Classically-inspired Mixers for QAOA Beat Goemans-Williamson’s Max-Cut at Low Circuit Depths

Reuben Tate, 1 Bryan Gard, 2 Greg Mohler, 2 and Swati Gupta 1, ∗

1 Georgia Institute of Technology, Atlanta, GA, USA
2 Georgia Tech Research Institute, Atlanta, GA 30332, USA
(Dated: December 22, 2021)

We generalize the Quantum Approximate Optimization Algorithm (QAOA) of Farhi et al. (2014) to allow for arbitrary separable initial states and corresponding mixers such that the starting state is the most excited state of the mixing Hamiltonian. We demonstrate this version of QAOA by simulating Max-Cut on weighted graphs. We initialize the starting state as a warm-start inspired by classical rank-2 and rank-3 approximations obtained using Burer-Monteiro’s heuristics, and choose a corresponding custom mixer. Our numerical simulations with this generalization, which we call QAOA-Warmest, yield higher quality cuts (compared to standard QAOA, the classical Goemans-Williamson algorithm, and a warm-started QAOA without custom mixers) for an instance library of 1148 graphs (upto 11 nodes) and depth \( p = 8 \). We further show that QAOA-Warmest outperforms the standard QAOA of Farhi et al. in experiments on current IBM-Q hardware.

I. INTRODUCTION

In order to realize a quantum advantage, many researchers have been considering the usage of NISQ devices 1, 2 for the purposes of solving difficult problems in combinatorial optimization. Of particular interest is the Quantum Approximate Optimization Algorithm (QAOA), a hybrid quantum-classical algorithm developed by Farhi et al. that can be applied to a large variety of combinatorial optimization problems 3. We study the use of QAOA to solve one of the most famous NP-hard combinatorial optimization problems, called Max-Cut. Given a weighted graph \( G = (V,E) \), with vertex set \( V = [n] \), edge set \( E \subseteq \binom{V}{2} \) and weights \( w : E \rightarrow \mathbb{R} \), the Max-Cut problem is to find a partition of \( V \) into two disjoint sets \( S, V \setminus S \subseteq V \), such that the total weight of the edges across the partition, i.e. \( \text{cut}(S) := \sum_{e \in E} w_e \cdot 1[e \in S \times (V \setminus S)] \), is maximized. The Max-Cut of \( G \) is denoted by \( \text{Max-Cut}(G) = \max_{S \subseteq V} \text{cut}(S) \).

In the standard QAOA algorithm, qubits are initialized in the \( |+ \rangle \) state along the \( x \)-axis of the Bloch sphere, tensorized \( n \) times, and QAOA’s mixing Hamiltonian rotates each qubit about this same axis. It is not difficult to show that the initial state of standard QAOA is the highest energy eigenstate of the mixing Hamiltonian; a necessary condition needed to prove QAOA’s convergence towards the optimal cut 4.

In recent work, Tate et al. 4 and Egger et al. 5 explored using classical algorithms to specify the initial state for QAOA. Following the classical optimization literature (6–8), we will refer to these classically-inspired initializations as warm starts. The approach by Tate et al. 4 considered warm-starting QAOA using rank-2 and rank-3 Burer-Monteiro locally optimal solutions for Max-Cut; however, their method plateaus even at low circuit depths of \( p = 1 \) for some initializations and is unable to improve the cut quality for some instances. Egger et al. 5 similarly considered warm-starts for quadratic unconstrained binary optimization (QUBO) problems; however, they only focus on rank-1 solutions from classical linear programming relaxations for QUBO. Another method by Egger et al. initializes the quantum state via a particular cut in the graph; this QAOA variant can recover the value of the cut but not much is known regarding the convergence properties of this approach as the circuit depth increases.

In this work, we further explore modifying the ansatz of QAOA. Farhi et al. 9 consider having separate variational parameters for each vertex and edge. Hadfield et al. 10 and Wang et al. 11 consider versions of QAOA that are suitable for combinatorial optimization problems with both hard constraints (that must be satisfied) and soft constraints (for which we want to minimize violations). Zhu et al. 12 modify QAOA such that the ansatz is expanded in an iterative fashion with the mixing Hamiltonian being allowed to change between iterations.

Our work generalizes the standard QAOA algorithm by considering any separable configuration of qubits states (motivated by warm-starts) and mod-
ifying the mixing Hamiltonian to keep it in the eigenspace of those states. This mixer modification, which we call QAOA-warmest, causes the initial state to be the highest energy eigenstate of the mixer. QAOA-warmest allows us to begin QAOA with a biased superposition of cuts.

We show in numerical simulation that when the initial state is classically-inspired (e.g. using Burer and Monteiro’s relaxation \cite{13}), QAOA-Warmest outperforms the classical Goemans-Williamson algorithm and standard QAOA even at low-circuit depths. This result does not hold for any random initializations on the Bloch sphere, but rather for classically-inspired warm-start solutions. Additionally, we conjecture that QAOA-Warmest is able to converge to the optimal Max-Cut with infinite circuit depth. We partially resolve the conjecture for separable states where each qubit’s position on the Bloch sphere has positive $x$-component and lies on the intersection of the Bloch sphere and the $xz$-plane; we show that an adiabatic argument (similar to the one used for standard QAOA) indeed applies for these initial states.

A. Background

1. The Quantum Approximate Optimization Algorithm

First, we review the Quantum Approximate Optimization Algorithm (QAOA) in the context of the Max-Cut problem. QAOA is performed on $n$ qubits (with $n$ being the number of vertices in the input graph) with the final measurement being a bitstring of length $n$ corresponding to a cut $(S, V \setminus S)$ with $S = \{i : i$-th bit is 0$\}$.

For depth-$p$ QAOA, the quantum circuit alternates between applying a cost Hamiltonian $H_C = \frac{1}{2} \sum_{(i,j) \in E} w_{ij} (1 - \sigma_i^x \sigma_j^x)$ and a mixing Hamiltonian $H_B = \sum_{i \in [n]} \sigma_i^z$. Here, $\sigma^x, \sigma^y, \sigma^z$ are the standard Pauli operators and $\sigma_i^k$ is the Pauli operator applied to qubit $i$ for $k \in \{x, y, z\}$. The cost and mixing Hamiltonians are applied a total of $p$ times to generate a variational state,

$$|\psi_p(\gamma, \beta)\rangle = e^{-i\beta_p H_B} e^{-i\gamma_p H_C} \cdots e^{-i\beta_1 H_B} e^{-i\gamma_1 H_C} |s_0\rangle,$$

(1)

where $|s_0\rangle$ is the initial state and $\gamma = (\gamma_1, \ldots, \gamma_p), \beta = (\beta_1, \ldots, \beta_p)$ are variational parameters to be optimized. For standard QAOA, the initial state is given by $|s_0\rangle = |+\rangle^\otimes n$ which corresponds to an equal superposition of all $2^n$ possible cuts in the graph.

Finally, sampling from $|\psi_p(\gamma, \beta)\rangle$ yields a bitstring corresponding to a cut in the graph. If, for each $p$, we choose $\gamma, \beta$ optimally, then the expected value of the cut tends to the Max-Cut as $p \to \infty \ [3]$.

2. Classical Optimization Algorithms

We next review some classical optimization algorithms for Max-Cut. It is useful to reformulate the Max-Cut problem as follows:

$$\text{Max-Cut}(G) = \frac{1}{4} \max_{x \in \{\pm 1\}^n} \sum_{(i,j) \in E} w_{ij} (x_i - x_j)^2.$$

Given a solution $x$ to the maximization above, this yields the cut $(S, V \setminus S)$ with $S = \{i : x_i = 1\}$.

For positive integer $k$, the above formulation can then be relaxed as follows:

$$\max \frac{1}{4} \sum_{(i,j) \in E} w_{ij} \|x_i - x_j\|^2$$

subject to $\|x_i\| = 1, \ \forall i \in V,$

$$x_i \in \mathbb{R}^k, \ \forall i \in V.$$

Note that any solution to the above objective takes the form of $n$ points on the $(k-1)$-dimensional unit sphere.

When $k = n$, the problem can be reformulated as a convex semidefinite program (SDP) which is the formulation used by the seminal Goemans-Williamson (GW) algorithm \cite{14}, which yields a 0.878-approximation to Max-Cut for graphs with non-negative edge weights.

Burer and Monteiro proposed solving the relaxation for $1 < k < n$ (which we denote by BM-MC$_k$), using parametric forms for points on the $k-1$-dimensional sphere \cite{13}. This leads to a program that is no longer convex and thus one can not expect to easily find the global optima. However, high-quality local optima can be found by utilizing first and second order optimization methods. In general, the Burer-Monteiro technique has been found to work well in practice, even when $k = 2 \ [15]$.

II. RESULTS

A. Custom Mixers

For any separable initial state, we show how to construct a custom mixing Hamiltonian with the
property that the initial state is the most excited state of the mixer. Moreover, such a customized mixer has a geometric interpretation: the mixer is a collection of independent single-qubit rotations about the initial Bloch sphere positions of each qubit. Although our paper focuses on Max-Cut, the approach below, which we call QAOA-Warmest, can be applied to any suitable combinatorial optimization problem. In what follows, we describe QAOA-Warmest in detail and discuss some of its properties.

1. Description of Custom Mixer

The custom mixing Hamiltonian $H_B$ is defined so that in the $k$th stage of the QAOA circuit, one independently rotates the $j$th qubit about the axis determined by its original position by angle $2\beta_j$. We explain how to construct such an $H_B$ below.

Consider any unentangled state $|s_0\rangle$ composed of a completely separable state of $n$ qubits on the Bloch sphere, i.e., $|s_0\rangle$ can be written in the form:

$$|s_0\rangle = \bigotimes_{j=1}^{n} |s_{0,j}\rangle,$$

where for $j = 1, \ldots, n$,

$$|s_{0,j}\rangle = \cos(\theta_j/2) |0\rangle + e^{i\phi_j} \sin(\theta_j/2) |1\rangle.$$

Here, $\theta_j$ and $\phi_j$ can be interpreted as the polar and azimuthal angle of the $j$th qubit on the Bloch sphere respectively. The position of the $j$th qubit on the Bloch sphere can also be described in Cartesian coordinates $\hat{n}_j = (x_j, y_j, z_j)$ via the following transformation from spherical to Cartesian coordinates:

$$x_j = \sin \theta_j \cos \phi_j,$$
$$y_j = \sin \theta_j \sin \phi_j,$$
$$z_j = \cos \theta_j.$$

Given the Cartesian coordinates, one can define the custom mixing Hamiltonian $H_B$ as follows:

$$H_B = \bigoplus_{j=1}^{n} H_{B,j},$$

where $H_{B,j} = x_j \sigma^x + y_j \sigma^y + z_j \sigma^z$.

To develop a geometrical understanding of the custom mixer, consider the operator $R_{\hat{n},j}(\alpha)$ that rotates the $j$th qubit by angle $\alpha$ about the $\hat{n}$-axis for some unit vector $\hat{n} = (x, y, z)$; such as operation can be written as:

$$R_{\hat{n},j}(\alpha) = \exp(-i\frac{\alpha}{2}(x \sigma^x + y \sigma^y + z \sigma^z)).$$

Recall that for the $k$th of the $p$ stages of the QAOA circuit (where $p$ is the circuit depth), one applies the unitary operator $e^{-i\beta_k H_B}$ with $\beta_k$ being a variational parameter (to be optimized); this operator, $e^{-i\beta_k H_B}$, can be written as $\prod_{j=1}^{n} R_{\beta_k,j}(2\beta_k)$, i.e., rotations about each qubit’s original position.

The standard QAOA [3], is therefore, a special case of our custom mixer approach as the standard starting state has each qubit on the $x$-axis (with Cartesian coordinates $(x_j, y_j, z_j) = (1, 0, 0)$) and with the standard mixer $H_B = \sum_{j=1}^{n} \sigma_j^z$, the unitary operator $e^{-i\beta_k H_B}$ corresponds to rotations (by $2\beta_k$) about the $x$-axis. Additionally, we also recover the results of Egger et al. [3] by restricting the initialization to the $xz$-plane with $x > 0$.

2. Eigenstates of Custom Mixers

As previously discussed, convergence guarantees that the standard QAOA finds the Max-Cut with infinite circuit depth and requires the initial state of QAOA to also be the most excited state of the mixing Hamiltonian. We show that this is also the case for custom mixers. We first show that the result holds for a single qubit, and then generalize to the Kronecker sums of matrices. We include proofs of all the statements in Appendix A.

**Lemma 1.** Let $|s\rangle = \cos(\theta/2) |0\rangle + e^{i\phi} \sin(\theta/2) |1\rangle$ be a single-qubit quantum state and let $\hat{n} = (x, y, z)$ be the Cartesian coordinates of that qubit on the Bloch sphere. Let $U = x \sigma^x + y \sigma^y + z \sigma^z$. Then $|s\rangle$ is the most-excited eigenstate of $U$.

We can then formulate the most-excited eigenstate of $U$ using the following relation between eigenvalues of matrices involved in a Kronecker sum and the resultant matrix.

**Theorem 2.** (Theorem 13.16 in [10]) Let $A \in \mathbb{C}^{n \times n}$ have eigenvalues $\lambda_1, \ldots, \lambda_n$ and let $B \in \mathbb{C}^{m \times m}$ have eigenvalues $\mu_1, \ldots, \mu_m$. Then the Kronecker sum $A \oplus B$ has $mn$ eigenvalues given by $\{\lambda_i + \mu_j : i \in [n], j \in [m]\}$. Moreover, if $x_1, \ldots, x_p (p \leq n)$ are linearly independent eigenvectors of $A$ corresponding to $\lambda_1, \ldots, \lambda_n$ and $z_1, \ldots, z_q (q \leq m)$ are linearly independent eigenvectors of $B$ corresponding to $\mu_1, \ldots, \mu_q$, then, for all $i \in [p]$ and $j \in [q]$, we have that $x_i \odot z_j$ are linearly independent eigenvectors of $A \oplus B$ corresponding to $\lambda_i + \mu_j$.

Suppose for each $j = 1, \ldots, n$ we have a matrix $A_j$ with real eigenvalues and suppose the largest
eigenvalue of $A_j$ is $\lambda_j$ with corresponding eigenvector $v_j$. As a consequence of Theorem 2, we have that the largest eigenvalue of $\bigoplus_{j=1}^n A_j$ is $\sum_{j=1}^n \lambda_j$ with one corresponding eigenvector being $\bigoplus_{j=1}^n v_j$.

Letting $A_j = H_{B,j}$ and $v_j = |s_{0,j}\rangle$, we see that $|s_0\rangle = \bigotimes_{j=1}^n |s_{0,j}\rangle$ is a highest-energy eigenstate for $H_B = \bigoplus_{j=1}^n H_{B,j}$.

3. Eigenvalue Gap For Custom Mixers

Let $H_B$ and $H_C$ be the mixing and the cost Hamiltonian for QAOA respectively. It is known that if the Quantum Adiabatic Algorithm is run for long enough time $T$ with time-varying Hamiltonian $H(t) = (1 - t/T)H_B + (t/T)H_C$ starting with the highest-energy eigenstate of $H(0) = H_B$, then one can arrive at the highest-energy eigenstate of $H(T) = H_C$, i.e. the optimal solution, provided that the gap between the largest and second-largest eigenvalue of $H(t)$ is strictly positive for all $t < T$. This translates to finding an optimal solution when running QAOA as we let the circuit depth $p$ tend to infinity. Farhi et al. showed that this eigenvalue gap was strictly positive for standard QAOA, thus guaranteeing convergence to the optimal solution. In particular, they applied the following Perron-Frobenius theorem to irreducible stoquastic matrices.

**Theorem 3.** \cite{7} Let $A$ be an irreducible matrix whose entries are all real and non-negative. Let $r$ be the spectral radius of $A$, i.e., $r = \max\{ |\lambda| : \lambda$ is eigenvalue of $A \}$. Then $r$ is an eigenvalue of $A$ and furthermore, it has algebraic multiplicity of 1.

If the eigenvalues of an $n \times n$ matrix $A$ are real (e.g. if $A$ is Hermitian), then its eigenvalues (with multiplicity) can be ordered as $\lambda_1 \leq \cdots \leq \lambda_n$; if $A$ is also irreducible and has real, non-negative entries then Theorem 3 ensures a gap between the two largest eigenvalues (otherwise, if $\lambda_{n-1} = \lambda_n$, then the algebraic multiplicity of $\lambda_n$ would be at least 2, contradicting the statement of the theorem.) This observation still holds if we relax the non-negativity condition to allow negative entries along the diagonal as seen in the following lemma.

**Lemma 4.** Let $A$ be an irreducible stoquastic Hermitian matrix. Then the difference between the largest and second-largest eigenvalue of $A$ is strictly positive.

If the custom mixer $H_B$ has the form $\sum_{j=1}^n (x_j \sigma_j^x + z_j \sigma_j^z)$ with $x_j \in \mathbb{R}^+$ and $z_j \in \mathbb{R}$ for $j = 1, \ldots, n$, then one can show that $H(t)$ is an irreducible, stoquastic matrix. Thus by Lemma 4, the eigenvalue gap of $H(t)$ is strictly positive meaning that one can achieve the optimal solution as the circuit depth $p \to \infty$ in QAOA-Warmest. Geometrically, this special case corresponds to an initial separable state whose qubits lie in the $xz$-plane on the Bloch sphere with $x > 0$. The stoquasticity and irreducibility of this special case is formalized in the following two propositions respectively.

**Proposition 1.** Let $n$ be a positive integer. For each $j = 1, \ldots, n$ let $x_j$ be any non-negative real number and let $z_j$ be any real number. Let $H_B = \sum_{j=1}^n (x_j \sigma_j^x + z_j \sigma_j^z)$ and let $H_C$ be the problem Hamiltonian for QAOA. Then $H(t) = (1 - t/T)H_B + (t/T)H_C$ is stoquastic for all $0 \leq t < T$.

**Proposition 2.** Let $n$ be a positive integer. For each $j = 1, \ldots, n$ let $x_j$ be a positive real number and let $z_j$ be any real number. Let $H_B = \sum_{j=1}^n (x_j \sigma_j^x + z_j \sigma_j^z)$ and let $H_C$ be the problem Hamiltonian for QAOA. Then $H(t) = (1 - t/T)H_B + (t/T)H_C$ is irreducible for all $0 \leq t < T$.

Recall that for standard QAOA, we have that $H_B = \sum_{j=1}^n (1 \cdot \sigma_j^x + 0 \cdot \sigma_j^z)$. Thus, the fact that standard QAOA converges to the optimal cut as $p \to \infty$ is a special case of Propositions 1 and 2.

Although we conjecture that QAOA-Warmest has non-zero spectral gap in general (excluding some degenerate cases) and thus achieves optimality as the circuit depth $p$ tends to infinity, the analysis of convergence for the special case seen earlier does not apply for arbitrary mixers since such mixers do not necessarily have real entries and thus the Perron-Frobenius theorem cannot be applied.

We include numerical simulations in further support of our conjecture that QAOA-Warmest must tend to optimality as the circuit depth increases.

We first visualized how the gap between the two largest eigenvalues changes as $t$ varies in $H(t) = (1 - t/T)H_B + (t/T)H_C$ from $0 \leq t/T \leq 1$. See Figure 4 for an example plot. Similar plots were created for every graph considered in our numerical experiments (see Section 4.2). In these numerical simulations, we did not encounter any instance where the two highest eigenvalues were equal for $0 \leq t/T < 1$. 

---

1 Stoquastic matrices are square matrices with real entries so that all of the off-diagonal entries are non-negative. Let $A$ be an $n \times n$ square matrix. Construct a directed graph $G_A$ with vertex set $[n]$ where the edge $(i,j)$ is included if and only if $A_{ij} > 0$. If $G_A$ is strongly connected, then we say that $A$ is irreducible. Otherwise, we say that $A$ is reducible.
FIG. 1. Figure showing the change in the two largest eigenvalues throughout the evolution of the time-dependent Hamiltonian $H(t)$ for two different 5-node graphs. The red and blue curves represent the largest and second-largest eigenvalues respectively. The black dashed lines are the secant lines for both curves. The plot on top shows a case where the two eigenvalues are well-separated whereas the plot on the bottom shows that the eigenvalues appear to almost cross, but not quite.

We can further show a theoretical upper bound on the top two eigenvalues of time-varying Hamiltonian, for any $t/T \in (0, 1)$. Pictorially, these upper bounds are the black secant lines that bound the two eigenvalue curves in Figure 1. This is a consequence of Weyl’s theorem: letting $\lambda_i(A)$ denote the $i$th largest eigenvalue of $A$ (starting at $i = 1$), we have that if $M = N + R$, where $N$ and $R$ are $n \times n$ Hermitian matrices, then for $i = 1, \ldots, n$,

$$\lambda_i(N) + \lambda_n(R) \leq \lambda_i(M) \leq \lambda_i(N) + \lambda_1(R).$$

Letting $i = 2, N = (1 - t/T)H_B, R = (t/T)H_C$, we have that,

$$\lambda_2(H(t)) \leq \lambda_2((1 - t/T)H_B) + \lambda_1((t/T)H_C) = (1 - t/T)\lambda_2(H_B) + (t/T)\lambda_1(H_C) = (1 - t/T)\lambda_2(H_B) + (t/T)\lambda_2(H_C).$$

where $\lambda_1(H_p) = \lambda_2(H_p)$ is a consequence of there always being at least two Max-Cuts by symmetry. The right side of the inequality is a linear function and the inequality is tight at $t/T = 0$ and $t/T = 1$.

A similar argument shows the same properties hold for the second-largest eigenvalue of the time-varying Hamiltonian.

Unfortunately, this upper-bound on the second-largest eigenvalue is not enough to prove the existence of a gap since the largest eigenvalue could possibly dip below this bound (as seen by the example in Figure 1).

4. Warm-Starts

We now discuss the notion of warm-starting the quantum circuit for QAOA-Warmest. This method involves using classical methods to modify the initial quantum state $|s_0\rangle$ found in Equation 1. In standard QAOA, the initial state $|s_0\rangle$ corresponds to an unbiased superposition of cuts. However, one can instead consider an initial quantum state that implicitly favors some cuts over the others in an effort to incorporate information from a classical optimization solver, e.g., using warm-starts.

In this work, we find a warm-start by using a locally optimal rank-$k$ solution $x^*$ (with $k \in \{2, 3\}$) obtained by the Burer-Monteiro relaxation for Max-Cut on a $(k - 1)$-sphere. We initialize the quantum state by first performing a random vertex-at-top ro-
tation of $x^*$ (a global rotation to $x^*$ so that a random vertex $v$ coincides with $(0,0,1)$) and apply the natural mapping from the rotated solution to the Bloch sphere to obtain a separable, unentangled state $|s_0\rangle$. We refer to this variant of the QAOA algorithm initialized with such warm-starts and the standard mixers as QAOA-Warm.

In contrast to [4], we replace the standard mixers in QAOA by custom mixers that depend on the initial warm-started state, as described in Section II A. The premise of our work is that warm-starts with custom mixers indeed can outperform existing QAOA-variants as well as the classical benchmark of Goemans-Williamson algorithm (verified in simulations). Before we discuss these improvements, we would like to comment on whether custom mixers are alone responsible for improvement in performance of QAOA, or if the warm-starts also play a role in this improvement. Indeed, in Figure 2, we show that warm-starts with custom mixers give a much higher performance than custom mixers used with random initializations, i.e., by independently choosing points on the surface of the Bloch sphere.

### B. Numerical Simulations and Experiments

In this section, we demonstrate that QAOA-Warmest outperforms Goemans-Williamson Max-Cut algorithm as well as QAOA variants such as standard QAOA [3], QAOA-warm [4]. In addition, we consider the effects of noise on QAOA and its variants. For our simulations, we use CI-QuBe library instances [18] which contains graphs up to 11 nodes using a variety of random graph models (Erdős-Rényi, Barabasi Albert, Dual of Barabasi-Albert, Watts-Strogatz, Newman-Watts-Strogatz, and random regular graphs) and edge weight distributions. These instances, which we refer to as $G$, have a varied distribution of Max-Cuts, which is important when testing heuristics and algorithms for solving this problem. We consider comparisons with respect to a recent warm-starts approach of Egger et al. [5] in Appendix C.

We find that QAOA on graph instances with both positive and negative edge-weights has a much higher variance in approximation ratios achieved, and therefore we will present results for positive weight only and mixed weight graphs separately. This is not surprising since mixed-weight graphs are difficult to approximate; in fact, the only classical approximation factor known for mixed weight graphs is an $O(1/\log(n))$ approximation, when the sum of all the edge-weights is positive [19].

In our simulations, for each instance, we first find five locally approximate solutions to BM-MC$_2$ and keep the best (in terms of the BM-MC$_2$ objective value). We do the same for BM-MC$_3$. Next, for both the best BM-MC$_2$ and best BM-MC$_3$ solution, we perform 5 different vertex-at-top rotations and 5 different uniform rotations, yielding 20 different initial warm-started quantum states per instance. We run QAOA-Warm and QAOA-Warmest using all 20 of these warm-started states and, for each combination of rank and rotation scheme, record which one performed the best in terms of approximation ratio. Finally, we run standard QAOA on the instance. For each run for each variant of QAOA, we initialize the variational parameters $\gamma$ and $\beta$ close to zero.

#### 1. Choice of Rank and Rotation

Figure 3 demonstrates the average approximation ratio achieved by QAOA-Warmest for various combinations of the rank used for BM-MC$_k$ and the rotation scheme applied to the BM-MC$_k$ solution. We find that vertex-at-top rotations perform better than uniform rotations, especially in the context of rank-3 solutions. The data is inconclusive in regards to if rank-2 or rank-3 solutions are better for QAOA-Warmest, both are promising. Finally, we remark that in the case of positive-weighted graphs, any choice of rank or rotation scheme gave at least a 0.998 average approximation ratio across the instances.

To give fair comparison against QAOA-Warm [4] (and also for comparisons with [5]), the remainder of this section will assume that we are using rank-2 initializations and vertex-at-top rotations unless otherwise stated, since these were the recommended setting in [4].

---

2 One can also consider a random uniform rotation of the solution where a global rotation is performed so that a uniformly selected point on the sphere gets mapped to $(0,0,1)$. Tate et al. [4] show that this rotation scheme is not as effective as vertex-at-top rotations; we show similar results in Section II A.

3 https://github.com/swati1729/CI-QuBe
2. Comparing QAOA-Warmest to Other Methods

In Figure 4, we show the proportion of graphs where each Max-Cut algorithm (GW and variants of QAOA) performs the best for varying values of depth \( p = 1, 2, 4, 8 \). We observe that for nearly all instances, QAOA-Warmest beats every other QAOA variant considered and eventually performs at least as well as GW as the circuit depth increases. A breakdown of percentage of instances where each method performed is included as Table 12 in Appendix C. We note that at \( p = 8 \), QAOA-Warmest beats GW on all but three instances but this is easily rectified with a suitable vertex-at-top rotation.

Also at \( p = 8 \), QAOA-Warmest outperforms standard QAOA on all but two instances but the gap in approximation ratio is less than 0.02. More information regarding these five instances can be found in Appendix C.

We next report the improvement in approximation ratios when considering standard-QAOA, QAOA-warm, and QAOA-warmest with circuit depths \( p = 1, 8 \). For convenience, for any Max-Cut algorithm, we define the approximation error (AE) by \( \text{AE} = 1 - \text{AR} \) where AR is the (expected) approximation ratio. Additionally, we will refer to \( \log_{10} (\text{AE}) \) as the log-error. Figure 5 gives a comparison of log-errors achieved for various instances. All points below the \( x = y \) solid line indicate instances where QAOA-Warmest beats either QAOA-Standard or QAOA-Warm. Note that due to the plots being log-scaled, being below -2 on each axis corresponds to having an approximation ratio of at least 0.99. For both plots, we see that higher approximation ratios can be achieved for positive-weighted graphs (cross-marks) and that QAOA-Warmest performs significantly better for most instances. When comparing QAOA-Warmest and standard QAOA at various circuit depths (red v/s blue), we see that the performance for both standard QAOA and QAOA-Warmest improves at \( p = 8 \); however, this phenomenon is not that apparent for QAOA-Warm (which is known to plateau for small instances).

Finally, we show the trend in approximation quality with increase in the number of nodes \( n \) and the depth of the circuit \( p \), in Figure 6. We see that, across all node sizes, that circuit depth plays an important role in how good an approximation ratio one can expect to achieve using QAOA-Warmest. It is clear that QAOA-Warmest has superior (median) performance compared to the other algorithms for every combination of circuit depth and node-size. We remark that in contrast, an increased circuit depth resulted in only a marginal improvement in the approximation ratio for QAOA-Warm, bolstering our claim that custom mixers are crucial to the improvement in performance of QAOA.

3. QAOA-Warm With Noise

In addition to the theoretical (noise-less) behavior of QAOA-Warmest, we also demonstrate its performance with several example cases using noise models and experiments on IBM-Q hardware. We show in Figure 7 the performance of QAOA-Warmest and standard QAOA on an instance generated via a construction by Karloff [20]; this unweighted graph is chosen due to the fact that it achieves a GW approximation ratio of 0.912 (see Appendix D). In contrast, both QAOA-Warmest and standard QAOA are able to outperform this approximation ratio, under ideal, noiseless conditions. However, note that QAOA-Warmest outperforms standard QAOA for
FIG. 4. For each Max-Cut algorithm (Goemans-Williamson, standard QAOA, QAOA-Warmest with custom mixers, and QAOA-Warm), we use a pie chart to depict the percentage of instances for which it did the best (in terms of approximation ratio). The top row considers only graphs in our graph ensemble \( G \) with positive edge weights whereas the bottom row includes all graphs in the ensemble. The columns correspond to circuit depths \( p = 1, 2, 4, 8 \) from left to right. It should be noted that the green region contains instances for which QAOA-Warmest performs (roughly) at least as well as the other algorithms, i.e., if the approximation ratio achieved by QAOA-Warmest is within 0.01 of the best algorithm. A more detailed breakdown of the results depicted in this figure can be found in Table 12.

FIG. 5. For both plots, we compare the log-error of QAOA-Warmest to both QAOA-Warm (right) and standard QAOA (left). Note that all axes are log-scaled. Each marker in the plot corresponds to a combination of instance (from our graph ensemble \( G \)) and circuit depth (either \( p = 1 \) or \( p = 8 \)) with the shape of the marker being used to denote if the instance has only positive edge weights or not. Points below the black line correspond to instances where QAOA-Warmest performs better than the other algorithm being compared.

all QAOA depths and outperforms GW after \( p > 1 \). We also consider a noise model utilizing Qiskit’s built in modules and use calibration data in order to simulate IBM’s Guadalupe device. We note that QAOA-Warmest outperforms standard QAOA for all noisy simulations, using the same fixed noise model. While we use a noise model derived from a physical quantum device, we do not consider the limited qubit connectivity of the device. Therefore, the additional required swap operations to accommodate the limited connectivity would necessarily lead to worse results if this specific graph is run on the Guadalupe device.

In addition to this device focused noise simulation, we also run QAOA-Warmest on a native hardware graph matching IBM’s Guadalupe device. In general, the connectivity of the graph and its matching to physical qubit hardware connectivity plays a key
FIG. 6. Trends in median log-error of standard QAOA (dotted), QAOA-Warm (dashed), and QAOA-Warmest (solid) as one varies the number of nodes and circuit depths.

FIG. 7. Performance of QAOA-Warmest and standard QAOA as a function of QAOA depth for an ideal (dashed) and noisy simulation (dotted). For the chosen 20-node graph, GW achieves an approximation ratio of 0.912, while in the ideal case, QAOA-Warmest outperforms GW for $p > 1$ while standard QAOA requires $p > 4$.

role in performance due to the overhead of inserting swap operations in order to compensate for limited connectivity. Therefore, the simplest graph is a so-called native graph, which is a graph with exactly the same connectivity as the underlying physical qubit device. This graph is shown in Figure 8. We assign randomly chosen weights to each edge chosen from a uniform distribution $[-10, 10]$. Finding the maxcut solution to this graph can still be done by brute force and for a fixed choice of randomly weighted edges, we find the Max-Cut value to be approximately 33.96209.

We show the results of QAOA-Warmest and standard QAOA in an ideal simulation and on hardware in Figure 9. The color scale is shared across all plots, showing that QAOA-Warmest is able to find larger cut values as compared to QAOA-Standard, both in simulation and on actual hardware. For hardware results, we apply the efficient SPAM noise mitigation strategy based on a CTMP strategy [21, 22].

In order to demonstrate the scaling of QAOA-Warmest, we also show results for depths $p = 0, 1, 2$ in ideal simulation, noisy simulation, and on hardware, as shown in Figure 10. We define $p = 0$ to simply mean the preparation and measurement of the initial state. In the case of QAOA-Warmest, this directly demonstrates the ability of the QPU to create and measure the classically suggested cut. For both the ideal and noisy simulation, we use IBM’s Qiskit software package [23]. In the case of the noisy simulation, we exercise the capability of Qiskit to pull calibration data directly from the Guadalupe device and use it to construct a noise model for use in the simulator. In principle, this combination of actual hardware calibration and noise simulation should predict the behavior of the device. However, the noise models themselves have inherent assumptions that the noise itself is uncorrelated and only directly models effects such as single and two-qubit gate errors, finite qubit lifetime and dephasing time, and readout noise. While these serve as a good starting point to model the noise in a quantum device, as shown in the Figure 10, there

4 The warm-starts come from rank-2 or rank-3 solutions whereas as the GW algorithm uses rank-n solutions. Moreover, the way cuts are determined are different (hyperplane rounding vs quantum measurement) so we should expect there to be a difference in approximation ratios.
is significant disagreement between the noise simulation and the actual hardware results. This disagreement is mainly attributed to the assumptions mentioned earlier, specifically the assumption of uncorrelated noise, where physical hardware experience significant crosstalk. For more detailed description of the noise model, see Appendix B.

In addition, Figure 10 shows results for two different choices of the state initialization for QAOA-Warmest. The left plot shows the result of applying a uniform rotation in the classical preprocessing stage whereas the right shows the result of using the best vertex-at-top rotation amongst the 16 possible vertices, i.e., the rotation that gives the largest approximation ratio at \( p = 0 \). These two plots clearly show the importance of initializing the initial quantum state in an optimal way. Another important point shown in these plots is that small scale QAOA problems on 16 nodes, are nearly exactly solved when a suitable vertex-at-top rotation is chosen. When the best vertex-at-top rotation is used, the use of QAOA actually shows a decrease in solution quality on hardware. This is due to the inherent noise on the device and the fact that the solution quality is nearly optimal in the initial state. The presence of noisy two-qubit gates in further layers of the algorithm (32 CNOT gates per layer), overwhelm the small benefit of the algorithm itself for these small problems. A remaining goal then is to find native graphs on hardware for large systems, while also offering sufficiently low error rates, in order to demonstrate improved solutions with an optimally chosen initial quantum state and increased algorithmic depth (\( p > 0 \)).

III. DISCUSSION

Our experimental results suggest that our QAOA-Warmest method combined with initializations obtained by classical means can outperform both the standard QAOA and the Goemans-Williamson algorithm at relatively shallow circuit depths. Conversely, not all initializations on the Bloch sphere are useful; in particular random initializations underperform compared to classically obtained initializations. Moreover, adversarial initializations could be chosen if one wanted QAOA to perform poorly (i.e. by putting qubits near the poles of the Bloch sphere that correspond to the minimum cut). Over-
all, finding a suitable initialization is needed in order to see success in QAOA-Warmest. In the case of classically-inspired initializations such as those via Burer-Monteiro Max-Cut relaxations which are (classically) invariant under global rotations, this also includes picking a suitable rotation scheme before embedding the solution into a quantum state.

According to a paper by Farhi, Gamarnik, and Gutmann [24], QAOA needs to “see the whole graph” (i.e. have a high enough circuit depth) in order to achieve desireable results. Their results rely on the fact that local changes in the graph (e.g. modifying an edge weight) give uncorrelated results in regards to measured qubits that are sufficiently far away from such a local change. In other words, standard QAOA cannot distinguish between graphs whose local subgraph-structure is identical. It should be noted that the circuit used in QAOA-Warmest also suffers from such a locality property; however, if we consider the entirety of the QAOA-Warmest procedure, including the preprocessing stage of computing warm-starts, then this procedure can possibly distinguish between graphs with identical local subgraph structure since the initial state is sensitive to the global structure of the graph (when using BM-MC\textsubscript{k} relaxations). This suggests that certain negative theoretical results seen for standard QAOA may not necessarily hold for QAOA-Warmest since the distinguishability arguments used would no longer apply.

As mentioned previously, we conjecture that QAOA-Warmest converges to the optimal cut as the circuit depth \( p \to \infty \). We hope to resolve this conjecture in future work. Our work suggests that it may be possible to achieve convergence guarantees and on methods that lead to good practical results.

IV. METHODS

Numerical simulations were performed using both custom and pre-packaged codes in the Tensorflow
Numerical experiments in Section II B 2 were performed on the high performance computing cluster at the School of Industrial and Systems Engineering at Georgia Institute of Technology. Jobs were sent to various servers in the cluster as they became available; a listing of the servers and their specifications can be found in Table I.

Classical optimization was performed using standard optimizers available in Python, including ADAM [25], L-BFGS-B [26], and COBYLA [27]. For hardware results, we exclusively used IBM’s Guadalupe device, Qiskit software and the COBYLA optimizer. The Guadalupe device is a 16 qubit superconducting hardware with a heavy hexagonal connectivity. This device typically has average single qubit gate errors of $3.7204 \times 10^{-4}$, two-qubit gate errors of $1.075 \times 10^{-2}$ and measurement error of $1.776 \times 10^{-2}$, representing a quantum device of comparable quality to the state of the art. For our hardware runs, we seed the classical optimization process with ideal parameters $(\gamma^*, \beta^*)$ found in simulation and perform 20 optimization steps each with 8192 shots on hardware. Noisy simulations using hardware informed noise models were performed with 200 optimization steps and 3000 shots. These simulations were performed on GTRI’s Icehammer cluster using a node with a Xeon-Gold6242R processor with 80 cores and 376 GB of memory.

V. DATA AVAILABILITY

The graph instances used in the numerical simulation are accessible at the following github repository: https://github.com/swati1729/CI-QuBe. Additional experimental data is available via request from the corresponding author.

ACKNOWLEDGEMENT

This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Contract No. HR001120C0046. This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725. The authors would also like to thank Jai Moondra and Hassan Mortagy on their careful comments and feedback on an initial version of this work.

[1] Preskill, J. Quantum computing in the NISQ era and beyond. Quantum 2, 79 (2018).
[2] Harrow, A. W. & Montanaro, A. Quantum computational supremacy. Nature 549, 203–209 (2017).
[3] Farhi, E., Goldstone, J. & Gutmann, S. A quantum approximate optimization algorithm. arXiv preprint arXiv:1411.4028 (2014).
[4] Tate, R., Farhadi, M., Herold, C., Mohler, G. & Gupta, S. Bridging classical and quantum with SDP initialized warm-starts for QAOA. arXiv preprint arXiv:2010.14021 (2020).
[5] Egger, D. J., Mareček, J. & Woerner, S. Warm-starting quantum optimization. Quantum 5, 479 (2021). URL https://doi.org/10.22331/q-2021-06-17-479
[6] Bertsimas, D., King, A. & Mazumder, R. Best subset selection via a modern optimization lens. The Annals of Statistics 44, 813–852 (2016).
[7] Marcucci, T. & Tedrake, R. Warm start of mixed-integer programs for model predictive control of hybrid systems. IEEE Transactions on Automatic Control (2020).
[8] Ralphs, T. & Güzelsoy, M. Duality and warm starting in integer programming. In Proceedings of 2006 NSF Design, Service, and Manufacturing Grantees and Research Conference (2006).
[9] Farhi, E., Goldstone, J. & Gutmann, S. Quantum algorithms for fixed qubit architectures. arXiv preprint arXiv:1703.06199 (2017).
[10] Hadfield, S. et al. From the quantum approximate optimization algorithm to a quantum alternating operator ansatz. Algorithms 12 (2019). URL https://www.mdpi.com/1999-4893/12/2/34
[11] Wang, Z., Rubin, N. C., Dominy, J. M. & Rieffel, E. G. xy mixers: Analytical and numerical results for the quantum alternating operator ansatz. Phys. Rev. A 101, 012320 (2020). URL https://link.aps.org/doi/10.1103/PhysRevA.101.012320
[12] Zhu, L. et al. An adaptive quantum approximate optimization algorithm for solving combinatorial problems on a quantum computer. arXiv preprint arXiv:2005.10258 (2020).
[13] Burer, S. & Monteiro, R. D. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. Mathematical Program-
TABLE I. This table details the specifications of the server groups in the high performance computing cluster at Georgia Institute of Technology that were used in our numerical experiments.

| Number of Servers | Cores Per Server | CPU             | RAM     | Speed  |
|--------------------|------------------|-----------------|---------|--------|
| 1                  | 12               | E5-2630         | 128 GB  | 2.30 GHz |
| 1                  | 32               | Opteron 6274    | 128 GB  | 2.20 GHz |
| 4                  | 12               | E5-2630         | 128 GB  | 2.30 GHz |
| 2                  | 12               | X5660           | 72 GB   | 2.80 GHz |
| 2                  | 12               | X5660           | 148 GB  | 2.80 GHz |
| 4                  | 12               | E5645           | 96 GB   | 2.40 GHz |

[14] Goemans, M. X. & Williamson, D. P. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM (JACM)* **42**, 1115–1145 (1995).

[15] Rank-2 relaxation heuristics for max-cut and other binary quadratic programs. *SIAM Journal on Optimization* **12**, 503—521 (2001).

[16] Laub, A. J. *Matrix Analysis for Scientists and Engineers*. URL [https://archive.siam.org/books/textbooks/OT91sample.pdf](https://archive.siam.org/books/textbooks/OT91sample.pdf).

[17] Frobenius, G. Ueber matrizen aus nicht negativen elementen. *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften* 456–477 (1912).

[18] Tate, R. & Gupta, S. Ci-qube. *GitHub repository* (2021).

[19] Charikar, M. & Wirth, A. Maximizing quadratic programs: extending grothendieck’s inequality. In *45th Annual IEEE Symposium on Foundations of Computer Science*, 54–60 (2004).

[20] Karloff, H. How good is the Goemans–Williamson MAX-CUT algorithm? *SIAM Journal on Computing* **29**, 336–350 (1999).

[21] Bravyi, S., Sheldon, S., Kandala, A., Mckay, D. C. & Gambetta, J. M. Mitigating measurement errors in multiqubit experiments. *Phys. Rev. A* **103**, 042605 (2021). URL [https://link.aps.org/doi/10.1103/PhysRevA.103.042605](https://link.aps.org/doi/10.1103/PhysRevA.103.042605).

[22] Barron, G. S. & Wood, C. J. Measurement error mitigation for variational quantum algorithms (2020). 2010.08520.

[23] Abraham, H. *et al.* Qiskit: An open-source framework for quantum computing (2019).

[24] Farhi, E., Gamarnik, D. & Gutmann, S. The quantum approximate optimization algorithm needs to see the whole graph: A typical case. *arXiv preprint arXiv:2004.09002* (2020).

[25] Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[26] Fletcher, R. *Practical Methods of Optimization (2nd edition)* (John Wiley and Sons, New York, NY, USA, 1987).

[27] Powell, M. A direct search optimization method that models the objective and constraint functions by linear interpolation. *Advances in Optimization and Numerical Analysis* **275**, 51–67 (1994).

[28] Kaveh, A. & Rahami, H. A unified method for eigen-decomposition of graph products. *Communications in Numerical Methods in Engineering with Biomedical Applications* **21**, 377–388 (2005).

[29] Spacapan, S. Connectivity of cartesian products of graphs. *Applied Mathematics Letters* **21**, 682–685 (2008).

[30] Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, New York, 2011).

[31] Brouwer, A. E., Cioabă, S. M., Ihringer, F. & McGinnis, M. The smallest eigenvalues of hamming graphs, johnson graphs and other distance-regular graphs with classical parameters. *Journal of Combinatorial Theory, Series B* **133**, 88–121 (2018).

[32] Knuth, D. *Combinatorial matrices* (1991).
Appendix A: Proofs

a. Proof of Lemma 7

Proof. We have the following relationship between the Cartesian and spherical coordinates: $x = \cos \phi \sin \theta, y = \sin \phi \sin \theta, z = \cos \theta$. Thus, the matrix $U = x\sigma^x + y\sigma^y + z\sigma^z$ is given by

$$U = \begin{bmatrix} \cos \theta & \sin \theta e^{-i\phi} \\ \sin \theta e^{i\phi} & -\cos \theta \end{bmatrix}.$$ 

One can show that the matrix can be diagonalized as $U = PDP^{-1}$ where $P = [v_1 \ v_2], D = \text{diag}(1, -1), v_1 = \cos(\theta/2) |0\rangle + \sin(\theta/2)e^{i\phi} |1\rangle, v_2 = -\sin(\theta/2) |0\rangle + \cos(\theta/2)e^{i\phi} |1\rangle$. Thus, $v_1 = \cos(\theta/2) |0\rangle + \sin(\theta/2)e^{i\phi} |1\rangle$ is the highest-energy eigenstate of $U$. \qed

b. Proof of Lemma 8

Proof. Since $A$ is stoquastic, then all of the off-diagonal elements are already non-negative; however, the diagonal elements may be negative. Observe that for large enough $k$, we have that $A + kI$ is a matrix with all non-negative entries. Note that $A + kI$ is Hermitian (since $A$ and $I$ are) and thus the eigenvalues of $A + kI$ are real. If we apply the Perron-Frobenius theorem to $A + kI$, one observes that the gap between the largest and second-largest eigenvalue is strictly positive.

One can show that the eigenvalues of $A + kI$ can be obtained by shifting all of the eigenvalues of $A$ by $k$ (i.e., of $\lambda$ is an eigenvalue of $A$, then $\lambda + k$ is an eigenvalue of $A + kI$). Moreover, the multiplicities of these shifted eigenvalues are preserved. Thus, the gap between the largest and second-largest eigenvalue of $A + kI$ (which is strictly positive) is equal to the gap between the largest and second-largest eigenvalue of $A$. \qed

c. Proof of Proposition 1

We first prove a technical lemma that is needed in order to prove the proposition.

Lemma 5. If $A$ and $B$ are $n \times n$ and $m \times m$ stoquastic matrices respectively, then so is $A \oplus B$.

Proof. By definition, $A \oplus B = A \otimes I_m + I_n \otimes B$. Since we know the sum of stoquastic matrices is stoquastic, it suffices to show that $A \otimes I_m$ and $I_n \otimes B$ is stoquastic.

Observe that,

$$A \otimes I_m = \begin{bmatrix} A_{11}I_m & \cdots & A_{1n}I_m \\ \vdots & \ddots & \vdots \\ A_{n1}I_m & \cdots & A_{nn}I_m \end{bmatrix}.$$ 

Note that for $i \neq j$, the $ij$th block in the block matrix above is $A_{ij}I_m$, which contains only non-negative entries since $A_{ij}$ is an off-diagonal element of $A$ and $A$ is stoquastic. Now consider the entries in the $ij$th block where $i = j$. Note that if there is an off-diagonal entry of $A \otimes I_m$ that is part of the $i$th block, then it is also an off-diagonal entry of that block but all the off-diagonal entries of the $i$th block $(A_{ii}I)$ are zero. Thus, we have shown that every off-diagonal element is non-negative, thus $A \otimes I_m$ is stoquastic.

Next, observe that

$$I_n \otimes B = \begin{bmatrix} 1B & 0B & \cdots & 0B \\ 0B & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0B \\ 0B & \cdots & 0B & 1B \end{bmatrix},$$

which makes it clear that the off-diagonal elements of $I_n \otimes B$ are either 0 or the off-diagonal elements of $B$ which are non-negative (by stoquasticity of $B$) and thus $I_n \otimes B$ is stoquastic. \qed

d. Proof of Proposition 2

Proof. By construction $H_C$ (and thus $(t/T)H_C$) is a diagonal matrix (as $|b\rangle$ is an eigenvector of $H_C$ for each $n$-length bitstring $b$). If $H_B$ were stoquastic, then $(1 - t/T)H_B$ is stoquastic (as $1 - t/T \geq 0$ for $0 \leq t \leq T$) and thus $H(t) = (1 - t/T)H_B + (t/T)H_C$ is stoquastic (as adding a diagonal matrix to a stoquastic matrix yields a stoquastic matrix). Thus, it remains to show that $H_B$ is stoquastic.

Let $H_{B,j} = x_j \sigma^x + z_j \sigma^z$. Expanding $\sigma^x$ and $\sigma^z$, we have that

$$H_{B,j} = \begin{bmatrix} 0 & x_j \\ x_j & 0 \end{bmatrix},$$

which is clearly stoquastic as we assumed that $x_j \geq 0$. As $H_B = \bigoplus_{j=1}^n H_{B,j}$, the result now follows from Lemma 5. \qed

We are now ready to prove Proposition 3.

Proof. First, we recall the definition of irreducible matrix. Let $A$ be an $n \times n$ square matrix. Construct a directed graph $G_A$ with vertex set $[n]$ where the edge $(i, j)$ is included if and only if $A_{ij} > 0$. If $G_A$ is strongly connected, then we say that $A$ is irreducible. Otherwise, we say that $A$ is reducible.

For any square matrix $M$, let $G_M$ be the corresponding directed graph as described above. Observe that $H_C$ (and hence $(t/T)H_C$) is a diagonal
The irreducibility of the matrix, thus, by the definition of irreducibility, the irreducibility of \((1-t/T)H_B + (t/T)H_C\) is the same as \((1-t/T)H_B\). Similarly, scaling a matrix by a non-zero constant does not affect its irreducibility, so it suffices to prove the irreducibility of \(H_B\).

Observe that \(\sigma^x, \sigma^z\) are symmetric and thus it is not very difficult to show that \(H_B\) is also symmetric. This means, for the purposes of showing irreducibility, \(G_{H_B}\) is effectively an undirected graph and we just need to show that it is connected. One can write \(H_B\) as \(H_B = \bigoplus_{j=1}^n (x_j\sigma^x + z_j\sigma^z)\) where \(\bigoplus\) denotes the Kronecker sum. According to [28], this means that \(G_{H_B}\) can be written as the Cartesian graph product of the graphs \(H_1, H_2, \ldots, H_n\) where \(H_j = G_{A_j}\) with \(A_j = x_j\sigma^x + z_j\sigma^z\). Observe, that each of the \(H_j\)'s are connected if and only if \(x_j \neq 0\) which is true by assumption. Since each of the \(H_j\)'s are connected, then it is also the case that \(G_{H_B}\) is connected as well (see Theorem 1 of [29]) which finishes the proof.

### Appendix B: Noise Simulations

In Figure [11] we consider 20 instances of Erdos-Renyi graphs, with edge probabilities of 50% and 8 nodes. In this case, we also choose random negative and positive edges weights from a uniform distribution defined on \([-1, 1]\). We show results for the ideal, noiseless case as well as the noisy case when 3% phase noise is present. For this noisy case, we consider only one simple source of noise, phase damping, which is present for every single qubit gate operation.

Phase noise is a type of noise that is likely to be a dominant source of error in ion trap systems and therefore is a reasonable first step to investigate for QAOA with noise. Generically, this noise is due to interactions with the environment which is composed of many subsystems. Each interaction itself is weak, but the result of many such interactions, while not likely to cause energy transitions, does introduce a loss of phase coherence. We can describe this process with a set of (non-unique) Kraus operators [30], given by,

\[
M_0 = \sqrt{1 - q} \mathbb{I},\ M_1 = \sqrt{q} |0\rangle \langle 0|,\ M_2 = \sqrt{q} |1\rangle \langle 1|,\]

where \(q\) is the probability of a dephasing event occurring. These Kraus operators then have the effect on the state evolution as,

\[
S(\rho) = \sum_{k=0}^{2} M_k \rho M_k^\dagger = (1 - q) \rho + q |0\rangle \langle 0| + q |1\rangle \langle 1|.\]

If we associate the probability of a dephasing event occurring during a time interval \(\Delta t\), then \(q = \Gamma \Delta t\) with \(\Gamma\) being the characteristic dephasing rate and we can write \(n\) applications of the noisy channel, \(S(\rho, t)\) in matrix form as,

\[
S(\rho, t) = \begin{pmatrix} \rho_{00} & e^{-\Gamma t} \rho_{01} \\ e^{-\Gamma t} \rho_{10} & \rho_{11} \end{pmatrix}.\]

We can then see that this dephasing process, preserves population as \(\rho_{00}, \rho_{11}\) are preserved but exponentially suppresses coherences at a rate defined by \(\Gamma\), which also defines the dephasing time, \(T_2 = 1/\Gamma\).

In addition to modelling phase noise, we also include several other noise models in the results shown in Fig. 5 of the main text. In total, these noisy simulations utilize noise models that incorporate gate error probability of each gate, the gate length of all gates, the \(T_1\) and \(T_2\) times of all qubits, as well as the readout error probability. One can define similar Kraus operators for these noise channels as well [30]. However, while these are a comprehensive treatment of quantum noise they do not accurately capture crosstalk and other correlated noise sources.

### Appendix C: Additional Numerical Results

#### 1. Detailed Performance Comparison

In Section II B 2, Figure 4 depicts, for each Max-Cut algorithm (standard QAOA, QAOA-Warm, QAOA-Warmest, GW), for which percentage of instances that algorithm achieved the best approximation ratio. Below, in Table 12 we have a more detailed comparison which considers, for each instance and circuit depth, the top two algorithms.

#### 2. Interesting Instances

In Section II B 2, Figure 4 shows that at circuit depth \(p = 8\), there are five instances for which QAOA-Warmest did not achieve the highest approximation ratio compared to the other algorithms considered.
FIG. 11. Comparison between standard QAOA mixer to using warm starts with custom mixers. We show the noiseless (left) and noisy (right) case. In both cases, the custom mixer significantly outperforms the standard mixer. Shaded regions indicate the distribution of results for 20 randomly chosen 8 node graphs with positive and negative weights.

FIG. 12. For each Max-Cut algorithm (Goemans-Williamson, standard QAOA, QAOA-Warmest, and QAOA-Warm), we report the percentage of instances for which it did the best and second-best (in terms of approximation ratio). The upper portion of the table considers only graphs in our graph ensemble $G$ with positive edge weights whereas the bottom portion includes all graphs in the ensemble. Each row corresponds to a different circuit depth ($p = 1, 2, 4, 8$) for which the comparisons were made. As an example, the first percentage reported in the top-left corner of the table indicates that amongst all positive-weighted graphs in $G$, 0.69% of them have the property that, amongst the four algorithms considered, (depth-1) QAOA-Warmest has the best approximation ratio and (depth-1) standard QAOA has the second-best approximation ratio. We say there is a tie (last column) if the top two algorithms have approximation ratios that differ by no more than 0.01; it should be noted that QAOA-Warmest is a part of every tie and that every instance that is part of a tie is not counted in the columns to the left of the Tie column. For the column labeled *, we report, for each circuit depth, the percentage of instances for which QAOA-Warmest was within 0.01 approximation ratio of the best algorithm; these numbers correspond to the green region of the pie charts in Figure 4.

Of these five instances, standard QAOA was the best algorithm for precisely two of these (instances #778 and #1820). For the remaining three instances (#1698, #1889, #2010), GW was the best algorithm; however, all three of these instances have the property that there is a single negative edge-weight whose magnitude is much smaller than the other edge weights in the graph and additional numerical simulations show that a suitable vertex-at-top rotation (selecting the vertex that is incident to the large-magnitude negative edge weight) that allows QAOA-Warmest to outperform GW.

Table 13 gives detailed statistics for the approximation ratios achieved by each of the Max-Cut algorithms considered for these five instances.
## Instances Where Depth-8 Standard QAOA Performs Best

| Instance ID | QAOA-Warmest | QAOA-Warm | Standard QAOA | GW |
|-------------|--------------|-----------|---------------|----|
| 778         | 0.900        | 0.774     | 0.919         | 0.890 |
| 1820        | 0.789        | 0.622     | 0.799         | 0.766 |

## Instances Where GW Performs Best

| Instance ID | QAOA-Warmest | QAOA-Warm | Standard QAOA | GW | QAOA-Warmest (modified) |
|-------------|--------------|-----------|---------------|----|-------------------------|
| 1698        | 0.510        | 0.583     | -0.980        | 0.630 | 0.983                   |
| 1889        | -0.094       | 0.064     | -2.080        | 0.759 | 0.966                   |
| 2010        |              |           | -2.080        | 0.759 | 0.966                   |

FIG. 13. These tables report the approximation ratios achieved for the five instances (amongst those in our instance library \( \mathcal{G} \)) for which depth-8 QAOA-Warmest did not obtain the best approximation ratio when compared to depth-8 QAOA-Warm, depth-8 standard QAOA, and GW. The top and bottom tables are for instances in which standard QAOA and GW performed the best respectively. The instances in the bottom table have the property that there exists exactly one negative edge weight whose magnitude is much larger than the other edge weights. For the bottom table, in the last column, we also include the approximation ratio for QAOA-Warmest in the case where a more suitable vertex-at-top rotation is used; i.e., we take one of the vertices incident to the large-magnitude negative edge and rotate it to the top.

### 3. Comparison with Egger et al.

Figure 14 compares the approximation ratios achieved by QAOA-Warmest and a variant of QAOA considered by Egger et al.\[5\]. In the context of the Max-Cut problem, Egger et al. considered an approach which takes a good starting cut \((S, V \setminus S)\) (obtained via GW or possibly other means) and uses this cut to construct an initial quantum state \(|s_0\rangle\). With this modified initial quantum state and an appropriate modification of the mixing Hamiltonian, Egger et al. show that their variation of QAOA is able to recover the cut at circuit depth \(p = 1\), i.e., there is a choice of variational parameters \(\gamma_1\) and \(\beta_1\) such that the only cut obtained at those parameters is precisely \((S, V \setminus S)\).

To give a fair comparison for Egger et al.’s approach, we consider 10 cuts generated by the GW algorithm and take the best 5. Due to the size of the instances we consider, usually at least one of the best 5 cuts would be optimal and hence Egger et al.’s approach would essentially already start with an optimal solution which is not interesting. For this reason, in Figure 14 we only consider those instances (22.7% of the instance library) for which neither QAOA-Warmest nor Egger et al.’s approach starts with the optimal solution.

For Egger et al.’s approach, we consider two different choices for initialization of the variational parameters: near the origin and the choice of parameters that recovers the value of cut used to initialize the QAOA variant (i.e. \(\beta_1 = \pi/2\) with the remaining parameters being set to zero). In both cases, Figure 14 demonstrates that QAOA-Warmest typically has the superior performance.

There are a total of 169 instances for which the approximation ratio achieved by depth-8 QAOA-Warmest beats GW (by at least 0.001) and the approximation ratio achieved by GW beats Egger et al.’s approach (with initialization \(\beta_1 = \pi/2\) with the remaining parameters being set to zero) at depth-8 (by at least 0.001). For these instances, the median gap in approximation ratio between QAOA-Warmest and GW was 0.0543 and the median gap in approximation ratio between GW and Egger et al.’s approach is 0.0649.

### Appendix D: Twenty-Node Graph

In this section, we discuss why the graph considered in Figure 8 is interesting. The graph used is a 20-node instance, where Goemans-Williamson achieves an approximation ratio of 0.912. We briefly summarize the construction and properties of this graph.

Recall that the worst-case approximation ratio for the Goemans-Williamson (GW) ratio is 0.878. Karloff\[20\] showed that the 0.878 bound for GW is tight by constructing a family of graphs whose approximation ratio approaches 0.878 as graph size increases. The construction for this family of instances is as follows: consider non-negative integers \(b \leq t \leq m\) and let \(J(m, t, b)\) denote the graph with vertex set \([m]\), i.e., the vertices are all \(t\)-element subsets of \([m]\); two distinct vertices/subsets \(S\) and \(T\) of \(J(m, t, b)\) are adjacent if and only if they have exactly \(b\) elements in common, i.e. \(|S \cap T| = b\). Karloff proved the approximation ratio for GW on
some of these instances:

**Theorem 6.** [20] Let \( m \) be an even positive integer and \( G = J(m, m/2, b) \). If \( 0 \leq b \leq m/12 \), then the approximation ratio for Goemans-Williamson on \( G \) is given by

\[
\frac{1}{\pi} \arccos \left( \frac{4b}{m} - 1 \right) \frac{1}{1 - \frac{2b}{m}}.
\]

Non-trivial instances using Karloff’s construction arises once \( b \geq 1 \); however, in order for the hypotheses of Theorem 6 to be satisfied, this requires \( m \geq 12 \) which implies that one needs to consider instances with at least \( \binom{m}{6} = 924 \) nodes. Performing reliable experiments on such large instances is not feasible for current quantum hardware.

We noticed that smaller instances with a weak GW approximation ratio could be constructed if the following conjecture by Karloff was true.

**Conjecture 7.** (Conjecture 2.12 in [20]) If \( m \) is an even positive integer with \( 0 \leq b < m/4 \), then the smallest eigenvalue of the adjacency matrix of \( J(m, m/2, b) \) is

\[
\beta_1 = \left( \frac{m}{2} \right)^2 \left( \frac{4b}{m} - 1 \right).
\]

Karloff [20] proved a special case of this conjecture in the case where \( 0 \leq b \leq m/12 \) which was instrumental in showing Theorem 6 (with the same inequality in the hypotheses).

We found that in 2018, Brouwer et al. in fact proved the following theorem below; substituting \( t = m/2 \) into Theorem 8 and performing a few simple calculations, they also found that Conjecture 7 follows as a corollary.

**Theorem 8.** (Theorem 3.10 in [31]) Let \( 0 \leq b < t \) and let

\[
\beta_s = \sum_{r=0}^{s} (-1)^{s-r} \binom{s}{r} \left( \frac{t-r}{t-b} \right) \left( \frac{m-t-s+r}{m-2t+b} \right).
\]

Then \( \beta_1 \) is the smallest eigenvalue of the adjacency matrix of \( J(m, t, b) \) if and only if \( (t-b)(m-1) \geq t(m-t) \). In this case, \( \beta_1 \) is also the second largest in absolute value among the eigenvalues of the adjacency matrix of \( J(m, t, b) \).

Conjecture 7 allows us to relax the inequality in Theorem 6 to \( 0 \leq b < m/4 \); thus, it can be applied with \( m = 6 \) and \( b = 1 \) to obtain the graph \( J(6, 3, 1) \) (used in Figure 7) with \( \binom{6}{6} = 20 \) nodes where GW achieves an approximation ratio of

\[
\frac{1}{\pi} \arccos \left( \frac{4}{6} - 1 \right) \frac{1}{1 - \frac{2}{6}} = 0.912.
\]

5 The statement of Theorem 8 has been modified in order to be consistent with the notation used in Karloff’s construction [20].

6 The eigenvalues of the adjacency matrix of \( J(m, t, b) \) are \( \beta_0, \beta_1, \ldots, \beta_t \) (each with positive multiplicity) [22].
Table II shows all non-trivial instances under 1000 nodes that use Karloff’s [20] construction for which we can calculate the approximation ratio using Theorem 6 and Conjecture 7. These instances may be of interest to those working with near-term quantum devices.

| Instance  | # Nodes | # Edges | Approx. Ratio          |
|-----------|---------|---------|------------------------|
| J(6, 3, 1) | 20      | 90      | 0.912260171954089      |
| J(8, 4, 1) | 70      | 560     | 0.8888888888888888     |
| J(10, 5, 1)| 252     | 3150    | 0.881040955873917      |
| J(10, 5, 2)| 252     | 12600   | 0.940157028081625      |
| J(12, 6, 1)| 924     | 16632   | 0.878735432638524      |
| J(12, 6, 2)| 924     | 103950  | 0.912260171954089      |

TABLE II. Small (< 1000 nodes) instances using Karloff’s [20] construction. For each instance, we include the number of nodes, edges, and the theoretical expected approximation ratio one would obtain using Goemans-Williamson algorithm on that instance.