

Warm-Started QAOA with Custom Mixers Provably Converges and Computationally Beats Goemans-Williamson's Max-Cut at Low Circuit Depths

Reuben Tate¹, Jai Moondra², Bryan Gard³, Greg Mohler³, and Swati Gupta⁴

¹CCS-3 Information Sciences, Los Alamos National Laboratory, Los Alamos, NM 87544, USA

²Georgia Institute of Technology, Atlanta, GA 30332, USA

³Georgia Tech Research Institute, Atlanta, GA 30332, USA

⁴Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02142, USA

We generalize the Quantum Approximate Optimization Algorithm (QAOA) of Farhi et al. (2014) to allow for arbitrary separable initial states with corresponding mixers such that the starting state is the most excited state of the mixing Hamiltonian. We demonstrate this version of QAOA, which we call *QAOA-warmest*, by simulating Max-Cut on weighted graphs. We initialize the starting state as a *warm-start* using 2 and 3-dimensional approximations obtained using randomized projections of solutions to Max-Cut's semi-definite program, and define a warm-start dependent *custom mixer*. We show that these warm-starts initialize the QAOA circuit with constant-factor approximations of 0.658 for 2-dimensional and 0.585 for 3-dimensional warm-starts for graphs with non-negative edge weights, improving upon previously known trivial (i.e., 0.5 for standard initialization) worst-case bounds at $p = 0$. These factors in fact lower bound the approximation achieved for Max-Cut at higher circuit depths, since we also show that QAOA-warmest with any separable initial state converges to Max-Cut under the adiabatic limit as $p \rightarrow \infty$. However, the choice of warm-starts significantly impacts the rate of conver-

gence to Max-Cut, and we show empirically that our warm-starts achieve a faster convergence compared to existing approaches. Additionally, our numerical simulations show 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 and Quantinuum hardware.

1 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.¹ Given a weighted

Swati Gupta: email of the corresponding author is swatig@mit.edu

¹Although this work focuses on Max-Cut, our approach can be applied to any suitable combinatorial

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. $\text{cut}(S) := \sum_{e \in E} w_e \cdot \mathbf{1}[e \in S \times (V \setminus S)]$, 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. The promise of this approach lies in Farhi et al.’s [3] result that establishes a connection between standard QAOA and quantum adiabatic computing, showing that with increased circuit depth, standard QAOA converges to the Max-Cut. This connection relies on the initial state of standard QAOA being the highest energy eigenstate of the mixing Hamiltonian.

We propose to initialize the circuit with separable initial states (other than $|+\rangle^{\otimes n}$) generated using classical relaxations of the Max-Cut problem; following classical optimization literature, we refer to these states as *warm-starts*. We modify the mixing Hamiltonians in the QAOA ansatz in a way dependent on the initial state, so that the adiabatic assumptions hold; we call these mixing Hamiltonians *custom mixers*, and call the new QAOA variant *QAOA-warmest*. We show that like QAOA, QAOA-warmest converges to the Max-Cut with increased circuit depth (Proposition 1).

We further note that the convergence *rate* of QAOA-warmest is heavily dependent on the separable initial state used. In this work, we focus on two approaches for warm-starting that generate classically-inspired separable initial states which (empirically) converge faster compared to existing initializations: (1) low-dimensional projections of optimal solutions to the Goemans-Williamson (GW) semidefinite program (SDP) [5], and (2)

optimization problem by converting a generic QUBO instance into a Max-Cut instance [4], and then applying the same techniques. This conversion requires only one additional qubit; however, such a conversion is not necessarily approximation-preserving.

locally optimal solutions to low-dimensional Burer-Monteiro [6] relaxations of the Max-Cut problem. Using a diverse library of 1148 graphs (up to 11 nodes), we show through numerical simulations that QAOA-warmest that uses our two types of warm-starts outperforms the classical Goemans-Williamson algorithm and standard QAOA even at low-circuit depths around $p = 4$. Such a low-depth for this performance does not hold for any random initialization on the Bloch sphere. We further tested QAOA-warmest in the presence of noise using Qiskit’s built-in modules and hardware calibration data [7], and on IBM’s 16-qubit Guadalupe device and Quantinuum’s 20-qubit devices. Our simulations demonstrate that QAOA-warmest is robust and still yields high (instance-specific) approximation ratios in such noisy regimes.

Additionally, we provide theoretical performance guarantees at depth $p = 0$ in the case that projected GW solutions are used. Specifically, for positive-weighted graphs, we show (Theorem 2) that, in expectation, the value of the cut obtained from hyperplane rounding of the projected GW solutions is at least 0.878 of the optimal cut value (just like GW). We further show that quantum measurement of the projected solutions (i.e. depth-0 QAOA-warmest) yields at least a $\frac{3}{4}(0.878) \approx 0.658$ -approximation and $\frac{2}{3}(0.878) \approx 0.585$ -approximation for dimensions $k = 2$ and $k = 3$ respectively. These guarantees also serve as a lower bound for what can be achieved with higher depth since, like standard QAOA, our approach also has monotonically increasing expected cut-quality with increased circuit depth for any instance.

This paper presents the first warm-start approach for QAOA (for Max-Cut) which simultaneously (1) provides a nontrivial constant-factor approximation ratio at depth $p = 0$, (2) satisfies provable convergence to Max-Cut as circuit depth increases, and (3) enjoys a fast convergence rate, as shown through numerical simulations and experiments on quantum hardware. Some of the previous approaches [8, 9] consider warm-start initializations using perturbations of a single-cut (obtained for ex-

ample, using GW algorithm). While such initializations theoretically yield higher approximation ratio of 0.878 using quantum sampling at depth $p = 0$, this warm-start has a much slower convergence rate in simulations (see Table 1). Therefore, we believe that our approach is promising for near-term NISQ devices.

1.1 Related Work

Since Farhi et al.'s [3] seminal paper in 2014, many have researched and analyzed the QAOA algorithm. Many empirical studies have been performed including the effects of different parameter initialization strategies [10–14], the performance of QAOA in the context of devices with superconducting qubits with limited connectivity [15–17], the use of machine learning to determine (for specific instances) whether Max-Cut QAOA would yield better cuts compared to the classical GW algorithm [18], the effectiveness of different encoding schemes for objectives involving higher-order terms with more than 2 qubits [19], the effect of various graph properties on the performance of QAOA [20], and the use of QAOA to generate highly squeezed states which are useful in the context of quantum metrology [21].

Many researchers have also analyzed QAOA from a more theoretical perspective. Shaydulin et al. [22] gives a series of results regarding the classical symmetries in the objective function and how those symmetries are reflected in the QAOA dynamics. In their seminal paper, Farhi et al. show that depth-1 QAOA achieves an approximation ratio of 0.694 for Max-Cut on 3-regular graphs [3]. Wurst and Love show that at depth-2, this approximation ratio improves to 0.7559 for Max-Cut QAOA on 3-regular graphs [23]. For quantum devices with limited connectivity, Farhi et al. [24] show that a variation of Max-Cut QAOA on 3-regular graphs on a device whose native graph is a square grid achieves an approximation ratio of 0.5293 without the use of swap operations. Others have also proven limitations of the standard QAOA algorithm

as well: Bravyi et al. [25] show that, for all $d \geq 3$, there exists a sequence of d -regular bipartite graphs such that depth- p QAOA with $p < (1/3 \log_2 n - 4)d^{-1}$ on such instances produces a cut (in expectation) whose value is at most $\frac{5}{6} + \frac{\sqrt{d-1}}{3d}$, meaning that, in the worst case, constant-depth QAOA for Max-Cut is inferior to the classical Goemans-Williamson algorithm as $\lim_{d \rightarrow \infty} \left(\frac{5}{6} + \frac{\sqrt{d-1}}{3d}\right) = \frac{5}{6} \approx 0.833 < 0.878$. Farhi et al. [26] show a similar result when QAOA is applied to the Max Independent Set problem² and Bravyi et al. [27] give similar results for a recursive variant of QAOA applied to the Max- k -Cut problem.³ Hastings [28] defines a notion of a "local" classical algorithm and shows that, for triangle-free d -regular graphs with $2 \leq d \leq 1000$, there exists⁴ a d -dependent local classical algorithm for Max-Cut with a provably better approximation ratio compared to depth-1 QAOA. Marwaha [29] extended this result, showing that for each $2 \leq d \leq 500$, there exists local classical algorithms that yields a better expected cut compared to depth-2 QAOA for all d -regular graphs with girth greater than 5. Barak and Marwaha [30] have continued this line of research, showing that for every one-local algorithm (classical or quantum), that the maximum cut achieved is at most $\frac{1}{2} + \frac{1/\sqrt{2}}{d}$ of the maximum cut for d -regular graphs with girth greater than 5 and that there exists a k -local algorithm with approximation ratio $\frac{1}{2} + \frac{2/\pi}{d} - O\left(\frac{1}{\sqrt{k}}\right)$ for d -regular graphs with girth greater than $2k + 1$. However, general approximation guarantees obtained by the QAOA algorithm still remain elusive.

The above results also suggest that in order

²Given a graph $G = (V, E)$, the goal of the Max Independent Set problem is to find $S \subseteq V$, with $|S|$ as large as possible, such that for all vertices $u, v \in S$, the edge $(u, v) \in E$.

³In the Max- k -Cut problem, one is given a weighted graph $G = (V, E)$ with weights $w : E \rightarrow \mathbb{R}$ and the goal is to partition the vertices into k disjoint groups so that the sum of weights of edges across partitions is maximized

⁴Numerical evidence suggests that the results of Hastings [28] and Marwaha [29] hold for all d ; however, no formal proof is given.

to realize some kind of quantum advantage, more information beyond the standard QAOA algorithm might be needed. Recent work has considered modifications and variations of the QAOA algorithm itself, as in this work, which we discuss next.

The closest works related to this paper are those by Tate et al. [31] and Egger et al. [9], where the authors have explored multiple approaches for warm-starting QAOA. Tate et al. [31] considered warm-starting QAOA using 2 and 3-dimensional Burer-Monteiro locally optimal solutions for Max-Cut; however, their method plateaus even at low circuit depths of $p = 1$ for some initializations, and it is unable to improve the cut quality for some instances. For the Max-Cut problem, Egger et al. [9] constructed the initial quantum state by (non-trivially) mapping a single specific cut $(S, V \setminus S)$ (obtained via the Goemans-Williamson algorithm or possibly other means) to an initial quantum state. Egger et al. also modified the mixing Hamiltonian so that at depth $p = 1$ (with the right choice of QAOA variational parameters), the cut $(S, V \setminus S)$ is recovered; however, there is no evidence to suggest that such an approach will converge to the optimal solution for depth $p \rightarrow \infty$. They further proposed a different continuous warm-started QAOA for Quadratic Unconstrained Binary Optimization (QUBO) problems, which enjoys convergence guarantees. For certain classes of QUBO's, the binary variables can be relaxed to be in the interval $[0, 1]$ to obtain a convex quadratic program whose optimal solution can then be mapped to the Bloch sphere. Our mixer is a generalization of the mixer used in this particular approach by Egger et al.; however, the initialization scheme is quite different. In particular, their continuous warm-started QAOA approach cannot be directly applied to Max-Cut as the corresponding relaxed quadratic program is not convex.

Recent work by Cain et al. [8] further explored convergence properties of warm-starts, when augmented with standard mixers. They showed using a perturbative approach that standard QAOA with single-cut warm-starts

(i.e., each qubit is initialized at $|0\rangle$ or $|1\rangle$) does not show any improvement in the approximation ratio, even when the circuit depth is increased. This result is interesting in the context of our work, since we find that (1) using custom mixers, one can guarantee convergence as long as the initialization is not at a single-cut, and (2) warm-starts that are perturbations of single-cut initializations (e.g., each qubit is initialized at $R_Y(\theta^*)|0\rangle$ or $R_Y(\pi - \theta^*)$ for some small θ^* where $R_Y(\theta)$ is a single-qubit rotation about the y -axis by angle θ) converge very slowly, in computations, even with custom mixers. Our work therefore complements the existing work and clarifies what kind of warm-starts are useful.

Additionally, other works have explored different kind of modifications of the QAOA ansatz. Farhi et al. [24] considered having separate variational parameters for each vertex and edge. Hadfield et al. [32] and Wang et al. [33] considered 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. [34] modify QAOA such that the ansatz is expanded in an iterative fashion with the mixing Hamiltonian being allowed to change between iterations. Bravyi et al. [25] proposed a recursive QAOA approach that decreases the instance size at each iteration; Egger et al. [9] also consider a similar recursive version of their approaches. Bärtschi et al. [35] and Jiang et al. [36] consider modifications of QAOA inspired by Grover's (quantum) algorithm [37] for fast database search. For the scope of this work, we do not consider these alternate approaches; however, it may serve as an interesting direction for future work as QAOA-warmest can likely be used in conjunction with these other approaches.

1.2 Background

1.2.1 The Quantum Approximate Optimization Algorithm

First, we review 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 $S = \{i : i\text{th bit is } 0\}$.

For depth- p QAOA, the quantum circuit alternates between applying a cost Hamiltonian $H_C = \frac{1}{2} \sum_{(i,j) \in E} w_{ij} (1 - \sigma_i^z \sigma_j^z)$ and a mixing Hamiltonian $H_B = \sum_{i \in [n]} \sigma_i^x$. Here, $\sigma^x, \sigma^y, \sigma^z$ are the standard Pauli operators and σ_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,

$$\begin{aligned} & |\psi_p(\gamma, \beta)\rangle \\ &= e^{-i\beta_p H_B} e^{-i\gamma_p H_C} \dots e^{-i\beta_1 H_B} e^{-i\gamma_1 H_C} |s_0\rangle, \end{aligned} \quad (1)$$

where $|s_0\rangle$ is the initial state and $\gamma = (\gamma_1, \dots, \gamma_p), \beta = (\beta_1, \dots, \beta_p)$ 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 bit-string corresponding to a cut in the graph. We let $F_p(\gamma, \beta)$ denote the expected cut value from sampling $|\psi_p(\gamma, \beta)\rangle$, i.e.,

$$F_p(\gamma, \beta) = \langle \psi_p(\gamma, \beta) | H_C | \psi_p(\gamma, \beta) \rangle.$$

If, for each p , we choose γ, β optimally, then the expected value of the cut tends to the Max-Cut as $p \rightarrow \infty$ [3].

In practice, the optimal choice of γ, β is unknown in advance and thus a hybrid classical-quantum hybrid loop is often used to find values of γ and β that yield high expectation values. The initialization of γ and β in these optimization routines has been investigated [10–14]; however, for simplicity, we initialize γ and β near the origin for our experiments as discussed in Section 4.

1.2.2 Classical Optimization Algorithms

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

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

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:

$$\begin{aligned} & \max \quad \frac{1}{4} \sum_{(i,j) \in E} w_{ij} \|x_i - x_j\|^2 \\ & \text{subject to} \quad \|x_i\| = 1, \quad \forall i \in V, \\ & \quad \quad \quad x_i \in \mathbb{R}^k, \quad \forall i \in V. \end{aligned} \quad (3)$$

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 [5], 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 [6]. 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 $k = 2$ [38]; however, in theory, Mei et al. [39] show that hyperplane rounding of a locally optimal BM-MC $_k$ solution achieves a $0.878(1 - \frac{1}{k-1})$ fraction of the optimal cut, yielding a 0-approximation and a 0.439-approximation for dimensions $k = 2$ and $k = 3$ respectively.

For ease of notation, given a (feasible) BM-MC $_k$ solution x , we let

$$\text{HP}(x) = \sum_{(i,j) \in E} \frac{1}{\pi} \arccos(x_i \cdot x_j),$$

denote the expected cut value obtained from performing hyperplane rounding on x [5] and we let

$$\text{BM-MC}_k(x) = \sum_{(i,j) \in E} \frac{1}{4} \|x_i - x_j\|^2,$$

denote the BM-MC $_k$ objective at x . Lastly, we say that the solution x is κ -approximate

if $\text{HP}(x) \geq \kappa \text{Max-Cut}(G)$ and similarly, x is considered κ -close if $\text{BM-MC}_k(x) \geq \kappa \text{Max-Cut}(G)$.

1.2.3 Approximation Ratio

For this work, we adopt the same definition and notation for the (instance-specific) approximation ratio (AR) as discussed in Tate et al. [31]. Specifically, for a graph G and algorithm \mathcal{A} , the (instance-specific) AR is given by,

$$\alpha_{\mathcal{A},G} = \frac{\mathcal{E}_{\mathcal{A},G} - \text{MIN-CUT}(G)}{\text{MAX-CUT}(G) - \text{MIN-CUT}(G)}; \quad (4)$$

where $\mathcal{E}_{\mathcal{A},G}$ is the expected cut value (either from hyperplane rounding of a classical solution or quantum measurement of a quantum state) and $\text{MIN-CUT}(G)$ is the (possibly trivial) minimum cut in the graph G . This definition is chosen as a means of "normalizing" the approximation ratio to lie in the interval $[0, 1]$, even in the case of graphs with both positive and negative edge weights. For positive-weighted graphs, we have that $\text{MIN-CUT}(G) = 0$, in which case, the definition of $\alpha_{\mathcal{A},G}$ reduces to the "typical" definition of approximation ratio often seen in the classical optimization literature. The worst-case AR $\alpha_{\mathcal{A}}$ for an algorithm \mathcal{A} is defined as the worst instance-specific AR across all instances, i.e., $\alpha_{\mathcal{A}} = \min_G \alpha_{\mathcal{A},G}$; alternatively, we say such an algorithm \mathcal{A} provides an $\alpha_{\mathcal{A}}$ -approximation.

It should be noted that in our numerical simulations, we can calculate $\mathcal{E}_{\mathcal{A},G}$ exactly. Please refer to [31] for more details.

2 Custom Mixers

In this section, we provide a general framework that shows, for *any* separable initial state, how to construct a custom mixing Hamiltonian for QAOA so that the warm-start is the most excited state of the mixer. This property will be useful in showing convergence results for QAOA with custom mixers. We refer to such variants of QAOA as QAOA-warmest (compared to QAOA-warm,

proposed by Tate et al. [31], which uses the standard mixer, $H_B = \sum_{i \in [n]} \sigma_i^x$, together with a separable quantum initial state).

Consider a separable state $|s_0\rangle$ on n qubits; if the j th qubit's Cartesian coordinates on the Bloch sphere are given by (x_j, y_j, z_j) , then the corresponding custom mixing Hamiltonian H_B is given by,

$$H_B = \bigoplus_{j=1}^n H_{B,j}, \quad (5)$$

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

Geometrically, the custom mixer rotates qubits about their original position on the Bloch sphere (details included in Appendix B). Note that the standard mixers in QAOA [3] are, therefore, a special case of our custom mixers since each qubit is initialized at $|+\rangle$, i.e., the the x -axis (with Cartesian coordinates $(x_j, y_j, z_j) = (1, 0, 0)$) and the unitary operator $e^{-i\beta_k H_B}$ for the mixer corresponds to rotations (by $2\beta_k$) about the x -axis. When the initial state is composed by qubits restricted to the xz -plane with $x > 0$, then this custom mixer recovers the one considered by Egger et al [9].

2.1 Convergence to Max-Cut

In this section, we show that the expected cut obtained by QAOA-warmest converges to Max-Cut as the circuit depth goes to infinity.

Theorem 1. *Let $|s_0\rangle$ be any separable initial state whose qubits do not lie at the poles of the Bloch sphere. Running QAOA with a warm-start $|s_0\rangle$, its corresponding custom mixer (5), with the choice of optimal variational parameters, yields a distribution of cuts whose expected value reaches Max-Cut as the circuit depth p tends to infinity, i.e.,*

$$\lim_{p \rightarrow \infty} \max_{\gamma, \beta} F_p(\gamma, \beta) = \text{Max-Cut}(G).$$

We will first show that Theorem 1 holds when the warm-start is initialized in the xz -plane of the Bloch sphere, with $x > 0$. We will then show the main result by showing equivalence of the distribution of cuts obtained by

QAOA-warmest when a phase is added to each qubit (i.e., the initial state is not restricted to 0 phase).

Proposition 1. *Let $|s_0\rangle$ be any separable initial state such that all qubits lie at the intersection of the Bloch sphere and the xz -plane with positive x -coordinate. Running QAOA with initial state $|s_0\rangle$ and its corresponding custom mixer yields that*

$$\lim_{p \rightarrow \infty} \max_{\gamma, \beta} F_p(\gamma, \beta) = \text{Max-Cut}(G),$$

i.e., the expected cut value of QAOA-warmest with optimal variational parameters will yield the optimal cut value as the circuit depth p tends to infinity.

Egger et al. [9] use the same custom mixers⁵ for warm-starts restricted to the xz -plane with $x > 0$, for convex quadratic programs, and state that convergence holds, without proof. While straightforward calculations show that the initial state is the highest-energy eigenstate of the mixer (a condition needed in order to apply the adiabatic theorem and guarantee convergence), a complete proof requires careful inspection of the eigenvalues of the time-dependent Hamiltonian $H(t) = (1-t/T)H_B + (t/T)H_C$. We show that there is a non-zero gap between the largest and the second largest eigenvalues of the time-varying Hamiltonian, using the Perron-Frobenius theorem for irreducible stoquastic Hermitian matrices. These details can be found in Appendix C.

This analysis can not be directly applied for arbitrary separable states since the corresponding mixers do not necessarily have real entries, and therefore, the Perron-Frobenius theorem cannot be applied. However, instead of calculating the eigenvalue gaps directly, we show that the phase in the warm-start is not reflected in the distribution of cuts obtained using QAOA-warmest, which would then complete the proof of Theorem 1.

⁵Although [9] use the same construction of the mixer, they use significantly different warm-starts, which ultimately play a role in the rate of convergence of the overall approach.

Proposition 2. *Let $|s_0\rangle$ be any separable initial state and let H_B be its corresponding custom mixer. Let $|s'_0\rangle$ be the state $|s_0\rangle$ where each qubit's phase is set to 0, and let H'_B be the corresponding custom mixer. Then, for any fixed choice of variational parameters (γ, β) , the distribution of cuts obtained from QAOA-warmest with initial state $|s_0\rangle$ and mixer H_B is the same as the distribution of cuts from obtained QAOA-warmest with initial state $|s'_0\rangle$ and mixer H'_B .*

Proposition 2 can be proved by first decomposing the general custom mixer (corresponding to $|s_0\rangle$) in terms of the custom mixer corresponding to the initial state with the phase removed (i.e. $|s'_0\rangle$). Using this decomposition, we show that any effect caused by the change in rotation direction (due to switching from the custom mixer for $|s'_0\rangle$ to the custom mixer for $|s_0\rangle$) exactly cancels out with the effect of introducing a phase to $|s'_0\rangle$ to obtain $|s_0\rangle$. The details of the proof can be found in Appendix A.

The convergence given by Theorem 1 for QAOA-warmest is especially interesting considering that many previous warm-started QAOA approaches lack such guarantees. The QAOA-warm approach by Tate et al. [31] considered arbitrary separable states but with the standard mixer; they showed examples where QAOA-warm plateaued and did not converge to the Max-Cut. Cain et al. [8] considered the case where the initialization is a single bitstring/cut with the standard mixer, and showed that such an initialization also does not converge to Max-Cut. One of the approaches by Egger et al. consider a mixer that is neither the standard mixer nor the custom mixer approach presented above. Their mixer is used in conjunction with what we call a perturbed single-cut initialization (see Section 3) which recovers a *particular* cut (obtained via GW or other means) at depth 1. However, this initialization comes with no guarantees on convergence, and our results also do not apply to these different mixers.

Although Theorem 1 applies to any warm-start with aligned custom mixers, we will show that there is a significant difference in the *rate*

of convergence to Max-Cut which depends on the type of warm-start chosen. Theoretically characterizing this rate of convergence still remains an open question, but we show that there is a stark observable difference in rate of convergence through numerical simulations (See Table 3 for a summary, and Section 4.2 for more details). We discuss the choice of warm-starts next.

3 Warm-Starts

We now discuss the notion of warm-starting the quantum circuit for QAOA by biasing the initial quantum state $|s_0\rangle$ in Equation (1) to certain cuts, as opposed to taking an equal superposition of all cuts as in standard QAOA. Many initialization schemes have been used for warm-starting QAOA (see Table 1 for a summary); we focus on two that produce classical solutions which can easily be mapped to a separable quantum state in a way that roughly approximates the corresponding classical distribution of cuts.

Both approaches consider the k -dimensional relaxation (3) for Max-Cut. When $k = n$, this is the GW semidefinite program. Recall that the Goemans-Williamson algorithm [5] rounds an optimal solution $u_1, \dots, u_n \in \mathbb{R}^n$ to this SDP to a cut $v_1, \dots, v_n \in \{-1, 1\}$ using a random hyperplane. Since we wish to produce a separable quantum state, we instead round $u_1, \dots, u_n \in \mathbb{R}^n$ to vectors $v_1, \dots, v_n \in \mathbb{R}^k$, $k \in \{2, 3\}$ and map these to the Bloch sphere for each of the n qubits.

Alternately, we can solve the relaxation itself for $k \in \{2, 3\}$ [6], and map the locally-optimal solution vectors (called BM-MC solutions) directly to the corresponding Bloch spheres. While BM-MC solutions do not enjoy even the trivial 0.5-approximation guarantees [39], they are much faster to compute in practice and therefore scale better for larger graphs: e.g., Burer et al. [38] show that GW took over 1.5 days to complete on a 20,000 node instance, whereas a $k = 2$ BM-MC solution was found in a little over a second; repeated runs of BM-MC₂ over the course of a couple minutes on the same graph yielded cuts

that were at least as good as those obtained by GW [38].

Lastly, we compare the two warm-start techniques above to perturbed single-cut initializations, another warm-start technique that has been considered in previous literature [8, 9]. We show that, theoretically, such initializations give better depth-0 guarantees for QAOA (compared to the two warm-start techniques previously discussed); however, we later show (Section 4) that such initializations yield a comparatively much slower rate of convergence with increased circuit depth.

3.1 SDP-based Relaxations

Recall that a solution to the Goemans-Williamson (GW) SDP relaxation consists of n unit vectors $\{u_i \in \mathbb{R}^n : i \in V\}$, and their rounding algorithm uses a random hyperplane to obtain an approximation for the Max-Cut on the given graph G . We propose to create a warm-start to QAOA by instead rounding each of these vectors to \mathbb{R}^k (with $k \in \{2, 3\}$) and then mapping the rounded vectors to the Bloch sphere.⁶

Specifically, given $u \in \mathbb{R}^n$ for some positive integer n , and a linear subspace A of \mathbb{R}^n , let $\Pi_A(u)$ denote the (Euclidean) projection of u on A . Given $\Pi_A(u) \neq 0$, define

$$\Lambda_A(u) = \frac{\Pi_A(u)}{\|\Pi_A(u)\|_2},$$

as the *unit-scale* projection of u on A . This corresponds to normalizing $\Pi_A(u)$ so it is a unit vector.

To motivate warm-starts using solutions for the Goemans-Williamson SDP, consider the following two-step process for rounding an optimal SDP solution $\{u_i : i \in [n]\}$ to a cut:

- Choose a uniformly random linear subspace A of \mathbb{R}^n of dimension $k \in \{2, 3\}$, and consider the unit-scale projections $\Lambda_A(u_i) \in \mathbb{R}^k$, $i \in [n]$.

⁶An iterative rounding of the SDP solution was independently shown by Parekh and Thompson, a couple of months before our update on arXiv, for a more general setting of the Quantum Max-Cut [40].

Initializations	Mixers	Citation	Worst case AR at $p = 0$	Converges to Max-Cut?	Comment
Equal	Standard mixer	[3]	0.5	Yes	Standard and custom mixers are
	Custom mixer				
BM-MC $_k$	Standard mixer (QAOA-warm)	[31]	0 ($k = 2$) 0.333 ($k = 3$)	No	equivalent for this initialization
	Custom mixers (QAOA-warmest)	This work	0 ($k = 2$) 0.333 ($k = 3$)	Yes	Converges quickly (see Section 4.2)
Projected GW	Standard mixer (QAOA-warm)	This work	0.658 ($k = 2$), 0.585 ($k = 3$), (Corollary 3)	No	
	Custom mixers (QAOA-warmest)	This work	0.658 ($k = 2$), 0.585 ($k = 3$), (Corollary 3)	Yes	Converges quickly (see Section 4.2)
Perturbed Single Cut (with regularization angle θ^*)	Standard mixer (QAOA-warm)	[8] ($\theta^* = 0$) case	$0.878 \cos^{2n}(\theta^*/2)$, (Observation 4)	No	
	Custom mixers (QAOA-warmest)	Appendix I of [9]	$0.878 \cos^{2n}(\theta^*/2)$, (Observation 4)	Yes	Converges slowly for small θ^* (see Section 4.2)
	Other custom mixers	Section 2.3 of [9]	0.878	No	AR result occurs at $(\gamma_1, \beta_1) = (0, \pi/2)$

Table 1: A summary of results for different variants of QAOA (for Max-Cut) based on various combinations of initializations (equal superposition, BM-MC $_k$, projected GW, and perturbed single-cut initializations) and mixing Hamiltonian (standard, custom mixers (Section 2), and the mixer proposed by Egger et al. [9] for perturbed single-cut initializations). For various combinations, we state what is known regarding the convergence and worst-case approximation ratio (for depths $p \geq 0$) of the corresponding QAOA variant for graphs with non-negative edge weights.

- Use Goemans-Williams hyperplane rounding on vectors $\Lambda_A(u_i), i \in [n]$ to get a (random) cut M'

That is, round the vectors $u_i \in \mathbb{R}^n, i \in [n]$ to a random k -dimensional subspace first, and then subsequently use the Goemans-Williamson hyperplane rounding on these k -dimensional vectors.

The following theorem shows that this two-step rounding is equivalent to the Goemans-Williamson hyperplane rounding on vectors $u_i, i \in [n]$; we include the proof in Appendix A.

Theorem 2. *Suppose we are given unit vectors $u_1, \dots, u_n \in \mathbb{R}^n$ that form an optimal solution to the SDP relaxation for Max-Cut on some graph $G = (V, E)$ with n vertices and non-negative weights on the edges. Suppose the GW random hyperplane rounding on u_1, \dots, u_n obtains a (random) cut M of value X , and the two-step rounding described above produces a (random) cut M' of value Y . Then,*

1. *The random variables $X = Y$. In particular, $\mathbb{E}(X) = \mathbb{E}(Y)$ and therefore, in expectation, M' provides a 0.878-approximation to Max-Cut on G .*

2. *Furthermore, the two-step rounding procedure produces a cut of value $(0.878 - \epsilon)$ times the Max-Cut value with high probability if performed independently $\frac{\log n}{\epsilon}$ times for any constant $\epsilon \in (0, 1/2)$. Further, if $A_1, \dots, A_{\frac{\log n}{\epsilon}}$ are the intermediate k -dimensional subspaces in these $\frac{\log n}{\epsilon}$ runs, there is at least some A_i (with high probability) such that performing the hyperplane rounding on A_i produces a (random) cut of average value at least $(0.878 - \epsilon)$ times the Max-Cut.*

The last part of Theorem 2 illustrates that we can obtain a high-quality k -dimensional projected GW solution in regards to hyperplane rounding; however, it is natural to ask if any kind of guarantee can be preserved when mapping the solution to a quantum state. By adapting a theorem by Tate et al. [31], we can answer the question in the affirmative as seen in Corollary 3 below.

Corollary 3. *Let G be a graph with non-negative edge weights and let x be a corresponding κ -approximate projected GW solution in \mathbb{R}^3 with respect to hyperplane round-*

ing.⁷ Let $R_U(x)$ denote random uniform rotation applied to x , i.e., a global rotation where a uniformly selected point on the sphere gets mapped to $(0, 0, 1)$. Then initialization of QAOA with $R_U(x)$ has an (worst-case) approximation ratio of $\frac{2}{3}\kappa$ at $p = 0$, i.e., only using quantum sampling with initial state creation and no algorithmic depth for QAOA. Similarly, if x is a κ -approximate projected GW solution in \mathbb{R}^2 , initialization of QAOA with $R_U(x)$ is a $\frac{3}{4}\kappa$ -approximate solution at $p = 0$.

If x is chosen such that it is κ -approximate with $\kappa = 0.878 - \varepsilon$ (such an x is easily found via Theorem 2), then, for small ε , this yields (worst-case) approximation ratios (for depth-0 QAOA-warmest) of $\frac{3}{4}(0.878 - \varepsilon) \approx 0.658$ and $\frac{2}{3}(0.878 - \varepsilon) \approx 0.585$ for 2-dimensional and 3-dimensional projections respectively.

A proof of Corollary⁸ 3 can be found in Appendix A. This motivates us to use the projected vectors $\Lambda_A(u_i), i \in [n]$ to warm-start QAOA. We then use the same scheme as proposed by Tate et al. [31] to rotate and map the solution $x^* = \{\Lambda_A(u_i)\}_{i \in [n]}$ to the Bloch sphere. More specifically, we first perform either (1) a random vertex-at-top rotation of x^* (a global rotation to x^* so that a random vertex v coincides with $(0, 0, 1)$) or (2) a uniformly random rotation R_U , and apply the natural mapping from the rotated solution to the Bloch sphere to obtain a separable, unentangled state $|s_0\rangle$ [31]. Figure 1 illustrates the two-step rounding procedure and this warm-start.

3.2 Burer-Monteiro's Relaxation

For faster computational performance, we find a warm-start by using a locally optimal k -dimensional solution x^* (with $k \in \{2, 3\}$) obtained by the Burer-Monteiro relaxation for Max-Cut on a $(k - 1)$ -sphere. Similar to pro-

⁷The theorem also holds more generally for any feasible BM-MC₂ or BM-MC₃ solution.

⁸Note that Theorem 2 in [31] works with the assumption that x is κ -close whereas Corollary 3 assumes that x is κ -approximate (see Section 1.2.2).

jected GW solutions, we use the same rotation and quantum-mapping scheme to obtain an unentangled initial quantum state.

Theorem 2 of Tate et al. [31] shows that the same approximation ratios of $\frac{3}{4}\kappa$ and $\frac{2}{3}\kappa$ (for $k = 2$ and $k = 3$ solutions respectively) can be achieved for depth-0 QAOA-warmest if the solution is instead κ -close, i.e., $\text{BM-MC}_k(x^*) \geq \kappa \text{Max-Cut}(G)$. Using the same analysis as Goemans and Williamson [5], it is straightforward to show that a κ -close solution must also be 0.878κ -approximate (i.e. $\text{HP}(x) \geq 0.878\kappa \text{Max-Cut}(G)$). For locally optimal⁹ BM-MC_k solutions, Mei et al. [39] shows that such solutions are κ -close (and hence 0.878κ -approximate) with $\kappa = 1 - \frac{1}{k-1}$; thus, hyperplane rounding of such solutions yields (worst-case) approximation ratios of $0.878(0) = 0$ and $0.878(\frac{1}{2}) = 0.439$ while depth-0 QAOA-warmest can obtain (worst-case) approximation ratios of $\frac{3}{4}(0) = 0$ and $\frac{2}{3}(\frac{1}{2}) = \frac{1}{3}$ for $k = 2$ and $k = 3$ solutions respectively.

Experiments in Appendix F.3 compare the (instance-specific) approximation ratios achieved by hyperplane rounding of both types of warm-start initializations discussed; in particular, projected GW SDP solutions do well (compared to n -dimensional GW SDP solutions) but approximate BM-MC_k solutions degrade in performance as the number of nodes increases.

3.3 Perturbed Single Cut Initializations

For the purposes of comparison in our numerical simulations (Section 4), we briefly review one more type of warm-start which we refer to as perturbed single-cut initializations that other researchers [8, 9] have used for QAOA; more details regarding this approach can be

⁹When x^* is a globally optimal BM-MC_k solution, it immediately follows that x^* is 1-close and 0.878 -approximate and hence (worst-case) approximation ratios of $\frac{3}{4}$ and $\frac{2}{3}$ (for $k = 2$ and $k = 3$ solutions respectively) are achieved for depth-0 QAOA-warmest; however, since the Burer-Monteiro relaxation is non-convex, finding such globally optimal solutions becomes intractable, especially as the number of nodes increases.

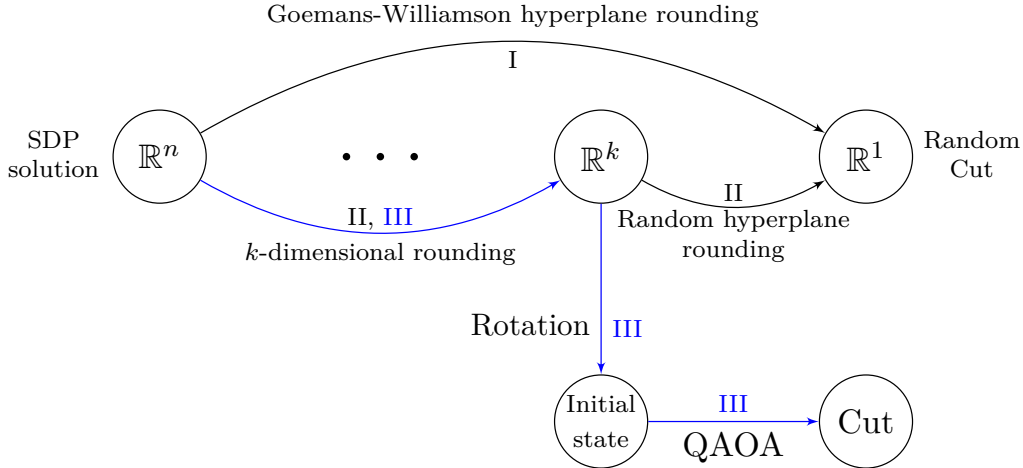


Figure 1: A schematic for GW warm-starts. There are three procedures to obtain a cut from an SDP solution. The first is to use Goemans-Williamson hyperplane rounding (on the top labelled I). The second (labelled II) is to do a two-step rounding through an intermediate state in \mathbb{R}^k , $k \in \{2, 3\}$. We prove that this two-step rounding procedure is equivalent to Goemans-Williamson hyperplane rounding in Theorem 2. The third procedure is our proposed warm-start of QAOA using the SDP solution (highlighted in blue, labelled III). This procedure involves rounding the SDP solution to \mathbb{R}^k first, then rotating this solution using uniform or vertex-at-top rotations and mapping to the Bloch sphere to get an initial state for QAOA, and finally running QAOA on this initial state.

found in (Appendix D). Given a regularization angle $\theta^* \in [0, \pi/2]$ and a cut $(S, V \setminus S)$, one can initialize the initial quantum state so that qubits lie along the xz -plane of the Bloch sphere with vertices in S and $V \setminus S$ being initialized at an angle θ^* away from the north and south poles of the Bloch sphere respectively; such a regularization angle aims to circumvent issues regarding reachability[9].¹⁰

Though not stated in the related works, if the warm-starts are initialized close to single-cuts, then they retain the classical approximation factors of the respective single-cut:

Observation 4. *Let $G = (V, E)$ be a graph, and suppose a cut $(S, V \setminus S)$ is an α -approximation to $\text{Max-Cut}(G)$, and is used to initialize a warm-start using regularization angle of $\theta^* \in [0, \pi/2]$. Then, a quantum measurement (i.e. depth-0 QAOA) of this state yields an approximation ratio of at least $\alpha \cdot \cos(\theta^*/2)^{2|V|}$.*

¹⁰It can be shown that if the initial qubit position lies on the poles of the Bloch sphere, e.g. if the initial qubit position is $|0\rangle$ (north pole), then it is guaranteed that that qubit will be measured to be $|0\rangle$ as the end of the QAOA algorithm if custom mixers (as described in Section 2) are used.

It is easy to see that such warm-starts approach an approximation ratio (at $p = 0$) of 0.878, when initialized randomly with the distribution of cuts obtained using the Goemans-Williamson algorithm, as the regularization angle θ^* approaches zero. We discuss a proof in the Appendix A to show this. Although Proposition 4 suggests it may be the superior warm-start technique for small regularization angle θ^* ; convergence is either lacking or slow empirically depending on the mixer used (see Table 1).

4 Numerical Simulations and Experiments

In this section, we demonstrate the importance of using a suitable warm-start and show that with such warm-starts, QAOA-warmest outperforms the Goemans-Williamson Max-Cut algorithm as well as standard QAOA [3] and QAOA-warm [31]. In particular, we show that QAOA-warmest does at least as well as these other algorithms at depths $p \geq 4$ for nearly every instance in our instance library. We consider comparisons with respect to a recent warm-starts approach of Egger et al. [9]

in Appendix F.

As discussed earlier (Section 3), for positive-weighted graphs, perturbed single-cut initializations have better depth-0 guarantees compared to the projected GW warm-starts that we propose. While possibly more advantageous at extremely low depths ($p = 0, 1$), we show that QAOA with perturbed single-cut initializations empirically converge to an optimal cut very slowly with increased circuit depth; for small enough regularization angle θ^* , the convergence towards an optimal cut is not even perceivable at the circuit depths tested. Meanwhile, with suitable warm-starts, the convergence is much quicker: for over 98% of instances tested, we found that depth-8 QAOA-warmest with BM-MC₂ initializations yields nearly optimal cuts.

Lastly, we consider the effects of noise on QAOA and its variants in the context of actual quantum devices (i.e. the IBM-Q Guadalupe and Quantinuum H1-1 devices). We show that for QAOA and its variants, the noise from these devices flattens the landscape without significantly altering the location of local extrema. Additionally, we show that even with noise, QAOA-warmest (with suitable warm-start) maintains a significant fraction of its expected solution quality, which suggests it may be useful for near-term NISQ devices (potentially with some noise mitigation) [1].

4.1 Simulation Details

For our simulations, we use the CI-QuBe library¹¹ [41] 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 \mathcal{G} , have a varied distribution of various graph properties, which is important when testing heuristics and algorithms for solving this problem.

In our simulations, for each instance, we first find five locally approximate solutions to BM-MC₂ and keep the best (in terms of

¹¹<https://github.com/swati1729/CI-QuBe>

the BM-MC₂ objective value). We do the same for BM-MC₃. Similarly, for each instance, we solve the GW SDP, perform 5 projections to random 2-dimensional subspaces, and keep the best (in terms of the BM-MC₂ objective); this process is repeated (using the same GW SDP solution) with projections to 3-dimensional subspaces. Next, for both the best BM-MC₂ and best BM-MC₃ solution, and for both of the best projected GW solutions (in 2 and 3 dimensions), we perform 5 different vertex-at-top rotations and 5 different uniform rotations, yielding 40 different initial warm-started quantum states per instance. We run QAOA-warm and QAOA-warmest using all 40 of these warm-started states and, for each combination of dimension and rotation scheme, record which one performed the best in terms of (instance-specific) approximation ratio (as defined in Section 1.2.3). Finally, we run standard QAOA on the instance.

For each run for each variant of QAOA, we initialize the variational parameters γ and β close to zero¹² and each run terminates when the difference in successive values of $F_p(\gamma, \beta)$ in the optimization loop is less than $\bar{W} \times 10^{-6}$ where \bar{W} is the sum of the absolute values of the edge weights.

To simplify the results, the figures and tables in this section will only consider runs of QAOA-warmest that use BM-MC₂ initializations with vertex-at-top rotations. This choice is due to runtime considerations and to allow for easier comparisons with previous related literature [9, 31]; more details on the results and the choice of this decision can be found in Appendix F.

Additionally, for conciseness, in this section we will use “approximation ratio” to mean the instance-specific approximation ratio as described in Section 1.2.3.

¹²For standard QAOA, many optimizers will immediately terminate if initialized exactly at the origin due to the presence of a saddle point. Instead, each variational parameter $(\gamma_1, \dots, \gamma_p, \beta_1, \dots, \beta_p)$ is initialized by sampling uniformly from the interval $[-0.0001, 0.0001]$.

	1st Best	QAOA-warmest				Standard QAOA			GW			Tie
	2nd Best	*	Standard	Warm	GW	Warmest	Warm	GW	Warm	Warmest	Standard	
Positive Weighted Graphs	p=1	90.3 %	0.69%	18.85%	15.22%	0.0%	0.0%	0.0%	0.17%	9.51%	0.0%	55.53%
	p=2	98.1%	0.69%	20.24%	25.08%	0.0%	0.0%	0.0%	0.0%	1.9%	0.0%	52.07%
	p=4	100%	8.65%	17.64%	20.58%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	53.11%
	p=8	100%	25.77%	5.01%	2.94%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	66.26%
All Graphs	p=1	90.6%	0.35%	17.96%	17.43%	0.0%	0.0%	0.0%	0.53%	8.84%	0.0%	54.86%
	p=2	98.1%	0.44%	20.17%	24.86%	0.0%	0.0%	0.0%	0.26%	1.68%	0.0%	52.56%
	p=4	99.6%	9.11%	19.2%	18.93%	0.0%	0.0%	0.0%	0.08%	0.26%	0.0%	52.38%
	p=8	99.6%	27.16%	7.96%	3.18%	0.17%	0.0%	0.0%	0.26%	0.0%	0.0%	61.23%

Table 2: 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). Both QAOA-warm and QAOA-warmest use BM-MC₂ warm-starts. There is a tie (last column) if the top two algorithms have approximation ratios that differ by no more than 0.01. QAOA-warmest is a part of every tie. Each instance is either accounted for in “Tie” or the other columns. 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.

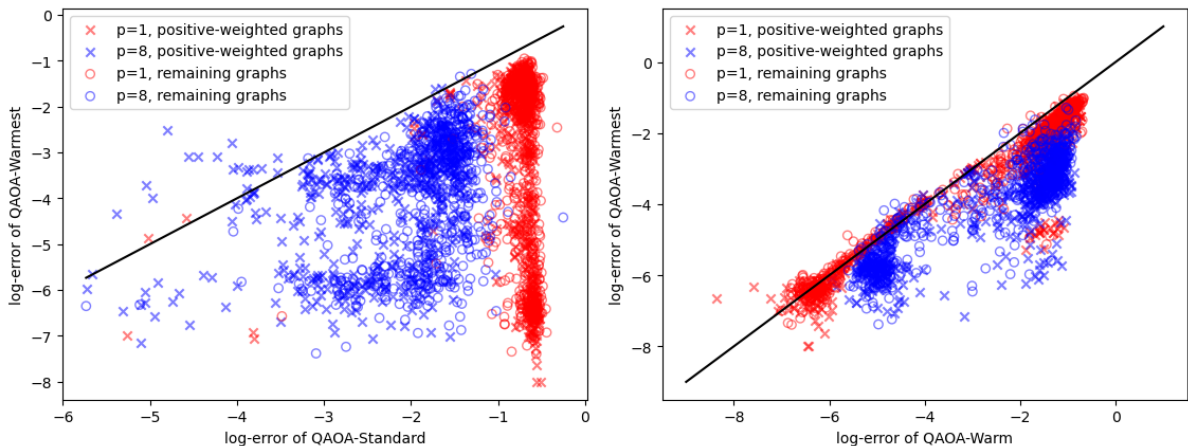


Figure 2: For both plots, we compare the log-error of QAOA-warmest to both QAOA-warm (right) and standard QAOA (left). BM-MC₂ warm-starts are used for both approaches. Each marker in the plot corresponds to a combination of instance (from our graph ensemble \mathcal{G}) and circuit depth (either $p = 1$ 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.

4.2 Comparing QAOA-warmest to Other Methods

In Table 2, 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 or performs as well as every other QAOA variant considered and eventually performs at least as well as GW as the circuit depth increases. 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 ap-

proximation ratio is less than 0.02. More information regarding these five instances can be found in Appendix F.

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 $AE = 1 - AR$ where AR is the approximation ratio. Additionally, we will refer to $\log_{10}(AE)$ as the log-error. Figure 2 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 standard QAOA 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 in performance with increased circuit depth for small instances).

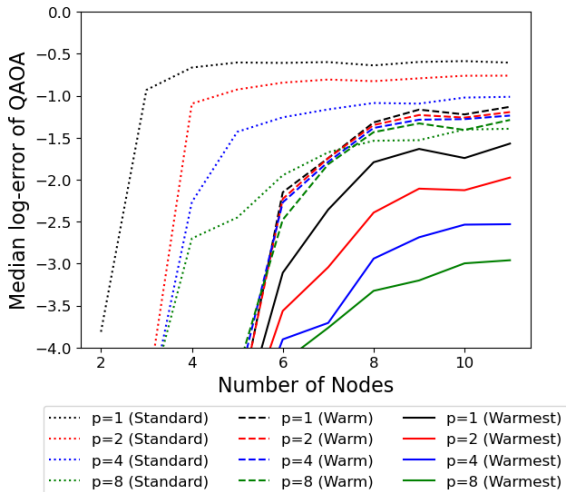


Figure 3: Trends in median log-error of standard QAOA (dotted), QAOA-warm (dashed), and QAOA-warmest (solid) as one varies the number of nodes and circuit depths; the median is taken across graphs in instance library \mathcal{G} . BM-MC₂ warm-starts are used for both QAOA-warm and QAOA-warmest.

Next, we show the trend in approximation quality with increase in the number of nodes n and the depth of the circuit p , in Figure 3. 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.

In Figure 4, we compare the convergence rates of standard QAOA and QAOA-warmest with various initializations: BM-MC_k, perturbed single-cut initializations (as described in Section 3.3), and uniform random initializations¹³ with custom mixers for a random 10-node instance in our graph library. Consistent with what is seen in the other figures, we see that standard QAOA, QAOA-warmest, and uniform random initializations quickly achieve high approximation ratios at relatively low circuit depths, with the BM-MC₂ initialization doing the best amongst the three across all circuit depths tested. On the other hand, perturbed single-cut initializations do not converge as quickly; in particular, when θ^* is small ($\theta^* \in \{0.1, 0.01\}$) hardly any improvement in the approximation ratio is observed at all. For larger regularization angles ($\theta^* = 0.5$), we do see worse performance at low depths ($p = 0, 1$) as well as a noticeable increase in performance with increased circuit depth; however, the amount of this increase is small compared to achieved by QAOA-warmest which begins to outperform the perturbed single-cut initialization (with $\theta^* = 0.5$) for $p \geq 2$. We find similar qualitative results for most other instances in our graph ensemble \mathcal{G} .

Table 3 provides a more aggregated view of the convergence of QAOA-warmest with dif-

¹³Here, a uniform random initialization refers to a separable state that is randomly created by (independently) picking a position on the surface of the Bloch sphere uniformly at random for each qubit, and then tensorizing the qubits. Although the phase of each qubit does not effect the expected cut value obtained from QAOA-warmest (as discussed in Section 2), it should be noted that the distribution of initializations obtained from sampling from the surface of the Bloch sphere and then removing the phases is different than the distribution of initializations obtained from sampling uniformly from the portion of the great-circle formed by the intersection of the Bloch sphere surface with the xz -plane with positive x -coordinate. Nonetheless, we have verified (via numerical simulation) that QAOA-warmest performs similarly between the two randomization schemes.

ferent choices of initializations across the entire instance library \mathcal{G} . For each combination of initialization method and circuit depth, the table states the percentage of instances in the library which achieved an instance-specific approximation ratio of 99.0% or higher. The data for the perturbed single-cut initializations were obtained as follows: for each instance, we obtained an optimal solution to the GW SDP relaxation, we performed 100 hyperplane roundings on the optimal SDP solution to obtain 100 cuts, we discarded all cuts whose value is more than $0.98 \cdot \text{MAX-CUT}(G)$, and we used the best remaining cut to create a perturbed single-cut initialization with regularization angle $\theta^* = 0.1$ (as described in Section 3.3); the discarding of cuts with very high values were done in order to ensure that, in the case of a high instance-specific approximation ratio, such a ratio can be partly attributed to the quantum circuit and not just the initial cut itself. When using the BM-MC₂ initializations (with vertex-at-top rotations), there are steady improvements with increased circuit depths with 42.3% of the instances achieving an instance-specific AR of 99.0% at depth-0; this percentage increases to 98.1% at depth-8. With the standard QAOA initialization, $|+\rangle^{\otimes n}$, none of the instances achieve an instance-specific AR of 99.0% or more; it is not until depth $p = 8$ that we see a considerable fraction of the graphs (39.4% respectively) achieving such an AR. As for the perturbed single-cut initialization with small regularization angle, we find that nearly none of the instances achieve an instance-specific AR of 99.0% with the exception of a few instances at depth $p = 8$.

4.3 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. For both the ideal and noisy simulation, we use IBM’s Qiskit software package [7]. In the case of the noisy simulation, we exercise the capability of Qiskit

	$p = 0$	$p = 1$	$p = 2$	$p = 4$	$p = 8$
BM-MC ₂	42.3%	57.8%	75.0%	91.9%	98.1%
Standard Initialization $ +\rangle^{\otimes n}$	0%	0.6%	2.4%	8.7%	39.4%
Single Cut Initialization with $\theta^* = 0.1$	0%	0%	0%	0%	0.7%

Table 3: The percentage of instances for which QAOA-warmest achieves an instance-specific AR of 99.0% for each combination of circuit depth and initialization method (standard initialization, BM-MC₂ initialization, perturbed single-cut initialization with $\theta^* = 0.1$).

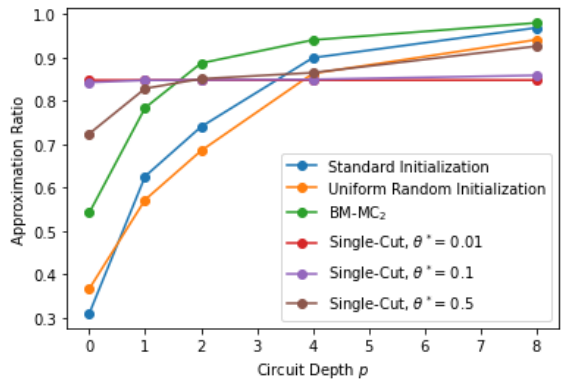


Figure 4: Instance specific approximation ratios achieved by QAOA-warmest with various types of initializations: standard initialization (equivalent to standard QAOA), BM-MC₂ initialization, perturbed single-cut initialization, and uniform random initializations) for a randomly selected 10-node instance. For QAOA-warmest, we used a BM-MC₂ initialization with a vertex-at-top rotation; here we intentionally chose the worst vertex (i.e. the one with worst AR at depth-0) to better illustrate the convergence rate. For the perturbed single-cut initialization, we chose a cut that obtains an instance-specific approximation ratio of 0.848 and created initial quantum states using regularization angles of $\theta^* = 0.01, 0.1$, and 0.5 radians. For the uniform random initializations, five such initializations were created and only the best one was kept (i.e. the one with best AR at depth-0).

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.

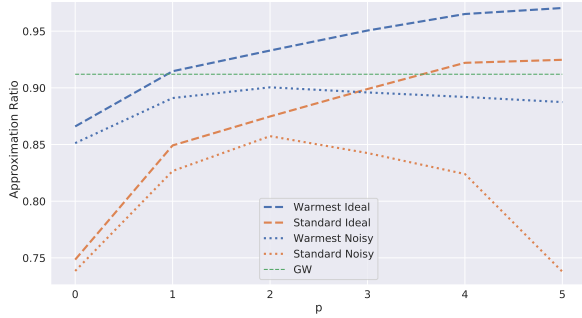


Figure 5: Performance of QAOA-warmest (with BM-MC₂ warm-starts) and standard QAOA as a function of QAOA depth for an ideal (dashed) and noisy simulation (dotted). For the chosen 20-node graph, GW achieves an approximation ratio of 0.912, while in the ideal case, QAOA-warmest outperforms GW for $p \geq 2$ while standard QAOA requires $p > 4$. The noise simulation is based on calibration data from IBM-Q’s Guadalupe device.

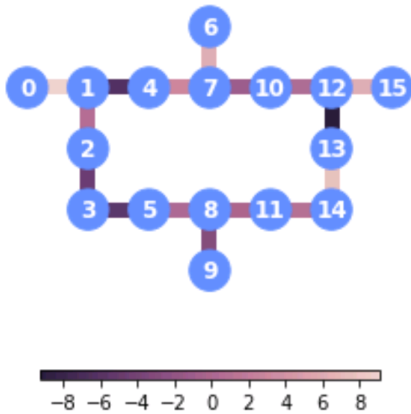


Figure 6: IBMQ Guadalupe device which shows the physical connectivity of qubits. We choose a native graph which matches this connectivity and random weights as indicated by color.

While these serve as a good starting point to model the noise in a quantum device, as shown in the Figure 8, there 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 an example demonstration of how typical noise models are utilized, see Appendix E.

We show in Figure 5 the performance of QAOA-warmest and standard QAOA on an instance generated via a construction by Karloff [42]; this unweighted graph is cho-

sen due to the fact that it is a small graph that achieves a GW approximation ratio of 0.912 (see Appendix G) which is close to the lower bound of 0.878 provided by the GW algorithm. 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 all QAOA depths and outperforms GW after $p > 1$. We also consider a noise model utilizing Qiskit’s built in modules [7] 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.

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 role in performance due to the overhead of inserting swap operations in order to compensate for limited connectivity [43]. 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 6. We assign randomly chosen weights to each edge chosen from a uniform distribution $[-10, 10]$. Finding the Max-Cut 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 7. The color scale is shared across all plots, showing that QAOA-warmest is able to find larger cut values as compared to standard QAOA, both in simulation and on actual hardware. For hardware results, we apply the efficient SPAM noise mitigation strategy based on a CTMP strategy [44, 45].

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 Fig-

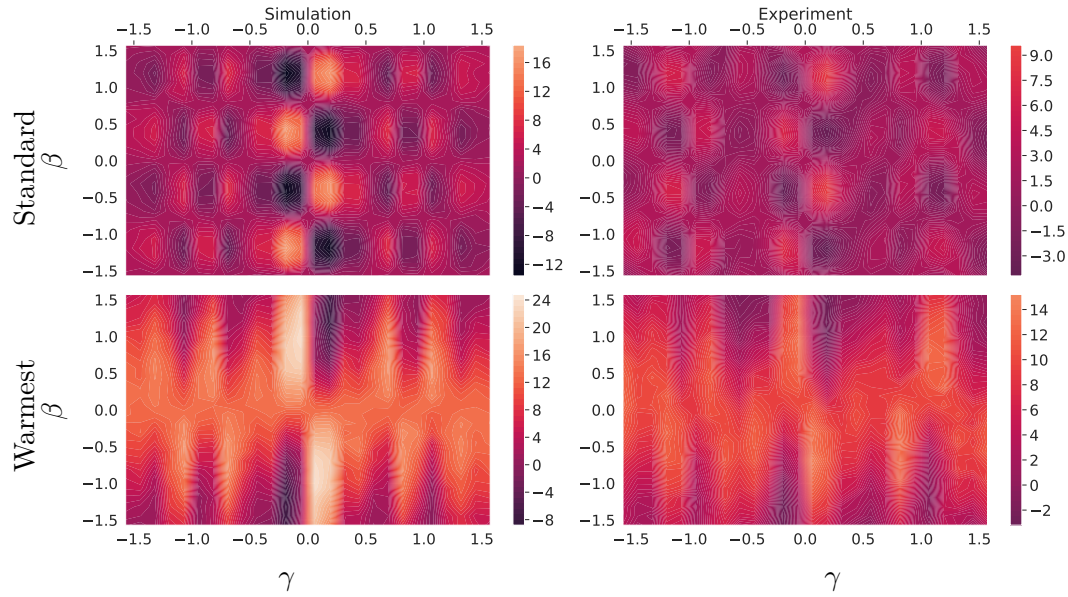


Figure 7: Performance of QAOA-warmest (with BM-MC₂ warm-starts) compared to standard QAOA in an ideal simulation and on IBM-Guadalupe hardware. Each subfigure is a scan of $p = 1$ parameters β vs γ , brighter regions indicating values which result in a larger cut. All figures share the same absolute color scale.

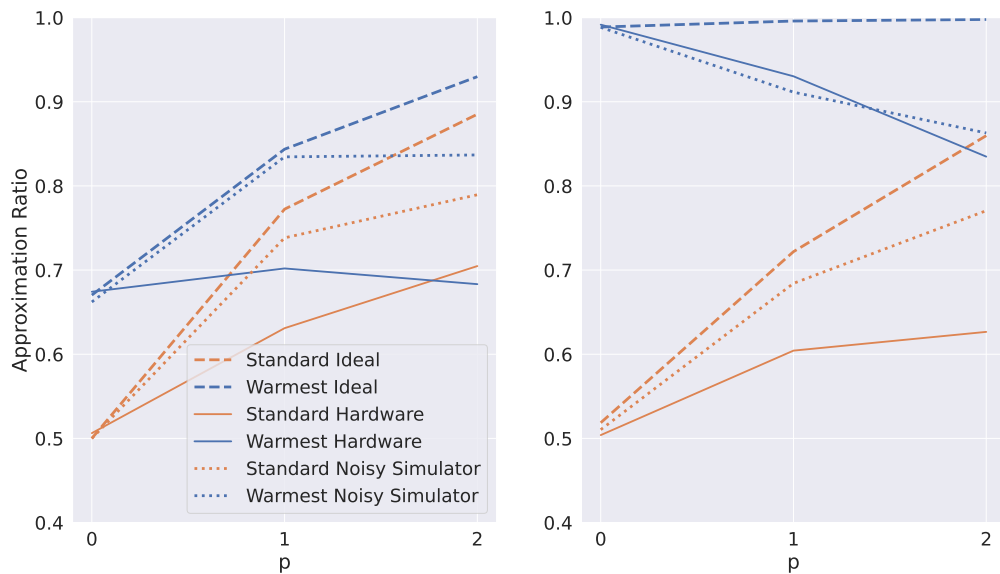


Figure 8: Performance of QAOA-warmest (with BM-MC₂ warm-starts) compared to standard QAOA in an ideal simulation (dashed), noisy simulation (dotted) and on IBM Guadalupe hardware (solid). Each subplot considers a different native hardware graph with randomly selected weights as well as a different choice of initialization procedures. (Left) For a random initialization of the classically informed QAOA-warmest start rotation. (Right) For an efficiently selected, optimal choice for the classically informed QAOA-warmest start rotation.

ure 8. We define $p = 0$ to simply mean the preparation and measurement of the ini-

tial state.¹⁴ In the case of QAOA-warmest,

¹⁴The warm-starts come from $k = 2$ or $k = 3$ solu-

this directly demonstrates the ability of the QPU to create and measure the classically suggested cut.

In addition, Figure 8 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 ($p > 0$).

Finally, we show results for QAOA-warmest run on Quantinuum hardware in Figure 9. This 20-ion linear trap allows for arbitrary qubit connectivity and thus has no overhead associated with mapping a specific graph to the hardware. In this case, we again consider the 20-node Karloff instance graphs used in Figure 5, but here we use the GW warmest start initialization. Notably we utilize a uniform rotation which gives a large initial approximation ratio (at $p = 0$) and while hard-

tions whereas as the GW algorithm uses n -dimensional 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.

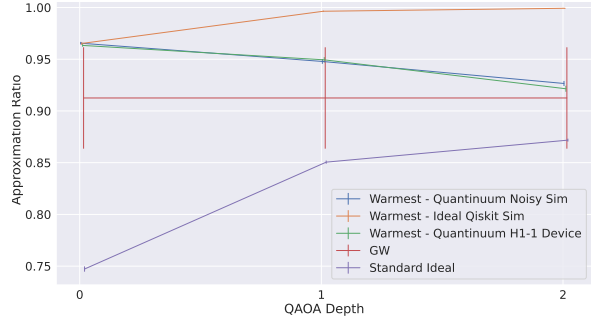


Figure 9: QAOA-warmest performance on Quantinuum simulators and hardware. The 20-node Karloff instance considered here is directly mapped to the fully-connected Quantinuum 20-ion hardware. In contrast to Figure 5, here we use a GW warmest start and find that this particular initialization outperforms GW on average for $p \leq 2$.

ware cannot improve on this initial state, the degradation is small considering that each p layer requires 90 two-qubit ZZ interactions (among many other single qubit operations). We also note the close agreement of the noisy simulator to the actual hardware results. In order to reduce the cost of these hardware runs, we only consider a single objective function evaluation (with 1000 shots), using noiseless simulations to find the optimal γ_i, β_i at each p depth. Even with these considerations, we see that the GW Warmest initialization outperforms the average GW performance on hardware up to $p \leq 2$. These results indicate that current quantum hardware is very close to demonstrating improvement over the Goemans-Williamson algorithm for Max-Cut on known hard instance graphs when using the QAOA-warmest initialization procedure, and it already outperforms the average performance of GW on this particular graph.

5 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 ini-

Instances Where Depth-8 Standard QAOA Performs Best

Instance ID	QAOA-warmest	QAOA-warm	Standard QAOA	GW
778	0.9550	0.8980	0.9635	0.9504
1820	0.9483	0.9078	0.9508	0.9429

Instances Where GW Performs Best

Instance ID	QAOA-warmest	QAOA-warm	Standard QAOA	GW	QAOA-warmest (modified)
1698	0.9899	0.9950	0.9677	0.9968	0.9998
1889	0.9867	0.9886	0.9461	0.9899	0.9995
2010	0.9762	0.9797	0.9330	0.9948	0.9992

Table 4: These tables reports 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.

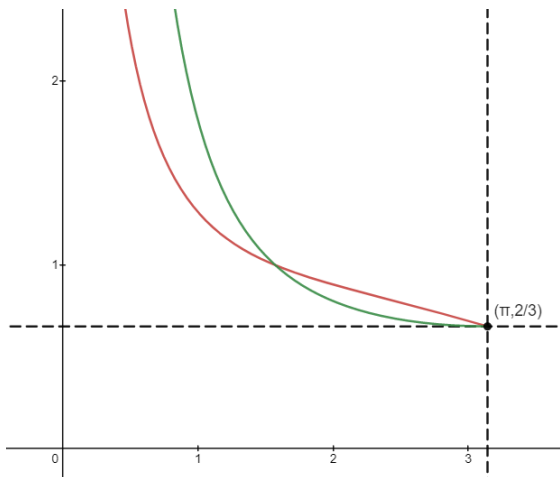


Figure 10: In red, the function $\frac{1}{2} \frac{(1 - \frac{\cos \theta}{k})}{\theta/\pi}$ that is minimized in the proof of Corollary 3 with $k = 3$. In green, a similar function $\frac{1}{2} \frac{(1 - \frac{\cos \theta}{k})}{1 - \cos(\theta)}$ that is minimized in the proof Theorem 2 of [31] (where the BM-MC $_k$ objective is compared to the maximum cut instead of the expected cut value from hyperplane rounding) with $k = 3$. Over the interval $[0, \pi]$, both achieve a minimum value of $2/3$ at $\theta = \pi$. The corresponding plots for $k = 2$ are similar but instead both functions reach a minimum value of $3/4$ at $\theta = \pi$.

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). Overall, finding a suitable initialization is needed in order to see

success in QAOA-warmest. In the case of classically-inspired initializations (e.g. Burer-Monteiro Max-Cut relaxations or projected GW SDP solutions) 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 [26], QAOA needs to “see the whole graph” (i.e. have a high enough circuit depth) in order to achieve desirable 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 $_k$ relaxations or projected GW SDP solutions). This suggests that certain negative theoretical results

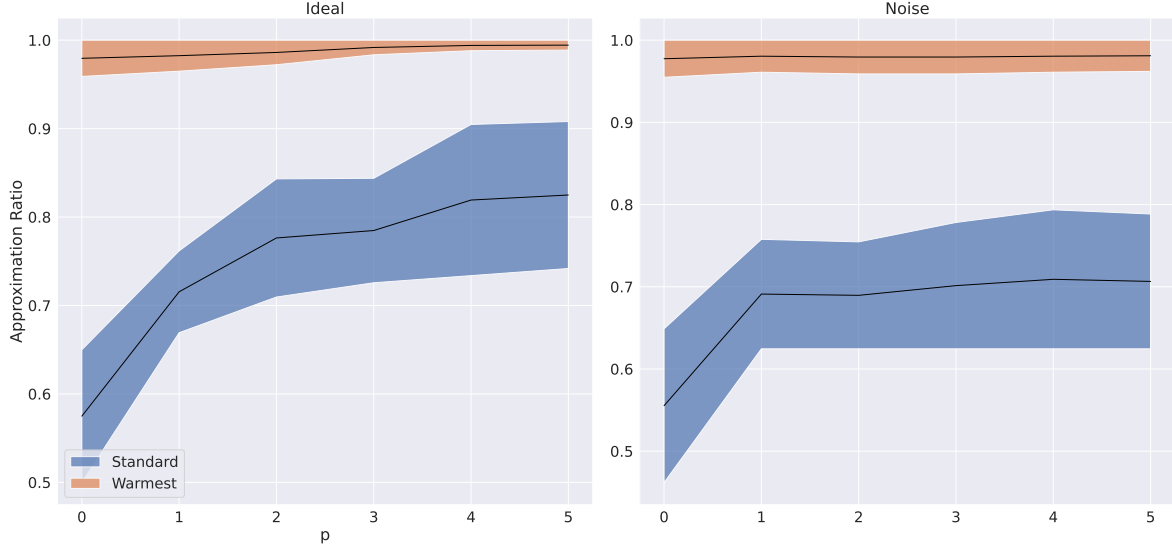


Figure 11: Comparison between standard QAOA mixer to using BM-MC₂ warm-starts with custom mixers. We show the noiseless (left) and noisy (right) case. In both cases, the custom mixer significantly outperforms the standard mixer. Shaded regions indicate the distribution of results for 20 randomly chosen 8 node graphs with positive and negative weights.

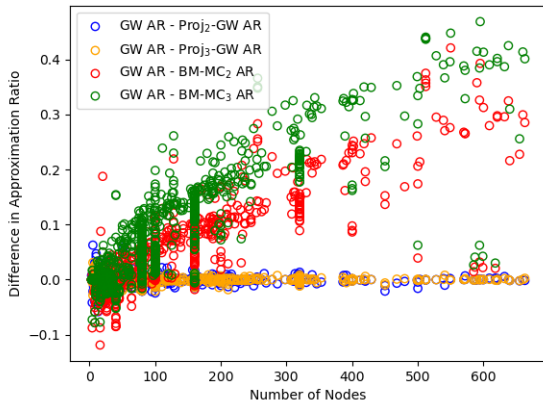


Figure 12: Difference in approximation ratio between n -dimensional GW hyperplane rounding and hyperplane rounding of various warm-start initializations (k -dimensional projected GW SDP solutions (Proj _{k} -GW) and approximate BM-MC _{k} solutions for $k = 2, 3$) as the number of nodes varies. For each instance and each dimension k , we obtained 5 random projections and 5 approximate BM-MC _{k} solutions, and then kept the best one (of 5) in regards to the BM-MC _{k} objective. Each circle in the figure corresponds to an instance from (a subset of) the MQLib library [4]; see Appendix F.3 for details.

seen for standard QAOA may not necessarily hold for QAOA-warmest since the distinguishability arguments used would no longer

apply.

The approximation guarantees for our warm-starts at $p = 0$ and convergence to Max-Cut (under adiabatic limit) combined with superior empirical performance provide strong evidence for quantum advantage of this approach at low circuit depths compared to existing classical methods, especially the Goemans-Williamson approximation. An interesting open question would be to quantify the approximation bounds obtained by QAOA-warmest for finite circuit depth greater than zero.

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

Table 5: 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.

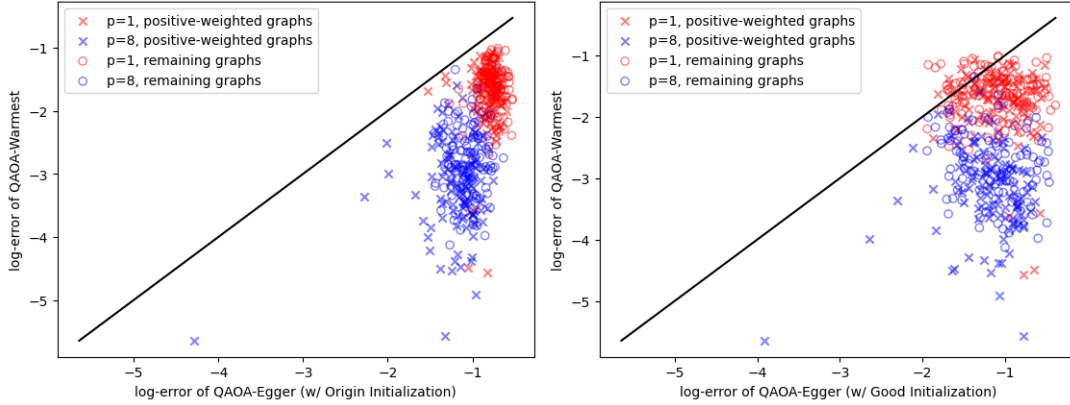


Figure 13: For both plots, we compare the log-error of QAOA-warmest (with BM-MC_2 warm-starts) to the variant of QAOA proposed by Egger et al. [9] for Max-Cut. For Egger et al.’s approach, we consider two different initializations: initializing the variational parameters to the origin (left) and initializing the parameters in a way that recovers the cut used to initialize the quantum state (right). Each marker in the plot corresponds to a combination of instance (from our graph ensemble \mathcal{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.

6 Methods

Numerical simulations were performed using both custom and pre-packaged codes in the Tensorflow [46] and Qiskit [7] software packages. Numerical experiments in Section 4.2 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 5.

Classical optimization was performed using standard optimizers available in PYTHON, including ADAM [47], L-BFGS-B [48], and COBYLA [49]. For hardware results, we first describe the usage of IBM’s Guadalupe device along with 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 3.7204×10^{-4} , two-qubit gate errors of 1.075×10^{-2} and measurement error of 1.776×10^{-2} , representing a quantum device of comparable quality to the state of the art. For the QAOA-Warmest runs shown in Fig 8, the run time on hardware can be estimated by the `schedule` method of Qiskit for

each corresponding circuit. The run times for $p = 0, 1, 2$ are 7182 ns, 14968 ns, 17379 ns, respectively. Standard QAOA runs have comparable run times as they differ only by single qubit gates as compared to QAOA-Warmest. For these hardware runs, we seed the classical optimization process with ideal parameters (γ^*, β^*) 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. Secondly, for our results on Quantinuum hardware, we only evaluate a single point in parameter space (at γ^*, β^*) at each p depth with 1000 shots. This choice was motivated in order to reduce the cost of hardware runs and do not represent any device or fundamental limitation. The Quantinuum H1-1 20 qubit device reports average single qubit gate errors of 5×10^{-5} , two qubit gate errors of 3×10^{-3} and SPAM errors of 3×10^{-3} . Noisy simulation of the Quantinuum device were also performed through the cloud, provided by the Quantinuum service.

7 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.

8 Author Contributions

Reuben Tate developed the theory for custom mixers, with guidance and helpful discussions with the team including Jai Moondra, Bryan Gard, Greg Mohler and Swati Gupta. Swati Gupta introduced the idea of iterative rounding of Goemans-Williamson SDP solution to lower dimensions. Jai Moondra developed theoretical guarantees for projected GW warm-starts with guidance from the team. All authors contributed to the design of the numerical simulations, and these were implemented by Reuben Tate. Bryan Gard designed and implemented the experiments on both IBM-Q and Quantinuum hardware. All authors contributed to the editing and writing of the manuscript to enable the best presentation of results.

9 Competing Interests

The authors declare that there are no competing financial or non-financial interests.

Acknowledgement

A part of this work was done when authors R. Tate and S. Gupta were at the Georgia Institute of Technology. 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. We acknowledge the use

of IBM Quantum services for this work. The views expressed are those of the authors and do not reflect the official policy or position of IBM or the IBM Quantum team. The authors would also like to thank Hassan Mortagy for his careful comments and feedback on an initial version of this work.

References

- [1] John Preskill. “Quantum computing in the NISQ era and beyond”. *Quantum* **2**, 79 (2018).
- [2] Aram W. Harrow and Ashley Montanaro. “Quantum computational supremacy”. *Nature* **549**, 203–209 (2017).
- [3] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. “A quantum approximate optimization algorithm” (2014).
- [4] Iain Dunning, Swati Gupta, and John Silberholz. “What works best when? A systematic evaluation of heuristics for Max-Cut and QUBO”. *INFORMS Journal on Computing* **30** (2018).
- [5] Michel X Goemans and David P Williamson. “Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming”. *Journal of the ACM (JACM)* **42**, 1115–1145 (1995).
- [6] Samuel Burer and Renato DC Monteiro. “A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization”. *Mathematical Programming* **95**, 329–357 (2003).
- [7] Héctor Abraham, AduOfiei, Rochisha Agarwal, Ismail Yunus Akhalwaya, Gadi Aleksandrowicz, et al. “Qiskit: An open-source framework for quantum computing” (2019).
- [8] Madelyn Cain, Edward Farhi, Sam Gutmann, Daniel Ranard, and Eugene Tang. “The QAOA gets stuck starting from a good classical string” (2022).
- [9] Daniel J. Egger, Jakub Mareček, and Stefan Woerner. “Warm-starting quantum optimization”. *Quantum* **5**, 479 (2021).
- [10] Stefan H Sack, Raimel A Medina, Richard Kueng, and Maksym Serbyn.

- “Recursive greedy initialization of the quantum approximate optimization algorithm with guaranteed improvement”. *Physical Review A* **107**, 062404 (2023).
- [11] Stefan H Sack and Maksym Serbyn. “Quantum annealing initialization of the quantum approximate optimization algorithm”. *quantum* **5**, 491 (2021).
- [12] Leo Zhou, Sheng-Tao Wang, Soonwon Choi, Hannes Pichler, and Mikhail D Lukin. “Quantum approximate optimization algorithm: Performance, mechanism, and implementation on near-term devices”. *Physical Review X* **10**, 021067 (2020).
- [13] Ruslan Shaydulin, Phillip C Lotshaw, Jeffrey Larson, James Ostrowski, and Travis S Humble. “Parameter transfer for quantum approximate optimization of weighted maxcut”. *ACM Transactions on Quantum Computing* **4**, 1–15 (2023).
- [14] Alexey Galda, Xiaoyuan Liu, Danylo Lykov, Yuri Alexeev, and Ilya Safro. “Transferability of optimal QAOA parameters between random graphs”. In 2021 IEEE International Conference on Quantum Computing and Engineering (QCE). Pages 171–180. IEEE (2021).
- [15] Johannes Weidenfeller, Lucia C Valor, Julien Gacon, Caroline Tornow, Luciano Bello, Stefan Woerner, and Daniel J Egger. “Scaling of the quantum approximate optimization algorithm on superconducting qubit based hardware”. *Quantum* **6**, 870 (2022).
- [16] Phillip C Lotshaw, Thien Nguyen, Anthony Santana, Alexander McCaskey, Rebekah Herrman, James Ostrowski, George Siopsis, and Travis S Humble. “Scaling quantum approximate optimization on near-term hardware”. *Scientific Reports* **12**, 12388 (2022).
- [17] Gian Giacomo Guerreschi and Anne Y Matsuura. “QAOA for max-cut requires hundreds of qubits for quantum speed-up”. *Scientific reports* **9**, 1–7 (2019).
- [18] Charles Moussa, Henri Calandra, and Vedran Dunjko. “To quantum or not to quantum: towards algorithm selection in near-term quantum optimization”. *Quantum Science and Technology* **5**, 044009 (2020).
- [19] Colin Campbell and Edward Dahl. “QAOA of the highest order”. In 2022 IEEE 19th International Conference on Software Architecture Companion (ICSA-C). Pages 141–146. IEEE (2022).
- [20] Rebekah Herrman, Lorna Treffert, James Ostrowski, Phillip C Lotshaw, Travis S Humble, and George Siopsis. “Impact of graph structures for QAOA on maxcut”. *Quantum Information Processing* **20**, 1–21 (2021).
- [21] Gopal Chandra Santra, Fred Jendrzejewski, Philipp Hauke, and Daniel J Egger. “Squeezing and quantum approximate optimization” (2022).
- [22] Ruslan Shaydulin, Stuart Hadfield, Tad Hogg, and Ilya Safro. “Classical symmetries and the quantum approximate optimization algorithm”. *Quantum Information Processing* **20**, 1–28 (2021).
- [23] Jonathan Wurtz and Peter Love. “Maxcut quantum approximate optimization algorithm performance guarantees for $p > 1$ ”. *Physical Review A* **103**, 042612 (2021).
- [24] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. “Quantum algorithms for fixed qubit architectures” (2017).
- [25] Sergey Bravyi, Alexander Kliesch, Robert Koenig, and Eugene Tang. “Obstacles to variational quantum optimization from symmetry protection”. *Physical Review Letters* **125**, 260505 (2020).
- [26] Edward Farhi, David Gamarnik, and Sam Gutmann. “The quantum approximate optimization algorithm needs to see the whole graph: A typical case” (2020).
- [27] Sergey Bravyi, Alexander Kliesch, Robert Koenig, and Eugene Tang. “Hybrid quantum-classical algorithms for approximate graph coloring”. *Quantum* **6**, 678 (2022).
- [28] Matthew B Hastings. “Classical and quantum bounded depth approximation algorithms” (2019).

- [29] Kunal Marwaha. “Local classical max-cut algorithm outperforms $p = 2$ QAOA on high-girth regular graphs”. *Quantum* **5**, 437 (2021).
- [30] Boaz Barak and Kunal Marwaha. “Classical algorithms and quantum limitations for maximum cut on high-girth graphs” (2021).
- [31] Reuben Tate, Majid Farhadi, Creston Herold, Greg Mohler, and Swati Gupta. “Bridging classical and quantum with SDP initialized warm-starts for QAOA”. *ACM Transactions on Quantum Computing* (2022).
- [32] Stuart Hadfield, Zihui Wang, Bryan O’Gorman, Eleanor G. Rieffel, Davide Venturelli, and Rupak Biswas. “From the quantum approximate optimization algorithm to a quantum alternating operator ansatz”. *Algorithms* **12** (2019).
- [33] Zihui Wang, Nicholas C. Rubin, Jason M. Dominy, and Eleanor G. Rieffel. “ xy mixers: Analytical and numerical results for the quantum alternating operator ansatz”. *Phys. Rev. A* **101**, 012320 (2020).
- [34] Linghua Zhu, Ho Lun Tang, George S. Barron, F. A. Calderon-Vargas, Nicholas J. Mayhall, Edwin Barnes, and Sophia E. Economou. “Adaptive quantum approximate optimization algorithm for solving combinatorial problems on a quantum computer”. *Phys. Rev. Research* **4**, 033029 (2022).
- [35] Andreas Bärttschi and Stephan Eidenbenz. “Grover mixers for QAOA: Shifting complexity from mixer design to state preparation”. In 2020 IEEE International Conference on Quantum Computing and Engineering (QCE). Pages 72–82. IEEE (2020).
- [36] Zhang Jiang, Eleanor G Rieffel, and Zihui Wang. “Near-optimal quantum circuit for grover’s unstructured search using a transverse field”. *Physical Review A* **95**, 062317 (2017).
- [37] Lov K Grover. “A fast quantum mechanical algorithm for database search”. In Proceedings of the twenty-eighth annual ACM symposium on Theory of computing. Pages 212–219. (1996).
- [38] Yin Zhang, Samuel Burer, and Renato D. C. Monteiro. “Rank-2 relaxation heuristics for max-cut and other binary quadratic programs”. *SIAM Journal on Optimization* **12**, 503—521 (2001).
- [39] Song Mei, Theodor Misiakiewicz, Andrea Montanari, and Roberto Imbuzeiro Oliveira. “Solving sdps for synchronization and maxcut problems via the grothendieck inequality”. In Conference on learning theory. Pages 1476–1515. PMLR (2017).
- [40] Ojas Parekh and Kevin Thompson. “An optimal product-state approximation for 2-local quantum hamiltonians with positive terms” (2022). [arXiv:2206.08342](https://arxiv.org/abs/2206.08342).
- [41] Reuben Tate and Swati Gupta. “Ciqube”. GitHub repository (2021). url: <https://github.com/swati1729/CI-QuBe>.
- [42] Howard Karloff. “How good is the Goemans–Williamson MAX-CUT algorithm?”. *SIAM Journal on Computing* **29**, 336–350 (1999).
- [43] Matthew P Harrigan, Kevin J Sung, Matthew Neeley, Kevin J Satzinger, Frank Arute, Kunal Arya, Juan Atalaya, Joseph C Bardin, Rami Barends, Sergio Boixo, et al. “Quantum approximate optimization of non-planar graph problems on a planar superconducting processor”. *Nature Physics* **17**, 332–336 (2021).
- [44] Sergey Bravyi, Sarah Sheldon, Abhinav Kandala, David C. McKay, and Jay M. Gambetta. “Mitigating measurement errors in multiqubit experiments”. *Phys. Rev. A* **103**, 042605 (2021).
- [45] George S. Barron and Christopher J. Wood. “Measurement error mitigation for variational quantum algorithms” (2020).
- [46] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefow-

- icz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. “TensorFlow: Large-scale machine learning on heterogeneous systems” (2015).
- [47] Diederik P. Kingma and Jimmy Ba. “Adam: A method for stochastic optimization” (2014).
- [48] Roger Fletcher. “Practical methods of optimization (2nd edition)”. *John Wiley and Sons*. New York, NY, USA (1987).
- [49] M.J.D. Powell. “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).
- [50] Alan J. Laub. “Matrix analysis for scientists and engineers”. *Volume 91*. Siam. (2005).
- [51] Georg Frobenius. “Ueber matrizen aus nicht negativen elementen”. *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften* Pages 456–477 (1912).
- [52] A. Kaveh and H. Rahami. “A unified method for eigendecomposition of graph products”. *Communications in Numerical Methods in Engineering with Biomedical Applications* **21**, 377–388 (2005).
- [53] Simon Špacapan. “Connectivity of cartesian products of graphs”. *Applied Mathematics Letters* **21**, 682–685 (2008).
- [54] Jacek Gondzio and Andreas Grothey. “Solving nonlinear financial planning problems with 109 decision variables on massively parallel architectures”. *WIT Transactions on Modelling and Simulation* **43** (2006).
- [55] Fan RK Chung. “Spectral graph theory”. *Volume 92*. American Mathematical Soc. (1997).
- [56] M. A. Nielsen and I. L. Chuang. “Quantum computation and quantum information: 10th anniversary edition”. *Cambridge University Press*, New York. (2011).
- [57] Vincent R. Pascuzzi, Andre He, Christian W. Bauer, Wibe A. de Jong, and Benjamin Nachman. “Computationally efficient zero-noise extrapolation for quantum-gate-error mitigation”. *Physical Review A* **105**, 042406 (2022).
- [58] Ewout Van Den Berg, Zlatko K Mineev, Abhinav Kandala, and Kristan Temme. “Probabilistic error cancellation with sparse pauli-lindblad models on noisy quantum processors”. *Nature Physics* Pages 1–6 (2023).
- [59] Nathan Krislock, Jérôme Malick, and Frédéric Roupin. “BiqCrunch: A semidefinite branch-and-bound method for solving binary quadratic problems”. *ACM Transactions on Mathematical Software* **43** (2017).
- [60] Andries E. Brouwer, Sebastian M. Cioabă, Ferdinand Ihringer, and Matt McGinnis. “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).
- [61] Donald Knuth. “Combinatorial matrices”. *Selected Papers on Discrete Mathematics* (2000).

A Proofs

Proof of Proposition 2

Proof. Let $|s_0\rangle = \bigotimes_{j=1}^n |s_{0,j}\rangle$ with $|s_{0,j}\rangle = \cos(\theta_j/2) |0\rangle + e^{i\phi_j} \sin(\theta_j/2) |1\rangle$ be an arbitrary separable initial state. As a consequence of Proposition 1, it suffices to show that QAOA-warmest with this initial state $|s_0\rangle$ yields the same expected cut value (at the same variational parameters) as another separable initial state $|s_0'\rangle$ where each qubit of $|s_0'\rangle$ lies in the xz -plane of the Bloch sphere with positive x -coordinate.

We consider the state $|s_0\rangle' = \bigotimes_{j=1}^n |s_{0,j}\rangle'$ where $|s_{0,j}\rangle' = \cos(\theta_j/2)|0\rangle + \sin(\theta_j/2)|1\rangle$. Geometrically, going from $|s_0\rangle$ to $|s_0\rangle'$ has the effect of dropping the phase for all qubits so that they lie in the xz -plane of the Bloch sphere with positive x -coordinate (assuming that none of the qubits are at the poles).

It suffices to show that we can drop the phase for single qubit of $|s_0\rangle$ (say qubit k) without changing the expected cut value; the argument can then be easily repeated for the remaining qubits to show that $|s_0\rangle$ and $|s_0\rangle'$ yield identical expected cut values. In this case, we consider the initial state $|\widehat{s}_0\rangle = \bigotimes_{j=1}^n |\widehat{s}_{0,j}\rangle$ where $|\widehat{s}_{0,k}\rangle = |s_{0,k}\rangle'$ and $|\widehat{s}_{0,j}\rangle = |s_{0,j}\rangle$ for $j \neq k$ (i.e. only the position of qubit k is modified). Letting $R_{x,k}(\theta)$, $R_{y,k}(\theta)$, $R_{z,k}(\theta)$ represent the standard rotation operator of the k th qubit (about axes x, y, z respectively) about the Bloch sphere by angle θ , we can also write

$$|\widehat{s}_0\rangle = R_{z,k}(-\phi_k)|s_0\rangle, \quad (6)$$

i.e., $|\widehat{s}_0\rangle$ can be obtained from $|s_0\rangle$ by rotating around the z -axis (of the Bloch sphere) by the appropriate amount.

Let H_B and \widehat{H}_B be the corresponding custom mixers for $|s_0\rangle$ and $|\widehat{s}_0\rangle$ respectively. Let $U_B(\beta_\ell) = \exp(-i\beta_\ell H_B)$ and $\widehat{U}_B(\beta_\ell) = \exp(-i\beta_\ell \widehat{H}_B)$. For convenience, let $U_C(\gamma_\ell) = \exp(-i\gamma_\ell H_C)$. We can write $U_B(\beta_\ell) = U_{B,\neq k}(\beta_\ell)U_{B,k}(\beta_\ell)$ where $U_{B,k}(\beta_\ell)$ is the portion of $U_B(\beta_\ell)$ that acts on qubit k and $U_{B,\neq k}(\beta_\ell)$ is the portion that acts on the remaining qubits; we can similarly write $\widehat{U}_B(\beta_\ell) = U_{B,\neq k}(\beta_\ell)\widehat{U}_{B,k}(\beta_\ell)$ (the part of the mixer that does not affect the k th qubit remains the same). Geometrically, the operation $U_{B,k}(\beta_\ell)$ corresponds to rotating qubit k around its original position on the Bloch sphere by angle $2\beta_\ell$ so,

$$U_{B,k} = \left[R_{z,k}(\phi_k)R_{y,k}(\theta_j)R_{z,k}(2\beta_\ell) \right. \\ \left. R_{y,k}(-\theta_j)R_{z,k}(-\phi_j) \right].$$

The equation above yields the following key

relation between $U_{B,k}$ and $\widehat{U}_{B,k}$:

$$R_{z,k}(-\phi_k)U_{B,k}R_{z,k}(\phi_k) \\ = R_{y,k}(\theta_j)R_{z,k}(2\beta_\ell)R_{y,k}(-\theta_j) = \widehat{U}_{B,k}. \quad (7)$$

For convenience, we will let

$$U(\gamma, \beta) = \prod_{\ell=1}^p \left[U_B(\beta_\ell)U_C(\gamma_\ell) \right],$$

and

$$\widehat{U}(\gamma, \beta) = \prod_{\ell=1}^p \left[\widehat{U}_B(\beta_\ell)U_C(\gamma_\ell) \right],$$

i.e., U_B and \widehat{U}_B correspond to the QAOA-warmest circuit (excluding the initial state) for $|s_0\rangle$ and $|\widehat{s}_0\rangle$ respectively.

The claim amounts to showing (up to some global phase) the following:

$$\langle s_0 | U(\gamma, \beta)^\dagger H_C U(\gamma, \beta) | s_0 \rangle \\ = \langle \widehat{s}_0 | \widehat{U}(\gamma, \beta)^\dagger H_C \widehat{U}(\gamma, \beta) | \widehat{s}_0 \rangle,$$

for any circuit depth p and any variational parameters $\gamma = (\gamma_1, \dots, \gamma_p)$ and $\beta = (\beta_1, \dots, \beta_p)$; and in particular, QAOA-warmest gives the same expected cut value for both $|s_0\rangle$ and $|\widehat{s}_0\rangle$.

First we observe that,

$$U(\gamma, \beta) | s_0 \rangle \\ = \prod_{\ell=1}^p \left[U_B(\beta_\ell)U_C(\gamma_\ell) \right] | s_0 \rangle \\ = \prod_{\ell=1}^p \left[U_{B,\neq k}(\beta_\ell)U_{B,k}(\beta_\ell)U_C(\gamma_\ell) \right] | s_0 \rangle \\ = \prod_{\ell=1}^p \left[U_{B,\neq k}(\beta_\ell)R_{z,k}(\phi_k) \right. \\ \left. R_{z,k}(-\phi_k)U_{B,k}(\beta_\ell)R_{z,k}(\phi_k) \right. \\ \left. R_{z,k}(-\phi_k)U_C(\gamma_\ell) \right] | s_0 \rangle \\ = \prod_{\ell=1}^p \left[U_{B,\neq k}(\beta_\ell)R_{z,k}(\phi_k)\widehat{U}_{B,k}(\beta_\ell) \right. \\ \left. R_{z,k}(-\phi_k)U_C(\gamma_\ell) \right] | s_0 \rangle \quad (\text{by Equation 7})$$

$$\begin{aligned}
&= \prod_{\ell=1}^p \left[R_{z,k}(\phi_k) U_{B,\neq k}(\beta_\ell) \widehat{U}_{B,k}(\beta_\ell) \right. \\
&\quad \left. U_C(\gamma_\ell) R_{z,k}(-\phi_k) \right] |s_0\rangle \quad (\text{commutativity}) \\
&= \prod_{\ell=1}^p \left[R_{z,k}(\phi_k) \widehat{U}_B(\beta_\ell) U_C(\gamma_\ell) R_{z,k}(-\phi_k) \right] |s_0\rangle \\
&\quad (\text{combine } \widehat{U}_{B,k} \text{ and } U_{B,\neq k}) \\
&= R_{z,k}(\phi_k) \prod_{\ell=1}^p \left[\widehat{U}_B(\beta_\ell) U_C(\gamma_\ell) \right] R_{z,k}(-\phi_k) |s_0\rangle \\
&\quad (\text{telescoping}) \\
&= R_{z,k}(\phi_k) \prod_{\ell=1}^p \left[\widehat{U}_B(\beta_\ell) U_C(\gamma_\ell) \right] |\widehat{s}_0\rangle \\
&\quad (\text{by Equation 6}) \\
&= R_{z,k}(\phi_k) \widehat{U}(\gamma, \beta) |\widehat{s}_0\rangle.
\end{aligned}$$

We now finally show that QAOA-warmest initialized with $|s_0\rangle$ and $|\widehat{s}_0\rangle$ yield the same value; in particular the extraneous $R_z(\phi_k)$ term from the previous calculations will not effect the measurement due to commutativity with the cost Hamiltonian:

$$\begin{aligned}
&\langle s_0 | U(\gamma, \beta)^\dagger H_C U(\gamma, \beta) | s_0 \rangle \\
&= \langle \widehat{s}_0 | \widehat{U}(\gamma, \beta)^\dagger R_z(\phi_k)^\dagger H_C R_z(\phi_k) \widehat{U}(\gamma, \beta) | \widehat{s}_0 \rangle \\
&= \langle \widehat{s}_0 | \widehat{U}(\gamma, \beta)^\dagger H_C \widehat{U}(\gamma, \beta) | \widehat{s}_0 \rangle,
\end{aligned}$$

where the last equality follows since H_C commutes with $R_z(\phi_k)$. This completes the proof. \square

Proof of Theorem 2. Given a subspace A of \mathbb{R}^n , if $A = \text{span}(v)$ for some unit vector $v \in \mathbb{R}^n$, we abuse the notation and denote $\Pi_v(u) = \Pi_{\text{span}(v)}(u)$ and $\Lambda_v(u) = \Lambda_{\text{span}(v)}(u)$. We need two lemmas before we prove the theorem.

Lemma 5. *Let u, v be unit vectors in \mathbb{R}^n and let A denote a linear subspace of \mathbb{R}^n of dimension k such that $v \in A$. If $\Pi_A(u) \neq 0$ and $\Pi_v(u) \neq 0$, then*

$$\Lambda_v(u) = \Lambda_v(\Lambda_A(u)).$$

That is, unit-scale projection of u on v is equivalent to first unit-scale projecting u to A and projecting this projection $\Lambda_A(u)$ on v .

Proof. Let $\{v_1, \dots, v_k\}$ be an orthonormal basis for A . Let $\alpha_i = u^\top v_i$. Then

$$\Pi_A(u) = \sum_{i \in [k]} \alpha_i v_i, \quad \Lambda_A(u) = \frac{\sum_{i \in [k]} \alpha_i v_i}{\sqrt{\sum_{i \in [k]} \alpha_i^2}}$$

Since $v \in A$, write $v = \sum_{i \in [k]} \beta_i v_i$. Then we have

$$\Pi_v(\Lambda_A(u)) = \frac{\sum_{i \in [k]} \alpha_i \beta_i}{\sqrt{\sum_{i \in [k]} \alpha_i^2}} v,$$

so that

$$\begin{aligned}
\Lambda_v(\Lambda_A(u)) &= \frac{\Pi_v(\Lambda_A(u))}{\|\Pi_v(\Lambda_A(u))\|_2} \\
&= \frac{\sum_{i \in [k]} \alpha_i \beta_i}{\left| \sum_{i \in [k]} \alpha_i \beta_i \right|} v \\
&= \frac{\Pi_v(u)}{\|\Pi_v(u)\|_2} v \\
&= \Lambda_v(u).
\end{aligned}$$

\square

Note that the above lemma is deterministic statement; we have not used any randomness so far.

Let us consider what happens if we select a linear subspace A of \mathbb{R}^n of dimension k uniformly randomly from \mathbb{R}^n (one way to ensure it is chosen uniformly randomly is to select unit vectors $v_i \in \mathbb{R}^n, i \in [k]$ recursively so that v_i is chosen uniformly randomly in the space orthogonal to v_1, \dots, v_{i-1}). Once we have A , let us select a vector $v \in A$ uniformly randomly again. Is this equivalent to choosing a vector $v \in \mathbb{R}^n$ uniformly randomly? By symmetry, it is, since the former experiment is not biased in favor of any direction. We omit the formal proof and state it as a lemma here:

Lemma 6. *Let E denote the experiment of choosing a unit vector v chosen uniformly randomly from \mathbb{R}^n . Let E' denote the experiment of choosing a linear subspace A of dimension k uniformly randomly from \mathbb{R}^n , and then choosing a unit vector v' uniformly randomly from*

A. Then $E' = E$, i.e., they correspond to the same probability space.

We are ready for the proof of Theorem 2.

Proof. Let $U_n = \{v \in \mathbb{R}^n : \|v\|_2 = 1\}$ be the set of unit vectors in \mathbb{R}^n . Recall that for a given probability space, a random variable is a real-valued function on the sample space, or that $X, Y : U_n \rightarrow \mathbb{R}$. From Lemma 6, the two experiments correspond to the same probability space. Therefore, it is enough to show that $X(v) = Y(v)$ for all $v \in U_n$.

One key observation is that rounding on a random hyperplane is equivalent to unit-scale projecting to a uniformly random vector v : indeed, let v be the vector normal to the uniform hyperplane, then any unit vector u is rounded to 1 if $u \cdot v > 0$ and to -1 if $u \cdot v < 0$. That is, u is rounded to $\frac{u \cdot v}{|u \cdot v|} = \Lambda_v(u)^\top v$.

Therefore,

$$X(v) = \sum_{ij \in E(G)} w_{ij} \mathbf{1} \left[\Lambda_v(u_i)^\top v \cdot \Lambda_v(u_j)^\top v < 0 \right].$$

Similarly, for a given A such that $v \in A$, we have $Y(v)$ is equal to:

$$\sum_{ij \in E(G)} w_{ij} \mathbf{1} \left[\Lambda_v(\Lambda_A(u_i))^\top v \cdot \Lambda_v(\Lambda_A(u_j))^\top v < 0 \right].$$

Since $\dim(A) = k$ is a constant and A is chosen uniformly randomly, $\Pi_A(u_i) \neq 0$ for each i with probability 1. From Lemma 5, we have $\Lambda_v(\Lambda_A(u)) = \Lambda_v(u)$ for all unit vectors u and for all A .

Therefore, we have,

$$\begin{aligned} Y(v) &= \sum_{ij \in E(G)} w_{ij} \mathbf{1} \left[\Lambda_v(u_i)^\top v \cdot \Lambda_v(u_j)^\top v < 0 \right] \\ &= X(v). \end{aligned}$$

Since X and Y have the same distribution, the same approximation guarantee holds for X, Y . This proves part 1 of the theorem.

We prove part 2 next. Let C denote the maximum cut value on graph G , and denote $\alpha = 0.878$ for convenience. Then, part 1

shows that $\mathbb{E}Y \geq \alpha C$. We first show that $\Pr(Y > (1 - \epsilon)\mathbb{E}Y) \geq \alpha\epsilon$ using Markov inequality:

$$\begin{aligned} \Pr(Y > (1 - \epsilon)\mathbb{E}Y) &= 1 - \Pr(Y \leq (1 - \epsilon)\mathbb{E}Y) \\ &= 1 - \Pr(C - Y \geq C - (1 - \epsilon)\mathbb{E}Y) \\ &\geq 1 - \frac{\mathbb{E}(C - Y)}{C - (1 - \epsilon)\mathbb{E}Y} \\ &= \frac{\epsilon\mathbb{E}Y}{C - (1 - \epsilon)\mathbb{E}Y} \\ &\geq \frac{\frac{\mathbb{E}Y}{\alpha} - (1 - \epsilon)\mathbb{E}Y}{\alpha\epsilon} \\ &= \frac{\alpha\epsilon}{1 - \alpha(1 - \epsilon)} \\ &\geq \alpha\epsilon. \end{aligned}$$

Suppose that $\frac{\log n}{\epsilon}$ independent cuts are produced by applying the two-step rounding procedure $\log n$ times. Then the probability that all of these cuts have value less than $(1 - \epsilon)\mathbb{E}Y$ is at most

$$(1 - \alpha\epsilon)^{\frac{\log n}{\epsilon}} \leq (e^{-\alpha\epsilon})^{\frac{\log n}{\epsilon}} = \frac{1}{n^\alpha},$$

where we have used the standard inequality $\exp(-x) \geq 1 - x$. Since $\alpha \in (0, 1)$, this probability goes to 0 as n goes to ∞ .

We prove the second claim of part 2. Given a k -dimensional subspace A of \mathbb{R}^n , let w_A denote the average cut value after Goemans-Williamson hyperplane rounding is performed on A , i.e.

$$w_A = \frac{\int_{v \in U_A} (\text{cut value along } v) dv}{\int_{v \in U_A} dv},$$

where U_A is the set of all unit vectors in A . Notice that

$$\mathbb{E}Y = \frac{\int_A w_A dA}{\int_A dA}.$$

For the first step of the two-step rounding procedure (i.e., the step selecting a random subspace of dimension k), let Z denote the random variable that takes value w_A when subspace A is selected. We need to show that for $\frac{\log n}{\epsilon}$ i.i.d. random variables $Z_1, \dots, Z_{\frac{\log n}{\epsilon}}$,

there is at least some Z_i such that $Z_i \geq (1 - \epsilon)\mathbb{E}Y$. Since the random subspace is selected uniformly randomly, we have that

$$\mathbb{E}Z = \frac{\int_A w_A dA}{\int_A dA} = \mathbb{E}Y.$$

A similar Markov inequality analysis on Z then gives the result. \square

Proof of Corollary 3

Proof. Let $F'_0 = F'_0(\gamma, \beta)$ be the expected value of MAX-CUT obtained by quantum sampling (i.e., QAOA for $p = 0$). Then,

$$\begin{aligned} & \frac{F'_0}{\text{MAX-CUT}(G)} \\ & \geq \kappa \cdot \frac{F'_0}{\text{HP}(x)} \\ & \quad (\text{since } \text{HP}(x) \geq \kappa \text{MAX-CUT}(G)) \\ & \geq \kappa \min_{(i,j) \in E} \frac{w_{ij} \mathbb{E}[\mathbf{1}[i \ \& \ j \ \text{have different spins}]]}{\frac{w_{ij}}{\pi} \arccos(x_i \cdot x_j)} \\ & \quad \left(\frac{a+c}{b+d} \geq \min\left(\frac{a}{b}, \frac{c}{d}\right) \text{ for } a, b, c, d \geq 0 \right) \\ & = \kappa \min_{(i,j) \in E} \frac{\mathbb{E}[\mathbf{1}[i \ \& \ j \ \text{have different spins}]]}{\frac{1}{\pi} \arccos(x_i \cdot x_j)}. \end{aligned}$$

For $i, j \in [n]$, let θ_{ij} denote the angle between x_i and x_j . We can write $\mathbb{E}[\mathbf{1}[i \ \& \ j \ \text{have different spins}]] = f_k(\theta_{ij})$, that is, this expectation is solely a function of the angle between the adjacent vertices and the dimension k considered. In particular, in Theorem 2 of Tate et al. [31], they show for $k \in \{2, 3\}$ that

$$f_k(\theta_{ij}) = \frac{1}{2} \left(1 - \frac{\cos \theta_{ij}}{k} \right).$$

Additionally, $\arccos(x_i \cdot x_j) = \theta_{ij}$. Using

these substitutions, we have

$$\begin{aligned} & \frac{F'_0}{\text{MAX-CUT}(G)} \\ & \geq \kappa \min_{(i,j) \in E} \frac{\mathbb{E}[\mathbf{1}[i \ \& \ j \ \text{have different spins}]]}{\frac{1}{\pi} \arccos(x_i \cdot x_j)} \\ & = \kappa \min_{(i,j) \in E} \frac{f_k(\theta_{ij})}{\theta_{ij}/\pi} \\ & \geq \kappa \min_{\theta \in [0, \pi]} \frac{f_k(\theta)}{\theta/\pi} \\ & = \kappa \min_{\theta \in [0, \pi]} \frac{\frac{1}{2} \left(1 - \frac{\cos \theta}{k} \right)}{\theta/\pi}. \end{aligned}$$

For $k \in \{2, 3\}$, it is straightforward to verify that the minimum in the last line above is achieved at $\theta = \pi$ (see Figure 10) which leads to the ratio $\frac{F'_0}{\text{MAX-CUT}(G)}$ being at least $\frac{3}{4}\kappa$ and $\frac{2}{3}\kappa$ respectively for $k = 2, 3$. \square

Proof of Observation 4

Proof. We prove something more general than what is stated in Proposition 4. Instead of considering a *particular* cut where hyperplane-rounding yields an approximation ratio of α , we will instead consider a *random* cut (obtained via hyperplane rounding where the hyperplane is selected uniformly at random). We will show that in expectation (considering both the randomness of the hyperplane rounding and the randomness of the quantum sampling), that the expected cut value is at least $0.878 \cos^{2|V|}(\theta^*/2)$.

Let X denote the random variable corresponding to the cut value obtained from the cut obtained by quantum measurement of a perturbed single-cut initialization obtained by GW hyperplane rounding as described in Section 3.3. For any $S \subseteq V$, let $\text{GW}(S)$ denote the event that the cut $(S, V \setminus S)$ was obtained from the GW hyperplane rounding step. Similarly, let $\text{QM}(S)$ denote the event that quantum measurement of the initial state resulted in the cut $(S, V \setminus S)$.

First, observe that if the cut $(S, V \setminus S)$ is used to initialize the quantum state with regularization angle θ^* , then the probability of quantum measurement getting the same cut

is $\cos^{2|V|}(\theta^*/2)$; this is because each vertex independently has probability $\cos^2(\theta^*/2)$ of remaining on the same side of the cut used to initialize the quantum state. Using this fact, we find that,

$$\begin{aligned}
& \mathbb{E}[X] \\
&= \sum_{S \subseteq V} \mathbb{E}[X \mid \text{GW}(S)] \Pr(\text{GW}(S)) \\
&= \sum_{S \subseteq V} \left(\sum_{T \subseteq V} \mathbb{E}[X \mid \text{GW}(S), \text{QM}(T)] \right. \\
&\quad \left. \cdot \Pr(\text{QM}(T) \mid \text{GW}(S)) \cdot \Pr(\text{GW}(S)) \right) \\
&\geq \sum_{S \subseteq V} \left(\mathbb{E}[X \mid \text{GW}(S), \text{QM}(S)] \right. \\
&\quad \left. \cdot \Pr(\text{QM}(S) \mid \text{GW}(S)) \cdot \Pr(\text{GW}(S)) \right) \\
&= \sum_{S \subseteq V} \left(\text{cut}(S) \cdot \cos^{2|V|}(\theta^*/2) \cdot \Pr(\text{GW}(S)) \right) \\
&= \cos^{2|V|}(\theta^*/2) \sum_{S \subseteq V} \left(\text{cut}(S) \Pr(\text{GW}(S)) \right) \\
&\geq \cos^{2|V|}(\theta^*/2) \cdot 0.878 \text{Max-Cut}(G),
\end{aligned}$$

and thus $\frac{\mathbb{E}[X]}{\text{Max-Cut}(G)} \geq 0.878 \cos^{2|V|}(\theta^*/2)$ as desired. In the above formulas, we used the fact that $\mathbb{E}[X \mid \text{GW}(S), \text{QM}(S)] = \text{cut}(S)$ and that the sum $\sum_{S \subseteq V} (\text{cut}(S) \Pr(\text{GW}(S)))$ is simply the expected cut value of the GW algorithm, which we know is at least 0.878 of the optimal cut value for graphs with non-negative weights [5]. \square

B Detailed Description of Custom Mixer

Consider any separable state $|s_0\rangle$ on 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 $j = 1, \dots, n$,

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

Here, θ_j and ϕ_j can be interpreted as the polar and azimuthal angle respectively of the j th qubit on the Bloch sphere. 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:

$$\begin{aligned}
x_j &= \sin \theta_j \cos \phi_j, \\
y_j &= \sin \theta_j \sin \phi_j, \\
z_j &= \cos \theta_j.
\end{aligned}$$

Recall (from Section 2) the custom mixing Hamiltonian H_B is then constructed 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 α about the \hat{n} -axis for some unit vector $\hat{n} = (x, y, z)$; such an operation can be written as:

$$R_{\hat{n}_j}(\alpha) = \exp\left(-i\frac{\alpha}{2}(x\sigma_j^x + y\sigma_j^y + z\sigma_j^z)\right).$$

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 β_k being a variational parameter (to be optimized); this operator, $e^{-i\beta_k H_B}$, can be written as $\prod_{j=1}^n R_{\hat{n}_j,j}(2\beta_k)$, i.e., 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 $2\beta_k$.

C Convergence of Custom Mixers in xz-plane

In order to prove Proposition 1 using the adiabatic theorem, we need to show that (1) $|s_0\rangle$ is indeed the highest-energy eigenstate of the corresponding custom mixer H_B , and (2) the difference between the largest and the second-largest eigenvalues of $H(t) = (1 - t/T)H_B +$

$(t/T)H_C$ is strictly positive. We divide this section into two parts to prove these two statements.

C.1 Eigenstates of Custom Mixers

We show first that $|s_0\rangle$ is the highest energy eigenstate of the corresponding custom mixer for a single qubit, and then generalize to the Kronecker sums of matrices (Proposition 3).

Lemma 7. *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 .*

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 . \square

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 8. *(Theorem 13.16 in [50]) Let $A \in \mathbb{C}^{n \times n}$ have eigenvalues $\lambda_1, \dots, \lambda_n$ and let $B \in \mathbb{C}^{m \times m}$ have eigenvalues μ_1, \dots, μ_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, \dots, x_p ($p \leq n$) are linearly independent eigenvectors of A corresponding to $\lambda_1, \dots, \lambda_n$ and z_1, \dots, z_q ($q \leq m$) are linearly independent eigenvectors of B corresponding to μ_1, \dots, μ_q , then, for all $i \in [p]$ and $j \in [q]$, we have that $x_i \otimes z_j$ are linearly independent eigenvectors of $A \oplus B$ corresponding to $\lambda_i + \mu_j$.*

By applying Theorem 8 with the summands $H_{B,j}$ of the mixing Hamiltonian H_B , we get the desired result as shown in Proposition 3.

Proposition 3. *Let $|s_0\rangle$ be any separable initial state and let H_B be its corresponding custom mixer. Then $|s_0\rangle$ is the highest-energy eigenstate of H_B .*

Proof. Suppose for each $j = 1, \dots, n$ we have a matrix A_j with real eigenvalues and suppose the largest eigenvalue of A_j is λ_j with corresponding eigenvector v_j . As a consequence of Theorem 8, 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 $\bigotimes_{j=1}^n v_j$. Letting $A_j = H_{B,j}$ and $v_j = |s_{0,j}\rangle$ and applying Lemma 7, we see that $|s_0\rangle = \bigotimes_{j=1}^n |s_{0,j}\rangle$ is a highest-energy eigenstate for $H_B = \bigoplus_{j=1}^n H_{B,j}$. \square

C.2 Eigenvalue Gap For Custom Mixers

It is known that if the Quantum Adiabatic Algorithm is run for large 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 [3], thus guaranteeing convergence to the optimal solution. In particular, they applied the following Perron-Frobenius theorem to irreducible stoquastic¹⁵ matrices.

Theorem 9. [51] *Let A be an irreducible matrix whose entries are all real and non-negative. Let r be the spectral radius of A , i.e.,*

¹⁵Stoquastic 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.

$r = \max\{|\lambda| : \lambda \text{ is eigenvalue of } A\}$. 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 $\lambda_1 \geq \dots \geq \lambda_n$; if A is also irreducible and has real, non-negative entries then Theorem 9 ensures a gap between the two largest eigenvalues (otherwise, if $\lambda_1 = \lambda_2$, then the algebraic multiplicity of λ_1 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 10. *Let A be an irreducible stoquastic Hermitian matrix. Then the difference between the largest and second-largest eigenvalue of A is strictly positive.*

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 λ 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 . \square

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 $j = 1, \dots, n$, then one can show that $H(t)$ is an irreducible, stoquastic matrix. Thus by Lemma 10, the eigenvalue gap of $H(t)$ is strictly positive meaning that one can achieve

the optimal solution as the circuit depth $p \rightarrow \infty$ 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. We include their proofs at the end of this section, and go on to prove our main result first.

Proposition 4. *Let n be a positive integer. For each $j = 1, \dots, n$ 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 $0 \leq t \leq T$.*

Proposition 5. *Let n be a positive integer. For each $j = 1, \dots, n$ let x_j be a positive real number and let z_j be any real 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 $0 \leq t < T$.*

We can now prove the convergence for custom mixers of the special 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 $j = 1, \dots, n$ and their corresponding initializations as described in Proposition 1 in Section 2.

Proof. By construction, the corresponding custom mixers will have 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 $j = 1, \dots, n$. The result then follows from Proposition 3, Lemma 10 (which is applicable due to Proposition 4 and Proposition 5), and the adiabatic theorem. \square

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 $p \rightarrow \infty$ is a special case of Propositions 4 and 5.

Proof of Proposition 4 We first prove a technical lemma that is needed in order to prove the proposition.

Lemma 11. *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 ii th block, then it is also an off-diagonal entry of that block but all the off-diagonal entries of the ii 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

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

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. \square

We are now ready to prove Proposition 4.

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 σ^x and σ^z , we have that

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

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 11. \square

Proof of Proposition 5

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 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 positive constant does not affect its irreducibility, so it suffices to prove the irreducibility of H_B .

Observe that σ^x, σ^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 [52], this means that G_{H_B} can be written as the Cartesian graph product of the graphs H_1, H_2, \dots, 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 [53]) which finishes the proof. \square

D Details on Perturbed Single-Cut Initialization

In a perturbed single-cut initialization scheme with cut $(S, V \setminus S)$ and regularization angle θ^* , the quantum state is given by,

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

where,

$$|s_{0,j}\rangle = \begin{cases} R_Y(\theta^*) |0\rangle, & j \in S \\ R_Y(\pi - \theta^*) |0\rangle, & j \notin S \end{cases}$$

where $R_Y(\theta)$ is a single-qubit rotation about the y -axis by angle θ . Geometrically, qubits lie on the xz -plane of the Bloch sphere (with $x > 0$) so that they lie at an angle θ^* away from either the north or south pole of the Bloch sphere, depending on which side of the cut $(S, V \setminus S)$ the vertex is on.

In one of their approaches for Max-Cut, Egger et al. [9] use a mixer for QAOA that is different than both the standard mixer and the custom mixers described in Section 2. They show that their mixer has the property that, when a {perturbed single-cut initialization (based on a cut $(S, V \setminus S)$) with regularization parameter $\theta^* = \pi/3$ is used, that measurement of the depth-1 QAOA with variational parameters $(\gamma_1, \beta_1) = (0, \frac{\pi}{2})$ produces exactly the cut $(S, V \setminus S)$ that was used to initialize the initial quantum state. The drawback is that, with such a mixer proposed by Egger et al., no convergence guarantees are known and experiments suggest that, unlike standard QAOA, the optimal cut value is not achieved in expectation with increased circuit depth.

Cain et al. [8] consider the case where $\theta^* = 0$ and the standard mixer is used. They find that such an approach performs very poorly; in particular, no convergence towards the optimal cut is found with increased circuit depth either.

One can also consider using a perturbed single-cut initialization together with the custom mixers proposed in Section 2; this idea was very briefly explored in the appendices of

Egger et al.'s work [9]. From Proposition 4 and Theorem 1, it is clear that this approach (with non-negative weighted graphs) yields an approximation ratio approaching 0.878 for $\theta^* \rightarrow 0$ for depth-0 QAOA and that such an approach converges to the optimal cut with infinite circuit depth.

Recall (Section 3.1) that QAOA with a (2-dimensional) projected GW initialization has a depth-0 approximation ratio of 0.658. One may be led to believe that, when custom mixers are used, that a perturbed single-cut initialization (with small regularization angle θ^*) is the better choice due to its (theoretically) better approximation ratio at depth-0. However, as seen empirically in Section 4, this is not the case: when θ^* is small, the convergence rate of QAOA with single-cut initializations is (empirically) incredibly slow across all instances. For small θ^* , QAOA with custom mixers geometrically performs rotations around axes that are near the poles of the Bloch sphere about the qubits' initial positions; it is possible that this geometric interpretation is responsible for the slow convergence for small θ^* .

D.1 Relation to QUBO Approach

Egger et al. [9] consider a warm-start approach (which they call continuous warm-started QAOA) for QUBO's of the form

$$\min_{y \in \{0,1\}^n} y^T M y,$$

where $M \in \mathbb{R}^{n \times n}$ is a real-symmetric matrix. They then consider the relaxation

$$\min_{y \in [0,1]^n} y^T M y,$$

i.e. the binary variables are relaxed to lie in the interval $[0, 1]$. For certain matrices¹⁶ M , this yields a convex quadratic program which can be easily solved to (global) optimality [54]; in this case, Egger et al. [9] find and use the

¹⁶In particular, Egger et al. [9] consider matrices of the form $M = N + D$ where $N \in \mathbb{R}^{n \times n}$ is a (symmetric) positive-semidefinite matrix and $D \in \mathbb{R}^{n \times n}$ is a diagonal matrix.

globally optimal solution y^* of the relaxation to produce a separable quantum initial state. We consider this approach in the context of Max-Cut, specifically in the case of graphs with non-negative edge weights.

One can formulate Max-Cut on a graph $G = (V, E)$ with edge weights $w : E \rightarrow \mathbb{R}$ as follows [4]. Simply construct the QUBO matrix M by setting $M_{ij} = w_{ij}$ for $i \neq j$ and $M_{ii} = -\sum_{j=1}^n w_{ij}$ for $i \in \{1, \dots, n\}$. If x^* is an optimal solution to Max-Cut (using the formulation in Equation 2), then there is a corresponding y^* that is an optimal solution of the QUBO such that $x_i^* = 2y_i^* - 1$ for $i = 1, \dots, n$.

Observe that $M = -L$ where L Laplacian matrix of the graph G which is known to be positive-semidefinite (for graphs with non-negative edge weights) [55]. Since L is positive-semidefinite, then the function $f(x) = x^T L x$ is convex in x (as the Hessian of f , $\nabla^2 f(x) = 2L$, is positive-semidefinite). Thus, $x^T M x = -f(x)$ is generally not convex, and hence solving $\min_{x \in [0,1]^n} x^T M x$ to global optimality (as is done by Egger et al. [9]) is non-trivial in the case of Max-Cut. However, we can still consider locally optimal solutions to the relaxation. Observe,

$$\min_{y \in [0,1]^n} y^T M y = \max_{y \in [0,1]^n} y^T L y,$$

i.e., the QUBO relaxation amounts to maximizing a convex function over a polytope, in which case, all strictly local maxima lie on the vertices of the polytope.¹⁷ The vertices of the polytope $[0, 1]^n$ correspond to cuts in the graph, thus, using the strictly locally optimal solution to the relaxation of the QUBO corresponding to Max-Cut degenerates to solutions corresponding to a single-cut; this means that, for Max-Cut (with non-negative edge-weights), this QUBO approach is (effectively) a single-cut initialization approach as described in Section 3.3.

¹⁷To see this, suppose by means of contradiction that y^* was a strict local maximum that did not lie at a vertex of the polytope. Then there exists $z \in \mathbb{R}^n$ such that $y^* - z$ and $y^* + z$ lie in the polytope such that $f(y^*) > f(y^* - z)$ and $f(y^*) > f(y^* + z)$. By convexity, $f(y^*) = f(\frac{1}{2}(y^* - z) + \frac{1}{2}(y^* + z)) \leq \frac{1}{2}f(y^* - z) + \frac{1}{2}f(y^* + z) < \frac{1}{2}f(y^*) + \frac{1}{2}f(y^*) = f(y^*)$, a contradiction.

E Noise Simulations

In Figure 11, we consider 20 instances of Erdős-Rényi 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. This choice is made for simplicity and the following discussion is used as an example of generic noise channel modelling.

As an example, we consider the modelling of phase noise using Kraus operators which equivalently can be described using noise channels. 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[56], 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,

$$\begin{aligned} S(\rho) &= \sum_{k=0}^2 M_k \rho M_k^\dagger \\ &= (1-q)\rho + q|0\rangle\langle 0| \rho |0\rangle\langle 0| \\ &\quad + q|1\rangle\langle 1| \rho |1\rangle\langle 1|. \end{aligned}$$

If we associate the probability of a dephasing event occurring during a time interval Δt , then $q = \Gamma \Delta t$ with Γ 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 ρ_{00}, ρ_{11} are preserved

but exponentially suppresses coherences at a rate defined by Γ , 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 Figures 5, 8, and 9 of the main text. The specifications for these noise models are generated from Qiskit’s **Noise-Model.from_backend** method while using the “*ibmq_guadalupe*” device as the targetted backend. 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. Each gate error consists of a **depolarizing_error()** followed by a **thermal_relaxation_error()** error channel. One can define similar Kraus operators for these noise channels as well [56]. However, while these are a comprehensive treatment of quantum noise they do not accurately capture crosstalk and other correlated noise sources.

As mentioned in the main text, we apply a SPAM mitigation strategy to our hardware results from IBM. Typically these SPAM errors can vary significantly across qubits on a single device. Specifically, typical readout errors on the *ibmq_guadalupe* device range from [1.07%, 12.95%]. Since the overall result is limited by the largest, worst case error, for simplicity, we only mitigate SPAM errors due to the fact that these errors are roughly an order of magnitude larger than gate errors on the same device as well as requiring only two additional circuit runs (independent of circuit width or depth) and do not implement any gate error mitigation strategies, which typically require significantly larger overhead [57, 58].

F Additional Numerical Results

F.1 GW vs BM-MC Warm-Starts

As described in Section 3, we consider two approaches for generating warm-starts: projected GW solutions and locally optimal BM-MC_k solutions, with the former approach

having better theoretical guarantees in regard to solution quality. However, numerical simulations displayed in Figure 14 show both approaches on the instance library \mathcal{G} (graphs with at most 11 nodes) achieve similar expected cut values at depth $p = 1$ QAOA-warmest; in particular, the difference in (instance-specific) approximation ratio is less than 0.04 for nearly all instances. This similarity in solution quality is even more pronounced at depth $p = 8$.

Since both warm-start approaches yield similar results (in numerical simulations) and since the Burer-Monteiro approach scales better in regards to runtime (Section 3), the results in Section 4 assumes that locally optimal BM-MC_k solutions are used to produce the warm-starts for QAOA-warm and QAOA-warmest.

F.2 Choice of Dimension and Rotation

Table 6 demonstrates the average (instance-specific) approximation ratio achieved by QAOA-warmest for various combinations of the dimension 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 $k = 3$ solutions. The data is inconclusive in regards to if $k = 2$ or $k = 3$ solutions are better for QAOA-warmest, both are promising. Finally, we remark that for depth-8 QAOA-warmest, any choice of dimension or rotation scheme gave at least a 0.996 average approximation ratio across the instances.

To give fair comparison against QAOA-warm [31] (and also for comparisons with Egger et al.’s approach [9]), the results in Section 4 assumes that we are using $k = 2$ initializations and vertex-at-top rotations unless otherwise stated, since these were the recommended setting in [31].

F.3 GW and BM-MC Scaling

In Section 3, two methods of creating a warm-start initialization are discussed: projecting GW SDP solutions and finding approximate BM-MC_k solutions. Figure 12 uses instances

		depth $p = 1$		depth $p = 8$		
		vert.	uniform	vert.	uniform	
all graphs	$k = 2$	0.9858	0.9758	$k = 2$	0.9988	0.9977
	$k = 3$	0.9854	0.9535	$k = 3$	0.9988	0.9960
positive-weight graphs	$k = 2$	0.9867	0.9789	$k = 2$	0.9991	0.9990
	$k = 3$	0.9864	0.9581	$k = 3$	0.9993	0.9981

Table 6: Multiple tables comparing the average (instance-specific) approximation ratio achieved during QAOA-warmest when utilizing different combinations of dimension and rotations during the preprocessing stage. For the top row of tables, these averages were computed using all the graphs in our graph library \mathcal{G} whereas for the bottom row, we restrict our attention to only those graphs in \mathcal{G} with positive edge weights.

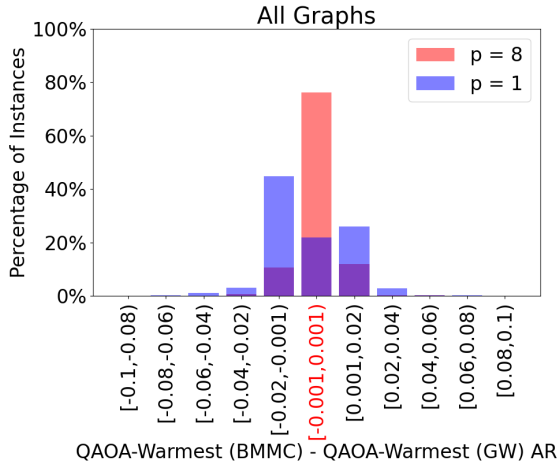


Figure 14: Histogram showing the difference in (instance-specific) approximation ratio (AR) when using QAOA-warmest (on instance library \mathcal{G}) with various warm-start strategies: projected GW and locally optimal BM-MC warm-starts. The blue and red bars correspond to depth $p = 1$ and $p = 8$ QAOA-warmest respectively with the purple regions indicating an overlap in the histograms. For both approaches, $k = 3$ solutions and vertex-at-top rotations are used to produce the figure; the results are similar if one instead uses a different combination of dimension and/or rotation scheme.

from the MQLib [4] library¹⁸ to compare these two methods at different dimensions ($k = 2, 3$) with respect to the (instance-specific) approximation ratio they achieve with hyperplane rounding; these approximation ratios are compared against hyperplane rounding of the n -dimensional GW solution. It is clear from the figure that the projecting of GW solutions preserves the approximation ratio (from hyperplane rounding); this is consistent with the results of Theorem 2. On the other hand, while BM-MC $_k$ solutions preserve the approximation ratio (from hyperplane rounding) for small graphs, the gap in approximation ratios (compared to n -dimensional GW hyperplane rounding) grows as the number of nodes increases.

F.4 Interesting Instances

In Section 4.2, Table 2 shows that at circuit depth $p = 8$, there are five instances for which QAOA-warmest did not achieve the highest (instance-specific) approximation ratio compared to the other algorithms considered.

Of these five instances, standard QAOA was

¹⁸MQLib [4] is a diverse library of Max-Cut instances; the results of Figure 12 may differ if different types/families of graphs are used (e.g. random Erdős-Rényi graphs). For each MQLib instance, we used the exact Max-Cut solver BiqCrunch [59] to try to find the optimal cut. Figure 12 uses all positive-weighted MQLib instances up to 663 nodes for which BiqCrunch was able to find the optimal cut within 24 hours.

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 larger 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) allows QAOA-warmest to outperform GW.

Table 4 gives detailed statistics for the approximation ratios achieved by each of the Max-Cut algorithms considered for these five instances.

F.5 Comparison with Egger et al.

Figure 13 compares the (instance-specific) approximation ratios achieved by QAOA-warmest and a variant of QAOA considered by Egger et al. [9]. 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 γ_1 and β_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 13, we only consider those instances (22.7% of the instance library) for which neither QAOA-warmest or 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 vari-

ational parameters: (1) near the origin and (2) 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 13 demonstrates that QAOA-warmest typically has the superior performance.

There are a total of 163 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 0.0466 and the median gap in approximation ratio between GW and Egger et al.’s approach is 0.0458.

G Twenty-Node Graph

In this section, we discuss why the graph considered in Figure 6 is interesting. The graph used is a 20-node instance, where Goemans-Williamson achieves an (instance-specific) 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) algorithm is 0.878. Karloff [42] showed that the 0.878 bound for GW is tight by constructing a family of graphs whose (instance-specific) approximation ratios 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 $\binom{[m]}{t}$, 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 12. [42] *Let m be an even positive integer and $G = J(m, m/2, b)$. If $0 \leq b \leq m/12$, then the (instance-specific) approxima-*

Instance	# Nodes	# Edges	Approx. Ratio
$J(6, 3, 1)$	20	90	0.9123
$J(8, 4, 1)$	70	560	0.8889
$J(10, 5, 1)$	252	3150	0.8810
$J(10, 5, 2)$	252	12600	0.9402
$J(12, 6, 1)$	924	16632	0.8787
$J(12, 6, 2)$	924	103950	0.9123

Table 7: Small (< 1000 nodes) instances using Karloff's [42] 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.

tion ratio for Goemans-Williamson on G is given by

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

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

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

Conjecture 13. (Conjecture 2.12 in [42]) *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 $\binom{m/2}{b}^2 \left(\frac{4b}{m} - 1\right)$.*

Karloff [42] proved a special case of this conjecture in the case where $0 \leq b \leq m/12$ which was instrumental in showing Theorem 12 (with the same inequality in the hypotheses).

In 2018, Brouwer et al. [60] proved the following theorem below; substituting $t = m/2$ into Theorem 14 and performing a few simple calculations, they also found that Conjecture 13 follows as a corollary.¹⁹

¹⁹The statement of Theorem 14 has been modified in order to be consistent with the notation used in Karloff's construction [42].

Theorem 14. (Theorem 3.10 in [60]) *Let $0 \leq b < t$ and let²⁰*

$$\beta_s = \sum_{r=0}^s (-1)^{s-r} \binom{s}{r} \binom{t-r}{t-b} \binom{m-t-s+r}{m-2t+b}.$$

Then β_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, β_1 is also the second largest in absolute value among the eigenvalues of the adjacency matrix of $J(m, t, b)$.

Conjecture 13 allows us to relax the inequality in Theorem 12 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 5) with $\binom{6}{6/2} = 20$ nodes where GW achieves an (instance-specific) approximation ratio of

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

Table 7 shows all non-trivial instances under 1000 nodes that use Karloff's [42] construction for which we can calculate the approximation ratio using Theorem 12 and Conjecture 13. These instances may be of interest to those working with near-term quantum devices.

²⁰The eigenvalues of the adjacency matrix of $J(m, t, b)$ are $\beta_0, \beta_1, \dots, \beta_t$ (each with positive multiplicity) [61].