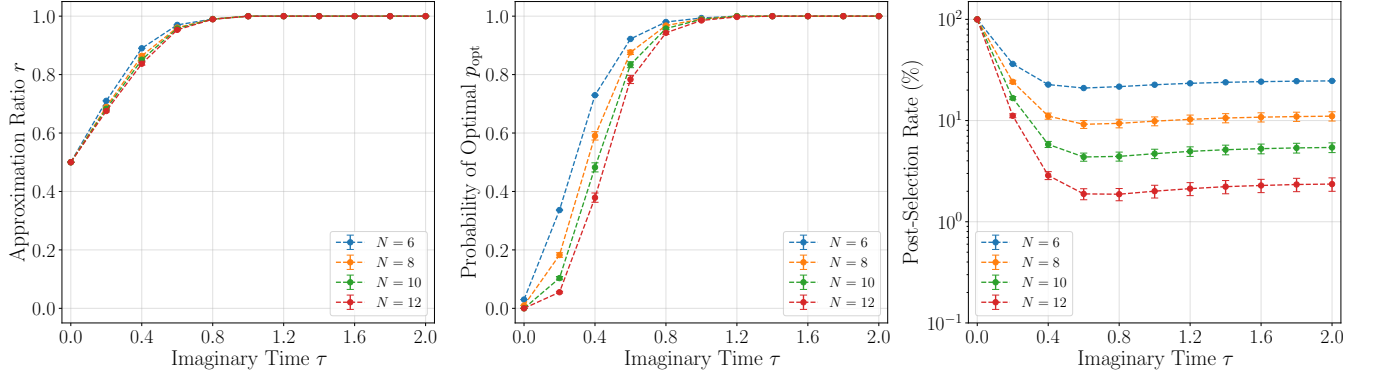


FIG. 5: The averaged performance of pure ITE-BE method on 40 randomly generated bipartite 3-regular graphs for $N = 6, 8, 10$ and 12 . The ITE-BE method converges faster than non-bipartite graphs and the success rate reaches a plateau after imaginary time $\tau \sim 0.6$.

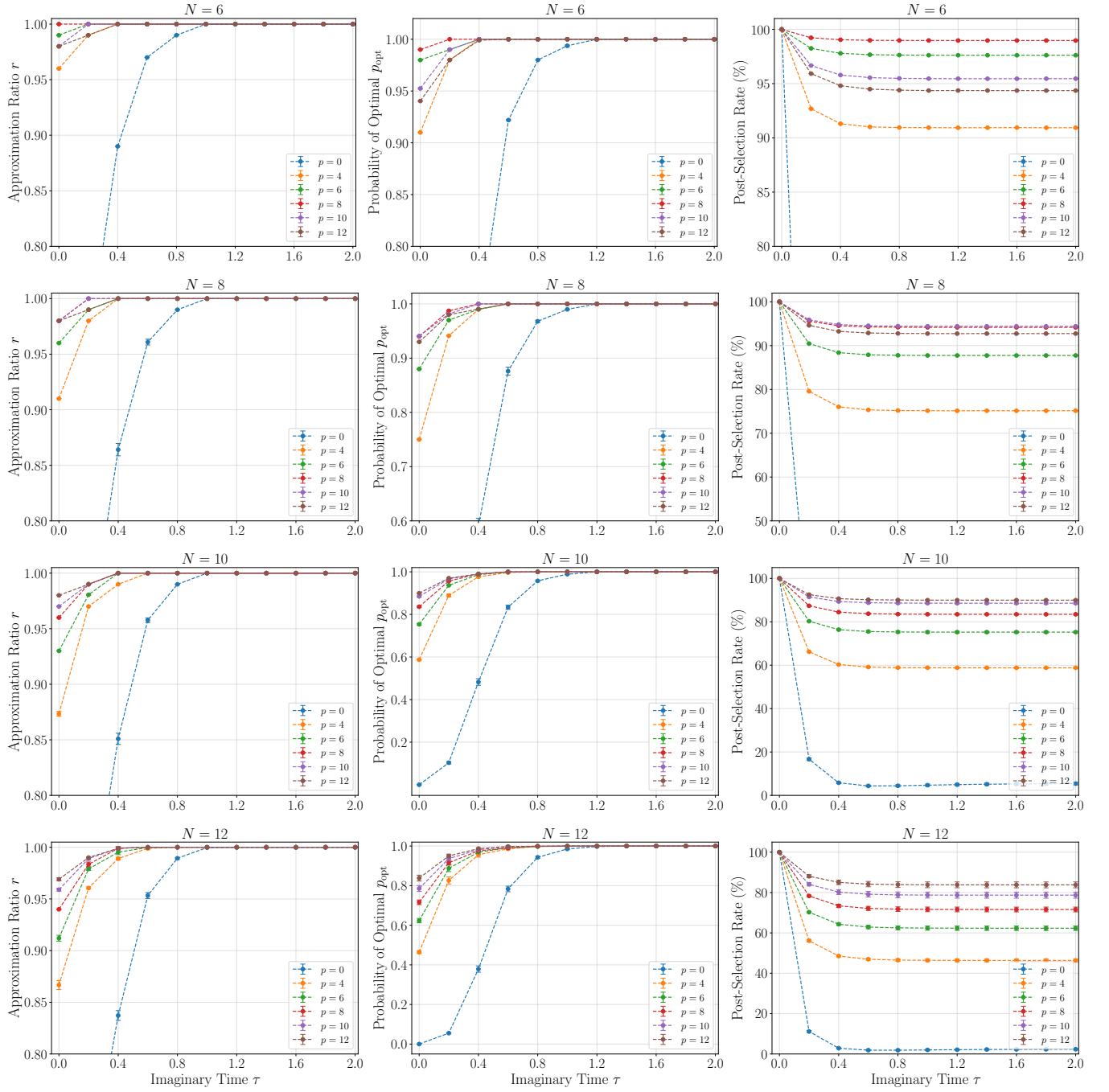


FIG. 6: The averaged performance of QAOA+ITE-BE method on 40 randomly generated bipartite 3-regular graphs for $N = 6, 8, 10$ and 12 . In this case, the QAOA initial state is very close to the optimal solution so the ITE-BE method have only smaller improvement. The success rate reaches a plateau after imaginary time $\tau \sim 0.4$. The $p = 0$ curves give the results from purely ITE-BE method on same bipartite graphs.

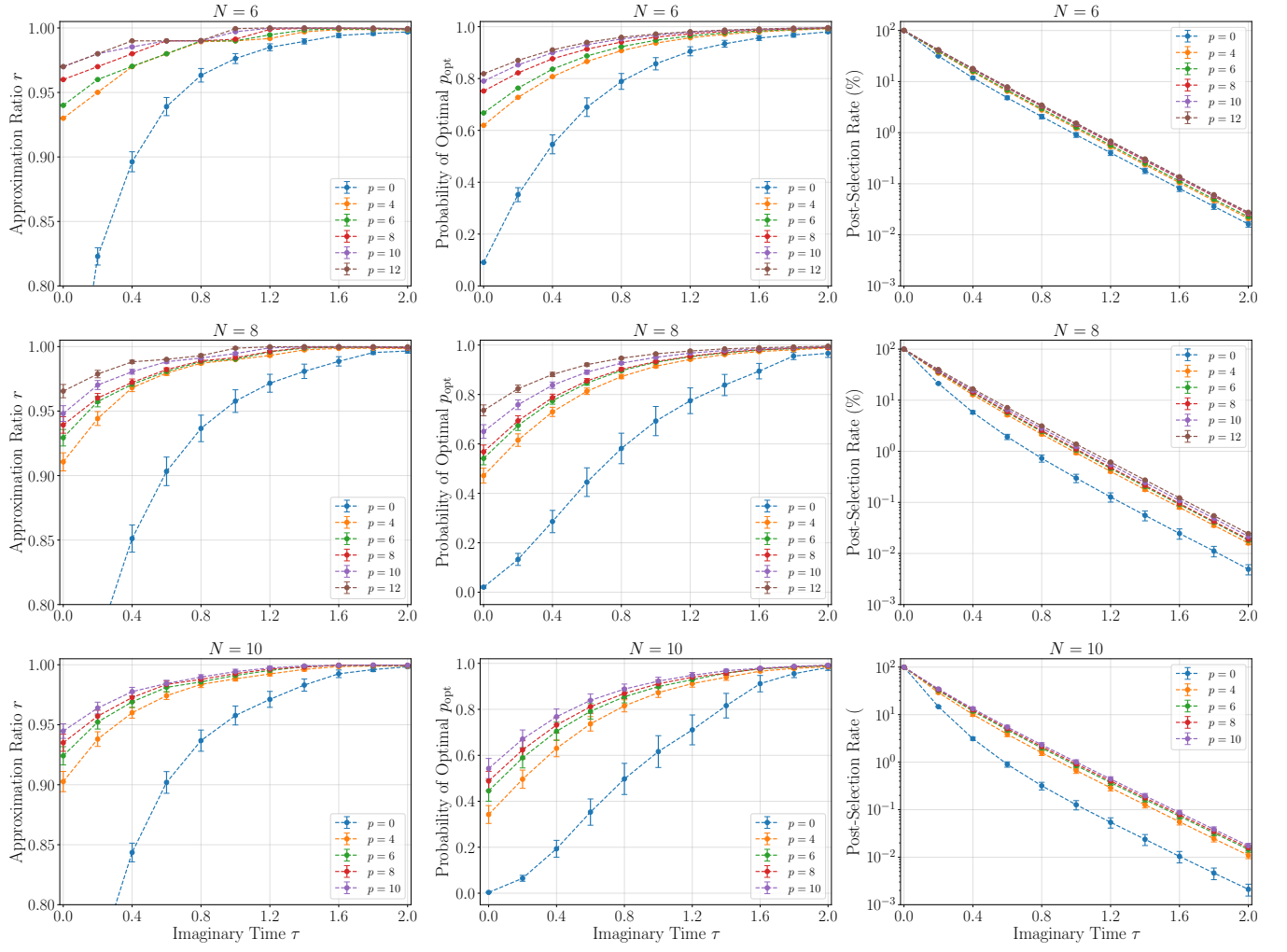


FIG. 7: The averaged performance of QAOA+ITE-BE method on 100 randomly generated u3R graphs for $N = 6, 8, 10$. The ITE-BE layer further enhances the performance of QAOA especially for large graphs. We also show results obtained from purely ITE-BE method ($p = 0$) on same u3R graph instances as a comparison.